1-1-2017

Stochastic Processes and Integrals

Jose L. Menaldi
Wayne State University, menaldi@wayne.edu

Stochastic Processes and Integrals [1]

Jose-Luis Menaldi[2]

Current Version: 12 December 2017[3]
First Version: – — 2014 [4]

[2]Wayne State University, Department of Mathematics, Detroit, MI 48202, USA (e-mail: menaldi@wayne.edu).

[3]Long Title. Stochastic Processes and Stochastic Integrals

[4]This book is being progressively updated and expanded. If you discover any errors or you have suggested improvements please e-mail the author.

Contents

Preface

This project has several parts, of which this book is the fourth one. The first part deals with measure and integration theory, while part two concerns basic function spaces (particularly the theory of distributions. Part three is dedicated to elementary probability (after measure theory), and stochastic ordinary differential equations are discussed in part five, with a clear emphasis on estimates. Each part was designed independent (as much as possible) of the others, but it makes a lot of sense to consider all five parts as a sequence.

This **part four** begins with a quick recall of basic probability, including conditional expectation, random processes, constructions of probability measures and ending with short comments on martingale in discrete time, in a way, this is an enlarged review of part three. Chapter 2 deals with stochastic processes in continuous times, martingales, Lévy processes, and ending with integer random measures. In Chapters 3 and 4, we introduce the stochastic calculus, in two iterations, beginning with stochastic integration and passing through stochastic differentials and ending with stochastic flows. Chapters 5 is more like an appendix, where Makrov process are discussed in a more 'analysis' viewpoint, which ends with a number of useful examples of transition functions.

Most of the style is formal (propositions, theorems, remarks), but there are instances where a more narrative presentation is used, the purpose being to force the student to pause and fill-in the details. Practically, there are no specific section of exercises, giving to the instructor the freedom of choosing problems from various sources (and according to a particular interest of subjects) and reinforcing the desired orientation. There is no intention to diminish the difficulty of the material to put students at ease, on the contrary, all points presented as blunt as possible, even some times shorten some proofs, but with appropriate references.

This book is written for the instructor rather than for the student in a sense that the instructor (familiar with the material) has to fill-in some (small) details and selects exercises to give a personal direction to the course. It should be taken more as Lecture Notes, addressed indirectly (via an instructor) the student. In a way, the student seeing this material for the first time may be overwhelmed, but with time and dedication the reader can check most of the points indicated in the references to complete some hard details, perhaps the expression of a guided tour could be used here. Essentially, it is known that a Proposition in one textbook may be an exercise in another, so that most of the exercises at

this level are hard (or simple), depending on the experience of the student.

The combination of parts IV and V could be regarded as an introduction to 'stochastic control', without making any precise application, i.e., in a neutral way, so that after a good comprehension of this material, the student is ready to fully understand most of the models used in stochastic optimal control theory. In a way, the purpose of these lecture notes is to develop a solid foundation on Stochastic Differential Equations so that *Stochastic Optimal Control* can be widely treated. A solid course in measure theory and Lebesgue spaces is a prerequisite, while some basic knowledge in functional spaces and probability is desired. Moreover, there is not effort to add "exercises" to either of these parts, however, the instructor may find appropriated problems in some of the references quoted in the text.

Michigan (USA), *Jose-Luis Menaldi*, June 2010

Introduction

The reader has several entry points to begin checking this book (as it sequel part five). Essentially, assuming a good background on measure theory (and some elementary probability) the reader may quickly review some basic probability in Chapter 1 and stochastic processes in Chapter 2. The heart of this book is in Chapters 3 and 4, which are dedicated to the theory of stochastic integration or *stochastic calculus* as commonly known. The last Chapter 5 is like a flash on the side, regarding an analytic view of Markov processes. In any case, it may be convenient for the reader to review certain points of 'real analysis', in particular, the interplay of measures, topology and integration, e.g., review Chapters 3 and 6 in our first part-book [123] and most of Chapters 1 and 2 in our second part-book [122].

Rationality for this book

In *Deterministic Control*, if time is regarded as either continuous or discrete then two models can be set, which combined yield the so called hybrid system. The state representation of the continuous model evolves following an ordinary differential equation (ODE) of the form

$$\dot{x}(t) = A(t)x(t) + B(t)v(t), \tag{1}$$

where $t \geq 0$ is the time, $x = x(t)$ is the state and $v = v(t)$ is the control. The state x (in $\mathbb{R}^n$) represents all variables needed to describe the physical system and the control v (in $\mathbb{R}^m$) contains all parameters that can be modified (as a controller's decision) as time passes. The matrices $A(t)$ and $B(t)$ are the coefficients of the system.

The first question one may ask is the validity of the model, which lead to the *identification* of the coefficients. Next, one may want to *control* the system, i.e., to start with an initial state $x(t_0) = x_0$ and to *drive* the system to a prescribed position $x(t_1) = x_0$. Variations of this question are well known and referred to as *controllability*.

Furthermore, another equation appear,

$$y(t) = C(t)x(t), \tag{2}$$

where $y = y(t)$ is the observation of the state and $C(t)$ is another coefficient. Clearly, y is in $\mathbb{R}^d$ with $d \leq n$. Thus, the problem is to reconstruct the state $\{x(t) : t_0 \leq t \leq t_1\}$ based on the observations $\{y(t) : t_0 \leq t \leq t_1\}$, which is called *observability*.

Another key question is the *stabilization* of the system, where one looks for a feedback, i.e., $v(t) = K(t)y(t)$ such that the closed system of ODE (1) and (2) is stable.

Variation of theses four basic questions: identification, controllability, observability and stabilization are solved in text books.

To each control (and state and observation) a cost (or profit) is associated with the intention of being minimized (or maximized), i.e., a performance index of the form

$$J = \int_0^T [y(t)]^* R(t) y(t) \mathrm{d}t + \int_0^T [v(t)]^* N(t) v(t) \mathrm{d}t \tag{3}$$

is to be optimized. This is called an *optimal control* problem.

Two methods are available to solve optimal control problems, namely, the Pontryagin maximum principle and the Bellman dynamic programming. The above (1), (2), (3) linear-quadratic model can be successfully solved by either method. The maximum principle transforms the given (infinite-dimensional optimization) problem into ODE with initial and terminal conditions and a finite-dimensional optimization problem, i.e., a Lagrange multiplier technique. The dynamic programming transforms the given problem into a non-linear partial differential equation (PDE). There is a vast bibliography under the subject *optimal control*, e.g. classic references such as the text book Bertsekas [10], and Fleming and Rishel [50] or more recently Bardi and Capuzzo-Dolcetta [2], among others.

The ODE defining the evolution equations (of the state and the observation) may be nonlinear and the performance index may have a more general form. Moreover, the state could be distribute, i.e., the evolution equation becomes a PDE. Again, there are many references on the subject.

Both, the maximum principle and the dynamic programming are innovations over the classic calculus of variations. The positive part of the maximum principle is the preservation of the equation type (i.e., if the evolution equation is an ODE then the maximum principle equation is an ODE), and the negative part is the open-loop solution (i.e., the optimal control is of the form $v = v(t)$). On the other hand, the positive part of the dynamic programming is the closed-loop or feedback control (i.e., the optimal control has the form $v = K(t, x(t))$), while the negative part is the new equation (i.e., if the evolution equation is an ODE then the dynamic programming equation is an PDE). It is clear that this material is built on the ODE theory.

In *Stochastic Control*, an *uncertainty* component is added to the previous model. The coefficients becomes *random* and the evolution equation includes a *noise*. Perhaps the most typical example is presented in *signal processing*, where

the signal (say x) has some noise. The ODE becomes stochastic

$$\dot{x}(t) = g(t, x(t), v(t)) + (\text{noise}). \tag{4}$$

Since Gauss and Poisson distributions are the main examples of continuous and discrete distributions, the driving noise is usually a Wiener process or a Poisson measure. Again, the four basic questions are discussed. Observability becomes *filtering*, which is very importance. Perhaps the most practical situation is the case with a linear state space and linear observation, which produces the celebrated Kalman filter. Clearly, an average performance index is used for the optimal stochastic control. Again, there is a vast bibliography on stochastic control from variety of points of view, e.g., Fleming and Soner [51], Morimoto [134], Oksendal and Sulem [139], Yong and Zhou [183], Zabczyk [184], among others.

It is clear that stochastic control is mainly based on the theory of stochastic differential equations, which begins with stochastic calculus, which is the main subject of this book.

Chapter 1

Probability Theory

A probability space $(\Omega, \mathcal{F}, P)$ is a measure space with $P(\Omega) = 1$, i.e., a nonempty set Ω (an abstract space) with a σ-algebra $\mathcal{F} \subset 2^\Omega$ of subsets of Ω and an σ-additive function P defined on $\mathcal{F}$. Usually, a measure μ is obtained from an outer measure μ^* by restriction to the measurable sets, and an outer measure is constructed from the expression

$$\mu^*(A) = \inf\big\{ \sum_{n=1}^{\infty} \mu(R_n) : A \subset \bigcup_n R_n,\ R_n \in \mathcal{R}\big\}.$$

Caratheodorys arguments shows that if μ is a σ-additive function defined on a semi-ring $\mathcal{R}$ (i.e., stable under the formation of finite unions and differences, such that the whole space Ω can be written as a countable union of sets in $\mathcal{R}$) then the outer measure defined by the above formula can be restricted to the (Caratheodorys) measurable sets to produce an extension of μ to the σ-algebra generated by $\mathcal{R}$. Also recall that if two measure μ and ν agree on a π-class $\mathcal{E}$ (i.e., containing the empty set and stable under the formation of finite intersections) then $\mu = \nu$ on the σ-algebra generated by $\mathcal{E}$. The reader interested in a guided tour to measure theoretic probability may take a look at the recent book by Pollard [146].

Thus, a probability measure on Ω is a σ-additive function defined on the σ-algebra $\mathcal{F}$ with values in $[0, 1]$ such that $\mu(\Omega) = 1$. A set A in $\mathcal{F}$ satisfying $P(A) = 0$ is called a negligible set or a null sets, or a set of probability zero (and it complement $A^c = \Omega \setminus A$ is a set of probability one or full probability). In probability, *almost surely* (a.s.) is used instead of almost everywhere (a.e.), a set of only one point (singleton) is called an *outcome*, an measurable set (i.e., an element in $\mathcal{F}$) is called an *event*. As discussed later, the integration with respect to the probability measure P is denoted by $\mathbb{E}\{\cdot\}$ and referred to as the expectation.

Random variable or measurable functions are discussed in Section 1, and the key instrument of probability, namely, the conditional expectation goes to Section 2. A first contact with random processes is addressed in Section 3,

while in Section 4 deals with the probability behind random processes. A short presentation on discrete martingales and Markov chains is given in Section 5.

1.1 Random Variables

Recall that a real-valued function x defined on a measurable space $(\Omega, \mathcal{F})$ is measurable if the pre-image $x^{-1}([a,b])$ is in $\mathcal{F}$ for any interval $[a,b]$, and in the probability context, measurable functions are called *random variable*. The σ-algebra $\mathcal{F}_x$ generated by a random variable x is the smallest σ-algebra for which x is measurable, i.e., generated by all sets of the form $x^{-1}(B)$, for any possible set B in a class $\mathcal{K}$ that generates the Borel σ-algebra $\mathcal{B}$ in the line $\mathbb{R}$.

If a probability measure P is defined on $(\Omega, \mathcal{F})$ and x is real-valued random variable then the mapping $B \mapsto P_x(B) = P(x^{-1}(B))$ is a probability measure defined on the Borel σ-algebra $\mathcal{B}$ is called the probability image of P via x, or simply the *law* or *distribution* of x under P. As usually, if two random variables x and y are almost surely equals then x and y should be considered equals, in other words, we work mainly with the vector space $L^0(\Omega, \mathcal{F}, P)$ of equivalence classes (under the a.s. equality) instead of the vector space $\mathcal{L}^0(\Omega, \mathcal{F}, P)$ of all real-valued random variables, and even a completion of the σ-algebra $\mathcal{F}$ is simplicity assumed. Indeed, we say that $x = y$ a.s. iff $x(\omega) = y(\omega)$ for any ω in $\Omega \setminus N$ with $P(N) = 0$; instead of saying that $x = y$ a.s. iff the set $N = \{\omega : x(\omega) \neq y(\omega)\}$ is measurable and $P(N) = 0$. This could be called *almost measurable function* or *almost random variables* to recall that random variables are properly defined only outside of a null event. Also, it is clear that random variables may take values in any measurable space $(E, \mathcal{E})$, but this is left for a later section.

Therefore, a *random variables* represents a measurement obtained while studying a natural object, which is technically viewed as a measurable function x (with values in E) on a probability space, and typically P the Lebesgue measure restricted to $\Omega = (0,1)$, or its infinite product in $[0,1]^\infty$. Usually, to simplify notation, the variable ω is not written explicitly, but the context determine when random elements are presented, e.g., if x is a random variable then the event $x^{-1}(B) = \{\omega \in \Omega : x(\omega) \in B\}$ is shorten to $\{x \in B\}$, and the probability of the event $x^{-1}(B)$ is written as $P(x \in B)$ or $P\{x \in B\}$.

The k-moment of a random variable x is defined by $\mathbb{E}\{x^k\}$, for $k = 1$ this is referred to as the mean and the expression $\mathbb{E}\{(x - \mathbb{E}\{x\})^2\}$ is called the variance, provided the expectation is finite. Recall that $\mathcal{L}^p(\Omega, \mathcal{F}, P)$, $p > 0$, denotes the of all random variables x such that $\mathbb{E}\{|x|^p\} < \infty$, and so, the vector space $L^p(\Omega, \mathcal{F}, P)$ of equivalence classes, which is a Banach space with the norm $x \mapsto (\mathbb{E}\{|x|^p\}^{1/p})$, for $1 \leq p \leq \infty$, and a complete metric for $0 \leq p < 1$.

1.1.1 Measurable Sets

Given a non empty set E (called space), recall that a *σ-algebra* (or *σ-field*) $\mathcal{E}$ is a class (or a subsets of 2^E, the family of subsets of E) containing $\emptyset$ which is

stable under the (formation of) complements and countable unions, i.e., (a) if $A \in \mathcal{E}$ then $A^c = E \setminus A \in \mathcal{E}$ and (b) if $A_i \in \mathcal{A}$, $i = 1, 2, \ldots$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. As mentioned early, the couple $(E, \mathcal{E})$ is called a *measurable space* and each element in $\mathcal{E}$ is called a measurable set. Moreover, the measurable space is said to be *separable* if $\mathcal{E}$ is countable generated, i.e., if there exists a countable class $\mathcal{K}$ such that $\sigma(\mathcal{K}) = \mathcal{E}$, usually, the class $\mathcal{K}$ is at least stable under the formation of finite intersections, and most desirable $\mathcal{K}$ is a (finite) semi-ring (i.e., stable under the formation of finite unions and differences, such that the whole space E can be written as a countable union of sets in $\mathcal{K}$). An atom of a σ-algebra $\mathcal{E}$ is a set A in $\mathcal{E}$ such that any other subset $B \subset A$ with B in $\mathcal{F}$ is either the empty set, $B = \emptyset$, or the whole E, $B = E$. Thus, a σ-algebra separates points (i.e., for any $x \neq y$ in E there exist two sets A and B in $\mathcal{E}$ such that $x \in A$, $y \in B$ and $A \cap B = \emptyset$) if and only if the only atoms of $\mathcal{E}$ are the singletons (i.e., sets of just one point, $\{x\}$ in $\mathcal{E}$).

Borel Sets

Recall that a *topology* on E is a class $\mathcal{T} \subset 2^E$ with the following properties: (1) $\emptyset, E \in \mathcal{T}$, (contain the empty set and the whole space) (2) if $U, V \in \mathcal{T}$ then $U \cap V \in \mathcal{T}$ (stable under finite intersections) and (3) if $U_i \in \mathcal{T}$ for an arbitrary set of indexes $i \in I$ then $\bigcup_{i \in I} U_i \in \mathcal{T}$ (stable under arbitrary unions). Every element of $\mathcal{T}$ is called *open* and the complement of an open set is called *closed*. A basis for a topology $\mathcal{T}$ is a class ${}_b\mathcal{T} \subset \mathcal{T}$ such that for any point $x \in E$ and any open set U containing x there exists an element $V \in {}_b\mathcal{T}$ such that $x \in V \subset U$, i.e., any open set can be written as a union of open sets in ${}_b\mathcal{T}$. Clearly, if ${}_b\mathcal{T}$ is known then also $\mathcal{T}$ is known as the smallest class satisfying (1), (2), (3) and containing ${}_b\mathcal{T}$. Moreover, a class ${}_{sb}\mathcal{T}$ containing $\emptyset$ and such that $\bigcup\{V \in {}_{sb}\mathcal{T}\} = E$ is called a sub-basis and the smallest class satisfying (1), (2), (3) and containing ${}_{sb}\mathcal{T}$ is called the *weakest topology* generated by ${}_{sb}\mathcal{T}$ (note that the class constructed as finite intersections of elements in a sub-basis forms a basis). A space E with a topology $\mathcal{T}$ having a countable basis ${}_b\mathcal{T}$ is commonly used. If the topology $\mathcal{T}$ is induced by a metric then the existence of a countable basis ${}_b\mathcal{T}$ is obtained by assuming that the space E is *separable*, i.e., there exists a countable dense set.

On a topological space $(E, \mathcal{T})$ the Borel σ-algebra $\mathcal{B} = \mathcal{B}(E)$ is defined as the σ-algebra generated by the topology $\mathcal{T}$. If the space E has a countable basis ${}_b\mathcal{T}$, then $\mathcal{B}$ is also generated by ${}_b\mathcal{T}$. However, if the topological space does not have a countable basis then we may have open sets which are not necessarily in the σ-algebra generated by a basis. The couple $(E, \mathcal{B})$ is called a Borel space, and any element of $\mathcal{B}$ is called a Borel set.

Sometimes, a measurable, a Borel or an open set is mentioned without making an explicit reference to the classes $\mathcal{E}$, $\mathcal{B}$ or $\mathcal{T}$. A Borel space $(E, \mathcal{B})$ presupposes a topological space $(E, \mathcal{T})$, which for us should separate points. When a measure (or probability) is defined, the concepts of null sets and almost everywhere (surely) make sense, and a measurable set is the union of a Borel set and a subset of a null set (so-called regular Borel measure). In most cases, the σ-algebra of measurable set is assumed to be completed, and the property that

for any measurable set A with $\mu(A) < \infty$ there exist an open set and a closed set such that $C \subset A \subset O$ with $\mu(C) = \mu(O)$ is desirable.

The classes $\mathcal{F}_\sigma$ (and $\mathcal{G}_\delta$) defined as the countable unions of closed (intersections of open) sets make sense an a topological space E. Moreover, any countable unions of sets in $\mathcal{F}_\sigma$ is again in $\mathcal{F}_\sigma$ and any countable intersections of sets in $\mathcal{G}_\delta$ is again in $\mathcal{G}_\delta$. In particular, if the singletons (sets of only one point) are closed then any countable set is an $\mathcal{F}_\sigma$. However, we can show (with a so-called category argument) that the set of rational numbers is not a $\mathcal{G}_\delta$ in $\mathbb{R} = E$.

In $\mathbb{R}$, we may argue directly that any open interval is a countable (disjoint) union of open intervals, and any open interval (a, b) can be written as the countable union $\bigcup_{n=1}^{\infty}[a + 1/n, b - 1/n]$ of closed sets, an in particular, this shows that any open set (in $\mathbb{R}$) is an $\mathcal{F}_\sigma$. In a metric space (Ω, d), a closed set F can be written as $F = \bigcap_{n=1}^{\infty} F_n$, with $F_n = \{x \in \Omega : \mathrm{d}(x, F) < 1/n\}$, which proves that any closed set is a $\mathcal{G}_\delta$, and by taking the complement, any open set in a metric space is a $\mathcal{F}_\sigma$.

Certainly, we can iterate these definitions to get the classes $\mathcal{F}_{\sigma\delta}$ (and $\mathcal{G}_{\delta\sigma}$) as countable intersections (unions) of sets in $\mathcal{F}_\sigma$ ($\mathcal{G}_\delta$), and further, $\mathcal{F}_{\sigma\delta\sigma}$, $\mathcal{G}_{\delta\sigma\delta}$, etc. Any of these classes are family of Borel sets, but in general, not every Borel set belongs necessarily to one of those classes.

Cartesian Product

Given a family of spaces E_i with a topology $\mathcal{T}_i$ for i in some arbitrary family of indexes I, the product topology $\mathcal{T} = \prod_{i\in I} \mathcal{T}_i$ (also denoted by $\otimes_i \mathcal{T}_i$) on the Cartesian product space $E = \prod_{i\in I} E_i$ is generated by the basis ${}_b\mathcal{T}$ of open *cylindrical sets*, i.e., sets of the form $\prod_{i\in I} U_i$, with $U_i \in \mathcal{T}_i$ and $U_i = \Omega_i$ except for a finite number of indexes i. Certainly, it suffices to take U_i in some basis ${}_b\mathcal{T}_i$ to get a basis ${}_b\mathcal{T}$, and therefore, if the index I is countable and each space E_i has a countable basis then so does the (countable!) product space E. Recall Tychonoff's Theorem which states that any (Cartesian) product of compact (Hausdorff) topological spaces is again a compact (Hausdorff) topological space with the product topology.

Similar to the product topology, if $\{(E_i, \mathcal{E}_i) : i \in I\}$ is a family of measurable spaces then the *product σ-algebra* on the product space $E = \prod_{i\in I} E_i$ is the σ-algebra $\mathcal{E} = \prod_{i\in I} \mathcal{E}_i$ (also denoted by $\otimes_i \mathcal{F}_i$) generated by all sets of form $\prod_{i\in I} A_i$, where $A_i \in \mathcal{E}_i$, $i \in I$ and $A_i = E_i$, $i \notin J$ with $J \subset I$, finite. However, only if I is finite or countable, we can ensure that the product σ-algebra $\prod_{i\in I} \mathcal{E}_i$ is also generated by all sets of form $\prod_{i\in I} A_i$, where $A_i \in \mathcal{E}_i$, $i \in I$. For a finite number of factors, we write $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2 \times \cdots \times \mathcal{E}_n$. However, the notation $\mathcal{E} = \otimes_{i\in I}\mathcal{E}_i$ is preferred (i.e., with $\otimes$ replacing $\times$), to distinguish from the Cartesian product (of classes, which is not used).

• *Remark* 1.1. It is not so hard to show that if E is a topological space such that every open set is a countable union of closed sets, then the Borel σ-algebra $\mathcal{B}(E)$ is the smallest class stable under countable unions and intersections which contains all closed sets.

As seen later, the particular case when all the spaces E_i in the Cartesian product are equals, the notation for the Cartesian product and product of topology and σ-algebras become E^I, $\mathcal{T}^I$ and $\mathcal{B}^T = \mathcal{B}^T(E)$. As mentioned above, for a countable index I we have $\mathcal{B}^I(E) = \mathcal{B}(E^I)$ (i.e., the cylindrical σ-algebra is equal to the Borel σ-algebra of the product topology), but this does not hold in general. In particular, if the index I is uncountable then a singleton may not be measurable. Certainly, the Cartesian product space E^I can be regarded as the space of functions from I into E, and a typical element in E^I written as $(e_i : i \in I)$ can also be interpreted as the coordinate mappings $(e_i) \mapsto e_i$ or $e \mapsto e(i)$, from E^I into E. In this respect, the cylindrical σ-algebra (or product σ-algebra) $\mathcal{B}^I(E)$ is the smallest σ-algebra for which all coordinate mappings are measurable.

1.1.2 Discrete RVs

Discrete random variables are those with values in a countable set, e.g., a discrete real-valued random variable x has values in some set $\{a_n : n = 1, 2, \ldots\} \subset \mathbb{R}$ almost surely, i.e., $P(x = a_n) > 0$ and $\sum_n P(x = a_n) = 1$. This means that the σ-algebra $\mathcal{F}_x$ generated by x is composed only by the atoms $x = a_n$, and the distribution of x is a probability measure P_x on $2^A \subset \mathcal{B}(\mathbb{R})$, with $A = \{a_1, a_2, \ldots\}$, some countable subset of real numbers.

Perhaps the simplest one is a deterministic random variable (i.e., constant function) $x(\omega) = x_0$ for every ω in Ω, whose distribution is the Dirac probability measure concentrated at x_0, i.e., $P_x(B) = 1$ if x_0 belongs to B and $P_x(B) = 0$ otherwise.

A Bernoulli random variable x takes only two values 1 with probability p and 0 with probability $q = 1 - p$, for some $0 < p < 1$. This yields the distribution $P_x(B) = 1$ if 1 and 0 belong to B, $P_x(B) = p$ if 1 belongs to B and 0 does not belong to B, $P_x(B) = 1 - p$ if 0 belongs to B and 1 does not belong to B, and $P_x(B) = 0$ otherwise. Iteration of this random variable (i.e., sequence of Bernoulli independent trials as seen in elementary probability) lead to the Binomial distribution P_x with parameters (n, p), $0 < p < 1$, which is defined on $A = \{0, 1, \ldots, n\}$ and $P_x(\{k\}) = p^k(1 - p)^{n-k}$, for any k in A.

The Geometric distribution with parameter $0 \leq c < 1$ and the Poisson distribution with parameter $\lambda > 0$ are both defined on $A = \{0, 1, 2, \ldots\}$, with $P_x(\{k\}) = (1 - c)c^k$ (Geometric, with the convention $0^0 = 1$), and $P_x(\{k\}) = \mathrm{e}^{-\lambda}\lambda^k/k!$ (Poisson, recall $k! = k(k-1)\ldots 1$), for any k in A.

For any random variable x, the characteristic function (or the Fourier transform) is defined by the complex-valued function

$$\Phi_x(t) = \mathbb{E}\{\mathrm{e}^{\mathrm{i}tx}\} = \sum_{n=0}^{\infty} \mathrm{e}^{\mathrm{i}tn} P(x = n), \quad \forall t \in \mathbb{R},$$

and if X is a random variable with nonnegative integer values then instead of working with its characteristic function Φ_x, we use the so-called (moment)

generating function

$$G_x(t) = \mathbb{E}\{t^x\} = \sum_{n=0}^{\infty} t^n P(x=n), \quad \forall t \in [-1,1],$$

from which all moments can be obtained, i.e., by calculating the derivatives, $G_x(1) = \mathbb{E}\{x\}$, $G_x(1) = \mathbb{E}\{x(x-1)\}$, and so on. Assuming analytic extension, it is clear that $G_x(\mathrm{e}^{\mathrm{i}t}) = \Phi_x(t)$. For the Binomial distribution with parameter (n,p) we have $G_x(t) = [1+p(t-1)]n$, for the Geometric distribution with parameter c we obtain $G_x(t) = (1-c)/(1-ct)$, and for the Poisson distribution with parameter we get $G_x(t) = \exp[(t-1)]$. Note that $\mathbb{E}\{x\} = \lambda$ (mean) and $\mathbb{E}\{(x-\lambda)^2\} = \lambda$ (variance) for a Poisson distributed random variable x.

1.1.3 Continuous RVs

In general, the cumulative distribution function of a real-valued random variable x is defined as $F_x(t) = P\{x \leq t\}$, for any t in $\mathbb{R}$. A probability measure is called diffuse if there is not atoms, i.e., $P\{x = t\} = 0$, for every t in $\mathbb{R}$. In term of the cumulative distribution function, this is equivalently to require that the function $t \mapsto F_x$ is continuous, i.e., $P\{x = t\} = F_x(t)F_x(t)$, where $F_x(t)$ is the left-hand limit. For a real valued random variable, we say that P_x or x has a density (with respect to the Lebesgue measure) if $t \mapsto F_x(t)$ is absolutely continuous and $F_x'(t) = f_x(t)$ defined almost every where for t in $\mathbb{R}$ is called the density function. A simple example is a random variable x with a uniform distribution on some Borel subset K of $\mathbb{R}$ with a positive and finite Lebesgue measure $|K| > 0$, which is defined as $f_x(t) = 1/|K|$ if t belongs to K and $f_x(t) = 0$ otherwise, typically K is a bounded interval.

Therefore, by taken the image of (or transporting) a probability we have established a clear connection between real-valued random variables of a particular distribution and probability measures on the real axis $\mathbb{R}$. As mentioned early, random variables represent measurements used to describe random phenomenons, and so, several distributions of interest appear. Two of them are important for us, first, the Gaussian (or normal) distribution with parameters m and $r > 0$, which is also denoted by $\mathcal{N}(m, r^2)$,

$$P\{x \leq t\} = F_x(t) = \int_{-\infty}^{t} \frac{1}{r\sqrt{2\pi}} \exp\Big(-\frac{|(x-m)/r|^2}{2}\Big)\mathrm{d}x, \quad t \in \mathbb{R},$$

and has mean $\mathbb{E}\{x\} = m$ and variance $\mathbb{E}\{(x-m)^2\} = r^2$. Second, the exponential distribution with parameter $\alpha > 0$, which has support in the semi-line $(0, \infty)$,

$$P\{x \leq t\} = F_x(t) = \int_{0}^{t} \alpha \exp\big(-\alpha x\big)\mathrm{d}x, \quad t \geq 0,$$

and has mean $\mathbb{E}\{x\} = 1/\alpha$ and variance $\mathbb{E}\{(x-1/\alpha)^2\} = 1/\alpha^2$.

The characteristic function of a Gaussian distributed real-valued random variable x is

$$\Phi_x(t) = \widehat{P_x}(t) = \mathbb{E}\big\{\mathrm{e}^{\mathrm{i}tx}\big\} = \exp\big(-rt^2/2 + \mathrm{i}mt\big), \quad t \in \mathbb{R},$$

while for a exponential distributed real-valued random variable x the Laplace transform is also defined

$$\widetilde{P_x}(t) = \mathbb{E}\{\mathrm{e}^{-tx}\} = \frac{\alpha}{\alpha + t}, \quad t \geq 0,$$

and $\Phi_x(t) = \alpha/(\alpha - \mathtt{i}t)$, for any t in $\mathbb{R}$.

1.1.4 Independent RVs

Perhaps the most important concept in probability is about *independence*, i.e., from the elementary idea of two measurable sets A and B (or events as they are called in probability) being independent (i.e., pairwise independent) if $P(A \cap B) = P(A)P(B)$ follows independence (i.e., mutually independence) of three or more events, and in general, independence of sub σ-algebras, and in particular, of measurable functions (i.e., random variables) via their generated sub σ-algebras. This yields

Definition 1.2. A family $\{\mathcal{A}_i : i \in I\}$ of non-empty classes $\mathcal{A}_i \subset 2^\Omega$ of subsets of Ω is called *mutually independent* if $P(A_{i_1} \cap \ldots \cap A_{i_n}) = P(A_{i_1}) \ldots P(A_{i_n})$, for any finite sequence $i_1, \ldots, i_n$ of distinct indexes in I and any choice of sets A_{i_1} in $\mathcal{A}_{i_1}$, ..., A_{i_n} in $\mathcal{A}_{i_n}$. □

In particular, the empty set $\emptyset$ (or a null set) and the whole space Ω (or a set of full probability) are independent of any other sets. Instead of mutually independent, usually just 'independent' is used, while a family $\{\mathcal{A}_i : i \in I\}$ is called *pairwise independent* if for any pair of indexes i and j of I, the family $\{\mathcal{A}_i, \mathcal{A}_j\}$ is independent, i.e., $P(A_i \cap A_j) = P(A_i)P(A_j)$ for any choice of sets A_i in $\mathcal{A}_i$ and A_j in $\mathcal{A}_j$. Recalling that a π-class (or π-system) is a class of sets stable under finite intersections, the about definition implies that if a family $\{\mathcal{A}_i : i \in I\}$ of non-empty classes is independent then the family $\{\sigma(\mathcal{A}_i) : i \in I\}$ is also independent, where $\sigma(\mathcal{A}_i)$ is the σ-algebra generated by the class $\mathcal{A}_i$. Thus, the σ-algebra generated by all null sets is independent of any other σ-algebra.

Similarly, a family $\{\mathcal{A}_i : i \in I\}$ of non-empty π-classes is called *conditional independent given* another non-empty π-class $\mathcal{B}$ if the family $\{\mathcal{A}_i \cap B : i \in I\}$ is independent for every B in $\mathcal{B}$, where $\mathcal{A}_i \cap B$ is the class of subset of the form $A_i \cap B$, with A_i in $\mathcal{A}_i$. Clearly, if $\mathcal{B}$ contains all classes $\mathcal{A}_i$ then conditional independent reduces to independent, and if $\mathcal{B}$ si the σ-algebra generated by all null sets then any family $\{\mathcal{A}_i : i \in I\}$ is independent given $\mathcal{B}$. In most fo the cases, the classes $\mathcal{A}_i$ and $\mathcal{B}$ are either σ-algebras or they reduce to only one element (as in the elementary case).

It should be clear that given a probability space $(\Omega, \mathcal{F}, P)$, it is not possible a priori to ensure the existence of independent random variables with a prescribed distribution. However, the typical (universal) probability space where realization are shown is the Lebesgue space on the interval $[0, 1]$. A well known example is to write any ω in $\Omega = [0, 1]$ in binary, i.e., $\omega = \sum_k 2^{-k}\omega_k$. Then the sequence of variables $\pi_n(\omega) = \omega_n$ for $n = 1, 2, \ldots$ are independent coin-tossing

variables each taking the values 0 or 1 with probability 1/2. Thus, given a mapping $i, j \mapsto k(i,j)$ which is injective from $\{1,2,\dots\} \times \{1,2,\dots\}$ into $\{1,2,\dots\}$, the expression $X_i = \sum_j 2^{-k(i,j)}\omega_{k(i,j)}$ for $i = 1,2,\dots$ defines an independent sequence of random variables, each with the same distribution as X, $X(\omega) = \omega$, i.e., each with the uniform distribution on $[0,1]$.

The construction of examples of independent sequences of random variables involve some conditions (infinitely divisible) on the probability space $(\Omega, \mathcal{F}, P)$, for instance if the σ-algebra $\mathcal{F} = \{\emptyset, F, \Omega \setminus F, \Omega\}$, with $P(F) > 0$, then any two independent sets A and B must be such that $A = \emptyset$ or $B = \emptyset$. There are many (classic) properties related to an independent sequence or series of random variables, commonly known as the (weak and strong) law of large numbers and the central limit theorem, e.g., the reader is referred to the classic probability books Doob [33], Feller [48] and Gnedenko [64], while an analytic view can be found in Dudley [37], Folland [52, Chapter 10], Halmos [67]), Stromberg [167] and Stroock [168].

In general, if S_i is a Borel space (i.e., a measurable space isomorphic to a Borel subset of $[0,1]$, for instance any complete separable metric space), P_i is a probability measure on the Borel σ-algebra $\mathcal{B}_i(S_i)$, for $i = 1,2,\dots$ then there exists a sequence $\{\xi_1, \xi_2, \dots\}$ of independent random variables defined on the universal Lebesgue probability space $[0,1]$ such that $P_i(B) = P(\{\omega : \xi_i(\omega) \in B\})$, for any B in $\mathcal{B}_i(S_i)$, $i = 1,2,\dots,$ i.e., the distribution of ξ_i is exactly P_i, e.g., see Kallenberg [88, Theorem 3.19, pp. 55–57].

There are several results regarding a sequence of independent events that are useful for us, e.g., the Borel-Cantelli Lemma and the Kolmogorov $0-1$ Law of which some details are given below.

Theorem 1.3 (Borel-Cantelli). *Let $\{A_i\}$ be a sequence of measurable sets, define the superior limit set $A = \bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i$. Then $\sum_{i=1}^{\infty} P(A_i) < \infty$ implies $P(A) = 0$. Moreover, if $\{A_i\}$ are also independent and $\sum_{i=1}^{\infty} P(A_i) = \infty$ then $P(A) = 1$.*

Proof. to check the first part, note that $A \subset \bigcup_{i=n}^{\infty} A_i$ and in view of the σ-sub-additivity, we have $P(A) \le \sum_{i=n}^{\infty} P(A_i)$. Since the series converges, the remainder satisfies $\sum_{i=n}^{\infty} P(A_i) \to 0$ as $n \to \infty$, i.e., $P(A) = 0$.

Now, using the complement, $A^c = \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} A_i^c$ and because A_i are independent, we obtain

$$1 - P(A) = P(A^c) = \lim_n P\Big(\bigcap_{i=n}^{\infty} A_i^c\Big) = \\ = \lim_n \lim_m \prod_{i=n}^{m} P\Big(\bigcap_{i=n}^{m} A_i^c\Big) = \lim_n \lim_m \prod_{i=n}^{m} \big(1 - P(A_i)\big).$$

Since $\ln(1-t) \le -t$ for every $0 \le t < 1$, we get

$$\sum_{i=n}^{m} \ln\big(1 - P(A_i)\big) \le -\sum_{i=n}^{m} P(A_i),$$

i.e.,

$$\prod_{i=n}^{m}\big(1-P(A_i)\big)\le\exp\big(-\sum_{i=n}^{m}P(A_i)\big),$$

which yields $P(A)=1$. □

As a corollary, we deduce a simple version of the $(0-1)$ *zero-one law*, i.e., if $\{A_n\}$ is a sequence of independent sets, then for $A\subset\bigcap_{i=n}^{\infty}A_i$ we have $P(A)=0$ or $P(A)=1$.

In general, this point can be better seen as follows. For a sequence $\{x_n\}$ of random variables define the sub σ-algebras:

$$\mathcal{F}_n^{\infty}=\sigma(x_k:k\ge n),\quad \mathcal{F}^n=\sigma(x_k:k\le n),\quad \mathcal{F}_\infty=\textstyle\bigcap_n\sigma(x_k:k\ge n),$$

where $\mathcal{F}_\infty$ is called the *tail* σ-algebra. It is clear that $\mathcal{F}_\infty\subset\mathcal{F}^\infty=\sigma\big(\bigcup_n\mathcal{F}^n\big)$.In the particular case of independent set of the form $A_n=x_n^{-1}(B_n)$, with B_n Borel sets, we note that the limit set $A\subset\bigcap_{i=n}^{\infty}A_i$ belongs to the tail σ-algebra $\mathcal{F}_\infty$.

Theorem 1.4 (Kolmogorov $0-1$ Law)**.** *Let $\{x_n\}$ be a sequence of independent random variables and $\mathcal{F}_\infty$ be the corresponding tail σ-algebra. Then, for each A in $\mathcal{F}_\infty$ we must have $P(A)=0$ or $P(A)=1$.*

Proof. By assumption, $\mathcal{F}_n^{\infty}$ and $\mathcal{F}^{n-1}$ are independent, i.e., if $A\in\mathcal{F}_n^{\infty}$ and $B\in\mathcal{F}^{n-1}$ we have $P(A\cap B)=P(A)\,P(B)$. Hence, $A\in\mathcal{F}_\infty\subset\mathcal{F}_n^{\infty}$ and $B\in\cup_n\mathcal{F}^n$ yield $P(A\cap B)=P(A)\,P(B)$, and by means of a monotone class argument, the last equality remains true for every $B\in\sigma\big(\bigcup_n\mathcal{F}^n\big)$. Since $\mathcal{F}_\infty\subset\sigma\big(\bigcup_n\mathcal{F}^n\big)$ we can take $A=B$ in $\mathcal{F}_\infty$ to have $P(A)=P(A)^2$, i.e., the desired result. □

As a consequence of the $0-1$ law, for any sequence $\{x_n\}$ of independent random variables, we have (1) since the set $\{\omega:\lim_n x_n(\omega)\ \text{exists}\}$ belongs to $\mathcal{F}_\infty$, the sequence x_n converges or diverges almost surely; (2) each random variable measurable with respect to $\mathcal{F}_\infty$, is indeed constant almost surely, in particular

$$\limsup_n x_n,\qquad \liminf_n x_n,\qquad \limsup_n\frac{1}{n}\sum_{i\le n}x_i,\qquad \liminf_n\frac{1}{n}\sum_{i\le n}x_i$$

are all constant almost surely.

It easy to realize that a family $\{x_n:n\in N\}$ of independent real valued random variable satisfies

$$\mathbb{E}\big\{\mathrm{e}^{\mathrm{i}\sum_{i\in J}t_jx_j}\big\}=\prod_{i\in J}\mathbb{E}\big\{\mathrm{e}^{\mathrm{i}t_jx_j}\big\},\quad\forall t_j\in\mathbb{R},$$

for any finite subset of index $J\subset N$, and it can be proved that the converse is also true.

1.1.5 Construction of RVs

It should be clear that random variables could take values in $\mathbb{R}^d$ instead of $\mathbb{R}$. For instance, for any d-dimensional vector m and invertible square matrix R, a d-dimensional Gaussian random variable x has a distribution P_x absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^d$,

$$P\{x \in B\} = \int_B [2\pi \det(RR^*)]^{-d/2} \exp\Big(\frac{|R^{-1}(x-m)|^2}{2}\Big)\mathrm{d}x,$$

for any Borel set B in $\mathbb{R}^d$, with mean $\mathbb{E}\{x\} = m$ and matrix-covariance RR^*, where $(\cdot)^*$ and $\det(\cdot)$ denote the adjoint and the determinant of a matrix. Its characteristic function is

$$\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}x\cdot\xi}\big\} = \mathrm{e}^{\mathtt{i}m\cdot\xi - |R\xi|^2/2}, \quad \forall \xi \in \mathbb{R}^d,$$

where $\cdot$ denotes the dot (scalar) product in $\mathbb{R}^d$. However, if π is a finite measure on $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$ then a random variable x with a composed Poisson distribution with parameter π is better expressed by its characteristic function

$$\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}x\cdot\xi}\big\} = \exp\Big[\int_{\mathbb{R}^d_*} \big(\mathrm{e}^{\mathtt{i}\zeta\cdot\xi} - 1\big)\pi(\mathrm{d}\zeta)\Big], \quad \forall \xi \in \mathbb{R}^d,$$

than by its actually distribution.

It is clear by now that modeling a random variable with a prescribed distribution is equivalent to choosing a particular probability measure on the space $\mathbb{R}^d$. One way of constructing a probability measure is by prescribing its characteristic function, classical Bochner's Theorem in $\mathbb{R}^d$ addresses this question

Theorem 1.5. *If $\Psi : \mathbb{R}^n \to \mathbb{C}$ is the characteristic function of a probability measure (space) $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P)$, i.e.,*

$$\Psi(\xi) = \int_{\mathbb{R}^n} \exp\big(\mathtt{i}(\xi, x)\big)\, P(\mathrm{d}x) = \mathbb{E}\big\{\exp\big(\mathtt{i}(\xi, \cdot)\big)\big\},$$

with $\mathtt{i} = \sqrt{-1}$, then (a) $\Psi(0) = 1$, (b) Ψ *is continuous and* (c) Ψ *is positive definite, i.e., for every natural number k, any ξ_i in $\mathbb{R}^n$ and any complex number z_i, $i = 1, \dots, k$ we have*

$$\sum_{i,j=1}^{k} \Psi(\xi_i - \xi_j) z_i \bar{z}_j \geq 0,$$

where $(\cdot,\cdot)$ denotes the scalar product in $\mathbb{R}^n$ and $\bar{z}$ is the conjugate of a complex number. Conversely, an arbitrary function $\Psi : \mathbb{R}^n \to \mathbb{C}$ satisfying the above properties (a), (b) *and* (c) *is the characteristic function of a probability measure P on $\mathbb{R}^n$.* □

The continuity follows from the dominated convergence theorem, and the equality

$$\sum_{i,j=1}^{k} \Psi(\xi_i - \xi_j) z_i \bar{z}_j = \int_{\mathbb{R}^d} \Big| \sum_{i=1}^{k} z_i \mathrm{e}^{\mathrm{i}\xi_i} \Big|^2 P(\mathrm{d}x) \geq 0, \quad \forall \xi_i, z_i,$$

shows that Ψ is positive definite. The converse is longer and it uses the fact that a nonnegative (tempered) distribution is indeeda measure, e.g., see Pallu de la Barrière [140, Theorem 7.1, pp. 157–159].

Bochner's Theorem 1.5 is used to construct a probability measure (or equivalent a random variable) in $\mathbb{R}^d$ with a composed Poisson distribution corresponding to a finite measure π on $\mathbb{R}^d_*$ as its parameter. Moreover, remarking that the characteristic function of a d-dimensional Gaussian random variable makes sense even if the square-matrix (parameter) R is not necessarily invertible, degenerate Gaussian distributions could be studied. Certainly, there are many other application of this results.

1.2 Conditional Expectation

The *conditional expectation* is intrinsically related to the concept of independence, and this operation is defined either as an orthogonal projection (over a subspace of functions measurable over a particular sub σ-algebra) or via Radon-Nikodym theorem. Moreover, the concepts of independent and conditional expectation are fundamental for probability theory and in fact, this is the main distinction with classical measure theory.

Definition 1.6 (conditional expectation)**.** Let x is an integrable random variable and $\mathcal{G}$ be a sub σ-algebra on a probability space $(\Omega, \mathcal{F}, P)$. An integrable random variable Y is called a *conditional expectation* of x given $\mathcal{G}$ if (a) y is $\mathcal{G}$-measurable and (b) $\mathbb{E}\{x\mathbb{1}_G\} = \mathbb{E}\{y\mathbb{1}_G\}$ for every set G in $\mathcal{G}$. The notation $y = \mathbb{E}\{x \mid \mathcal{G}\}$ is used, and if z is another random variable then $\mathbb{E}\{x \mid z\} = \mathbb{E}\{x \mid \sigma(z)\}$, where $\sigma(z)$ is the σ-algebra generated by z. However, if A is in $\mathcal{F}$ then $\mathbb{E}\{x \mid A\} = \mathbb{E}\{x\mathbb{1}_A\}/\mathbb{E}\{\mathbb{1}_A\}$ becomes a number, which is referred to as the conditional expectation or *evaluation* of x given A, provided that $P(A) > 0$. Even the evaluation $\mathbb{E}\{x \mid z = z_0\} = \mathbb{E}\{x \mid z^{-1}(z_0)\}$ for any value z_0 could be used. It is clear that this definition extends to one sided integrable (either the positive or the negative part is integrable) and σ-integrable (integrable on a each part of a countable partition of the whole space) random variables. □

In a sense we may say that conditional expectation is basic and fundamental to probability. A conditional expectation is related to the disintegration of probability measure, and it is a key concept to study martingales. Note first that if $x' = x$ almost surely then y is also a conditional expectation of x' given $\mathcal{G}$, and second, if y' is another conditional expectation of x given $\mathcal{G}$ then $\mathbb{E}\{(y - y')\mathbb{1}_G\} = 0$ for every G in $\mathcal{G}$, which yields $y = y'$ almost surely, because $y - y'$ is $\mathcal{G}$-measurable. This means that conditional expectation should be

properly considered as a operation on equivalence classes of functions, i.e., on the space $L^1(\Omega, \mathcal{F}, P)$. However, the conditional expectation is regarded as acting on the space of integrable random variables $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, where a choice of an element in the equivalence class have been made.

Definition 1.6 should be complemented with the following existence result:

Theorem 1.7. *If $\mathcal{G}$ is a sub σ-algebra on a given probability space $(\Omega, \mathcal{F}, P)$ then there exists a linear operator from $L^1(\Omega, \mathcal{F}, P)$ into $L^1(\Omega, \mathcal{G}, P)$ denoted by $\mathbb{E}\{\cdot \,|\, \mathcal{G}\}$ representing the conditional expectation, i.e., if x and y are integrable random variable satisfying $y = \mathbb{E}\{x \,|\, \mathcal{G}\}$ almost surely, then Y is a conditional expectation of x given $\mathcal{G}$.*

Proof. As mentioned early, the conditional expectation $\mathbb{E}\{x \,|\, \mathcal{G}\}$ given $\mathcal{G}$ is (uniquely determined up to null sets) a $\mathcal{G}$-measurable random variable satisfying

$$\int_A \mathbb{E}\{x \,|\, \mathcal{G}\}(\omega) P(\mathrm{d}\omega) = \int_A x(\omega) P(\mathrm{d}\omega), \quad \forall A \in \mathcal{G}. \tag{1.1}$$

Thus, the expression $A \mapsto \mathbb{E}\{\mathbb{1}_A x\}$ defines a signed measure on the measure space $(\Omega, \mathcal{G}, P)$, which is absolutely continuous with respect to P. Hence, the Radon-Nikodym theorem ensures the existence and uniqueness (up to null sets) of conditional expectations, i.e., given x and $\mathcal{G}$ there exists a null set N (which may depends on both x and $\mathcal{G}$) such that $\omega \to \mathbb{E}\{x \,|\, \mathcal{G}\}(\omega)$ is uniquely defined for ω in $\Omega \setminus N$. It should be understood that the conditional expectation acts on integrable random variables, which are identified almost surely, i.e., on the Banach space $L^1(\Omega, \mathcal{F}, P)$.

An alternative construction (without referring to the Radon-Nikodym theorem) is based on the orthogonal projection on the Hilbert space $L^2(\Omega, \mathcal{F}, P)$, i.e., the operation $x \mapsto \mathbb{E}\{x \,|\, \mathcal{G}\}$ is initially defined as the orthogonal projection on $L^2(\Omega, \mathcal{G}, P)$, which is considered as a closed subspace of $L^2(\Omega, \mathcal{F}, P)$. This mapping preserves the positive cone (i.e., if $x \geq 0$ then $\mathbb{E}\{x \,|\, \mathcal{G}\} \geq 0$), and so, a monotone extension yields a definition on the whole space $L^1(\Omega, \mathcal{F}, P)$, i.e., any nonnegative random variable x is written as the almost surely pointwise increasing limit $x = \lim_n x_n$ of a (almost surely monotone) sequence $\{x_n\}$ and the conditional expectation is defined by $\mathbb{E}\{x \,|\, \mathcal{G}\} = \lim_n \mathbb{E}\{x_n \,|\, \mathcal{G}\}$ as an almost surely pointwise increasing limit. □

Occasionally, the conditional expectation is used on σ-integrable variables. First, a random variable x is called σ-integrable with respect to a σ-algebra $\mathcal{G}$ if there exists a (increasing) sequence $\{\Omega_n\}$ in $\mathcal{G}$ such that $P(\Omega_n) \to 1$ and $\mathbb{E}\{|x \mathbb{1}_{G_n}|\} < \infty$, for every $n \geq 1$. Next, if x is a σ-integrable with respect to a σ-algebra $\mathcal{G}$ then $\mathbb{E}\{x \,|\, \mathcal{G}\}$ is defined as the unique (almost surely) $\mathcal{G}$-measurable random variable satisfying condition (1.1), for every A in $\mathcal{G}$ such that $\mathbb{E}\{|x \mathbb{1}_A|\} < \infty$.

1.2.1 Main Properties

Conditional expectation has properties similar to those of the integral, i.e., there are a couple of properties that are inherited from the integral:

(a) $x \le y$ a.s. implies $\mathbb{E}\{x \mid \mathcal{G}\} \le \mathbb{E}\{y \mid \mathcal{G}\}$ a.s.

(b) $\mathbb{E}\{y \mid \mathcal{G}\} = y$ a.s. if y is $\mathcal{G}$-measurable, in particular if Y is a constant function.

(c) If y is bounded and $\mathcal{G}$-measurable, then $\mathbb{E}\{xy \mid \mathcal{G}\} = y\mathbb{E}\{x \mid \mathcal{G}\}$ a.s.

(d) $\mathbb{E}\{x + y \mid \mathcal{G}\} = \mathbb{E}\{x \mid \mathcal{G}\} + \mathbb{E}\{y \mid \mathcal{G}\}$ a.s.

(e) If $A \in \mathcal{G}$ and if $x = y$ a.s. on A, then $\mathbb{E}\{x \mid \mathcal{G}\} = \mathbb{E}\{y \mid \mathcal{G}\}$ a.s. on A.

(f) If $A \in \mathcal{G}_1 \cap \mathcal{G}_2$ and $A \cap \mathcal{G}_1 = A \cap \mathcal{G}_2$ (i.e., if any subset of A is in $\mathcal{G}_1$ if and only if the subset is in $\mathcal{G}_2$), then $\mathbb{E}\{x \mid \mathcal{G}_1\} = \mathbb{E}\{x \mid \mathcal{G}_2\}$ a.s. on A.

(g) If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathbb{E}\{\mathbb{E}\{x \mid \mathcal{G}_1\} \mid \mathcal{G}_2\} = \mathbb{E}\{\mathbb{E}\{x \mid \mathcal{G}_2\} \mid \mathcal{G}_1\} = \mathbb{E}\{x \mid \mathcal{G}_1\}$ a.s.

(h) If x is independent of $\mathcal{G}$, then $\mathbb{E}\{x \mid \mathcal{G}\} = \mathbb{E}\{x\}$ a.s.

(i) If x is a fixed integrable random variable and $\{\mathcal{G}_i : i \in I\}$ denotes all possible sub σ-algebra on a probability space $(\Omega, \mathcal{F}, P)$ then the family $\{y_i : i \in I\}$ of random variables of the form $y_i = \mathbb{E}\{x \mid \mathcal{G}_i\}$ is uniformly integrable.

(j) Jensen's inequality for conditional expectations, i.e., if ϕ is a convex real-valued function, and x is an integrable random variable such that $\phi(x)$ is also integrable then $\phi\big(\mathbb{E}\{x \mid \mathcal{G}\}\big) \le \mathbb{E}\{\phi(x) \mid \mathcal{G}\}$ a.s.

Most of the above listed properties are immediate obtained from the definition and construction of the conditional expectation, in particular, from the inequality (a) follows that $-|x| \le x \le |x|$ yields $|y| \le \mathbb{E}\{|x| : \mathcal{G}\}$ with $y = \mathbb{E}\{x|\mathcal{G}\}$, which can be used to deduce (i). Indeed, the definition of conditional expectation implies that $\mathbb{E}\{|y|\mathbb{1}_{|y|>k}\} \le \mathbb{E}\{|x|\mathbb{1}_{|y|>k}\}$ and $kP\{|y| > k\} \le \mathbb{E}\{|y|\} \le \mathbb{E}\{|x|\}$, i.e., for k large, the probability $P\{|y| > k\}$ is small and therefore $\mathbb{E}\{|x|\mathbb{1}_{|y|>k}\}$ is small, which yields $\mathbb{E}\{|y|\mathbb{1}_{|y|>k}\}$ small. Similarly, expressing a convex function ϕ as the supremum of all linear functions it majorizes, the property (j) is obtained. Also, from the monotonicity (see also Vitali type Theorems)

Theorem 1.8 (Fatou Type)**.** *Let $\mathcal{G}$ be a sub σ-algebras on the probability space $(\Omega, \mathcal{F}, P)$ and let $\{x_n : n = 1, 2, \ldots\}$ be a sequence of nonnegative extended real valued random variables. Under these assumptions* $\liminf_{n\to\infty} \mathbb{E}\{x_n \mid \mathcal{G}\} \le \mathbb{E}\{\liminf_{n\to\infty} x_n|\mathcal{G}\}$*, a.s. Moreover, if the sequence $\{x_n\}$ is uniformly integrable then* $\limsup_{n\to\infty} \mathbb{E}\{x_n \mid \mathcal{G}\} \ge \mathbb{E}\{\limsup_{n\to\infty} x_n \mid \mathcal{G}\}$*, a.s.* □

Certainly, all these properties are valid (with obvious modifications) for σ-integrable random variable with respect to a σ-algebra $\mathcal{G}$.

1.2.2 Conditional Independence

Now, let us discuss the concept of conditional independence (for two events or σ-algebras or random variables) given another σ-algebra or random variable). If $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{C}$ is a sub σ-algebras of $\mathcal{F}$, then any two events (measurable sets) A and B are (conditional) independent given $\mathcal{C}$ if

$$\mathbb{E}\{\mathbb{1}_A \mathbb{1}_B \,|\, \mathcal{C}\} = \mathbb{E}\{\mathbb{1}_A \,|\, \mathcal{C}\}\, \mathbb{E}\{\mathbb{1}_B \,|\, \mathcal{C}\}, \quad \text{a.s.} \tag{1.2}$$

holds. Moreover, two sub σ-algebras $\mathcal{H}$ and $\mathcal{G}$ are (conditional) independent given $\mathcal{C}$ (relative to the probability P) if (1.2) is satisfied for any sets $A \in \mathcal{H}$, $B \in \mathcal{G}$. Particularly, if the sub σ-algebras are generated by a family of random variables, i.e., $\mathcal{H} = \sigma(x(i) : t \in I)$, $\mathcal{G} = \sigma(x(j) : j \in J)$ and $\mathcal{C} = \sigma(z(k) : k \in K)$, then (1.2) is equivalent to

$$\mathbb{E}\Big\{ \prod_i h_i(X(i)) \prod_j g_j(Y(j)) \prod_k c_k(Z(k)) \Big\} = \\ = \mathbb{E}\Big\{ \mathbb{E}\{\prod_i h_i(X(i)) \,|\, \mathcal{C}\}\, \mathbb{E}\{\prod_j g_j(Y(j)) \,|\, \mathcal{C}\} \prod_k c_k(Z(k)) \Big\},$$

where all products are extended to any finite family of subindexes and any real-valued bounded measurable functions h_i, g_j and c_k.

Certainly this concept extends to a family of measurable sets, a family of either sub σ-algebras or random variables, where mutually or pairwise (conditional independent given $\mathcal{C}$) are not the same.

In relation to orthogonality, remark that if $\mathcal{G}$ is a σ-algebra of $\mathcal{F}$ and x is an square integrable random variable with zero mean (i.e., $\mathbb{E}\{|x|^2\} < \infty$ and $\mathbb{E}\{x\} = 0$) then the conditional expectation $\mathbb{E}\{x \,|\, \mathcal{G}\}$ is the orthogonal projection of x onto the subspace $L^2(\Omega, \mathcal{G}, P)$ of $L^2(\Omega, \mathcal{F}, P)$. Similarly, two sub σ-algebras $\mathcal{H}$ and $\mathcal{G}$ are (conditional) independent given $\mathcal{C}$ (relative to the probability P) if and only if the subspace $\{x \in L^2(\Omega, \mathcal{G}, P) \cap L^2(\Omega, \mathcal{C}, P) : \mathbb{E}\{x\} = 0\}$ is orthogonal to $\{x \in L^2(\Omega, \mathcal{H}, P) \cap L^2(\Omega, \mathcal{C}, P) : \mathbb{E}\{x\} = 0\}$ in $L^2(\Omega, \mathcal{F}, P)$.

1.2.3 Regular Conditional Probability

A technical (but necessary) follow-up is the so-called *regular conditional probability* $P(B \,|\, \mathcal{G}) = \mathbb{E}\{1_B \,|\, \mathcal{G}\}$, which requires separability of the σ-algebra $\mathcal{F}$ or some topology on the abstract probability space Ω to define a function $(B, \omega) \mapsto P(B \,|\, \mathcal{G})(\omega)$ satisfying the σ-additivity property almost surely. The conditional probability is useful to establish that a family $\{\mathcal{A}_i : i \in I\}$ of non-empty π-classes is conditional independent given a σ-algebra $\mathcal{B}$ if and only if

$$P(A_{i_1} \cap \ldots \cap A_{i_n} \,|\, \mathcal{B}) = P(A_{i_1} \,|\, \mathcal{B}) \ldots P(A_{i_n} \,|\, \mathcal{B}), \quad \text{almost surely},$$

for any finite sequence $i_1, \ldots, i_n$ of distinct indexes in I and any choice of sets A_{i_1} in $\mathcal{A}_{i_1}$, $\ldots$, A_{i_n} in $\mathcal{A}_{i_n}$. It should be clear that the concept of independence makes sense only in the presence of a probability, i.e., a family of non-empty π-classes is independent with respect to a given probability.

Definition 1.9 (conditional probability)**.** A *transition kernel* $Q(\omega, A)$ on a probability space $(\Omega, \mathcal{F}, P)$ is a mapping from $\Omega \times \mathcal{F}$ into $[0,1]$ such that (a) for each A in $\mathcal{F}$ the function $\omega \mapsto Q(\omega, A)$ is a $\mathcal{F}$-measurable function and (b) for each ω in Ω the function $A \mapsto Q(\omega, A)$ is a probability measure on $(\Omega, \mathcal{F})$. A *regular probability measure* given a sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$ is a transition kernel denoted by $(\omega, A) \mapsto P\{A \,|\, \mathcal{G}\}(\omega)$ such that for any A in $\mathcal{F}$ the random variable $\omega \mapsto P\{A \,|\, \mathcal{G}\}(\omega)$ is a conditional expectation of $\mathbb{1}_A$, i.e., $\mathbb{E}\{\mathbb{1}_A \,|\, \mathcal{G}\} = P\{A \,|\, \mathcal{G}\}$, almost surely, which means that

$$P(A \cap B) = \int_B P\{A \,|\, \mathcal{G}\}(\omega) P(\mathrm{d}\omega), \quad \forall B \in \mathcal{G},$$

and $\omega \mapsto P\{A \,|\, \mathcal{G}\}(\omega)$ is $\mathcal{G}$-measurable. If the σ-algebra $\mathcal{G}$ is generated by a random variable Z then $P\{\mathbb{1}_A | Z\} = P\{\mathbb{1}_A | \sigma(Z)\}$ and $\mathbb{E}\{\mathbb{1}_A | Z\} = \mathbb{E}\{\mathbb{1}_A | \sigma(Z)\}$. In particular, if $Z = \mathbb{1}_G$ the characteristic function of some measurable set G then $\sigma(\mathbb{1}_G) = \{\emptyset, \Omega\} = \sigma(G)$ and $P\{A \,|\, \mathbb{1}_G\} = P\{A \,|\, \sigma(G)\}$. However, $P\{A \,|\, G\} = \mathbb{E}\{\mathbb{1}_A \,|\, G\} = P(A \cap G)/P(G)$ is a number that represents the *evaluation* of the conditional probability of A given G, provided $P(G) > 0$. □

Note that in the above definition, a kernel transition Q may be defined almost surely in the sense that there is a set N of probability zero such that the mapping $Q(\omega, A)$ is defined for any ω in $\Omega \setminus N$ and any A in $\mathcal{F}$ satisfying the measurability in ω and the σ-additivity in A. In general the mapping $(\omega, A) \mapsto \mathbb{E}\{\mathbb{1}_A | \mathcal{G}\}(\omega)$ satisfies the measurability in ω but, the σ-additivity is only satisfied almost surely, i.e., for each sequence $\{A_n\}$ of disjoint measurable sets with $A = \sum_n A_n$ there exists a set N of probability zero such that $\mathbb{E}\{\mathbb{1}_A \,|\, \mathcal{G}\}(\omega) = \sum_n \mathbb{E}\{\mathbb{1}_{A_n} | \mathcal{G}\}(\omega)$, for every ω in $\Omega \setminus N$. Now, we can prove the following result:

Theorem 1.10 (regular)**.** *Let $\mathcal{G}$ be sub σ-algebra on the probability space $(\Omega, \mathcal{F}, P)$, where Ω is a complete separable metric (Polish) space and $\mathcal{F} = \mathcal{B}(\Omega)$ is its Borel σ-algebra. Then there exists a regular conditional probability $P\{\cdot | \mathcal{G}\}$, i.e.,* (a) *for each A in $\mathcal{F}$ the function $\omega \mapsto P\{A \,|\, \mathcal{G}\}(\omega)$ is $\mathcal{G}$-measurable,* (b) *for every $A \in \mathcal{F}$ and $B \in \mathcal{G}$ we have*

$$P(A \cap B) = \int_B P\{A \,|\, \mathcal{G}\}(\omega) P(\mathrm{d}\omega),$$

and (c) *for each ω in Ω the function $A \mapsto P\{A \,|\, \mathcal{G}\}(\omega)$ is a probability measure on Ω and $P\{B \,|\, \mathcal{G}\}(\omega) = \mathbb{1}_B(\omega)$, for any ω in Ω and B in $\mathcal{G}_0$, where $\mathcal{G}_0$ is any finite-generated sub σ-algebra of $\mathcal{G}$.*

Proof. Because Ω is a Polish (complete separable metrizable) space its Borel σ-algebra $\mathcal{F}$ is separable, e.g., its is generated by the countable set $\mathcal{A}_0$ of all open balls with rational radii and centers in a countable dense set. Certainly, this countable set $\mathcal{A}_0$ generates an algebra $\mathcal{A}$, which is expressed a an increasing sequence of finite-generated algebras, and so, $\mathcal{A}$ is countable.

Also, any probability measure is regular in a Polish space, i.e., for every A in $\mathcal{A}$ there exists a an increasing sequence of compact sets $\{A_i\}$ such that

$\bigcup_i A_i = A$ and the monotone convergence implies that $P\{A_i \,|\, \mathcal{G}\} \to \mathbb{E}\{A \,|\, \mathcal{G}\}$ almost surely. These compact sets $\{A_i\}$ and the algebra $\mathcal{A}$ generate a countable algebra denoted by $\bar{\mathcal{A}}$. Hence, for a given finite-generated sub σ-algebra $\mathcal{G}_0$ of $\mathcal{G}$, we can choose a negligible set N such that the $\mathcal{G}$-measurable function $\omega \mapsto P\{F \,|\, \mathcal{G}\} = \mathbb{E}\{\mathbb{1}_F \,|\, \mathcal{G}\}$ satisfies, for every ω in $\Omega \setminus N$, the following conditions:

(a).- for every A in $\bar{\mathcal{A}}$ we have $P\{A \,|\, \mathcal{G}\}(\omega) \geq 0$,

(b).- we have $P\{B \,|\, \mathcal{G}\}(\omega) = \mathbb{1}_B(\omega)$ for every B in $\mathcal{G}_0$,

(c).- the function $A \mapsto P\{A \,|\, \mathcal{G}\}(\omega)$ is finitely additive on the algebra $\bar{\mathcal{A}}$,

(d).- for every A in $\mathcal{A}$ and the specify sequence $\{A_i\}$ chosen above we have $P\{A_i \,|\, \mathcal{G}\}(\omega) \to P\{A \,|\, \mathcal{G}\}(\omega)$.

Indeed, the above conditions are countable restriction on ω.

This conditions imply that

$$P\{A \,|\, \mathcal{G}\}(\omega) = \sup\big\{P\{K \,|\, \mathcal{G}\}(\omega) : K \subset A, K \in \bar{\mathcal{A}},\, K \text{ is compact}\big\},$$

which yields the σ-additivity of $P\{\cdot \,|\, \mathcal{G}\}(\omega)$ on $\mathcal{A}$. Indeed, by contradiction, if not, there exists $\delta > 0$ and a decreasing sequence $\{A_i\}$ in $\mathcal{A}$ such that $\bigcap_i A_i = \emptyset$ and $P\{A_i \,|\, \mathcal{G}\}(\omega) > \delta$. Then for each i there exists a compact set K_i in $\bar{\mathcal{A}}$ with $K_i \subset A_i$ and $P\{A_i \setminus K_i \,|\, \mathcal{G}\}(\omega)| < \delta 3^{-i}$. Therefore, for each n we have

$$P\{K_1 \cap \cdots \cap K_n \,|\, \mathcal{G}\}(\omega) \geq P\{C_n \,|\, \mathcal{G}\}(\omega) - \sum_{i=1}^{n} \delta 3^{-i} \geq \frac{\delta}{2},$$

which implies that $K_1 \cap \cdots \cap K_n$ is not empty, i.e., the sequence $\{K_i \cap K_1\}$ of compact subsets of K_1 has the finite intersection property. Since K_1 is compact, we must have $\bigcap_i K_i \neq \emptyset$, which contradict the fact that $\bigcap_i A_i = \emptyset$.

Finally, because $P\{\cdot \,|\, \mathcal{G}\}(\omega)$ is σ-additivity on $\mathcal{A}$, for every ω in $\Omega \setminus N$, it can be uniquely extended to a measure on $\mathcal{F} = \sigma(\mathcal{A})$. To complete the arguments, we redefine $P\{A \,|\, \mathcal{G}\}(\omega) = \mathbb{1}_A(\omega)$ for any ω in N. □

Note that the condition $P\{B \mid \mathcal{G}\}(\omega) = \mathbb{1}_B(\omega)$, for any ω in Ω and B in $\mathcal{G}_0$, any finite-generated sub σ-algebra of $\mathcal{G}$ is not really necessary, it suffices to impose only $P\{\Omega \mid \mathcal{G}\}(\omega) = 1$ and $P\{\emptyset \mid \mathcal{G}\}(\omega) = 0$ on the condition (b) of the construction given on the above proof to obtain a regular conditional probability.

Remark that in term of random variables, this result can be re-stated as follows: Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ be a sub σ-algebra, and x be a random variable with values in some Polish space E endowed with its Borel σ-algebra $\mathcal{E}$). Then, we can choose a regular conditional probability $P_x\{A \mid \mathcal{G}\}$ i.e., (a) for each A in $\mathcal{E}$ the function $\omega \mapsto P\{x^{-1}(A) \mid \mathcal{G}\}(\omega)$ is $\mathcal{G}$-measurable, (b) for every $A \in \mathcal{E}$ and $B \in \mathcal{G}$ we have

$$P(x^{-1}(A) \cap B) = \int_B P\{x^{-1}(A) \mid \mathcal{G}\}(\omega) P(\mathrm{d}\omega),$$

and (c) for each ω in Ω the function $A \mapsto P\{x^{-1}(A) \mid \mathcal{G}\}(\omega)$ is a probability measure on Ω and $P\{B \,|\, \mathcal{G}\}(\omega) = \mathbb{1}_B(\omega)$, for any ω in Ω and B in $\mathcal{G}_0$, where $\mathcal{G}_0$ is any finite-generated sub σ-algebra of $\mathcal{G}$.

1.3 Random Processes

Taking measurements of a random phenomenon as time goes by involves a family of random variables indexed by a parameter playing the role of the time, which is know as a random (or stochastic) process $X = \{X_t : t \in T\}$. Note the use of either X_t or $X(t)$ to indicated a random variable belonging to the family refereeing to the random process X. The so-called arrow of time yields a complete order (denoted by $\leq$ and $<$) on the index T, which can be considered discrete $T = \{t_0, t_1, \ldots\}$ (or simply $T = \{0, 1, 2, \ldots\}$) or continuous T is an interval in $\mathbb{R}$ (or simply $T = [0, \infty)$ or $T = [0, \infty]$ if necessary). Note that if T is the set of all nonnegative rational numbers then T is countable but not completely a discrete index of times, due to the order. Thus, a family $\mathbb{F}^X = \{\mathcal{F}_t^X : t \in T\}$ of increasing sub σ-algebras of $\mathcal{F}$ (so-called filtration) is associated with any random process X, where $\mathcal{F}_t^X$ is generated by the random variable x_s with $s \leq t$. This family $\mathbb{F}^X$ is called the history (or internal history) of X, or in general the filtration generated by X. A probability space with a filtration is called a filtered space $(\Omega, \mathbb{F}, P)$, where $\mathcal{F}_\infty$ is the minimum σ algebra containing all $\mathcal{F}_t$, for any $t \in T$, and usually, $\mathcal{F} = \mathcal{F}_\infty$. An important technical result on measurability affirms that any $\mathcal{F}_t$-measurable random variable Y should have the form $Y = f(X_{s_1}, \ldots, X_{s_k}, \ldots)$ for some sequence $\{s_k : k \geq 1\} \subset [0, t]$, where f is a Borel measurable function, and several concepts related to processes are attached to a filtration, e.g., adapted, predictable, optional, etc.

Typically, the random variables take values in some Borel space $(E, \mathcal{E})$, where E is an suitable subset of $\mathbb{R}^d$, usually $E = \mathbb{R}$. Mathematically, it is clear that a family of random variables X (with values in E and indexed by T) is equivalent to a random variable with values in the product space E^T, which means that not regularity is imposed on the path, i.e. the functions $t \mapsto x_t(\omega)$, considered for each fixed ω. In a way to be discussed later, if T is uncountable then the product space E^T is too big or equivalent, the cylindrical Borel σ-algebra $\mathcal{B}^T(E)$ is too small.

Realization of a stochastic process X refers to the construction of a probability space $(\Omega, \mathcal{F}, P)$ or better a filtered space $(\Omega, \mathbb{F}, P)$, where the stochastic process X is defined and satisfies some prescribed properties, such as the statistics necessary to describe X as a random variable with valued in the product space E^T and some pathwise conditions that make the mathematical analysis possible.

1.3.1 Discrete RPs

To motivate some delicate points in the theory of continuous time processes we discuss first *sequences of random variables*, i.e., random processes in discrete time. First, a filtered space is a (usually complete) probability space $(\Omega, \mathcal{F}, P)$ and an increasing sequence (so-called *filtration*) of sub σ-algebras $\mathbb{F} = (\mathcal{F}_n : n = 0, 1, \ldots)$, $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n$, for all $n = 1, 2, \ldots$, such that $\mathcal{F}_0$ contains all null sets of $\mathcal{F}$. A stochastic sequence (or process) $(X_n : n = 0, 1, \ldots)$ is a sequence of $\mathbb{R}$-valued (or $\mathbb{R}^d$-valued) random variables, 'identified' almost surely (i.e., P-equivalence

class). Its associated natural filtration is the sequence $(\mathcal{F}_n : n = 0, 1, \ldots)$ of sub σ-algebras generated by $\{X_0, X_1, \ldots, X_n\}$ and augmented with all null sets, i.e., $\sigma[X_0, X_1, \ldots, X_n]$ and all null sets. Given a filtered space, a stochastic sequence (or process) $(X_n : n = 0, 1, \ldots)$ is called *adapted* if every random variable X_n is $\mathcal{F}_n$-measurable. Also, it is called *predictable* if every random variable X_n is $\mathcal{F}_{n-1}$-measurable, for any $n = 1, 2, \ldots$, here X_0 is ignored or taken equal to zero. A stopping time η is a maps (identified almost surely) from Ω into the set $\{0, 1, \ldots, \infty\}$ such that $\{\eta \leq n\}$ (or equivalently $\{\eta = n\}$) belongs to $\mathcal{F}_n$ for any $n \geq 0$, where either $\mathcal{F}_\infty = \mathcal{F}$ or $\mathcal{F}_\infty$ is the minimal σ-algebra containing all $\mathcal{F}_n$, $n \geq 0$. For an given stopping time, the σ-algebra $\mathcal{F}_\eta$ is defined as the collection of all subsets A in $\mathcal{F}$ such that $A \cap \{\eta \leq n\}$ (or equivalently $A \cap \{\eta = n\}$) belongs to $\mathcal{F}_n$, for any $n \geq 0$. Note that a typical stopping time is the *hitting time* (or entry time) of a Borel subset B of $\mathcal{B}(\mathbb{R})$ (or $\mathcal{B}(\mathbb{R}^d)$) for a stochastic sequence $(X_n : n = 0, 1, \ldots)$, i.e., $\eta = \inf\{n \geq 0 : X_n \in B\}$, where $\eta = \infty$ if X_n does not belong to B for any $n \geq 0$.

In measure theory the construction of a finite product of measures requires some analysis, which does not extent to a countable product of measures. However, a construction of the direct product of probability spaces is possible (e.g., Halmos [67, Section VII.38, Theorem B, pp. 157–158]), namely, there exists a unique probability measure P on the (countable) product space $\Omega = \prod_n \Omega_n$ with the product σ-algebra $\mathcal{F}$ (generated by the collection of cylindrical (or cylinder) sets $C_n = \prod_{k=1}^n F_k \times \prod_{k=n+1}^\infty \Omega_k$, with F_k in $\mathcal{F}_k$,) such that $P(C_n) = \prod_{k=1}^n P_k(F_k)$ for every cylindrical set. Note that the countable assumption is really not an issue, it can be easily dropped.

A direct consequence of the above result is the construction of sequences of independent and identically distributed $\mathbb{R}^d$-valued random variables, i.e., given a distribution μ on $\mathbb{R}^d$ the exists a stochastic sequence $(Z_n : n = 0, 1, \ldots)$ on a complete probability space $(\Omega, \mathcal{F}, P)$ such that

$$P(Z_k \in B) = \mu(B), \quad \forall B \in \mathcal{B}(\mathbb{R}^d), \tag{1}$$

$$P(Z_k \in B_k,\ \forall k = 1, \ldots, n) = \prod_{k=1}^n P(Z_k \in B_k), \tag{2}$$

for every B_k in $\mathcal{B}(\mathbb{R}^d)$ and any $n \geq 1$, where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-algebra in $\mathbb{R}^d$. In this context, the series of partial sum $X_0 = 0$, $X_n = \sum_{k=1}^n Z_k$ is called a *random walk* in $\mathbb{R}^d$ or a d-dimensional random walk with incremental distribution μ.

Ionescu-Tulcea's theorem (e.g., Neveu [136, Section V.1, pp. 153–159], Shiryayev [160, Section II.9, Theorem 2, pp. 243–250]), is a generalization of the infinite product of probabilities, which is specially designed for construction of Markov chains (processes) from transition functions. To present this result on *product probability*, we need some notation. First, a *transition probability* between two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ is a function $Q : \Omega \times \mathcal{F}' \to [0, 1]$, $Q(\omega, F')$, which is measurable in ω and a probability in F'. Note two particular cases, (1) $Q(\omega, F') = P(F')$ a fixed probability on $(\Omega', \mathcal{F}')$ for every ω in Ω, and (2) $Q(\omega, F') = \mathbb{1}_{\{q(\omega) \in F'\}}$ where $q : \Omega \to \Omega'$ is a measurable function.

For $(\Omega_i, \mathcal{F}_i)$ a sequence of measurable spaces, the product σ-algebra $\mathcal{F} = \prod_{i=1}^{\infty} \mathcal{F}_i$ on the product space $\Omega = \prod_{i=1}^{\infty} \Omega_i$ is generated by the cylindrical sets

$$C_n = \prod_{i=1}^{n} F_i \times \prod_{i=n+1}^{\infty} \Omega_i, \quad \text{with} \quad F_i \in \mathcal{F}_i, \ \forall i, \quad n = 1, 2, \dots \tag{1.3}$$

For a fixed n, denote by $\mathcal{F}^n$ a sub σ-algebra of $\mathcal{F}$ generated by the n-cylindrical sets as above. It is clear that $\mathcal{F}^n$ can be identified with the σ-algebra $\prod_{i=1}^{n} \mathcal{F}_i$ of finite product space $\prod_{i=1}^{n} \Omega_i$, and that $\mathcal{F}$ is generated by the algebra $\cup_n \mathcal{F}^n$.

Let P_1 be a probability on $(\Omega_1, \mathcal{F}_1)$ and Q_k be a transition probability from finite product space $(\prod_{i=1}^{k-1} \Omega_i, \prod_{i=1}^{k-1} \mathcal{F}_i)$ into $(\Omega_k, \mathcal{F}_k)$, for $k \geq 2$. We desire to construct a probability P on the infinite product space $(\Omega, \mathcal{F})$ such that

$$P(C_n) = \int_{F_1} P_1(\mathrm{d}\omega_1) \int_{F_2} Q_2(\omega_1, \mathrm{d}\omega_2) \dots \int_{F_n} Q_n(\omega_1, \dots, \omega_{n-1}, \mathrm{d}\omega_n),$$

for any cylindrical set C_n as in (1.3). Note that if P_n denotes the restriction of P to $\prod_{i=1}^{n} \mathcal{F}_i$ (i.e., the finite-dimensional distributions of P) then the right-hand term prescribes a particular form for P_n, where a disintegration (by means of the transition probability Q_n) is assumed a priori. Comparing with Kolmogorov's extension theorem (see next subsections), here it is assumed that the finite-dimensional distributions enjoy a disintegration condition, instead of a topological assumption in the spaces Ω_i.

Now, for a fixed n, consider the following expression constructed backward by induction:

$$P(\omega_1, \dots, \omega_n; F) = \mathbb{1}_{F^n}(\omega_1, \dots, \omega_n), \quad F = F^n \times \prod_{i=n+1}^{\infty} \Omega_i, \ F^n \in \prod_{i=1}^{n} \mathcal{F}_i,$$

$$P(\omega_1, \dots, \omega_{k-1}; F) = \int_{\Omega_k} P(\omega_1, \dots, \omega_{k-1}, \omega_k; F)\, Q_k(\omega_1, \dots, \omega_{k-1}, \mathrm{d}\omega_k),$$

$$P(\omega_1; F) = \int_{\Omega_2} P(\omega_1, \omega_2; F)\, Q_2(\omega_1, \mathrm{d}\omega_2), \quad P(F) = \int_{\Omega_1} P(\omega_1; F)\, P_1(\mathrm{d}\omega_1).$$

A Fubini-Tonelli type theorem ensures that each step of the above construction is possible and that $P(\omega_1, \dots, \omega_k; F)$ is a transition probability from the (finite) product space $(\prod_{i=1}^{k} \Omega_i, \prod_{i=1}^{k} \mathcal{F}_i)$ into $(\Omega, \mathcal{F}^n)$, for any $k = n, \dots, 1$; and that $P(F)$ is a probability on $(\Omega, \mathcal{F}^n)$. It is also clear that for cylindrical sets as (1.3)

we have

$$P(C_n) = \int_{F_1} P_1(\mathrm{d}\omega_1) \int_{F_2} Q_2(\omega_1, \mathrm{d}\omega_2) \dots \int_{F_n} Q_n(\omega_1, \dots, \omega_{n-1}, \mathrm{d}\omega_n),$$

$$P(\omega_1, \dots, \omega_{k-1}; F) = \Big(\prod_{i=1}^{k-1} \mathbb{1}_{F_i}(\omega_i) \Big) \int_{F_k} Q_k(\omega_1, \dots, \omega_{k-1}, \mathrm{d}\omega_k) \times$$

$$\times \int_{F_{k+1}} Q_{k+1}(\omega_1, \dots, \omega_{k-1}, \omega_k, \mathrm{d}\omega_{k+1}) \dots \int_{F_n} Q_n(\omega_1, \dots, \omega_{n-1}, \mathrm{d}\omega_n),$$

$$P(\omega_1, \dots, \omega_n; C_n) = \prod_{i=1}^{n} \mathbb{1}_{F_i}(\omega_i),$$

and therefore, $P(\omega_1, \dots, \omega_n; F) = P(\omega_1, \dots, \omega_{n-1}; F)$ for any F in $\mathcal{F}^{n-1}$. This last property allows us to consider $n = 1, 2, \dots$ and to extend (uniquely) the definition of $P(\omega_1, \dots, \omega_n; F)$ to F in the algebra $\mathcal{F} = \bigvee_n \mathcal{F}^n$.

Theorem 1.11 (Ionescu-Tulcea)**.** *Under the above notation, the function*

$$P_n(\omega, F) = P(\omega_1, \dots, \omega_n; F), \quad \textit{with} \quad \omega = (\omega_1, \dots, \omega_n, \dots),$$

is a transition probability from $(\Omega, \mathcal{F}^n)$ *into* $(\Omega, \mathcal{F})$*. Moreover* $(\Omega, \mathcal{F}, P)$ *is a probability space on which* P_n *provides a regular conditional probability for the* σ*-algebra* $\mathcal{F}^n$*.*

Proof. Only a brief idea is given. The central point is show the σ-additivity of P_n on the algebra $\bigvee_n \mathcal{F}^n$ with $P_0 = P$, and then to use Caratheodory extension to have a probability on $\mathcal{F}$. To this purpose, suppose that there exists a decreasing sequence $\{A_k\}$ in $\bigcup_n \mathcal{F}^n$ such that $\bigcap_k A_k = \emptyset$ with $\lim_k P(A_k) \neq 0$. Then, the above construction of the P_1 show that there exists a ω_1^* such that $\lim_k P(\omega_1^*; A_k) \neq 0$, and by induction, we can construct a sequence $\omega^* = (\omega_1^*, \dots, \omega_n^*, \dots)$ such that $\lim_k P(\omega_1^*, \dots, \omega_n^*; A_k) \neq 0$. Since A_k belongs to some $\mathcal{F}^m$ with $m = m(k)$, from the construction of P we obtain $P(\omega_1^*, \dots, \omega_n^*; A_k) = \mathbb{1}_{A_k}(\omega^*)$ if $n \geq m(k)$. Hence ω belongs to A_k for every k, which is a contradiction. □

It is interesting to note that there is almost no difficulty to extend Tulcea's construction to a general product space with an index non necessarily countable. Indeed, we assume that P_s, with $s = (t_1, \dots, t_n)$, has the form

$$P_s(C_n) = \int_{F_1} P_{t_1}(\mathrm{d}\omega_1) \int_{F_2} Q_{t_1,t_2}(\omega_1, \mathrm{d}\omega_2) \dots \int_{F_n} Q_{t_1,\dots,t_n}(\omega_1, \dots, \omega_{n-1}, \mathrm{d}\omega_n),$$

for some family of transition probabilities $\{Q_s : s = (s', t),\ s' \in T^{n-1},\ n \geq 2,\ t \in T\}$ from $(\Omega^{s'}, \mathcal{F}^{s'})$ into $(\Omega_t, \mathcal{F}_t)$, and any cylindrical set $C_n = \prod_{t \in T} F_t$ with $F_t = \Omega_t$ if $t \neq t_i$ for every i, and $F_{t_i} \in \mathcal{F}_{t_i}$. Hence, we can construct a family of consistent probability on any countable product. Since only a countable number of finite-dimensional is involved in proving the σ-additivity, we have a probability in general product space Ω. Thus, the disintegration of the finite-dimensional

distributions in term of the transition probabilities $\{Q_s : s \in T^n,\ n \geq 2\}$ replace the extra condition on inner regular measures. Moreover, Tulcea's construction yields an expression for a regular conditional distribution on any countable subset of indexes.

1.3.2 Continuous RPs

On a given probability space $(\Omega, \mathcal{F}, P)$, the statistics of a stochastic processes $X = \{X(t), t \geq 0\}$ are represented by its finite-distributions, i.e., a family of probabilities $P_s(B) = P(X(s_1) \in B_1, \ldots, X(s_n) \in B_n)$, with $s = (s_1, \ldots, s_n)$ in $[0, \infty)^n$, $n = 1, 2, \ldots$, and B_i Borel (usually open or closed) subsets of $\mathbb{R}$. Thus, if a real-valued stochastic process X is interpreted as a family of random variables $X(t)$, $t \geq 0$, then X can also be regarded as a random variable with values in the product space $\mathbb{R}^{[0,\infty)}$ endowed with the cylindrical σ-algebra $\mathcal{B}^{[0,\infty)}$. To simplify notation, assume processes take values in E and the time t is in T, e.g., for a d-dimensional process in continuous time $E = \mathbb{R}^d$ and $T = [0, \infty)$. Thus, a point x in the product space E^T is denoted by $\{x_t : t \in T\}$, and a cylindrical set takes the from $B = \{B_t : t \in T\}$ with B_t a Borel subset of E satisfying $B_t = E$ except for a finite number of indexes t, and clearly, the cylindrical σ-algebra (which is not exactly the Borel σ-algebra generated by the open sets in the product topology) is generated by all cylindrical (or cylinder) sets.

If the index set T models the time then it should have an order (perhaps only partial) denoted by $\leq$ with the convention that $<$ means $\leq$ and $\neq$, when $T = [0, \infty)$ or $T = \{0, 1, 2, \ldots\}$ the order is complete. In any case, if a family of finite-dimensional distributions $\{P_s : s \in T^n,\ n = 1, 2, \ldots\}$ on a Borel subsets of $E = \mathbb{R}^d$ is obtained from a stochastic process, then they must satisfy a set of (natural) *consistency conditions*, namely

(a) if $s = (s_{i_1}, \ldots, s_{i_n})$ is a permutation of $t = (t_1, \ldots, t_n)$ then for any B_i in $\mathcal{B}(E)$, $i = 1, \ldots, n$, we have $P_t(B_1 \times \cdots \times B_n) = P_s(B_{i_1} \times \cdots \times B_{i_n})$,

(b) if $t = (t_1, \ldots, t_n)$ and $s = (s_1, \ldots, s_m)$ with $t_1 < \cdots < t_n < r < s_1 < \ldots < s_m$ and $A \times B$ in $\mathcal{B}(E^n) \times \mathcal{B}(E^m)$ then $P_{(t,r,s)}(A \times E \times B) = P_{(t,s)}(A \times B)$, for any $n, m = 0, 1, \ldots$.

The converse of this assertion is given by the following classic Kolmogorov (sometime called Daniell-Kolmogorov or Čentsov-Kolmogorov) construction or the *coordinate method* of constructing a process (see Kallenberg [88], Karatzas and Shreve [91], Malliavin [116], Revuz and Yor [151], among others, for a comprehensive treatment).

Theorem 1.12 (Kolmogorov). *Let $\{P_s : s \in T^n,\ n = 1, 2, \ldots\}$ be a consistent family of finite-dimensional distributions on a Borel subset E of $\mathbb{R}^d$. Then there exists a probability measure P on $(E^T, \mathcal{B}^T(E))$ such that the canonical process $X_t(\omega) = \omega(t)$ has $\{P_s\}$ as its finite-dimensional distributions.* □

Under the consistency conditions, an additive function can be easily defined on product space $(E^T, \mathcal{B}^T(E))$, the question is to prove its σ-additive property.

In this respect, we point out that one of the key conditions is the fact that the (Lebesgue) measure on the *state* space $(E, \mathcal{B}(E))$ is *inner regular* (see Doob [34, pp. 403, 777]). Actually, the above result remains true if E is a Lusin space, i.e., E is homeomorphic to a Borel subset of a compact metrizable space. Note that a Polish space is homeomorphic to a countable intersection of open sets of a compact metric space and that every probability measure in a Lusin space is inner regular, see Rogers and Williams [153, Chapter 2, Sections 3 and 6].

Note that a cylinder (or cylindical) set is a subset C of E^T such that ω belongs to C if and only if there exist an integer n, an n-uple $(t_1, t_2, \ldots, t_n)$ and $B \in \mathcal{B}(E^n)$ such that $(\omega(t_1), \omega(t_2), \ldots, \omega(t_n))$ belongs to B for any $i = 1, \ldots, n$. The class of cylinder sets with $t_1, \ldots, t_n$ fixed is equivalent to product σ-algebra in $E^{\{t_1,\ldots,t_n\}} \simeq E^n$ and referred to as a finite-dimensional projection. However, unless T is a finite set, the class of all cylinder sets is only an algebra. Based on cylinder sets, another way of re-phrasing the Kolmogorov's construction theorem is saying that any (additive) set function defined on the algebra of cylinder sets such that any finite-dimensional projection is a probability measure, has a unique extension to a probability measure on E^T. In particular, if $T = \{1, 2, \ldots\}$ then the above Kolmogorov's theorem shows the construction of an independent sequence of random variables with a prescribed distribution. In general, this is a realization of processes where the distribution at each time is given.

Note that a set of only one element $\{a\}$ is closed for the product topology of E^T and so it belongs to the Borel σ-algebra $\mathcal{B}(E^T)$ (generated by the product topology in E^T). However, the product σ-algebra $\mathcal{B}^T(E)$ (generated by cylinder sets) contains only sets that can be described by a countable number of restrictions on E, so that $\{a\}$ is not in $\mathcal{B}^T(E)$ if T is uncountable. Thus we see the importance of finding a subset Ω of E^T having the full measure under the outer measure P^* derived from P, which is itself a topological space.

1.3.3 Versions of RPs

To fully understand the previous sections in a more specific context, the reader should acquire some basic background on the very essential about probability, perhaps the beginning of books such as Jacod and Protter [83] or Williams [178], among many others, is a good example. This is not really necessary for what follows, but it is highly recommended.

On a probability space $(\Omega, \mathcal{F}, P)$, sometimes we may denote by $X(t, \omega)$ a stochastic process $X_t(\omega)$. Usually, equivalent classes are not used for stochastic process, but the definition of *separability* and *continuity* of a stochastic process have a natural extension in the presence of a probability measure, as *almost sure* (a.s.) properties, i.e., if the conditions are satisfied only for $\omega \in \Omega \setminus N$, where N is a null set, $P(N) = 0$. This is extremely important since we are actually working with a particular element of the equivalence class. Moreover, the concept of version is used, which is not exactly the same as equivalence class, unless some extra property (on the path) is imposed, e.g., separability or continuity. Actually, the member of the equivalence class used is ignored, but a *good* version is always needed. We are going to work mainly with d-dimensional

valued stochastic process with index sets equal to continuous times intervals e.g., a measurable and separable function $X : \Omega \times [0, +\infty] \to \mathbb{R}^d$.

It is then clear when two processes X and Y should be considered equivalent (or simply equal, $X = Y$), if

$$P(\{\omega : X_t(\omega) = Y_t(\omega),\ \forall t \in T\}) = 1.$$

This is often referred as X being *indistinguishable* from Y, or that $X = Y$ up to an *evanescent* set. So that any property valid for X is also valid for Y. When the index set is uncountable, this notion differs from the assertion X or Y is *a version* (or a modification) of the given process, where it is only required that

$$P(\{\omega : X_t(\omega) = Y_t(\omega)\}) = 1, \quad \forall t \in T, \tag{1.4}$$

which implies that both processes X and Y have the same family of finite-dimensional distributions. For instance, *sample path* properties such as (progressive) measurability and continuity depend on the version of the process in question.

Furthermore, the integrand of a stochastic integral is thought as an equivalence class with respect to a product measure in $(0, \infty) \times \Omega$ of the form $\mu = \mathrm{d}\alpha(t, \omega)P(\mathrm{d}\omega)$, where $\alpha(t, \omega)$ is an integrable nondecreasing process. In this case, two processes may belong to the same μ-equivalence class without being a version of each other. Conversely, two processes, which are versions of each other, may not belong to the same μ-equivalence class. However, any two indistinguishable processes must belong to the same μ-equivalence class.

The finite-dimensional distributions are not sufficient to determine the sample paths of a process, and so, the idea of separability is to use a countable set of time to determine the properties of a process.

Definition 1.13 (separable)**.** A d-dimensional *stochastic process* $X = \{X_t : t \in T\}$, $T \subset [0, +\infty)$ is *separable* if there exists a countable dense set of indexes $I \subset T$ (called separant) and a null set N such that for any t in T and any ω in $\Omega \setminus N$ there exists a sequence $\{t_n : n = 1, 2, \dots\}$ of elements in I which is convergent to t and such that $X(t_n, \omega)$ converges to $X(t, \omega)$. In other words, the stochastic process X can be considered either as a random variable in E^T or in the countable product E^I, without any loss. □

For instance, the reader may want to take a look at the book by Meyer [129, Chapter IV] to realize the complexity of this notion of separability.

The following result (see Doob [33, Theorem 2.4, pp. 60], Billingsley [14, Section 7.38, pp. 551-563] or Neveu [136, Proposition III.4.3, pp. 84-85]) is necessary to be able to assume that we are always working with a separable version of a process.

Theorem 1.14 (separability)**.** *Any d-dimensional stochastic process has a version which is separable i.e., if X is the given stochastic process indexed by some real interval T, then there exists a $\mathbb{R}^d$-valued stochastic process Y satisfying (1.4) and the condition of separability in Definition 1.13, which may be re-phrased as*

follows: there exist a countable dense subset I of T and a null measurable set N, $P(N) = 0$, such that for every open subset O of T and any closed subset C of $\mathbb{R}^d$ the set $\{\omega \in \Omega : Y(t,\omega) \in C,\ \forall t \in O \setminus I\}$ is a subset of N. □

By means of the above theorem, we will always assume that we have taken a (the qualifier a.s. is generally omitted) separable version of a (measurable) stochastic process provided we accept processes with values in $\bar{\mathbb{R}}^d = [-\infty, +\infty]^d$. Moreover, if we insist in calling stochastic process X a family of random variables $\{X_t\}$ indexed by t in T then we have to deal with the separability concept. Actually, the set $\{\omega : X_t(\omega) = Y_t(\omega),\ \forall t \in T\}$ used to define equivalent or indistinguishable processes may not be measurable when X or Y is not a measurable process. Even working only with measurable processes does not solve completely our analysis, e.g., a simple operation as $\sup_{t\in T} X_t$ for a family of uniformly bounded random variables $\{X_t\}$ may not yields a measurable random variable. The separability notion solves all these problems.

Furthermore, this generalizes to processes with values in a separable locally compact metric space (see Gikhman and Skorokhod [61, Section IV.2]), in particular, the above separable version Y may be chosen with values in $\mathbb{R}^d \cup \{\infty\}$, the one-point compactification of $\mathbb{R}^d$, and with $P\{Y(t) = \infty\} = 0$ for every t, but not necessarily $P\{Y(t) = \infty\ \forall t \in T\} = 0$. Thus in most cases, when we refer to a stochastic process X in a given probability space $(\Omega, \mathcal{F}, P)$, actually we are referring to a measurable and separable version Y of X. Note that in general, the initial process X is not necessarily separable or even measurable. By using the separable version of a process, we see that most of the *measurable* operations usually done with a function will make a proper sense. The construction of the separant set used (in the proof of the above theorem) may be quite complicate, e.g., see Neveu [136, Section III.4, pp. 81–88].

A stochastic process $\{X_t : t \in T\}$, $T \subset [0, +\infty)$ is *continuous* if for any $\omega \in \Omega$ the function $t \mapsto X_t(\omega)$ is continuous. On the other hand, a process X which is continuous in probability, i.e., for all $t \in T$ and $\varepsilon > 0$ we have

$$\lim_{s\to t} P(\{\omega \in \Omega : |X(s,\omega) - X(t,\omega)| \geq \varepsilon\}) = 0.$$

is called *stochastically continuous*. Similarly, we define left or right stochastically continuous. Actually, if the interval T is compact, then the process is uniformly stochastically continuous. In most of the cases, a stochastic process X will be (right or left) continuous in probability (see below) and then any dense set in T will be separant.

Most of the information of a stochastic process X is contained in the history σ-algebra, i.e., the family $\mathcal{F}_t$ or $\mathcal{F}(t)$ defined as the minimal sub σ-algebra of $\mathcal{F}$ that makes the random variables $\{X_s : s \leq t\}$ measurable. This is an increasing family of σ-algebra i.e., $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \leq t$, which is called the natural *filtration* associated with the stochastic process. Also, the processs X is called *progressively measurable* with respect to the natural filtration, i.e., the restriction of X to the set $\Omega \times [0, t]$ is measurable with respect to the product σ-algebra $\mathcal{F}_t \times \mathcal{B}([0, t])$, for any $t \geq 0$. Here, and in what follows, $\mathcal{B}(T)$ denotes the σ-algebra of Borel subsets of T, $T \subset \mathbb{R}$.

If the filtration is given a priori (independently of the stochastic process), then we will refer to as a stochastic process being *adapted* or *progressively measurable* with respect to the given filtration if the above conditions are satisfied. Moreover, we will see later that it is convenient to *normalize* the filtration to *standard* (or usual) conditions. As a caution, technical, we refers adapted as "adapted and measurable". However, note that sometimes it may be convenient to consider the notion of measurable independently of adapted, in this case, we may have a measurable process Y such that the mapping $\omega \mapsto Y(t,\omega)$ is $\mathcal{F}(t)$-measurable, but Y is not progressively measurable.

Note that the concept of stochastic continuity (or continuity in probability) is not a sample path (or pathwise) property, it does not depend on the particular version of the process involved. On the contrary, most of the smoothness properties such as separability, measurability or continuity are conditions on the sample paths and depend on the version of the process used to test the property.

It is known (e.g., see Da Prato and Zabczyk [28, p. 72–75], Gikhman and Skorokhod [61, Section IV.3]) that

Theorem 1.15 (measurability)**.** *Any (right or left) stochastically continuous d-dimensional stochastic process has a version which is measurable. Moreover, if the stochastic process is adapted then there is a version which is progressively measurable.* □

Sometimes we can take (a.s.) continuous modification of a given process on a bounded interval $[0,T]$

Theorem 1.16 (continuity)**.** *Let $\{X_t : t \in [0,T]\}$ be a d-dimensional stochastic process in a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$E|X_t - X_s|^\alpha \le C|t-s|^{1+\beta}, \quad \forall s,t \in [0,T], \tag{1.5}$$

for some positive constants α, β and C. Then there exists a continuous version $Y = \{Y_t : t \in [0,T]\}$ of X, which is locally Hölder continuous with exponent γ, for any $\gamma \in (0, \beta/\alpha)$ i.e., there exist a null set N, with $P(N) = 0$, an (a.s.) positive random variable $h(\omega)$ and a constant $K > 0$ such that for all $\omega \in \Omega \setminus N$, $s,t \in [0,T]$ we have

$$|Y_t(\omega) - Y_s(\omega)| \le K|t-s|^\gamma \quad \text{if } 0 < |t-s| < h(\omega). \ \square$$

The previous result is essentially based on the following arguments, e.g., Karatzas and Shreve [91, pp. 53–55]). Estimate (1.5) and the dyadic construction $\{X(k2^{-n}) : k = 0, 1, \dots, 2^n,\ n = 1, 2, \dots\}$ yields

$$P\{\max_{1\le k\le 2^n} |X(k2^{-n}) - X((k-1)2^{-n})| \ge 2^{-\gamma}\} \le$$

$$\le \sum_{k=1}^{2^n} P\{|X(k2^{-n}) - X((k-1)2^{-n})| \ge 2^{-\gamma}\} \le C2^{-n(\beta-\alpha\gamma)},$$

for any $\gamma > 0$ such that $\beta > \alpha\gamma$. Hence, the Borel-Cantelli lemma shows that there exists a measurable set Ω^* of probability 1 such that for any ω in Ω^* there is an index $n^*(\omega)$ with the property

$$\max_{1\le k\le 2^n} |X(k2^{-n},\omega) - X((k-1)2^{-n},\omega)| < 2^{-\gamma}, \quad \forall n \ge n^*(\omega).$$

This proves that for t of the form $k2^{-n}$ we have a uniformly continuous process which gives the desired modification. Certainly, if the process X itself is separable, then we get do not need a modification, we obtain an equivalent continuous process.

An interesting point in this result, is the fact that the condition (1.5) on the given process X can be verified by means of the so-called two-dimensional distribution of the process (see below). Moreover, the integrability of the process is irrelevant, i.e., (1.5) can be replaced by

$$\lim_{\delta\to 0} P\Big\{ \sup_{|t-s|<\delta} |X(t) - X(s)| > \varepsilon \Big\} = 0, \quad \forall \varepsilon > 0.$$

This condition is stronger that

$$\lim_{\delta\to 0} \sup_t P\Big\{ \sup_{|s|<\delta} |X(t) - X(t+s)| > \varepsilon \Big\} = 0, \quad \forall \varepsilon > 0,$$

which only yields almost surely continuity at every time t. In any case, if the process X is separable then the same X is continuous, otherwise, we construct a version Y which is continuous.

Recall that a real function on an interval $[0,T)$ (respectively $[0,\infty)$ or $[0,T]$) has only *discontinuities of the first kind* if (a) it is bounded on any compact subinterval of $[0,T)$ (respectively $[0,\infty)$ or $[0,T]$), (b) left-hand limits exist on $(0,T)$ (respectively $(0,\infty)$ or $(0,T]$) and (c) right-hand limits exist on $[0,T)$ (respectively $[0,\infty)$ or $[0,T)$). After a normalization of the function, this is actually equivalent to a right continuous functions having left-hand limits, these functions are called cad-lag.

It is interesting to note that continuity of a (separable) process X can be localized, X is called continuous (or a.s. continuous) at a time t if the set N_t of ω such that $s \mapsto X(s,\omega)$ is not continuous at $s = t$ has probability zero (i.e., N_t is measurable, which is always true if X is separable, and $P(N_t) = 0$). Thus, a (separable) process X may be continuous at any time (i.e., $P(N_t) = 0$ for every t in T) but not necessarily continuous (i.e., with continuous paths, namely $P(\cup_t N_t) = 0$). Remark that a cad-lag process X may be continuous at any (deterministic) time (i.e., $P(N_t) = 0$ for every t in T) without having continuous paths, as we will se later, a typical example is a Poisson process.

Analogously to the previous theorem, a condition for the case of a modification with only discontinuities of the first kind can be given (e.g., see Gikhman and Skorokhod [61, Section IV.4], Wong [179, Proposition 4.3, p. 59] and its references)

Theorem 1.17 (cad-lag). *Let $\{X_t : t \in [0,T]\}$ be a d-dimensional stochastic process in a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$\mathbb{E}\{|X_{t+h} - X_s|^\alpha |X_s - X_t|^\alpha\} \le Ch^{1+\beta}, \quad \forall\, 0 \le t \le s \le t+h \le T, \tag{1.6}$$

for some positive constants α, β and C. Then there exists a cad-lag version $Y = \{Y_t : t \in [0,T]\}$ of X. □

Note that sometimes, properties on the conditional distribution of $(X_{t+h} - X_s)$ given $(X_s - X_t)$ are such that the condition (1.6) is reduced to

$$\mathbb{E}\{|X_s - X_t|^\alpha\} \le C(s-t)^{\frac{1}{2}+\beta}, \quad \forall\, 0 \le t \le s \le T,$$

e.g., this is the case of a processes with independent increments. Similarly, as discussed later, for a Markov process with transition probability function $p(x,t,\mathrm{d}y)$, the expression

$$\int_{\mathbb{R}^d} |x-y|^\alpha p(x,t,\mathrm{d}y) \le Ct^{\frac{1}{2}+\beta}, \quad \forall\, 0 \le t \le T,$$

replaces (1.6), and certainly,

$$\int_{\mathbb{R}^d} |x-y|^\alpha p(x,t,\mathrm{d}y) \le Ct^{1+\beta}, \quad \forall\, 0 \le t \le T,$$

can be used instead of (1.5), to obtain path continuity.

Similarly, for processes of locally bounded variation we may replace the expression $|\cdot|$ in (1.5) by the *variation* to get a corresponding condition. In general, by looking at a process as a random variable in $\mathbb{R}^T$ we can use a complete separable metric space $D \subset \mathbb{R}^T$ to obtain results analogous to the above, i.e., if (1.5) holds for the metric $\mathrm{d}(X_t, X_s)$ instead of the Euclidean distance $|X_t - X_s|$, then the conclusions of Theorem 1.16 are valid with $\mathrm{d}(Y_t, Y_s)$ in lieu of $|Y_t - Y_s|$, e.g., see Durrett [40, p. 5, Theorem 1.6].

The statistics of a stochastic process are characterized by its *finite-dimensional* distributions, i.e., by the family of probability measures

$$P_s(B) = P(\{(X(s_1,\omega), \dots, X(s_n,\omega)) \in B\}), \quad \forall B \in \mathcal{B}(\mathbb{R}^n),$$

with $s = (s_1, \dots, s_n)$, $n = 1, 2, \dots$, for a real valued process $\{X(t,\omega) : t \in T\}$. This family of finite-dimensional distributions essentially determines a stochastic processes (i.e., modulo all possible version of a process), but not the process itself. The above results allow the verification of the (path) continuity properties of a given stochastic process in term of its two-dimensional distribution.

A typical (sample) *path* of a stochastic process is the function $X(\cdot,\omega)$ for each fixed ω, and so, a stochastic process (with prescribed finite-dimensional distributions) can always be constructed in the product space $\mathbb{R}^T$, endowed with the σ-algebra $\mathcal{B}^T(\mathbb{R})$ generated by the algebra of cylindrical sets, which may be smaller than the Borel σ-algebra $\mathcal{B}(\mathbb{R}^T)$. Thus we can view a stochastic

process X as probability measure P_X on $(\mathbb{R}^T, \mathcal{B}^T(\mathbb{R}))$, but in general the σ-algebra $\mathcal{B}^T(\mathbb{R})$ is not appropriated, it is *too small* comparatively with the *big* space $\mathbb{R}^T$ of all functions.

Note that the arguments in Theorems 1.15, 1.16 or 1.17 are such that if we begin with a separable process, then we find that the measurable, continuous or cad-lag version Y is actually indistinguishable from the initial process X, i.e., $P(\{\omega : X_t(\omega) = Y_t(\omega),\ \forall t \in T\}) = 1$.

1.3.4 Polish Spaces

As already mentioned, a Polish space (sometimes called Borel space) is a complete, separable and metric space endowed with its Borel σ-algebra. Usually, the metric that yields its (sequential) topology is complicated and really not so relevant, and its topology is characterized by other means, i.e., a Polish space is referred to as a complete, separable and metrizable space.

Pathwise properties of a stochastic process are described by sample spaces, i.e., where all paths exist almost surely. The following recasting of the concept of stochastic processes is necessary.

Definition 1.18 (process)**.** Given an index set T (usually $T \subset \mathbb{R}$), a measurable space $(E, \mathcal{E})$ (usually $E \subset \mathbb{R}^d$) and a probability space $(\Omega, \mathcal{F}, P)$, an E-valued *general stochastic process* is a measurable function X from $(\Omega, \mathcal{F})$ into $(E^T, \mathcal{E}^T)$, i.e. a family of E-valued random variables $\{X_t : t \in T\}$. Moreover, if E is a Hausdorff topological space, $\mathcal{E}$ is its Borel σ-algebra and there exits a topological sub-space B of the product space E^T (which is called *sample space* and endowed with its Borel σ-algebra $\mathcal{B}$) such that the restriction to B of the function $\omega \mapsto X(\cdot, \omega)$ (to emphasized, now denoted by $\bar{X}$) is a B-valued random variable, then $\bar{X}$ (or X) is called an E-valued *stochastic process with paths in* B. Usually B does not belong to the product σ-algebra $\mathcal{B}^T(E)$ (generated by all Borel cylindrical sets), and $\bar{X}$ (considered with values in $E^T \supset B$) is a version of the general process X. Actually $\bar{X}$ is identified with its P-equivalence class, and for each t in T, the *canonical* (coordinate, evaluation or projection) mapping $\bar{X} \mapsto \bar{X}_t$ from B into E is defined. The probability measure on B induced by $\bar{X}$ (denoted by P_X) is called the *law* of the process. Furthermore, if the index set $T = [0, \infty)$ then the minimal filtration satisfying the usual conditions (complete and right-continuous) $(\mathcal{F}_X(t) : t \geq 0)$ such that the E-valued random variables $\{\bar{X}_s : 0 \leq s \leq t\}$ are measurable is called the *canonical* filtration associated with the given process. On the other hand, given a family of finite-dimensional distributions on E^T of some (general) stochastic process X, a *realization* of a stochastic process X with paths in B and the prescribed finite-dimensional distributions is the probability space $(\Omega, \mathcal{F}, P)$ and the stochastic process $\bar{X}$ as above. □

In short, with $E = \mathbb{R}$, the above definition means that if there is a proper subset $\Omega \subset \mathbb{R}^T$ containing almost every paths of X, i.e., such that $P_X^*(\Omega) = 1$ (where P_X^* is the exterior probability measure defined for any subset of $\mathbb{R}^T$),

then the stochastic process X becomes a probability measure P on $(\Omega, \mathcal{B})$, where $\Omega \subset \mathbb{R}^T$ and

$$\mathcal{B} = \Omega \bigcap \mathcal{B}^T(\mathbb{R}) = \{\Omega \cap B : B \in \mathcal{B}^T(\mathbb{R})\}$$

is the restriction of $\mathcal{B}^T(\mathbb{R})$ to Ω with $P = P_X^*$, i.e., $P(\Omega \cap B) = P_X(B)$. It turn out that $\mathcal{B}$ contains only sets that can be described by a countable number of restrictions on $\mathbb{R}$, in particular a singleton (a one point set, which is closed for the product topology) may not be measurable. Usually, $\mathcal{B}$ is enlarged with all subsets of negligible (or null) sets with respect to P, and we can use the completion $\mathcal{B}^*$ of $\mathcal{B}$ as the measurable sets. Moreover, if Ω is an appropriate separable topological space by itself (e.g., continuous functions) so that the process have some regularity (e.g., continuous paths), then the Borel σ-algebra $\mathcal{B}(\Omega)$, generated by the open sets in Ω coincides with the previous $\mathcal{B}$. Note that another way to describe $\mathcal{B}$ is to see that $\mathcal{B}$ is the σ-algebra generated by sets (so-called cylinders in Ω) of the form $\{\omega \in \Omega : (\omega(s_1), \dots, \omega(s_n)) \in B\}$ for any $B \in \mathcal{B}(\mathbb{R}^n)$, with $s = (s_1, \dots, s_n)$, $n = 1, 2, \dots$.

At this point, the reader should be even more familiar with the topological aspect of real analysis. Perhaps some material like the beginning of the books by Billingsley [13], Pollard [145] and some points in Dudley [37] are necessary for the understanding of the next three sections.

Actually, we may look at an E-valued stochastic process $\{X(t) : t \in T\}$ as a random variable X with values in E^T endowed with the product Borel σ-algebra $\mathcal{B}^T(E)$ (generated by cylinder sets) Technically, we may talk about a random variable on a measurable space (without a given probability measure), however, the above Definition 1.18 assumes that a probability measure is given. If some information on the sample paths of the process is available (e.g., continuous paths) then the *big* space E^T and the *small* σ-algebra $\mathcal{B}^T(E)$ are adjusted to produce a suitable topological space $(\Omega, \mathcal{F})$ on which a probability measure can be defined.

When the index set T is uncountable, the σ-algebra $\mathcal{B}^T(E)$, $E \subset \mathbb{R}$ is rather small, since only a countable number of restrictions can be used to define a measurable set, so that a set of only one point $\{\omega\}$ is not measurable. This forces us to consider smaller sample spaces, where a topological structure is defined e.g., the space of continuous functions $C = C([0, \infty), E)$ from $[0, \infty)$ into E, with the uniform convergence over compact sets. The space $C([0, \infty), E)$ endowed with the natural metric

$$d_c(\omega, \omega') = \sum_{k=1}^{\infty} 2^{-k} \sup\{1 \wedge |\omega(t \wedge k) - \omega'(t \wedge k)| \quad : t \in [0, \infty)\}$$

becomes a complete separable metric space. Thus, the Borel σ-algebra on C coincides with the σ-algebra generated by the coordinate mappings

Another typical example and perhaps the most commonly used sample space is the $D = D([0, \infty), E)$ the space of right continuous functions ω from $[0, \infty)$ into E having left limits (refers to as cad-lag). Note that any function in

$D([0,\infty),E)$ is locally bounded and has at most countable many points of discontinuity. The space $D([0,\infty),E)$ can be endowed with a topology which makes it a complete separable metric space. This Skorokhod topology is given by the metric

$$d(\omega,\omega') = \inf\{p(\lambda) \vee \sum_{k=1}^{\infty} 2^{-k} q(\omega,\omega',\lambda,k) \quad : \lambda \in \Lambda\},$$

where Λ is the collection of strictly increasing functions λ mapping $[0,\infty)$ onto itself and such that

$$p(\lambda) = \sup\{|\ln(\lambda(s) - \lambda(t)) - \ln(s-t)| \quad : 0 \le t < s\}$$

is finite and

$$q(\omega,\omega',\lambda,k) = \sup\{1 \wedge |\omega(t \wedge k) - \omega'(\lambda(t) \wedge k)| \quad : t \in [0,\infty)\}.$$

We remark that the Skorokhod topology relative to $C([0,\infty),E)$ coincides with the locally uniform topology, so that C can be considered as a closed subspace of D. On the other hand, given an element ω in $D([0,\infty),E)$ and a positive number ε there exist times $0 = t_0 < t_1 < \cdots < t_n = 1/\varepsilon$ such that the oscillation of ω in each subinterval $[t_{i-1},t_i)$, $i = 1,\dots,n$ is not greater than ε, i.e., for ω_ε defined by $\omega_\varepsilon(t) = \omega(t_i)$ for any t in $[t_{i-1},t_i)$, we have $|\omega(t) - \omega_\varepsilon(t)| \le \varepsilon$. This is to say that any function in $D([0,\infty),E)$ can be approximated in the topology of $C([0,\infty),E)$ by right-continuous step functions, but it cannot be approximated in (the topology of) $D([0,\infty),E)$ by continuous functions. Clearly, the cadlag functions endowed with the locally uniform convergence (i.e., D with the topology of C) is not a separable topological space. The interested reader is referred to, e.g., Billingsley [13, Chapter 3, pp. 109–153] for a comprehensive study. Sometime it is convenient to define the sample spaces $D(]-\infty,+\infty[,E)$ and $C(]-\infty,+\infty[,E)$, and even to assume that E is only a Polish space (i.e., a complete and separable metric space). Some extra difficulties appear when E is not locally compact.

Any continuous function f with a compact support in $[0,\infty)$ (or in $]0,\infty[$, if necessary) defines a linear functional on $D([0,\infty),E)$, namely

$$\langle f,\omega\rangle = \int_0^\infty f(t)\omega(t)\mathrm{d}t,$$

which results continuous (with the Skorokhod topology). Hence, the Hausdorff topology generated by those linear functional is weaker than the Skorokhod topology and makes D a Lusin space (note that D is not a topological vector space, the addition is not necessarily a continuous operation).

Recall that if S is a metric space then $\mathcal{B}(S)$ denotes the σ-algebra of Borel subsets of S, i.e. the smallest σ-algebra on S which contains all open subsets of S. In particular $\mathcal{B}(E)$, $\mathcal{B}(D)$ and $\mathcal{B}(C)$ are the Borel σ-algebras of the metric space E, $D([0,\infty),E)$ and $C([0,\infty),E)$, respectively. Sometimes we may use

$\mathcal{B}$, when the metric space is known from the context. In particular, the Borel σ-algebra of $C = C([0,\infty), E)$ and $D = D([0,\infty), E)$ are the same as the σ-algebra generated by the coordinate functions $\{X_t(\omega) = \omega(t) : t\}$, i.e., a subset A of D belongs to $\mathcal{B}(D)$ if and only if $A \cap C$ belongs to $\mathcal{B}(C)$. Also, it is of common use the canonical right filtration (to be completed when a probability measure is given) $\bigcap_{s>t}\{\sigma\text{-algebra generated by } (X_r : r \le s)\}$. It can be proved that if $\{P_t : t \ge 0\}$ is a family of probability defined on $\mathcal{F}_t^0 = \sigma\{X_s : 0 \le s \le t\}$ such that the restriction of P_t to $\mathcal{F}_s^0$ coincides with P_s for every $s < t$, then there exists a probability P defined on $\mathcal{B}(D)$ such that P restricted to $\mathcal{F}_t^0$ agrees with P_t, e.g., see Bichteler [11, Appendix, Theorem A.7.1].

Again, the concept of continuous processes is reconsidered by means of sample spaces, i.e.,

Definition 1.19 (continuous)**.** An E-valued, usually $E \subset \mathbb{R}^d$, *continuous* stochastic process is a probability measure P on $(C([0,\infty), E), \mathcal{B})$ together with a measurable mapping (P-equivalence class) X from $C([0,\infty), E)$ into itself. If the mapping X is not mentioned, we assume that it is the *canonical* (coordinate, projection or identity) mapping $X_t(\omega) = \omega(t)$ for any ω in $C([0,\infty), E)$, and in this case, the probability measure $P = P_X$ is called the *law* of the process. Similarly, a *right continuous having left-hand limits* (cad-lag) stochastic process is a probability measure P on $(D([0,\infty), E), \mathcal{B})$ together with a measurable mapping X from $D([0,\infty), E)$ into itself. □

Note that a function X from $(C([0,\infty), E), \mathcal{B})$ into itself is measurable if and only if the functions $\omega \mapsto X(t,\omega)$ from $(C([0,\infty), E), \mathcal{B})$ into E are measurable for all t in $[0,\infty)$. Since $C([0,\infty), E) \subset D([0,\infty), E)$ as a topological space with the same relative topology, we may look at a continuous stochastic process as probability measure on D with support in the closed subspace C.

Thus, to get a continuous (or cad-lag) version of a *general* stochastic process X (see Definition 1.18) we need to show that its probability law P_X has support in $C([0,\infty), E)$ (or in $D([0,\infty), E)$). On the other hand, separability of a general stochastic process can be taken for granted (see Theorem 1.14), after a suitable modification. However, for general stochastic processes viewed as a collection of random variables defined almost surely, a minimum workable assumption is (right or left) stochastic continuity (i.e., continuous in probability). Clearly, stochastic continuity cannot be stated in term of random variable having values in some functional space, but rather as a function on $[0,\infty)$ with values in some probability space, such as $L^p(\Omega, P)$, with $p \ge 0$.

When two or more cad-lag processes are given, we may think of having several probability measures (on the suitable space), say $P_1, \ldots, P_n$, and we canonical process $X(t) = \omega(t)$. However, sometimes it may be convenience to fix a probability measure e.g., $P = P_1$, with a canonical process $X = X_1$ as a reference, and consider all the other processes $P_2, \ldots, P_n$ as either the probability measures $P_2, \ldots, P_n$ on $(D, \mathcal{B})$ or as measurable mapping $X_2, \ldots, X_n$, so that P_i is the image measure of P through the mapping X_i, for any $i = 2, \ldots, n$. On the space $(D, \mathcal{B})$ we can also define two more *canonical* processes, the *pure*

jumps process $\Delta X(t) = X(t) - X(t-)$, for $t > 0$ and the *left-limit* process

$$X(t-) = \begin{cases} X(0) & \text{if } t = 0, \\ \lim_{s \uparrow t} X(s) & \text{if } t > 0, \end{cases}$$

which may also be denoted by $X_-(t)$.

Processes X may be initially given in an abstract space $(\Omega, \mathcal{F}, P)$, but when some property on its sample path is given, such a continuity, then we may look at X as a random variable taking values in a suitable topological space (e.g. C or D). Then by taking the image measure of P through X, we may really forget about the initial space $(\Omega, \mathcal{F}, P)$ and refer everything to the sample space, usually C or D.

It is interesting to remark that $D([0,\infty), \mathbb{R}^d)$ is not a topological vector space, i.e., in the Skorokhod topology, we may have $\alpha_n \to \alpha$ and $\beta_n \to \beta$, but $\alpha_n + \beta_n$ is not converging to $\alpha + \beta$, unless α (or β) belongs to $C([0,\infty), \mathbb{R}^d)$. Moreover, the topology in $D([0,\infty), \mathbb{R}^d)$ is strictly stronger that the product topology in $D([0,\infty), \mathbb{R}^{d_1}) \times D([0,\infty), \mathbb{R}^{d_2})$, $d = d_1 + d_2$. The reader is referred to the book Jacod and Shiryaev [84, Chapter VI, pp. 288–347] for a comprehensive discussion.

1.3.5 Filtrations and Stopping Times

Typical construction in probability theory are on filtered space $(\Omega, \mathbb{F}, P)$, i.e., a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\mathbb{F} = (\mathcal{F}_t, t \geq 0)$ satisfying the usual conditions, namely, each $\mathcal{F}_t$ is a sub σ-algebra of $\mathcal{F}$, $\mathcal{F}_t \subset \mathcal{F}_s$ if $t \leq s$, $\mathcal{F}_0$ contains all null sets of $\mathcal{F}$ (completed), $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ (continuous from the right), and usually, also assuming that $\mathcal{F} = \mathcal{F}_\infty$, with $\mathcal{F}_\infty$ being the smallest σ-algebra containing all the $\mathcal{F}_t$ for $t \geq 0$. The filtration is a technical instrument constructed from (or together with) a given process, each $\mathcal{F}_t$ represents the information available at time t. Usually, the (*internal*) *history* of stochastic process $X = \{X(t), t \geq 0\}$ is defined as the filtration $\mathbb{H} = \{\mathcal{H}_t : t \geq 0\}$, with $\mathcal{H}_t$ being the σ-algebra generated by the random variables $\{X(s) : s \leq t\}$, and some more work is needed to obtain an adequate filtration satisfying the usual conditions.

In a filtered space $(\Omega, \mathbb{F}, P)$ the arrow of time is properly defined, if t is considered the present then $s < t$ is the past and $s > t$ is the future. Any $[0,\infty]$-valued random variable τ is not necessarily a good random time, only the so-called *stopping times* (or optimal times), i.e., satisfying $(\tau \leq t) = \{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ for every $t \geq 0$, preserves the past-present-future structure. Note that if $\{\tau_n : n \geq 1\}$ is a sequence of stopping times then $\sup_n\{\tau_n\}$, $\inf_n\{\tau_n\}$, $\limsup_n\{\tau_n\}$, and $\liminf_n\{\tau_n\}$ are also stopping times. For every stopping time τ, the σ-algebra $\mathcal{F}_\tau$ all sets F in $\mathcal{F}$ satisfying $F \cap (\tau \leq t)$ in $\mathcal{F}_t$ for every $t > 0$ is defined, and represents the information available at the random time τ. For instance, if τ and θ are stopping times then the sets $(\tau \leq \theta)$, $(\tau < \theta)$, and $(\tau = \theta)$ belong to $\mathcal{F}_{\tau \wedge \theta}$, and something (just for convenience) the notation $\mathcal{F}(t)$ or $\mathcal{F}(\tau)$ could be used.

A stochastic process $X = \{X(t), t \geq 0\}$ is called *adapted* to the filtration $\mathbb{F}$ if the random variable $X(t)$ is $\mathcal{F}_t$-measurable for each $t \geq 0$. Because the time is running continuously on $[0, \infty)$, the fact that an adapted process is not necessarily joint-measurable in (t, ω) cause some technical difficulties, and two σ-algebras are defined on the product space $[0, \infty) \times \Omega$, namely, (1) the *optional* σ-algebra $\mathcal{O}$ generated by the sets of the form $F_0 \times \{0\}$ and $F_s \times [s, t)$, where $F_s \in \mathcal{F}_s$ and $s < t$ in $[0, \infty)$, and (2) the *predictable* σ-algebra $\mathcal{P}$ generated by the sets of the form $F_0 \times \{0\}$ and $F_s \times (s, t]$, where $F_s \in \mathcal{F}_s$ and $s < t$ in $[0, \infty)$.

First remark that the notations $\mathcal{F}(t)$ or $\mathcal{F}_t$ is a matter of convenience, as well as $X(t)$ or X_t for the random processes. Now, let us give more details on stopping times. The problems of defining what is meant by a random time τ corresponding to the arrival time of an event whose arrival is determined by the preceding events and of defining the class $\mathcal{F}(\tau)$ of preceding events are solved by the following definition.

Definition 1.20. An *optional time* (*stopping* or *Markov time*) τ with respect to a filtration $\mathbb{F} = \{\mathcal{F}(t) : t \geq 0\}$ is a function from Ω into $[0, +\infty]$ satisfying

$$\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}(t) \quad \forall t \geq 0.$$

If an optional time τ is given, then $\mathcal{F}(\tau)$, respectively $\mathcal{F}(\tau-)$, is the σ-algebra of subsets A in $\mathcal{F}(+\infty)$ (or in $\mathcal{F}$) for which

$$A \cap \{\tau \leq t\} \in \mathcal{F}(t), \quad \text{respectively} \quad A \cap \{\tau < t\} \in \mathcal{F}(t),$$

for every $t \geq 0$. □

Sometime, optional times are defined as nonnegative random variables satisfying $\{\omega : \tau(\omega) < t\} \in \mathcal{F}(t)$ for every $t > 0$, e.g., see Karatzas and Shreve [91, Section 1.2, pp. 6-11]. Since $\{\tau \geq t\} = \cup_{n \geq 1}\{\tau > t - 1/n\}$ and $\mathcal{F}(t - 1/n) \subset \mathcal{F}(t)$, we see that stopping time is stronger than optional time. Conversely, under the right-continuity condition, i.e., $\mathcal{F}(t) = \mathcal{F}(t+)$, for every $t \geq 0$, the equality $\{\tau \leq t\} = \cap_{n \geq 1}\{\tau < t + 1/n\}$ shows that any optional time is also a stopping time. Thus, unless specially mentioned, we do not differentiate between optional and stopping times.

Most of the time we use the σ-algebra $\mathcal{F}(\tau)$, however, when dealing with jump processes we may need $\mathcal{F}(\tau-)$. Note that we have $\cap_{\varepsilon > 0}\mathcal{F}(\tau+\varepsilon) = \mathcal{F}(\tau+) = \mathcal{F}^+(\tau)$ for any optimal time τ. If τ_1 and τ_2 are two optional times with $\tau_1 \leq \tau_2$, the *stochastic interval* $[\![\tau_1, \tau_2]\!]$, is defined by

$$[\![\tau_1, \tau_2]\!] = \{(t, \omega) \in \mathbb{R}^+ \times \Omega : \tau_1 \leq t \leq \tau_2\}.$$

Similarly, we define the open stochastic interval $]\!]\tau_1, \tau_2[\![$ and the half-open ones $[\![\tau_1, \tau_2[\![$, and $]\!]\tau_1, \tau_2]\!]$. Several properties are satisfied by optional times, we will list some of them.

(a) If τ is optional, then τ is $\mathcal{F}(\tau)$-measurable.

(b) If τ is optional and if τ_1 is a random variable for which $\tau_1 \geq \tau$ and τ_1 is $\mathcal{F}(\tau)$ measurable, then τ_1 is optional.

(c) If τ_1 and τ_2 are optional, then $\tau_1 \vee \tau_2$ (max) and $\tau_1 \wedge \tau_2$ (min) are optional.

(d) If τ_1 and τ_2 are optional and $\tau_1 \leq \tau_2$, then $\mathcal{F}(\tau_1) \subset \mathcal{F}(\tau_2)$; if $\tau_1 < \tau_2$, then $\mathcal{F}(\tau_1+) \subset \mathcal{F}(\tau_2)$.

(e) If τ_1 and τ_2 are optional, then $\mathcal{F}(\tau_1) \cap \mathcal{F}(\tau_2) = \mathcal{F}(\tau_1 \wedge \tau_2)$. In particular, $\{\tau_1 \leq t\} \in \mathcal{F}(\tau_1 \wedge t)$.

(f) If τ_1 and τ_2 are optional, then the sets $\{\tau_1 < \tau_2\}$, $\{\tau_1 \leq \tau_2\}$ and $\{\tau_1 = \tau_2\}$ are in $\mathcal{F}(\tau_1 \wedge \tau_2)$.

(g) If τ_1 and τ_2 are optional and if $A \in \mathcal{F}(\tau_1)$, then $A \cap \{\tau_1 \leq \tau_2\} \in \mathcal{F}(\tau_1 \wedge \tau_2)$.

(h) Let τ_1 be optional and finite valued, and let τ_2 be random variable with values in $[0, +\infty]$. The optionality of $\tau_1 + \tau_2$ implies optionality of τ_2 relative to $\mathcal{F}(\tau_1 + \cdot)$. Moreover, the converse is true if $\mathcal{F}(\cdot)$ is right continuous i.e., if τ_2 is optional for $\mathcal{F}_{\tau_1}(\cdot) = \mathcal{F}(\tau_1 + \cdot)$, then $\tau_1 + \tau_2$ is optional for $\mathcal{F}(\cdot)$ and $\mathcal{F}(\tau_1 + \tau_2) = \mathcal{F}_{\tau_1}(\tau_2)$.

(i) Let $\{\tau_n : n = 1, 2, \dots\}$ be a sequence of optional times. Then $\sup_n \tau_n$ is optional, and $\inf \tau_n$, $\liminf_n \tau_n$, $\limsup_n \tau_n$ are optional for $\mathcal{F}^+(\cdot)$. If $\lim_n \tau_n = \tau = \inf_n \tau_n$, then $\mathcal{F}^+(\tau) = \cap_n \mathcal{F}^+(\tau_n)$. If the sequence is decreasing [resp., increasing] and $\tau_n(\omega) = \tau(\omega)$ for $n \geq n(\omega)$, then τ is optional and $\mathcal{F}(\tau) = \cap_n \mathcal{F}(\tau_n)$ [resp., $\mathcal{F}(\tau)$ is equal to the smaller σ-algebra containing $\cup_n \mathcal{F}(\tau_n)$].

There are many relations between optional times, progressively measurable stochastic processes and filtration, we only mention the following result (see Doob [34, pp. 419–423])

Theorem 1.21 (exit times)**.** *Let B be a Borel subset of $[0, T] \times \mathbb{R}^d$ and $\{X(t) : t \in [0, T]\}$ be a d-dimensional progressively measurable stochastic process with respect to a filtration $\mathbb{F}$ satisfying the usual conditions on a probability space $(\Omega, \mathcal{F}, P)$, Then the hitting, entry and exit times are optional times with respect to $\mathbb{F}$, i.e., for the hitting time*

$$\tau(\omega) = \inf\{t > 0 : (t, X(t, \omega)) \in B\},$$

where we take $\tau(\omega) = +\infty$ if the set in question is empty. Similarly, the entry time is define with $t > 0$ replaced by $t \geq 0$ and the exit time is the entry time of complement of B, with the convention of being equal to T if the set in question is empty. □

Note that the *last hitting time* of a Borel set B, which is defined by

$$\hat{\tau}(\omega) = \sup\{t > 0 : (t, X(t, \omega)) \in B\},$$

is not in general an optional time. However, if τ_c denotes the hitting time of B by the process $(t + c, X(t + c, \omega))$ then $\{\hat{\tau} > c\} = \{\tau_c < +\infty\}$ so that measurability properties for the last hitting time can be considered.

1.3.6 Random Fields

Sometimes, the index of a collection of E-valued random variables is not necessarily the time, i.e., a family $X = \{X_r : r \in R\}$ of random variables with R a topological space is called a random field with values in E and parameter R, typically, R is a subset of the Euclidean space $\mathbb{R}^n$ or $\mathbb{R}^n \times T$, where T represents the time. The sample spaces corresponding to random fields are $C(R, E)$ or other separable Fréchet spaces, and even the Polish space $D(R \times [0, \infty), E)$ i.e., continuous in R and cad-lag in $[0, \infty)$.

If $X = \{X_r : r \in R\}$ is a d-dimensional random field with parameter $R \subset \mathbb{R}^n$ then the probability distribution of X is initially on the product space E^R, and some conditions are needed to restrict Tulcea or Kolmogorov construction theorems 1.11 or 1.12 to a suitable sample space, e.g., getting continuity in the parameter. Similar to Theorem 1.16 we have

Theorem 1.22 (continuity)**.** *Let $\{X_r : r \in R\}$ be a d-dimensional random field with parameter $R \subset \mathbb{R}^n$ in a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$E|X_r - X_s|^\alpha \le C|r - s|^{n+\beta}, \quad \forall r, s \in R \subset \mathbb{R}^n, \tag{1.7}$$

for some positive constants α, β and C. Then there exists a continuous version $Y = \{Y_r : r \in R\}$ of X, which is locally Hölder continuous with exponent γ, for any $\gamma \in (0, \beta/\alpha)$ i.e., there exist a null set N, with $P(N) = 0$, an (a.s.) positive random variable $h(\omega)$ and a constant $K > 0$ such that for all $\omega \in \Omega \setminus N$, $s, t \in [0, T]$ we have

$$|Y_r(\omega) - Y_s(\omega)| \le K|r - s|^\gamma \quad \text{if } 0 < |r - s| < h(\omega). \ \square$$

There are other ways of continuity conditions similar to (1.7), e.g., instead of $|r - s|^{n+\beta}$ with $\beta > 0$ we may use

$$\sum_{i=1}^n |r_i - s_i|^{\beta_i}, \quad \text{with} \quad \sum_{i=1}^n \frac{1}{\beta_i} < 1 \tag{1.8}$$

forevery r, s in $R \subset \mathbb{R}^n$. For instance, the reader may check the books Billingsley [13, Chapter 3, pp. 109–153], Ethier and Kurtz [45, Chapter 3, pp. 95–154], Khoshnevisan [97, Chapter 11, pp. 412–454], or Kunita [104, Section 1.4, pp. 31–42] for a complete discussion.

1.4 Existence of Probabilities

The underlaying objective of this section is the construction of Lévy processes, but much more can be obtained form the construction of probability measures with a given characteristic function. Instead of changing the process, the image of the probability measure under a fixed (canonical) measurable function is studied via its characteristic function. This yields an alternative way for constructing probability measures with prescribed (or desired) properties on suitable Borel spaces.

1.4.1 Fourier Transform

First, recall the space $\mathcal{S}(\mathbb{R}^d)$ of rapidly decreasing smooth functions, i.e., functions φ having partial derivatives $\partial^\alpha\varphi$, with a multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$ of any order $|\alpha| = \alpha_1 + \cdots + \alpha_d$, such that the quantities

$$\mathtt{p}_{n,k}(\varphi) = \sup\{(1+|x|^2)^{k/2}|\partial^\alpha\varphi(x)| \,:\, x \in \mathbb{R}^d,\ |\alpha| \le n\}, \quad n,k = 0,1,\dots,$$

are all finite. Thus, the countable family of semi-norms $\{\mathtt{p}_{n,k}\}$ makes $\mathcal{S}(\mathbb{R}^d)$ a Fréchet space, i.e., metrizable locally convex and complete topological vector space.

The Fourier transform can be initially defined in various function spaces, perhaps the most natural we are $\mathcal{S}(\mathbb{R}^d)$. In its definition, the constant π can be placed conveniently, for instance, in harmonic analysis

$$(\mathfrak{F}f)(\xi) = \int_{\mathbb{R}^d} f(x)\,\mathrm{e}^{-2\pi\mathtt{i}x\cdot\xi}\,\mathrm{d}x, \quad \forall\xi \in \mathbb{R}^d,$$

where $x \cdot \xi$ is the Euclidean scalar product in $\mathbb{R}^d$, or

$$(Ff)(\xi) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(x)\,\mathrm{e}^{-\mathtt{i}x\cdot\xi}\,\mathrm{d}x,$$

is used, while

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x)\,\mathrm{e}^{\mathtt{i}x\cdot\xi}\,\mathrm{d}x, \tag{1.9}$$

is used in probability (so-called *characteristic function*), in any case, the constant π plays an important role in the inversion formula. In this section, we retain the expression (1.9), with the simplified notation either $\widehat{f}$ or $\mathfrak{F}f$. For instance, the textbook by Stein and Shakarchi [166] is an introduction to this topic.

Essentially, by completing the square, the following one-dimensional calculation

$$\begin{aligned}
&\int_{\mathbb{R}} \mathrm{e}^{-\pi\lambda x^2 - 2\pi\mathtt{i}x\cdot\xi}\,\mathrm{d}x = \mathrm{e}^{-\pi\xi^2/\lambda}\int_{\mathbb{R}} \mathrm{e}^{-\pi(x\sqrt{\lambda}+\mathtt{i}\xi/\sqrt{\lambda})^2}\,\mathrm{d}x,\\
&\partial_\xi \int_{\mathbb{R}} \mathrm{e}^{-\pi(x\sqrt{\lambda}+\mathtt{i}\xi/\sqrt{\lambda})^2}\,\mathrm{d}x = (\mathtt{i}/\lambda)\int_{\mathbb{R}} \partial_x \mathrm{e}^{-\pi(x\sqrt{\lambda}+\mathtt{i}\xi/\sqrt{\lambda})^2}\,\mathrm{d}x = 0,\\
&\int_{\mathbb{R}} \mathrm{e}^{-\pi\lambda x^2/2}\,\mathrm{d}x = (1/\sqrt{\lambda})\int_{\mathbb{R}} \mathrm{e}^{-\pi x^2/2}\,\mathrm{d}x = 1/\sqrt{\lambda},
\end{aligned}$$

shows that

$$\int_{\mathbb{R}} \mathrm{e}^{-\pi\lambda x^2 - 2\pi\mathtt{i}x\cdot\xi}\,\mathrm{d}x = (1/\sqrt{\lambda})\,\mathrm{e}^{-\pi\xi^2/\lambda}.$$

Using the product form the exponential (and a rotation in the integration variable), this yields

$$\int_{\mathbb{R}^d} \mathrm{e}^{-x\cdot ax+2\pi \mathtt{i} x\cdot\xi}\,\mathrm{d}x = \frac{\pi^{d/2}}{\sqrt{\det(a)}}\,\mathrm{e}^{-\xi\cdot a^{-1}\xi}, \quad \forall \xi\in\mathbb{R}^d, \tag{1.10}$$

for any (complex) symmetric matrix $a=(a_{ij})$ whose real part is positive definite, i.e., $\Re\{x\cdot ax\}>0$, for every x in $\mathbb{R}^d$. Therefore, in particular,

$$\mathfrak{F}(\mathrm{e}^{-\pi|x|^2})(\xi) = \mathrm{e}^{-\pi|\xi|^2}, \quad \forall \xi\in\mathbb{R}^d,$$

i.e., the function $x\mapsto \mathrm{e}^{-\pi|x|^2/2}$ is a fixed point for the Fourier transform. Moreover, this space $\mathcal{S}(\mathbb{R}^d)$ and its dual $\mathcal{S}'(\mathbb{R}^d)$ (the space of tempered distributions) are invariant under the Fourier transform.

For instance, an introduction at the beginning of the graduate level can be found in the book Pinsky [144], among others. It can be proved that the Fourier transform $\mathfrak{F}$ defined by (1.9) is a continuous linear bijective application from $\mathcal{S}(\mathbb{R}^d)$ onto itself. The expression

$$(\mathfrak{F}^{-1}\varphi)(x) = \int_{\mathbb{R}^d} \varphi(\xi)\,\mathrm{e}^{2\pi \mathtt{i} x\cdot\xi/(2\pi)}\,\mathrm{d}\xi, \quad \forall x\in\mathbb{R}^d.$$

defines the inverse of $\mathfrak{F}$, which is also continuous. It is clear that s change of variable yields the inverse for the expression (1.9). Certainly, there are many important properties of the Fourier transform that we do not mention. For instance, the interested reader may check the books by Duoandikoetxea [39] or Grafakos [65, 66] for a comprehensive study on Fourier analysis.

1.4.2 Bochner Type Theorems

At this point, the reader may revise some of the basic subjects treated in the book Malliavin [115]. In particular, a revision on measure theory, e.g., as in Kallenberg [88, Chapters 1 and 2, pp. 1–44], may be necessary.

To construct a probability measure from the characteristic function of a stochastic process (instead of a random variable) we need an infinite dimensional version of Bochner Theorem 1.5.

Theorem 1.23 (Bochner-Minlos). *Let Ψ be a complex-valued functional on the space of test functions $\mathcal{S}(\mathbb{R})$. If Ψ satisfies $\Psi(0)=1$, and Ψ is continuous and positive definite (i.e., $\varphi_n\to 0$ in $\mathcal{S}(\mathbb{R})$ implies $\Psi(\varphi_n)\to\Psi(0)=1$ in $\mathbb{C}$, and $\sum_{i,j=1}^k \Psi(\varphi_i-\varphi_j)z_i\bar{z}_j\geq 0$, for every φ_i in $\mathcal{S}(\mathbb{R})$, any $k\geq 1$, and any complex number z_i, $i=1,\ldots,k$, then there exists a (unique) probability measure P on the space of tempered distributions $\mathcal{S}'(\mathbb{R})$ having Ψ as its characteristic function, i.e.,*

$$\Psi(\varphi) = \int_{\mathcal{S}'(\mathbb{R})} \exp\big(\mathtt{i}\langle\omega,\varphi\rangle\big)\,P(\mathrm{d}\omega) = \mathbb{E}\big\{\exp\big(\mathtt{i}\langle\cdot,\varphi\rangle\big)\big\},$$

where $\langle\cdot,\cdot\rangle$ denote the paring between $\mathcal{S}'(\mathbb{R})$ and $\mathcal{S}(\mathbb{R})$, i.e., the $L^2(\mathbb{R})$ inner product.

Proof. In other words, any positive definite complex-valued tempered distribution is the Fourier transform of some finite measure on $\mathcal{S}'(\mathbb{R})$. In what follows, only a brief idea is given, for instance, the interested read may take a look at Holden et al. [74, Appendix A, pp. 193–197]) and the reference therein.

Indeed, several steps are necessary

Step 1 Begin with the (Schwartz) space of rapidly decreasing and smooth functions $\mathcal{S}(\mathbb{R})$ and its dual space of tempered distributions $\mathcal{S}'(\mathbb{R})$. These spaces are identified with the space of sequences s and its dual s', via Hermite functions, i.e., given a sequence in s we form a function in $\mathcal{S}(\mathbb{R})$ by using the terms as coefficients in the expansion along the orthonormal basis $\{\xi_n(x) : n \geq 1\}$ in $L^2(\mathbb{R})$, with

$$\xi_{n+1}(x) = \frac{\mathrm{e}^{-x^2/2}}{\pi^{1/4}\sqrt{n!}}\, p_n(\sqrt{2}x), \quad n = 1, 2, \dots,$$

where p_n is the Hermite polynomial of order n). Thus

$$s = \big\{a = \{a_k\}_{k=0}^{\infty} \,:\, \lim_k k^m a_k = 0,\ \forall m = 1, 2, \dots\big\}$$

is the Fréchet space of rapidly decreasing sequences.

Step 2 This Fréchet space of rapidly decreasing sequences is decomposed as $s = \bigcap_{m=0}^{\infty} s_m$ with s_m defined for every integer m as the space of all sequences $a = \{a_k\}_{k=0}^{\infty}$ satisfying

$$\|a\|_m = \Big[\sum_{k=0}^{\infty}(1 + k^2)^m |a_k|^2\Big]^{1/2} < \infty,$$

which is a Hilbert space.

(b) Its dual space is decomposed as $s' = \bigcup_{m=0}^{\infty} s'_m$, with $s'_m = s_{-m}$ and the natural paring between elements in s' and s (also between s'_m and s_m), namely,

$$\langle a', a\rangle = \sum_{k=0}^{\infty} a'_k a_k, \quad \forall a' \in s',\ a \in s.$$

Note that s' is the space of sequences $\{a_k\}$ with polynomial growth, i.e., $k^m a_k \to 0$ as $k \to \infty$, for same $m > 0$.

Step 3 Based on Bochner's result for finite dimensional spaces and Kolmogorov's extension, a probability measure with a prescribed characteristic function can be constructed in the space $\mathbb{R}^{\infty}$, the space of all sequences of real numbers. It takes some more effort to check that the probability measure is concentrated on the dual space s'. Indeed, use the continuity and the condition $\Psi(0) = 1$ to deduce that for any $\varepsilon > 0$ there exist $m > 0$ and $\delta > 0$ such that $\|a\|_m < \delta$ implies $|\Psi(a) - 1| < \varepsilon$. This yields

$$\int_{\mathbb{R}^{\infty}} \cos(\langle a', a\rangle)\, P(\mathrm{d}a') \geq 1 - \varepsilon - 2\delta^{-2}\|a\|_m^2, \quad \forall a \in s. \tag{1.11}$$

as desired.

Step 4 Now, for every sequence $b = \{b_k\}$, with $b_k > 0$ consider the (Gaussian) probability measure $\mu_{n,\sigma}$ on $\mathbb{R}^{n+1}$ defined by

$$\mu_{n,\sigma} = \prod_{k=0}^{n} (2\pi\sigma b_k)^{-1/2} \exp\big[-\frac{a_k^2}{2\sigma b_k}\big] \mathrm{d}a_k,$$

for any $\sigma > 0$. Recall that

$$\int_{\mathbb{R}^{n+1}} \cos(\langle a', a\rangle)\, \mu_{n,\sigma}(\mathrm{d}a) = \exp\big[-\frac{\sigma}{2} \sum_{k=0}^{n} b_k (a'_k)^2\big],$$

$$\int_{\mathbb{R}^{n+1}} \|a\|_m^2\, \mu_{n,\sigma}(\mathrm{d}a) = \sigma \sum_{k=0}^{n} (1+k^2)^m b_k,$$

and integrate (1.11) with respect to $\mu_{n,\sigma}$ on $\mathbb{R}^{n+1}$ to get

$$\int_{\mathbb{R}^\infty} \exp\big[-\frac{\sigma}{2} \sum_{k=0}^{n} b_k (a'_k)^2\big]\, P(\mathrm{d}a') \geq 1 - \varepsilon - 2\delta^{-2}\sigma \sum_{k=0}^{n} (1+k^2)^m b_k.$$

Hence, if $b_k = (1+k^2)^{-m-1}$ then $\sum_{k=0}^{n} (1+k^2)^m b_k = C < \infty$, which imply, by means of the monotone convergence,

$$\int_{\mathbb{R}^\infty} \exp\big[-\frac{\sigma}{2} \|a'\|_{-m-1}^2\big]\, P(\mathrm{d}a') \geq 1 - \varepsilon - 2\delta^{-2}\sigma C.$$

Finally, let σ vanish to get $P(s'_{m+1}) \geq 1 - \varepsilon$, which proves that $P(s') = 1$. □

At this point, we apply this previous result to various particular cases, which gives the existence of the so-called Lévy and Gaussian noises.

1.4.3 Lévy and Gaussian Noises

Certainly, the previous version of Bochner-Minlos' Theorem 1.23 extends to multi-dimensional case, i.e., $\mathcal{S}(\mathbb{R}^d)$ and vector-valued functions $\mathcal{S}(\mathbb{R}^d; \mathbb{R}^n)$. Thus, we can state the following very useful result regarding the construction of a Lévy martingale measures:

Theorem 1.24 (Lévy noise)**.** *Let $\mathcal{S}'(\mathbb{R}; \mathbb{R}^d)$ be the space of tempered distributions in $\mathbb{R}$ with values in $\mathbb{R}^d$. Suppose that σ is a (real-valued) square $d \times d$ matrix and that π is a Radon measure in $\mathbb{R}^d$ satisfying*

$$\int_{\mathbb{R}^d} \big(|y|^2 \wedge |y|\big)\pi(\mathrm{d}y) < \infty, \qquad \pi(\{0\}) = 0. \tag{1.12}$$

Then, there exists a unique probability measure P on $(\Omega, \mathcal{B})$, with $\Omega = \mathcal{S}'(\mathbb{R}; \mathbb{R}^d)$ and $\mathcal{B} = \mathcal{B}(\Omega)$ such that

$$\mathbb{E}\big\{ \exp\big[\mathtt{i}\langle \cdot, \varphi\rangle\big]\big\} = \exp\Big(-\frac{1}{2}\int_{\mathbb{R}} |\sigma\varphi(t)|^2 \mathrm{d}t\Big) \times \\ \times \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{\mathbb{R}^d} \big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)} - 1 - \mathtt{i}(\varphi(t), y)\big]\pi(\mathrm{d}y)\Big), \quad (1.13)$$

where $\mathbb{E}\{\cdot\}$ denotes the expectation with respect to P and $|\cdot|$ and $(\cdot,\cdot)$ are the Euclidean norm and scalar product, respectively. In particular, $\mathbb{E}\big\{\langle \cdot, \varphi\rangle\big\} = 0$, and if also

$$\int_{\mathbb{R}^m} |y|^2 \, \pi(\mathrm{d}y) < \infty, \quad (1.14)$$

then

$$\mathbb{E}\big\{\big|\langle \cdot, \varphi\rangle\big|^2\big\} = \int_{\mathbb{R}} \big|\sigma\varphi(t)\big|^2 \mathrm{d}t + \int_{\mathbb{R}} \mathrm{d}t \int_{\mathbb{R}^d} \big|(\varphi(t), y)\big|^2 \pi(\mathrm{d}y), \quad (1.15)$$

for any test function φ. □

Actually, if condition (1.12) is replaced by

$$\pi(\mathbb{R}^d) < \infty, \qquad \pi(\{0\}) = 0,$$

i.e., a finite measure on $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$, then

$$\mathbb{E}\big\{ \exp\big[\mathtt{i}\langle \cdot, \varphi\rangle\big]\big\} = \exp\Big(-\frac{1}{2}\int_{\mathbb{R}} |\sigma\varphi(t)|^2 \mathrm{d}t + \int_{\mathbb{R}} \mathrm{d}t \int_{\mathbb{R}^d} \big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)} - 1\big]\pi(\mathrm{d}y)\Big)$$

replaces (1.13). Thus, for $\sigma = 0$ this is a compound Poisson noise (which first moment is not necessarily finite) while, for $\pi = 0$ this is a Wiener (or white) noise (which has finite moments of all order).

Note that by replacing φ with $\lambda\varphi$, taking derivatives with respect to λ and setting $\lambda = 0$ we deduce the isometry condition (1.15), which yields an analogous equality for the scalar product $\mathbb{E}\big\{\langle \cdot, \varphi\rangle \, \langle \cdot, \psi\rangle\big\}$, with φ and ψ in $\mathcal{S}(\mathbb{R}; \mathbb{R}^d)$.

It should be clear that, from the calculation point of view, the Fourier transform for h in $\mathcal{S}(\mathbb{R}^d)$

$$\hat{h}(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} h(x) \mathrm{e}^{-\mathtt{i}(x,\xi)} \mathrm{d}x,$$

and its inverse

$$h(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \hat{h}(\xi) \mathrm{e}^{\mathtt{i}(x,\xi)} \mathrm{d}\xi,$$

are useful to estimate

$$\mathbb{E}\{\{h(\langle\cdot,\varphi_1\rangle,\dots,\langle\cdot,\varphi_d\rangle)\}\} = \\ = (2\pi)^{-d/2}\int_{\mathbb{R}^d}\hat{h}(\xi)\Psi(\xi_1\varphi_1+\ldots+\xi_d\varphi_d)\mathrm{d}\xi, \quad (1.16)$$

where Ψ is the characteristic function, i.e., the right-hand-side in (1.13). In particular, if for some $q>1$,

$$\int_{\mathbb{R}^m}|y|^{2p}\,\pi(\mathrm{d}y)<\infty, \quad \text{for any } p \text{ such that } 1\le p\le q, \quad (1.17)$$

then there exists a constant $c_q>0$ depending only on q and the dimension d such that

$$c_q\mathbb{E}\{|\langle\cdot,\varphi\rangle|^{2p}\}\le\Big(\int_{\mathbb{R}}|\sigma\varphi(t)|^2\mathrm{d}t\Big)^p+\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}|(\varphi(t),y)|^2\pi(\mathrm{d}y)\Big)^p+ \\ +\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}|(\varphi(t),y)|^{2p}\pi(\mathrm{d}y)\Big)^p, \quad (1.18)$$

i.e., the $2p$-moment is finite for any $p\le q$. Clearly, the assumption (1.17) imposed restrictions only the measure π for $|y|\ge 1$, and the expectation $\mathbb{E}$ could be written as $\mathbb{E}_{\sigma,\pi}$ to indicate the dependency on the data σ and π.

Also, from the finite-dimensional case, we know that the functions

$$\exp\big(-|x|^2/2\big),\quad \exp\big(\mathrm{e}^{\mathtt{i}(x\cdot b)}-1\big),\quad \exp\big(-\mathtt{i}(x\cdot b)\big),$$

for b fixed, are characteristic functions of the Gaussian, the Poisson and the Dirac distributions. Therefore, any matrix $a=(a_{ij})$ of the form

$$a_{ij}=\exp\Big\{-|\zeta_i-\zeta_j|^2/2+\mathrm{e}^{\mathtt{i}(\varsigma_i-\varsigma_j)}-1\Big\}$$

is a positive definite matrix. Thus, by approximating the integrals (by partial sums) in right-hand-side (called Ψ) of (1.13), we show that Ψ is indeed positive define.

Hence, we have constructed a d-dimensional *smoothed* (1-parameter) Lévy noise associated with (σ,π). Indeed, the canonical action-projection process, which is the natural paring

$$X(\varphi)=X(\varphi,\omega)=\langle\omega,\varphi\rangle,\quad \forall\varphi\in\mathcal{S}(\mathbb{R};\mathbb{R}^d),$$

can be regarded as a family of $\mathbb{R}$-valued random variables $X(\varphi)$ on the probability space $(\Omega,\mathcal{B}(\Omega),P)$, with $\Omega=\mathcal{S}'(\mathbb{R};\mathbb{R}^d)$ and P as above. Clearly, this is viewed as a generalized process and the actual Lévy noise is defined by $\dot{X}(\varphi)=-\langle\omega,\dot{\varphi}\rangle$.

Considering the space $L^2(P)$ and the vector-valued space $L^2_{\sigma,\pi}(\mathbb{R};\mathbb{R}^d)$ with the inner product defined by

$$\langle\varphi,\psi\rangle_{\sigma,\pi}=\int_{\mathbb{R}}\big(\sigma\varphi(t),\sigma\psi(t)\big)\mathrm{d}t+\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}(\varphi(t),y)\,(\psi(t),y)\,\pi(\mathrm{d}y),$$

we can view $\varphi \mapsto X(\varphi,\cdot)$ as an isometry from $L^2_{\sigma,\pi}(\mathbb{R};\mathbb{R}^d)$ into $L^2(P)$, initially defined on the test space $S(\mathbb{R};\mathbb{R}^d)$ and uniquely extended everywhere. Thus, the expression $\langle\omega,\varphi\rangle$ makes sense almost surely (passing to the limit) for φ in $L^2_{\sigma,\pi}(\mathbb{R};\mathbb{R}^d)$. Now, for a given test function φ we denote by $\varphi_{i,t}$ the test function with only one non-zero component, namely, the i-component which is given by the expression $\mathbb{1}_{(0,t]}$, i.e., $\varphi_{i,t} = (0,\dots,\mathbb{1}_{(0,t]},\dots,0)$. Thus, a d-dimensional Lévy (martingale) process $\ell_i(t) = X(\varphi_{i,t})$ for $i = 1,2,\dots,d$ (with diffusion matrix $\sigma^*\sigma/2$ and Levy measure π) is almost sure defined. Indeed, because the scalar product is preserved, the stochastic process ℓ has orthogonal increments. Moreover, the linearity in φ and the *product* (or integral and exponential) form of the characteristic function (1.13) show that the random variable $\langle\cdot,\varphi\rangle$ is independent of $\langle\cdot,\psi\rangle$ as long as φ and ψ have disjoint support. Thus, the stochastic process $(\ell(t) : t \ge 0)$ is stationary with independent increments. The existence of a cad-lag version follows from the estimate

$$\mathbb{E}\big\{|\ell_i(s+r)-\ell_i(t)|^2|\ell_i(t)-\ell_i(s)|^2\big\} = \\ = \mathbb{E}\big\{\big(\ell_i(s+r-t)\big)^2\big\}\,\mathbb{E}\big\{\big(\ell_i(t-s)\big)^2\big\} \le Cr^2,$$

for any i, $0 \le s \le t \le s+r \le T$, any $T > 0$ and some positive constant $C = C_T$.

On the other hand, we can impose less restrictive assumptions on the Radon measure π, i.e., to separate the small jumps from the large jumps so that only assumption

$$\int_{\mathbb{R}^d} \big(|y|^2 \wedge 1\big)\pi(\mathrm{d}y) < \infty, \qquad \pi(\{0\}) = 0. \tag{1.19}$$

is needed. For instance, the Cauchy process in $\mathbb{R}^d$, where $\sigma = 0$ and the Radon measure π has the form

$$\int_{\mathbb{R}^d} \varphi(y)\pi(\mathrm{d}y) = \lim_{\varepsilon\to 0}\int_{|y|\ge\varepsilon} \varphi(y)|y|^{-d-1}\mathrm{d}y,$$

π does not integrate the function $\varphi(y) = |y|$, and

$$\exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}\big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)} - 1 - \mathtt{i}(\varphi(t),y)\mathbb{1}_{|y|\le 1}\big]|y|^{-d-1}\mathrm{d}y\Big) = \\ = \exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d} 2\big[\cos(\varphi(t),y) - 1\big]|y|^{-d-1}\mathrm{d}y\Big),$$

replaces the second exponential in (1.13). Sometimes, we require a stronger (at the origin) integrability assumption on the Radon measure π, namely,

$$\int_{\mathbb{R}^m} \big(|y| \wedge 1\big)\pi(\mathrm{d}y) < \infty, \qquad \pi(\{0\}) = 0.$$

and the second exponential in (1.13) takes the form

$$\exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}\big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)} - 1\big]\pi(\mathrm{d}y)\Big),$$

for instance, the case of the Γ-process in $\mathbb{R}^d$, $d = 1$ with parameters $c, \alpha > 0$, where $\sigma = 0$ and the measure π is given by

$$\int_{\mathbb{R}} \varphi(y)\pi(\mathrm{d}y) = \lim_{\varepsilon \to 0} c \int_{\varepsilon}^{\infty} \varphi(y) y^{-1} \mathrm{e}^{-\alpha y} \mathrm{d}y,$$

π does not have a finite mass, and

$$\exp\Big(c \int_{\mathbb{R}} \mathrm{d}t \int_0^{\infty} \big[\mathrm{e}^{\mathbf{i}\varphi(t)y} - 1\big] y^{-1} \mathrm{e}^{-\alpha y} \mathrm{d}y\Big)$$

replaces the second exponential in (1.13).

The Lévy-Itô decomposition of sample functions (e.g., see Sato [157, Chapter 4, 119–144]) shows that the Lévy (martingale) process ℓ can be written as a continuous part (its Wiener process) and a purely discontinuous part (its Poisson jumps part). Alternatively, we can split the $\mathbb{R}^d$ space into $\mathbb{R}^n \times \mathbb{R}^m$, namely, $\omega = (\omega_{n\curvearrowleft}, \omega_{\curvearrowright m})$ where $\omega_{n\curvearrowleft}$ and $\omega_{\curvearrowright m}$ are tempered distributions in $\mathbb{R}$ with values in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Thus if $\varphi(t) = (\varphi_{n\curvearrowleft}(t), \varphi_{\curvearrowright m}(t))$, where $\varphi_{n\curvearrowleft}(t)$ and $\varphi_{\curvearrowright m}(t)$ denote test functions in $\mathbb{R}$ with values in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, then $\langle \omega, \varphi \rangle = \langle \omega_{n\curvearrowleft}, \varphi_{n\curvearrowleft} \rangle + \langle \omega_{\curvearrowright m}, \varphi_{\curvearrowright m} \rangle$. Hence, we have a $(n+m)$-dimensional *smoothed* (1-parameter) Wiener-Poisson (Lévy) noise, i.e.,

$$X_{n\curvearrowleft}(\varphi, \omega) = \langle \omega_{n\curvearrowleft}, \varphi_{n\curvearrowleft} \rangle, \qquad X_{\curvearrowright m}(\varphi, \omega) = \langle \omega_{\curvearrowright m}, \varphi_{\curvearrowright m} \rangle,$$

the projection on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Clearly, $X_{n\curvearrowleft}$ provides a Wiener process independent of the Poisson martingale measure obtained from $X_{\curvearrowright m}$.

Therefore, by considering the vector-valued space $L^2_{\sigma,\pi}(\mathbb{R}; \mathbb{R}^{n+m})$ where we have separate the first n components from the last m components, we can construct (almost surely defined) a n-dimensional Wiener process $w_i(t) = X(\varphi_{i,t})$ for $i = 1, 2, \ldots, n$ (with diffusion matrix $\sigma^*\sigma/2$) and a m-dimensional Poisson martingale measure $q_i(t) = X(\varphi_{i,t})$ for $i = n+1, n+2, \ldots, n+m$ (with Levy measure π, so that its jumps Δq_i form a Poisson point process). Indeed, the stochastic process

$$X_t = x + \big(w_1(t), \ldots, w_n(t), q_1(t), \ldots, q_m(t)\big), \quad \forall\, t \geq 0,\ x \in \mathbb{R}^{n+m} \tag{1.20}$$

(also denoted by X_t^x) has orthogonal increments, which implies that $(X_t : t \geq 0)$ is stationary with independent increments, i.e., a Lévy process in law. To take a cad-lag version (which results continuous in the first n components) under assumption (1.14), we may use the estimates

$$\mathbb{E}\big\{|w_i(t) - w_i(s)|^4\big\} = \mathbb{E}\big\{\big(w_i(t-s)\big)^4\big\} \leq C|t-s|^2,$$

$$\mathbb{E}\big\{|q_j(s+r) - q_j(t)|^2 |q_j(t) - q_j(s)|^2\big\} =$$

$$= \mathbb{E}\big\{\big(q_j(s+r-t)\big)^2\big\}\, \mathbb{E}\big\{\big(q_j(t-s)\big)^2\big\} \leq Cr^2,$$

for any i, j, $0 \leq s \leq t \leq s+r \leq T$, any $T > 0$ and some positive constant $C = C_T$. However, (for the Poisson point process) if only condition (1.19) holds then we can obtain suitable estimates using the equality (1.16). We have then described a way of constructing these processes.

1.4.4 Countably Hilbertian Spaces

Actually, the only properties used in Lévy's Theorem 1.24 is the fact that the complex-valued characteristic function Ψ is continuous (at zero suffices), positive definite and $\Psi(0) = 1$. Indeed, this generalizes to separable Hilbert spaces, e.g., see the book Da Prato and Zabczyk [28, Theorem 2.13, pp. 49–52], by adding an extra condition on Ψ. Recall that on a separable Hilbert space H, a mapping $S : H \to H$ is called a nuclear (or trace class) operator if for any (or some) orthonormal basis $\{e_i : i \geq 1\}$ in H the series $\sum_i |(Se_i, e_i)|$ is convergent. On the other hand, $\sigma : H \to H$ is called a Hilbert-Schmidt operator if for any (or some) orthonormal basis $\{e_i : i \geq 1\}$ in H the series $\sum_i (\sigma e_i, \sigma e_i)$ is finite.

Theorem 1.25 (Sazonov). *A complex-valued function Ψ on a separable Hilbert space H is the characteristic function of a probability measure P on $(H, \mathcal{B}(H))$ if and only if* (a) *Ψ is continuous,* (b) *is positive definite,* (c) *$\Psi(0) = 1$ and satisfies the following condition:*

(d) *for every $\varepsilon > 0$ there exists a nonnegative nuclear (or trace class) operator S_ε such that each h in H with $(S_\varepsilon h, h) \leq 1$ yields $1 - \Re\{\Psi(h)\} \leq \varepsilon$.* □

Let $\sigma_i : H_0 \to H_0$ $(i = 1, 2)$ be two (symmetric) Hilbert-Schmidt operators on a separable Hilbert space H_0 with inner product $(\cdot, \cdot)_0$ and norm $|\cdot|_0$. Now, on the Hilbert space $H = L^2(\mathbb{R}, H_0^2)$, $H_0^2 = H_0 \times H_0$, consider the *characteristic* function

$$\Psi(h_1, h_2) = \exp\Big(-\frac{1}{2} \int_{\mathbb{R}} |\sigma_1 h_1(t)|_0^2 \mathrm{d}t \Big) \times$$
$$\times \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_0} \big[\mathrm{e}^{\mathtt{i}(\sigma_2 h_2(t), \sigma_2 u)_0} - 1 - \mathtt{i}(\sigma_2 h_2(t), \sigma_2 u)_0 \big] \pi(\mathrm{d}u) \Big), \quad (1.21)$$

where π is a measure on $\mathcal{B}(H_0)$ such that

$$\int_{H_0} \big(|\sigma_2 u|_0^2 \wedge |\sigma_2 u|_0 \big) \pi(\mathrm{d}u) < \infty, \qquad \pi(\{0\}) = 0. \quad (1.22)$$

Under these assumptions the function Ψ is continuous on H, positive definite, $\Psi(0) = 1$ and the condition (d) of Theorem 1.25 is satisfied for a given $\varepsilon > 0$ with a trace class operator $S_\varepsilon : H \to H$ of the form

$$S_\varepsilon((b_k, b_\ell)e_j) = \begin{cases} (\sigma_1^* \sigma_1 b_k, \sigma_2^* \sigma_2 b_\ell) e_j & \text{if } j \leq n, \\ 0 & \text{otherwise,} \end{cases}$$

for any $k, \ell = 1, \ldots,$ and for some $n = n(\varepsilon)$, where $\{e_j : j \geq 1\}$ is an orthonormal basis in Lebesgue space $L^2(\mathbb{R})$ and σ_i^* is the adjoint of σ_i, $i = 1, 2$, while $\{b_k : k \geq 1\}$ and $\{(b_k, b_\ell) : k, \ell \geq 1\}$ are orthonormal basis in the spaces H_0 and H_0^2, this means that

$$\big(S_\varepsilon h, (b_k, b_\ell) e_j \big)_H = \int_{\mathbb{R}} \big[(\sigma_1 h_1(s), \sigma_1 b_k)_0 + (\sigma_2 h_2(s), \sigma_2 b_\ell)_0 \big] e_j(s) \mathrm{d}s,$$

for every $h = (h_1, h_2)$, with h_i in H_0, for any $k, \ell = 1, \dots,$ and $j = 1, \dots, n$ (otherwise, the left-hand term vanishes), where $(\cdot, \cdot)_H$ denotes the inner product in H.

Therefore Ψ is the characteristic function of a probability measure P on the Hilbert space H, i.e.,

$$\mathbb{E}\{e^{i(h,\cdot)_H}\} = \Psi(h_1, h_2), \quad \forall h = (h_1, h_2) \in H,$$

where $(\cdot, \cdot)_H$ denotes the inner product in $H = L^2(\mathbb{R}, H_0^2)$. Hence a cad-lag version of a Lévy process on $\mathbb{R}$ or $[0, \infty)$ with parameters $(\sigma_1, \sigma_2, \pi)$ and values in H_0 is obtained as previously discussed in $\mathbb{R}^n \times \mathbb{R}^m$. Thus, the Lévy measure $\pi(\sigma_2^* \sigma_2)^{-1}$ is defined on the Hilbert space image $H_2 = \sigma_2^* \sigma_2(H_0)$ and the probability P can be considered on canonical sample space $\Omega = D([0, \infty), H_1 \times H_2)$ or $\Omega = D([0, \infty), H_1) \times D([0, \infty), H_2)$, with $H_1 = \sigma_1^* \sigma_1(H_0)$, where the canonical process $X(\omega) = \omega(t)$ has Ψ as its characteristic function. Clearly, a drift can be added and the parameters $(\sigma_1, \sigma_2, \pi)$ can be time-dependent with suitable assumptions.

The above arguments extend to the case of a countably Hilbertian space (of which a typical example is the space $\mathcal{S}(\mathbb{R}^d)$ of rapidly decreasing smooth functions with its dual $\mathcal{S}'(\mathbb{R}^d)$ of tempered distributions), where the role the Hilbert-Schmidt operators σ_i is better understood.

A *countably Hilbertian space* K is a separable Fréchet (i.e., complete locally convex topological) space where the topology is given by an increasing sequence $\{\| \cdot \|_n : n \geq 0\}$ of compatible (i.e., any Cauchy sequence in two norms and convergent to zero in one norm results convergent to zero also in the other norm) Hilbertian norms. Moreover, a space K is called *nuclear* if for any $n \geq 0$ there exists $m > n$ such that the canonical injection from K_m into K_n is Hilbert-Schmidt, where K_n denote the completion of K with the Hilbertian norm $\| \cdot \|_n$. Thus K_n is a sequence of decreasing Hilbert spaces and $K = \cap_n K_n$. Next, if we identify K_0 with its dual space K_0' (by Riezs' representation theorem) and we denote the dual space K_n' by K_{-n} (with its dual Hilbertian norm $\| \cdot \|_{-n}$, $n \geq 1$) then K_{-n} is a sequence of increasing Hilbert spaces, the dual space K' is sequentially complete and $K' = \cup_n K_{-n}$.

Theorem 1.26 (Minlos)**.** *A complex-valued function Ψ on a countably Hilbertian nuclear space K is the characteristic function of a probability measure P on the dual space $(K', \mathcal{B}(K'))$ if and only if Ψ is continuous at 0 in K, positive definite and $\Psi(0) = 1$.* □

Note that if K is a countably Hilbertian nuclear space then so is $\mathcal{S}(\mathbb{R}^d, K)$ (for instance, regarding $\mathcal{S}(\mathbb{R}^d, K)$ as the tensor product $\mathcal{S}(\mathbb{R}^d, K) = \mathcal{S}(\mathbb{R}^d) \otimes K$) and $K = \mathcal{S}(\mathbb{R}^d; \mathbb{R}^m)$ with $K' = \mathcal{S}'(\mathbb{R}^d; \mathbb{R}^m)$ is a typical example. Also $C([0, \infty), X)$ is a Fréchet space if X is so. However, $D([0, \infty), X)$ is a Polish (not a topological vector) space X is so. If $(\cdot, \cdot)$ is continuous inner product in a countably Hilbertian nuclear space K (i.e., the inner product is continuous in K_n for some n) and H is the Hilbert space completion of K with respect to $(\cdot, \cdot)$ then H is called *rigged Hilbert space* in K, and we have the triplet $K \subset H \subset K'$. Certainly, any K_n can be used as H, but this is not necessary in general.

On the other hand, a set A in $D([0,\infty),K')$ (resp. $C([0,\infty),K')$) is relatively compact if and only if one of the following conditions is satisfied:

(1) For any k in K the set $\{\langle\omega(\cdot),k\rangle : \omega \in A\}$ is relatively compact in $D([0,\infty),\mathbb{R})$ (resp. $C([0,\infty),\mathbb{R})$).

(2) For every $T > 0$ there exists n such that A_T the restriction of A to $D([0,T],\mathbb{R})$ (resp. $C([0,T],\mathbb{R})$) is relatively compact in $D([0,T],K_{-n})$ (resp. $C([0,T],K_{-n})$).

Clearly, any k in K defines a measurable map π_k from $D([0,\infty),K')$ (resp. $C([0,\infty),K')$) into $D([0,\infty),\mathbb{R})$ (resp. $C([0,\infty),\mathbb{R})$), $\pi_k(t,\omega) = \langle\omega,k\rangle$. Then a sequence $\{\mu_i : i \geq 1\}$ is tight in $D([0,\infty),K')$ (resp. $C([0,\infty),K')$) if and only if for every k in K the sequence $\{\mu_i\pi_k^{-1} : i \geq 1\}$ is tight as a Borel probability measure in $D([0,\infty),\mathbb{R})$ (resp. $C([0,\infty),\mathbb{R})$). Moreover, if for every $T > 0$ there is n with the property that for every $\varepsilon > 0$ there exists $M > 0$ such that

$$\mu_i\big(\{\omega \in D([0,T],K') : \sup_{0\leq t\leq T} |\omega(t)|_{-n} \leq M\}\big) \geq 1-\varepsilon,$$

for every $i \geq 1$, then the sequence $\{\mu_i : i \geq 1\}$ regarded as Borel probability measure in $D([0,T],K_{-m})$ is tight, with $m \geq n$ such that the canonical injection from K_m into K_n (and so from K_{-n} into K_{-m}) is Hilbert-Schmidt.

Hence if $K \subset H_i \subset K'$, $i = 1,2$ are two rigged Hilbert spaces then there is a probability measure P on $\mathcal{S}'(\mathbb{R}^n; H_1 \times H_2)$ with characteristic function

$$\mathbb{E}\big\{\exp\big(\mathtt{i}[(\varphi_1,\cdot)_1 + (\varphi_1,\cdot)_2]\big)\big\} = \exp\Big(-\frac{1}{2}\int_{\mathbb{R}^n} |\varphi_1(t)|_1^2 \mathrm{d}t\Big)\times$$
$$\times \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_2} \big[\mathrm{e}^{\mathtt{i}(\varphi_2(t),u)_2} - 1 - \mathtt{i}(\varphi_2(t),u)_2\big]\pi(\mathrm{d}u)\Big), \quad (1.23)$$

where π is a Radon measure on H_2 satisfying

$$\int_{H_2} \big(|u|_2^2 \wedge |u|_2\big)\pi(\mathrm{d}u) < \infty, \qquad \pi(\{0\}) = 0, \quad (1.24)$$

and $(\cdot,\cdot)_i$, $|\cdot|_i$ denote the inner product and the norm in H_i, $i = 1,2$. By comparison with (1.21) and (1.22) we see that the nuclear (or trace class) operators σ_1, σ_2 are really part of the Hilbert space where the Lévy process takes values. Moreover, the parameter t may be in $\mathbb{R}^d$ and a Lévy noise is realized as a generalized process.

For instance, the reader is referred to the book by Kallianpur and Xiong [90, Chapters 1 and 2, pp, 1–83] for details on most of the preceding definitions.

1.5 Discrete Martingales

It may be worthwhile to recall that *independence* is stable under weak convergence, i.e., if a sequence $(\xi_1,\xi_2,\ldots)$ of $\mathbb{R}^d$-valued random variables converges

weakly (i.e., $\mathbb{E}\{f(\xi_n)\} \to \mathbb{E}\{f(\xi)\}$ for any bounded continuous function) to a random variable ξ then the coordinates of ξ are independent if the coordinates of ξ_n are so. On the other hand, for any sequence $(\mathcal{F}_1, \mathcal{F}_2, \ldots)$ of σ-algebras the *tail* or terminal σ-algebra is defined as $\mathcal{F}_{tail} = \cap_n \vee_{k \geq n} \mathcal{F}_k$, where $\vee_{k \geq n} \mathcal{F}_k$ is the smaller σ-algebra containing all σ-algebras $\{\mathcal{F}_k : k \geq n\}$. An important fact related to the independence property is the so-called Kolmogorov's zero-one law, which states that any tail set (that is measurable with respect to a tail σ-algebra) has probability 0 or 1.

Another typical application of Borel-Cantelli lemma is to deduce almost surely convergence from convergence in probability, i.e., if a sequence $\{x_n\}$ converges in probability to x (i.e., $P\{|x_n - x| \geq \varepsilon\} \to 0$ for every $\varepsilon > 0$) with a stronger rate, namely, the series $\sum_n P\{|x_n - x| \geq \varepsilon\} < \infty$, then $x_n \to x$ almost surely.

1.5.1 Main Properties

A key tool to study sequences of integrable random variables is the martingale concept.

Definition 1.27 (discrete martingale)**.** A stochastic sequence $(X_n : n = 0, 1, \ldots)$ is called a *martingale* relative to a filtration $(\mathcal{F}_n : n = 0, 1, \ldots)$ if

$$\mathbb{E}\{|X_n|\} < \infty, \ \forall n, \quad \text{and} \quad \mathbb{E}\{X_n \,|\, \mathcal{F}_{n-1}\} = X_{n-1}, \quad \text{a.s.}, \ n \geq 1.$$

A *super or sub martingale* is defined similarly, replacing the equal sign $=$ by the $\leq$ or $\geq$ signs, respectively. □

Note that X_n turns out to be $\mathcal{F}_n$-measurable and it is determined almost surely, actually we take X_n as a $\mathcal{F}_n$-measurable function defined everywhere. If only the complete probability space $(\Omega, \mathcal{F}, P)$ is given, then the filtration $(\mathcal{F}_n : n = 0, 1, \ldots)$ is naturally generated by the stochastic sequence $(X_n : n = 0, 1, \ldots)$, i.e., $\mathcal{F}_n$ is the smallest sub σ-algebra of $\mathcal{F}$ containing all null sets and rendering measurable the random variables $\{X_0, X_1, \ldots, X_n\}$. A super-martingale *decreases on average* while a sub-martingale *increases on average*. Since X_0 is integrable, we may focus our attention on sequences with $X_0 = 0$. A typical example of martingale is a real valued random walk or $\mathbb{R}^d$-valued random walk since (super-/sub-) martingales can be defined by coordinates when dealing with $\mathbb{R}^d$-valued random variables. Also, if φ is a convex and increasing real-valued function such that $\mathbb{E}\{\varphi(X_n)\} < \infty$ for some sub-martingale $(X_n : n = 0, 1, \ldots)$ then the stochastic sequence $(\varphi(X_n) : n = 0, 1, \ldots)$ is also a sub-martingale.

In most cases, the filtration $\mathcal{F}_n$ is generated by another sequence of random variables $\{Y_0, Y_1, \ldots\}$, i.e., $\mathcal{F}_n = \sigma[Y_0, \ldots, Y_n]$, which is regarded as the history. In this case, $X_n = h_n(Y_0, \ldots, Y_n)$ for some Borel function $h_n \colon \mathbb{R}^{n+1} \to \mathbb{R}$, e.g., see Karr [94].

Many important results are found in the study of martingales, related to estimates and representation, we will mention only some of them. For Doob's

upcrossing estimate, denote by $U_N(X,[a,b])$ the number of up-crossings of $[a,b]$ by time N for a fixed ω, i.e., the largest k such that $0 \le s_1 < t_1 < \cdots < s_k < t_k \le N$, $X_{s_i} < a$ and $X_{t_i} > b$, for any $i = 1, 2, \dots k$. Then for any super-martingale the estimate

$$(b-a)\,\mathbb{E}\{U_N(X,[a,b])\} \le \mathbb{E}\{(X_N - a)^-\} \tag{1.25}$$

holds. Note that the number of steps does not appear directly on the right-hand side, only the final variable X_N is relevant. To show this key estimate, by induction, we define $C_1 = \mathbb{1}_{X_0<a}$, i.e., $C_1 = 1$ if $X_0 < a$ and $C_1 = 0$ otherwise, and for $n \ge 2$,

$$C_n = \mathbb{1}_{C_{n-1}=1}\,\mathbb{1}_{X_{n-1}\le b} + \mathbb{1}_{C_{n-1}=0}\,\mathbb{1}_{X_{n-1}<a}$$

to construct a bounded nonnegative super-martingale $Y_n = \sum_{k=1}^n C_k(X_k - X_{k-1})$. Clearly, the sequence $(C_n : n = 1, 2, \dots)$ is predictable. Based on the inequality

$$Y_N \ge (b-a)\,U_N(X,[a,b]) - [X_N - a]^-,$$

for each ω, the estimate (1.25) follows.

The Doob's super-martingale convergence states that for a super martingale $(X_n : n = 0, 1, \dots)$ bounded in L^1, i.e., $\sup_n |X_n| < \infty$ the limits $X_\infty = \lim_n X_n$ exists almost surely. The convergence is in L^1 if and only if the sequence $(X_n : n = 0, 1, \dots)$ is uniformly integrable, and in this case we have $\mathbb{E}\{X_\infty \,|\, \mathcal{F}_n\} \le X_n$, almost surely, with the equality for a martingale. To prove this convergence, we express the set Ω_0 of all ω such that the limit $\lim_n X_n(\omega)$ does not exist in the extended real number $[-\infty, +\infty]$ as a countable union of subsets $\Omega_{a,b}$ where $\liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)$, for any rational numbers $a < b$. By means of the upcrossing estimate (1.25) we deduce

$$\Omega_{a,b} \subseteq \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \{\omega : U_n(X,[a,b]) > m\},$$
$$P\Big(\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \{\omega : U_n(X,[a,b]) > m\}\Big) = 0,$$

which yields $P(\Omega_0) = 0$. Thus the limit exists in $[-\infty, +\infty]$ and by Fatou's Lemma, it is finite almost surely.

If $p > 1$ and $(X_n : n = 0, 1, \dots)$ is a nonnegative sub-martingale bounded in L^p then Doob's L^p inequality reads as follows

$$\|\sup_n X_n\|_p \le p' \sup_n \|X_n\|_p, \qquad \text{with } 1/p + 1/p' = 1, \tag{1.26}$$

where $\|\cdot\|_p$ denotes the in $L^p = L^p(\Omega, \mathcal{F}, P)$. Note that $(p')^p \le 4$ for every $p \ge 2$. Indeed, if the set Ω_n^r of all ω where $\sup_{k\le n} X_k \ge r$ is expressed as the

disjoint union

$$\Omega_n^r = \bigcup_{k=0}^{n} \Omega_{n,0}^r \quad \text{with}$$
$$\Omega_{n,k}^r = \{X_0 < r\} \cap \{X_1 < r\} \cap \cdots \{X_{k-1} < r\} \cap \{X_k \geq r\},$$

and $\Omega_{n,0}^r = \{X_0 \geq r\}$, then we have $X_k \geq r$ on $\Omega_{n,k}^r$, which yields the Doob's maximal inequality

$$r\, P\big(\sup_n X_n \geq r\big) \leq \mathbb{E}\{X_n \mathbb{1}_{\sup_n X_n \geq r}\} \leq \mathbb{E}\{X_n\}.$$

Now, to deduce Doob sup-estimate (1.26) for sub-martingales bounded in L^p, with $p > 1$, first check the claim that for any two nonnegative random variables x and y,

$$\text{if} \quad r\, P(y \geq r) \leq \mathbb{E}\{x \mathbb{1}_{y \geq r}\} \quad \text{then} \quad \mathbb{E}\{y^p\} \leq (p')^p \mathbb{E}\{x^p\}. \tag{1.27}$$

by using Hölder inequality in the last equality of

$$\mathbb{E}\{y^p\} = p \int_0^\infty r^{p-1} P(y \geq r) \mathrm{d}r \leq p \int_0^\infty r^{p-2} \mathbb{E}\{x \mathbb{1}_{y \geq r}\} \mathrm{d}r =$$
$$= \frac{p}{p-1} \mathbb{E}\{x y^{p-1}\} = p' \big(\mathbb{E}\{x^p\}\big)^{1/p} \big(\mathbb{E}\{y^p\}\big)^{1/p'},$$

and replace y with $y \wedge k$ with $k \to \infty$ if necessary, to obtain (1.27). Next, choose $y = \sup_n X_n$ and $x = X_n$ to conclude.

1.5.2 Doob's decomposition

The Doob's decomposition gives a clean insight into martingale properties. Let $(X_n : n = 0, 1, \ldots)$ be a stochastic sequence of random variables in L^1, and denote by $(\mathcal{F}_n : n = 0, 1, \ldots)$ its natural filtration, i.e., $\mathcal{F}_n = \sigma[X_0, X_1, \ldots, X_n]$. Then there exists a martingale $(M_n : n = 0, 1, \ldots)$ relative to $(\mathcal{F}_n : n = 0, 1, \ldots)$ and a predictable sequence $(A_n : n = 0, 1, \ldots)$ with respect to $(\mathcal{F}_n : n = 0, 1, \ldots)$ such that

$$X_n = X_0 + M_n + A_n, \ \forall n, \quad \text{and} \quad M_0 = A_0 = 0. \tag{1.28}$$

This decomposition is unique almost surely and the stochastic sequence $(X_n : n = 0, 1, \ldots)$ is a sub-martingale if and only if the stochastic sequence $(A_n : n = 0, 1, \ldots)$ is monotone increasing, i.e., $A_{n-1} \leq A_n$ almost surely for any n. Indeed, define the stochastic sequences $(A_n : n = 1, \ldots)$ by

$$A_n = \sum_{k=1}^{n} \mathbb{E}\{X_k - X_{k-1} \,|\, \mathcal{F}_{k-1}\}, \quad \text{with} \quad \mathcal{F}_k = \sigma[X_0, X_1, \ldots, X_k]$$

and $(M_n : n = 1, \ldots)$ with $M_n = X_n - X_0 - A_n$ to obtain the decomposition (1.28). This implies that the only *deterministic* martingale is a constant.

Given a martingale $M = (M_n : n = 0, 1, \dots)$ with each M_n in L^2 and $M_0 = 0$, we may use the above decomposition to express the sub-martingale $M^2 = (M_n^2 : n = 0, 1, \dots)$ as $M^2 = N + A$, where $N = (N_n : n = 0, 1, \dots)$ is a martingale and $A = (A_n : n = 0, 1, \dots)$ is a predictable increasing sequence, both N and A null at $n = 0$. The stochastic sequence A is written as $\langle M \rangle$ and called the *angle-brackets* sequence of M. Note that

$$\mathbb{E}\{M_n^2 - M_{n-1}^2 \,|\, \mathcal{F}_{n-1}\} = \mathbb{E}\{(M_n - M_{n-1})^2 \,|\, \mathcal{F}_{n-1}\} = A_n - A_{n-1},$$

for every $n \geq 1$. Similarly, define the stochastic sequence (of quadratic variation)

$$[M]_n = \sum_{k=1}^{n} (M_k - M_{k-1})^2, \quad \forall n \geq 1,$$

and $[M]_0 = 0$. Then the stochastic sequence $V = (V_n : n = 1, 2, \dots)$,

$$V_n = M_n^2 - [M]_n = \sum_{k=1}^{n} 2M_{k-1}\, M_k$$

is a martingale. Note that $[M]$ is an adapted sequence while $\langle M \rangle$ is predictable, so the strength of the Doob's decomposition. It is clear that

$$\mathbb{E}\{|M_n|^2\} = \mathbb{E}\{\langle M \rangle_n\} = \mathbb{E}\{[M]_n\}, \quad \forall n \geq 1,$$

which combined with the p-estimate (1.26), $p = 2$, yields

$$\mathbb{E}\{\sup_{k \leq n} |M_k|^2\} \leq 4 \sup_{k \leq n} \mathbb{E}\{\langle M \rangle_k\}, \quad \forall n \geq 1.$$

Actually, this generalize into the following Davis-Burkhölder-Gundy inequality

$$c_p\, \mathbb{E}\{([M]_n)^{p/2}\} \leq \mathbb{E}\{\sup_{k \leq n} |M_k|^p\} \leq C_p\, \mathbb{E}\{([M]_n)^{p/2}\}, \tag{1.29}$$

valid for any $n \geq 1$ and $p > 0$ and some constants $C_p > c_p > 0$ independent of the martingale $(M_n : n = 0, 1, \dots)$. Even for $p = 1$, we may use $C_1 = 3$ in the right-hand side of (1.29). Moreover, the L^2-martingale $(M_n : n = 0, 1, \dots)$ may be only a *local martingale* (i.e., there exists a sequence of stopping times $\eta = (\eta_k : k = 0, 1, \dots)$ such that $M^{\eta,k} = (M_n^{\eta,k} : n = 0, 1, \dots)$, defined by $M_n^{\eta,k}(\omega) = M_{n \wedge \eta_k(\omega)}(\omega)$, is a martingale for any $k \geq 0$ and $\eta_k \to \infty$ almost surely), the time n may be replaced by a stopping time η (or ∞), the angle-brackets $\langle M \rangle$ can be used in lieu of $[M]$, and the above inequality holds true. All these facts play an important role in the continuous time case.

Let $X = (X_n : n = 0, 1, \dots)$ be a sub-martingale with respect to $(\mathcal{F}_n : n = 0, 1, \dots)$ and uniformly integrable, i.e., for every ε there exists a sufficiently large $r > 0$ such that $P(|X_n| \geq r) \leq \varepsilon$ for any $n \geq 0$. Denote by $A = (A_n : n = 0, 1, \dots)$ and $M = (M_n : n = 0, 1, \dots)$ the predictable and martingale sequences given in the decomposition (1.28), $X_n = X_0 + M_n + A_n$, for

all $n \geq 0$. Since X is a sub-martingale, the predictable sequence A is monotone increasing. The Doob's optional sampling theorem implies that the martingale M is uniformly integrable, moreover $A_\infty = \lim_n A_n$ is integrable and the families of random variable $\{X_\eta : \eta \text{ is a stopping}\}$ and $\{M_\eta : \eta \text{ is a stopping}\}$ are uniformly integrable. Furthermore, for any two stopping times $\eta \leq \theta$ we have

$$\mathbb{E}\{M_\theta \,|\, \mathcal{F}_\eta\} = M_\eta, \quad \text{a.s.} \qquad \text{and} \qquad \mathbb{E}\{X_\theta \,|\, \mathcal{F}_\eta\} \geq X_\eta, \quad \text{a.s.} \tag{1.30}$$

We skip the proof (easily found in the references below) of this fundamental results. Key elements are the convergence and integrability of the limit $M_\infty = \lim_n M_n$ (almost surely defined), which allow to represent M_n as $\mathbb{E}\{M_\infty \,|\, \mathcal{F}_n\}$. Thus, specific properties of the conditional expectation yield the result.

For instance, the reader is referred to the books Bremaud [19], Chung [23], Dellacherie and Meyer [32, Chapters I–IV], Doob [33, 35], Karlin and Taylor [92, 93], Nelson [135], Neveu [137], Rosenthal [154], Williams [178], among others.

1.5.3 Markov Chains

Martingales is a purely stochastic concept, in the sense that a deterministic martingale is necessarily a constant processes, and therefore of little or none interest. Preserving the expectation (as in the martingale condition) is good tool, but not an evolution-type property. In a deterministic setting, a differential or a difference equation is an excellent model to describe evolution, and this is view in a probabilistic setting as a Markov model, where the evolution is imposed on the probability of the process. The simplest case are the so-called Markov chains.

Let $\{X(t) : t \in T\}$, $T \subset \mathbb{R}$ be an E-valued stochastic process, i.e. a (complete) probability measure P on $(E^T, \mathcal{B}^T(E))$. If the cardinality of the state space E is finite, we say that the stochastic process takes finitely many values, labeled $1, \ldots, n$. This means that the probability law P on $(E^T, \mathcal{B}^T(E))$ is concentrated in n points. Even in this situation, when the index set T is uncountable, the σ-algebra $\mathcal{B}^T(E)$ is rather small, a set of a single point is not measurable). A typical path takes the form of a function $t \mapsto X(t, \omega)$ and cannot be a continuous function in t. As discussed later, it turn out that cadlag functions are a good choice. The characteristics of the stochastic processes $\{X(t) : t \in T\}$ are the functions $t \mapsto x_i(t) = P\{X(t) = i\}$, for any $i = 1, \ldots, n$, with the property $\sum_{i=1}^n x_i = 1$. We are interested in the case where the index set T is usually an interval of $\mathbb{R}$.

Now, we turn our attention where the stochastic process describes some evolution process, e.g., a dynamical system. If we assume that the dimension of X is sufficiently large to include all relevant information and that the index t represents the time, then the knowledge of $X(t)$, referred to as the state of the system at time t, should summarize all information up to the present time t. This translated mathematically to

$$P\{X(t) = j \mid X(r),\, r \leq s\} = P\{X(t) = j \mid X(s)\}, \tag{1.31}$$

almost surely, for every $t > s$, $j = 1, \dots, n$. At this point, the reader may consult the classic book Doob [33, Section VI.1, pp. 235–255] for more details. Thus, the evolution of the system is characterized by the transition function $p_{ij}(s,t) = P\{X(t) = j \mid X(s) = i\}$, i.e., a transition from the state j at time s to the state i at a later time t. Since the stochastic process is assumed to be cad-lag, it seems natural to suppose that the functions $p_{ij}(s,t)$ satisfies for every $i, j = 1, \dots, n$ conditions

$$\begin{aligned} &\sum_{j=1}^{n} p_{ij}(s,t) = 1, \ \forall t > s, \\ &\lim_{(t-s)\to 0} p_{ij}(s,t) = \delta_{ij}, \ \forall t > s, \\ &p_{ij}(s,t) = \sum_{k=1}^{n} p_{ik}(s,r) p_{kj}(r,t), \ \forall t > r > s. \end{aligned} \tag{1.32}$$

The first condition expresses the fact that $X(t)$ takes values in $\{1, \dots, n\}$, the second condition is a *natural regularity* requirement, and the last conditions are known as the Chapman-Kolmogorov identities. Moreover, if $p_{ij}(s,t)$ is smooth in s, t so that we can differentiate either in s or in t the last condition, and then let $r - s$ or $t - r$ approaches 0 we deduce a system of ordinary differential equations, either the *backward* equation

$$\begin{aligned} &\partial_s p_{ij}(s,t) = \sum_{k=1}^{n} \rho_{ik}^{+}(s)\, p_{kj}(s,t), \quad \forall t > s, \ i, j, \\ &\rho_{ij}^{+}(s) = \lim_{r\to s} \partial_s p_{ij}(s,r) \quad \forall s, \ i, j, \end{aligned} \tag{1.33}$$

or the *forward* equation

$$\begin{aligned} &\partial_t p_{ij}(s,t) = \sum_{k=1}^{n} p_{ik}(s,t)\, \rho_{kj}^{-}(t), \quad \forall t > s, \ i, j, \\ &\rho_{ij}^{-}(t) = \lim_{r\to t} \partial_t p_{ij}(r,t) \quad \forall t, \ i, j, \end{aligned} \tag{1.34}$$

The quantities $\rho_{ij}^{+}(s)$ and $\rho_{ij}^{-}(s)$ are the characteristic of the process, referred to as *infinitesimal rate*. The initial condition of (1.32) suggests that $\rho_{ij}^{-}(s) = -\rho_{ij}^{+}(t) = \rho_{ij}(t)$, if $s = t$. Since $\sum_{j=1}^{n} p_{ij}(s,t) = 1$ we deduce

$$\rho(t,i,j) \geq 0, \quad \forall i \neq j, \qquad \rho(t,i,i) = -\sum_{j\neq i} \rho(t,i,j). \tag{1.35}$$

Using matrix notation, $R(\cdot) = \{\rho_{ij}\}$, $P(s,t) = \{p_{ij}(s,t)\}$ we have

$$\begin{aligned} &\partial_s P(s,t) = -R(s)P(s,t), \quad \forall s < t, \\ &\partial_t P(s,t) = P(s,t)R(t), \quad \forall t > s, \\ &\lim_{t-s\to 0} P(s,t) = I, \quad \forall t > s. \end{aligned} \tag{1.36}$$

Conversely, given the integrable functions $\rho_{ij}(t)$, $i,j = 1,\dots,n$, $t \geq 0$ satisfying (1.35), we may solve the system of (non-homogeneous and linear) ordinary differential equations (1.33), (1.34) or (1.36) to obtain the transition (matrix) function $P(s,t) = \{p_{ij}(s,t)\}$ as the fundamental solution (or Green function). For instance, the reader may consult the books by Chung [23], Yin and Zhang [182, Chapters 2 and 3, pp. 15–50].

Since $P(s,t)$ is continuous in $t > s \geq 0$ and satisfies the conditions in (1.32), if we give an initial distribution, we can find a cad-lag realization of the corresponding Markov chain, i.e., a stochastic process $\{X(t) : t \geq 0\}$ with cad-lag paths such that $P\{X(t) = j \mid X(s) = i\} = p_{ij}(s,t)$, for any $i,j = 1,\dots,n$ and $t \geq 0$. In particular, if the rates $\rho_{ij}(t)$ are independent of t, i.e., $R = \{\rho_{ij}\}$, then the transition matrix $P(s,t) = \exp[(t-s)R]$. In this case, a realization of the Markov chain can be obtained directly from the rate matrix $R = \{\rho_{ij}\}$ as follows. First, let Y_n, $n = 0,1,\dots$ be a sequence of E-valued random variables with $E = \{1,\dots,n\}$ and satisfying $P(Y_n = j \mid Y_{n-1} = i) = \rho_{ij}/\lambda$, if $i \neq j$ with $\lambda = -\inf_i \rho_{ii}$, $i > 0$, and Y_0 initially given. Next, let $\tau_1, \tau_2, \dots$ be a sequence of independent identically distributed exponentially random variables with parameter λ i.e., $P(\tau_i > t) = \exp(-\lambda t)$, which is independent of $(Y_0, Y_1, \dots)$. If we define $X(t) = Y_n$ for t in the stochastic interval $[\![T_n, T_{n+1}[\![$, where $T_0 = 0$ and $T_n = \tau_1 + \tau_2 + \cdots + \tau_n$, then $X(t)$ gives a realization of the Markov chain with the above infinitesimal rates.

A typical setting includes $T = [0,\infty)$ and a denumerable state space E (with the discrete topology). This type of processes are very useful in modeling dynamical systems, referred to either as *queueing systems* or as *point processes* very well known in the literature, e.g., Bremaud [20], Kemeny and Snell [95], Kleinrock [98], Nelson [135].

This study is simplified if the time is discrete, i.e., the Markov chain X_n, $n = 0,1,\dots$, with values in subset E of $\mathbb{R}^d$ is defined by recurrence by

$$P\{X_{n+1} \in \cdot \mid X_n\} = P(X_n, \cdot), \quad n \geq 1,$$

where the *stochastic kernel* $P(x,A)$ satisfies:

(a) $P(x,\cdot)$ is a probability measure on $\mathcal{B}(E)$ for any x in E

(b) $P(\cdot,A)$ is measurable in E for any A in $\mathcal{B}(E)$.

The finite-dimensional distributions of a Markov chain is given by

$$P\{X_0 \in A_0,\, X_1 \in A_1, \dots\, X_n \in A_n\} = \\ = \int_{A_0} \nu(\mathrm{d}x_0) \int_{A_1} P(x_0, \mathrm{d}x_1) \cdots \int_{A_n} P(x_{n-1}, \mathrm{d}x_n), \quad (1.37)$$

for any $A_0, A_1, \dots, A_n$ in $\mathcal{B}(E)$, and where $\nu(\cdot)$ is the initial distribution. Thus, given the stochastic kernel we can use Kolmogorov's construction theorem (see Theorem 1.12 below) to obtain a Markov chain X_n for $n = 0,1,\dots$ satisfying the above equation (1.37). Moreover, we have

$$P\{X_n \mid X_0 = x\} = P^n(x,\cdot)$$

where $P^n(x, A)$ denote the n kernel convolutions, defined by induction by

$$P^n(x, A) = \int_E P^{n-1}(x, \mathrm{d}y) P(y, A). \tag{1.38}$$

The reader may consult the book by Chung [23] and Shields [159], among others, for a more precise discussion.

Before going further, let us mention a couple of classic simple processes which can be viewed as Markov chains with denumerable states, e.g., see Feller [48, Vol I, Sections XVII.2–5, pp. 400–411]. All processes below $\{X(t) : t \geq 0\}$ take values in $\mathbb{N} = \{0, 1, \ldots\}$, with an homogeneous transition given by $p(j, t-s, n) = P\{X(t) = j \mid X(r),\, 0 \leq r < s,\, X(s) = n\}$, for every $t > s \geq 0$ and j, n in $\mathbb{N}$. Thus, these processes are completely determined by the knowledge of the characteristics $p(t, n) = P\{X(t) = n\}$, for every $t \geq 0$ and n in $\mathbb{N}$, and a description on the change of values.

The first example is the Poisson process where there are only changes from n to $n+1$ (at a random time) with a fix rate $\lambda > 0$, i.e.,

$$\begin{aligned} \partial_t p(t, n) &= -\lambda \big[p(t, n) - p(t, n-1)\big], \\ \partial_t p(t, 0) &= -\lambda p(t, 0), \end{aligned} \tag{1.39}$$

for every $t \geq 0$ and n in $\mathbb{N}$. Solving this system we obtain

$$p(t, n) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad t \geq 0,\ n \in \mathbb{N},$$

which is the Poisson distribution.

The second example is a *pure birth* process where the only variation relative to the Poisson process is the fact that the rate is variable, i.e., the rate is $\lambda_n > 0$ when $X(t) = n$. This means that (1.39) becomes

$$\begin{aligned} \partial_t p(t, n) &= -\lambda_n p(t, n) + \lambda_{n-1} p(t, n-1), \\ \partial_t p(t, 0) &= -\lambda p(t, 0), \end{aligned} \tag{1.40}$$

for every $t \geq 0$ and n in $\mathbb{N}$. Certainly, this system can be solved explicitly, but the expression is rather complicate in general. If X represents the size of a population then the quantity λ_n is called the average rate of growth. An interesting point is the fact that $\{p(t, n) : n \in \mathbb{N}\}$ is indeed a probability distribution, i.e.,

$$\sum_{n=1}^{\infty} p(t, n) = 1$$

if and only if the coefficients λ_n increase sufficiently fast, i.e., if and only if the series $\sum_n \lambda_n^{-1}$ diverges.

The last example is the *birth-and-death* process, where the variation is the fact that either a change from n to $n+1$ (birth) with a rate λ_n or from n to

$n-1$, if $n \geq 1$ (death) with a rate μ_n may occur. Again, the system (1.39) is modifies as follows

$$\begin{aligned} \partial_t p(t,n) &= -(\lambda_n + \mu_n)p(t,n) + \lambda_{n-1}p(t,n-1) + \mu_{n+1}p(t,n+1), \\ \partial_t p(t,0) &= -\lambda p(t,0) + \mu_1 p(t,1), \end{aligned} \tag{1.41}$$

for every $t \geq 0$ and n in $\mathbb{N}$. As in the case of a pure birth process, some conditions are needed on the rates $\{\lambda_0, \lambda_1, \ldots\}$ and $\{\mu_1, \mu_2, \ldots\}$ to ensure that the birth-and-death process does not get trap in 0 or ∞ and the characteristics $\{p(t,n) : n \in \mathbb{N}\}$ is a probability distribution.

The reader may be interested in taking a look at the books Bensoussan [5], Bertsekas [9], Bremaud [20], Hernández-Lerma and Lasserre [69, 70, 71], Peskir and Shiryaev [143] (among many other books), to check some control problems for Markov chains.

Chapter 2

Stochastic Processes

For someone familiar with elementary probability theory this may be the beginning of the reading. Indeed, this chapter reinforces (or describes in more detail) some difficulties that appear in probability theory when dealing with general processes. Certainly, the whole chapter can be viewed as a detour (or a scenery view) of the main objective of this book. However, all this may help to retain a better (or larger) picture of the subject under consideration. Certainly, there are many books dealing with stochastic processes (at various level of difficulties) that the reader may consult, e.g., Bass [3], Bobrowski [16], Borodin and Salminen [18], Doob [33], Dynkin [41], Freedman [54], Friz and Victoir [56], Gihman and Skorohod [60], Itô [77], Karlin and Taylor [92, 93], Lamperti [108], Métivier [127], Pollard [145], Rao [150], Wentzell [177], Wong and Hajek [180], among many others.

First, rewind the scenario probability theory and more details on stochastic processes are given in Section 1 (where filtered probability spaces are discussed) and Section 2 (where Lévy processes are superficially considered). Secondly, a very light treatment of martingales in continuous time is given in Section 3; and preparing for stochastic modelling, Gaussian and Poisson noises are presented in Sections 4 and 5. Next, in Section 6, another analysis on Gaussian and compensated Poisson processes is developed. Finally, integer random measures on Euclidean spaces is property discussed.

2.1 Calculus and Probability

As mentioned early, a tern $(\Omega, \mathcal{F}, P)$ means an abstract probability space, i.e., $\mathcal{F}$ is a σ-algebra of subset in Ω and P is a σ-additive function defined on $\mathcal{F}$ such that $P(\emptyset) = 0$ and $P(\Omega) = 1$. A random variable x is a measurable function from Ω into some Borel space $(E, \mathcal{E})$, usually $E = \mathbb{R}^d$. Stochastic processes are initially a family of random variables with values in some $\mathbb{R}^d$, but they are implicitly considered as random variables in some topological space, the so-called sample space. This section is somehow a repetition of concepts already

discusses in the previous chapter, however, several viewpoints are possibles and many details are not really considered, neither in the previous chapter nor in this section.

2.1.1 Version of Processes

Let $(X_t = X(t,\omega) : t \geq 0)$ be a family of $\mathbb{R}^d$-valued random variables in a given (complete) probability space $(\Omega, \mathcal{F}, P)$. Certainly, measurability is a first difficulty encountered, i.e., there is not condition on the paths $t \mapsto X(t,\omega)$, for a fixed ω, which means that the process is indeed a random variable taken values in the product space $(\mathbb{R}^d)^{[0,\infty)}$, but its Borel σ-algebra is too small for practical purposes (e.g., a set of one point is not measurable). Most constructions in probability are based on the values of X_t for every t, so that there is not problem in considering two processes X and Y to be same process if they are *indistinguishable*, i.e., there exists a null set N such that $X_t(\omega) = Y_t(\omega)$, for every (t,ω) in $[0,\infty) \times (\Omega \setminus N)$, namely. $P\{\omega : X_t(\omega) = Y_t(\omega), \forall t\} = 1$. However, in many circumstances, only a *version* (or modification) of a process is known, i.e., the random variable X_t could be modified in a null set for each fixed $t \geq 0$, i.e., for each $t \geq 0$ the exists a null set N_t such that $X_t(\omega) = Y_t(\omega)$, for every ω in $\Omega \setminus N_t$, namely, $P\{\omega : X_t(\omega) = Y_t(\omega)\} = 1$, for every $t \geq 0$.

To make aware the reader of some difficulties that may arrive in the theory of general processes, we discuss some initial issues. Even if it is not explicitly stated, a *(stochastic) process* is a family of $\mathbb{R}^d$-valued random variables where some regularity in the t-variable index have been imposed. For instance:

(1) a stochastically (left or right) continuous process $(X_t : t \geq 0)$ satisfies $X_s \to X_t$ in probability as $s \to t$ ($s < t$ or $s > t$ for left or right continuous) for any $t \geq 0$, or

(2) a (left or right) continuous process has almost surely (left or right) continuous paths, i.e., $t \mapsto X_t(\omega)$ is (left or right) continuous for any ω outside of a null set, or even

(3) a separable process $(X_t : t \geq 0)$ has a countable dense subset of indexes (which is called a separant set) $I \subset [0,\infty]$ such that for some null set N and for every (t,ω) in $[0,\infty) \times (\Omega \setminus N)$ there exists a sequence $\{t_n\} \subset I$ satisfying $t_n \to t$ and $X_{t_n}(\omega) \to X_t(\omega)$.

Note that condition (2) implies condition (1), and that any countable dense set could serve as a separant set under condition (2). Also, (1) implies measurable, in the sense that any stochastically (left or right) continuous process has a version which is measurable, in the couple (t,ω). However, it is clear that not any process has a version which satisfies (1) or (2), but it can be proves that given a process $(X_t : t \geq 0)$ there is a version $(Y_t : t \geq 0)$ which is separable, i.e., Y satisfies (3) and $P\{X_t = Y_t\} = 1$ for every $t \geq 0$. Nevertheless, these conditions are essentially very different one from each other. Condition (3) is very useful, but very hard to manipulate. Condition (1) is intrinsic to the joint finite-dimensional distributions of the family of random variables $\{X(t) : t \geq 0\}$

and therefore remains valid for any version of the process $(X_t : t \geq 0)$, while condition (2) is attached to the particular version of the process, say a *pathwise* condition. In the first case (1), we are looking at the process as a function from $[0, \infty)$ into the set of $\mathbb{R}^d$-valued random variables, while in the second case (2), we have random variables with values in the space of (left or right) continuous $\mathbb{R}^d$-valued functions, almost surely. Both concept are intended to address the difficulty presented by the fact that the conditions

$$\text{(a)} \quad P\{X_t = Y_t\} = 0, \ \forall t \geq 0, \qquad \text{(b)} \quad P\{X_t = Y_t, \ \forall t \geq 0\} = 0,$$

are not equivalent, when t ranges on an uncountable set. If both processes $(X_t : t \geq 0)$ and $(Y_t : t \geq 0)$ are left or right continuous (or separable) then (a) and (b) are actually equivalent. Indeed, take a countable dense set I and consider the event $N = \bigcup_{t \in I} \{\omega : X_t(\omega) \neq Y_t(\omega)\}$ for two processes satisfying (a). Since the union is countable, $P(N) = 0$ and the continuity of their paths imply that $X_t(\omega) = Y_t(\omega)$ for any ω in $\Omega \setminus N$ and any t. If both processes are only separable then we take $I = I_X \cup I_Y$ (where I_X or I_Y are a separant set associated with X or Y) and proceed along the line of the previous argument.

On the other hand, if the processes are only stochastically right (or left) continuous then (a) and (b) may not be equivalent. However, a simple argument shows that given a separable stochastically right (or left) continuous process X then any countable dense set is separant. Indeed, for any countable dense set $I = \{t_1, t_2, \ldots\}$ we can find a sequence of positive numbers $\{\delta_1, \delta_2, \ldots\}$ such that $P\{|X(t) - X(t_n)| \geq 2^{-n}\} < 2^{-n}$ for any t in $[t_n, t_n + \delta_n]$. By the Borel-Cantelli lemma the set

$$N_t = \bigcap_m \bigcup_{n \geq m} \left\{\omega : \ |X(t, \omega) - X(t_n, \omega)| \geq 2^{-n}\right\}$$

has probability zero. Since $\mathbb{R} = \bigcup_n [t_n, t_n + \delta_n]$, for any t in $\mathbb{R}$ and any ω in $\Omega \setminus N_t$ there is a sequence of indexes in I such that $X(t_k, \omega)$ converges to $X(t, \omega)$. Because X is separable, there is countable dense set J and null set N, $P(N) = 0$ such that for any t in $\mathbb{R}$ and ω in $\Omega \setminus N$ the previous convergence holds with indexes in J. Therefore, for ω outside of the null set $\bar{N} = N \cup \bigcup_{t \in J} N_t$, there is a sequence of indexes in I such that $X(t_k, \omega)$ converges to $X(t, \omega)$. Moreover, for the given process X, this argument shows that there exists a separable process Y satisfying (a), but not necessarily (b). Indeed, it suffices to define $Y_t(\omega) = X_t(\omega)$ for any t and ω such that ω belongs to $\Omega \setminus N_t$ and $Y_t(\omega) = 0$ otherwise.

In a typical example we consider the Lebesgue measure on $[0, 1]$, two processes $X_t(\omega) = t$ for any t, ω in $[0, 1]$ and $Y_t(\omega) = t$ for $\omega \neq t$ and $Y_t(\omega) = 0$ otherwise. It is clear that condition (a) is satisfied, but (b) does not hold. The process X is continuous (as in (2), sometimes referred to as pathwise continuity), but Y is only stochastically continuous (as in (1), sometimes referred to as continuous in probability), since is clearly almost sure continuous. Also, note that a stochastic process $(X_t : t \geq 0)$ is (right or left) continuous if its restriction to a separant set is so.

Therefore, the intuitive idea that two processes are equals when their finite-dimensional distributions are the same translates into being version of each

other. However, some properties associate with a process are actually depending on the particular version being used, i.e., key properties like measurability on the joint variables (t, ω) or path-continuity depend on the particular version of the process. As mentioned early, these difficulties appear because the index of the family of random variables (i.e., the stochastic process) is uncountable. This is to say that the finite-dimensional distributions of a stochastic process are actually given on a countable family of index, and some kind of continuity (in probability, pathwise or separability) is used to completely determine the stochastic process, i.e., suitable versions of processes are always taken for granted.

2.1.2 Filtered Probability Space

Another key issue is the filtration, i.e., a family of sub σ-algebras $(\mathcal{F}_t : t \geq 0)$ of $\mathcal{F}$, such that $\mathcal{F}_s \subset \mathcal{F}_t$ for every $t > s \geq 0$. As long as the probability P is unchanged, we may complete the $\mathcal{F}$ and $\mathcal{F}_0$ with all the subsets of measure zero. However, in the case of Markov processes, the probability $P = P_\mu$ depends on the initial distribution μ and the *universally completed* filtration is used to properly express the *strong Markov property*. On the other hand, the right-continuity of the filtration, i.e., the property $\mathcal{F}_t = \mathcal{F}_{t+}$, for every $t \geq 0$, where $\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$, is a desirable condition at the point that by *filtration* we understand a right-continuous increasing family of sub σ-algebras $(\mathcal{F}_t : t \geq 0)$ of $\mathcal{F}$ as above.

Usually, the filtration $(\mathcal{F}_t : t \geq 0)$ is attached to a stochastic process $(X_t : t \geq 0)$ in the sense that the random variables $(X_s : s \leq t)$ are $\mathcal{F}_t$-measurable. The filtration generated by a process (or the *history* of the process, i.e, $\mathcal{F}_t = \mathcal{H}_t$ is the smaller sub σ-algebra of $\mathcal{F}$ such that all random variables $(X_s : s \leq t)$ are measurable) represents the *information* obtained by observing the process. The new information is related to the *innovation*, which is defined as the decreasing family of sub σ-algebras $(\mathcal{I}_t : t \geq 0)$, where $\mathcal{I}_t = \mathcal{F}_t^\perp$ is the smaller sub σ-algebra of $\mathcal{F}$ containing all set independent of $\mathcal{F}_t$, i.e., a bounded function f is $\mathcal{F}_t^\perp$-measurable if and only if $\mathbb{E}\{f\, g\} = \mathbb{E}\{f\}\,\mathbb{E}\{g\}$ for any integrable g in $\mathcal{F}_t$-measurable. Hence, another stochastic process $(Y_t : t \geq 0)$ is called *adapted* if Y_t is $\mathcal{F}_t$-measurable for every $t \geq 0$ and *non-anticipating* (or *non-anticipative*) if Y_t is independent of the innovation $\mathcal{I}$, which is equivalent to say that Y_t is $\mathcal{I}_t^\perp$-measurable or $\mathcal{F}_t^{\perp\perp}$-measurable, i.e., $\mathbb{E}\{\varphi(Y_t)\, g\} = \mathbb{E}\{\varphi(Y_t)\}\,\mathbb{E}\{g\}$ for any bounded real Borel measurable function φ and any integrable g satisfying $\mathbb{E}\{f\, g\} = \mathbb{E}\{f\}\,\mathbb{E}\{g\}$ for every integrable f which is $\mathcal{F}_t$-measurable. Note that the filtration $(\mathcal{F}_t : t \geq 0)$, the process or the concept adapted can be defined in a measurable space $(\Omega, \mathcal{F})$, but the innovation $(\mathcal{I}_t : t \geq 0)$ or the concept of non-anticipative requires a probability space $(\Omega, \mathcal{F}, P)$, which involves the regularity in the t-variable index discussed above.

Thus, for a *filtered* space $(\Omega, \mathbb{F}, P)$ or $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, we understand a probability space $(\Omega, \mathcal{F}, P)$ endowed with a filtration $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$ which is always right-continuous. As long as P is fixed, we may assume that $\mathcal{F}_0$ is complete (with respect to $\mathcal{F}$), even more that $\mathcal{F}_t = \mathcal{F}_t^{\perp\perp}$ for every $t \geq 0$ and $\mathcal{F} = \bigvee_{t\geq 0} \mathcal{F}_t$. Sometimes we may change the probability P, but the filtration

may change only when the whole measurable space is changed, except that it may be completed with all null sets as needed. This is referred to as the 'usual conditions' (completed and right-continuity), and when necessary, even $\mathcal{F} = \bigvee_{t\geq 0} \mathcal{F}_t$, with the notations either $\mathcal{F}_t$ or $\mathcal{F}(t)$.

A minimum condition required for a stochastic process is to be *measurable*, i.e., the function $(t,\omega) \mapsto X(t,\omega)$ is measurable with respect to the product σ-algebra $\mathcal{B} \times \mathcal{F}$, where $\mathcal{B} = \mathcal{B}([0,\infty[)$ is the Borel σ-algebra in $[0,\infty[= [0,\infty)$. When general processes are involved, subsets N of $\mathcal{B} \times \mathcal{F}$ with the property that $P\big(\bigcup_{t\geq 0}\{\omega : (t,\omega) \in N\}\big) = 0$ are called *evanescent* and two processes which differ in an evanescent set are considered equals (or undistinguished), every concepts and results are valid *except an evanescent set*, without special mention. As mentioned above, if the processes have some extra *path regularity*, such as separable or stochastically left (or right) continuous, then this is the same as *modification* or *version* of the process.

However, the standard technique is to make a regular (e.g., cad-lag) modification of a general process and then, to refer always to this version. Related to the adapted processes are the *progressively measurable* processes, which are stochastic processes such that the function $(t,\omega) \mapsto X(t,\omega)$ is measurable with respect to the product σ-algebra $\mathcal{B}([0,T]) \times \mathcal{F}$, when considered as a mapping from $\Omega \times [0,T]$, for every T in $[0,\infty[$. There are a couple of useful sub σ-algebras of $\mathcal{B} \times \mathcal{F}$:

(1) the *predictable* σ-algebra $\mathcal{P}$, generated by sets of the form $\{0\} \times F_0$ and $(s,t] \times F_s$ for any F_s in $\mathcal{F}_s$, any $t > s \geq 0$

(2) the *optional* (or *well measurable*) σ-algebra $\mathcal{O}$, generated by sets of the form $\{0\} \times F_0$ and $[s,t) \times F_s$ for any F_s in $\mathcal{F}_s$, any $t > s \geq 0$.

Note that sets of the form $(s,\infty[\times F_s / [s,\infty[\times F_s$ could be used as generator of the predictable/optional σ-algebras. For the sake of convenience and as long as no confusion may arrive, we may exchange the order of the variables t and ω so that $\Omega \times [0,\infty)$ or $[0,\infty) \times \Omega$ are regarded as the same. Clearly $\mathcal{P} \subset \mathcal{O} \subset \mathcal{B}([0,\infty[) \times \mathcal{F}$, where in general the inclusions are strict. It can be proved that $\mathcal{P}$ is the σ-algebra generated by continuous (or left continuous) adapted processes, and that $\mathcal{O}$ is generated by right continuous (or cad-lag) adapted processes.

Thus, a stochastic process X is called predictable (or optional) if the function $(t,\omega) \mapsto X(t,\omega)$ is measurable with respect to $\mathcal{P}$ (or $\mathcal{O}$). However, a $\mathcal{F}$-measurable function from Ω into $[0,\infty]$ is called an optional (or stopping) time if $\{\tau \leq t\}$ (or $\{\tau < t\}$ because $\mathcal{F}_t = \mathcal{F}_{t+}$) is in $\mathcal{F}_t$ for every $t \geq 0$ and $\mathcal{F}_\tau$ is the σ-algebra of all sets A in $\mathcal{F}_\infty = \bigvee_{t\geq 0} \mathcal{F}_t$ such that $A \cap \{\tau \leq t\}$ belongs to $\mathcal{F}_t$ for every $t \geq 0$. If τ and θ are optional times then stochastic intervals of the form $[0,\tau]$ and $(\theta,\tau]$ are predictable. A stopping time is called predictable if there exists a (announcing) sequence of stopping time $\{\tau_1 \leq \tau_2 \leq \tau_k < \tau\}$ convergent to τ. It can be proved that τ is optional (or predictable) if and only if the function $(t,\omega) \mapsto \mathbb{1}_{t\geq\tau}$ is an optional (or predictable) process. Note that if two processes X and Y are equals except in an evanescent set then X is predictable (or optional or progressively measurable or adapted) if and only if Y

is so. Hence, two such a processes are regarded as the same in practically all probabilistic aspects.

These measurability properties are not preserved when using versions of the same process. For instance, if X is a stochastically left continuous adapted process then for every $t, \varepsilon > 0$ there exists $\delta = \delta(t, \varepsilon)$ such that $P\{|X(t) - X(s)| \geq \varepsilon\} \leq \varepsilon$, for any s in $[t - \delta, t]$. Thus, for every sequence of partitions $\pi_n = \{0 = t_{0,n} < t_{1,n} < \cdots < t_{k,n} < \cdots\}$, with $\sup_k(t_{k,n} - t_{k-1,n})$ vanishing as $n \to \infty$, we can define

$$X_n(t,\omega) = \begin{cases} X(0,\omega) & \text{if } t = 0, \\ X(t_{k-1,n},\omega) & \text{if } t_{k-1,n} < t \leq t_{k,n}, \quad k \geq 1. \end{cases}$$

It is clear that X_n is predictable and so is the subset A of $\Omega \times [0,\infty)$, where the sequence $X_n(t,\omega)$ is convergent is also predictable. Therefore the limit

$$Y(t,\omega) = \begin{cases} \lim_n X_n(t,\omega) & \text{for } (t,\omega) \in A, \\ 0 & \text{otherwise,} \end{cases}$$

is also a predictable process. By Borel-Cantelli lemma the set

$$N_t = \bigcap_m \bigcup_{n \geq m} \{\omega : \exists\, k \text{ such that } t_{k-1,n} < t \leq t_{k,n}, \\ , |X(t,\omega) - X(t_{k,n},\omega)| \geq 2^{-n}\}$$

has probability zero for every $t > 0$. Hence, for any ω in $\Omega \setminus N_t$ the sequence $X_n(t,\omega)$ is convergent to $X(t,\omega)$, i.e., $P\{X(t) = Y(t)\} = 1$, for every $t \geq 0$. Thus any stochastically left continuous adapted process has a predictable version. It is clear that X and Y does not necessarily differ on an evanescent set, i.e., the complement of A is not an evanescent set.

To summing-up, in most cases the starting point is a filtered probability space $(\Omega, \mathbb{F}, P)$, where the filtration $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$ satisfies the usual conditions, i.e., $\mathcal{F}_0$ contains all null sets of $\mathcal{F}$ and $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$. An increasing family $\{\mathcal{F}_t^0 : t \geq 0\}$ of σ-algebras is constructed as the history a given process, this family is completed to satisfy the usual conditions, without any loss of properties for the given process. Thus other processes are called adapted, predictable or optional relative to the filtration $\mathbb{F}$, which is better to manipulate than using the original family $\{\mathcal{F}_t^0 : t \geq 0\}$. Therefore, together with the filtered space the predictable $\mathcal{P}$ and optimal $\mathcal{O}$ σ-algebras are defined on the product space $[0,\infty) \times \Omega$. Moreover, sometimes even the condition $\mathcal{F}_\infty = \bigvee_{t \geq 0} \mathcal{F}_t = \mathcal{F}$ may be imposed. It should be clear that properties related to filtered probability spaces depend on the particular version of the process under consideration, but they are considered invariant when the process is changed in an evanescent set.

2.2 Lévy Processes

There are several excellent books on Lévy processes at various levels, e.g., the reader may check Applebaum [1] and Sato [157] (among others) to appreciate

the differences and difficulties involved.

Random walks capture most of the relevant features found in sequences of random variables while Lévy processes can be thought are their equivalent in continuous times, i.e., they are stochastic processes with independent and stationary increments. The best well known examples are the Poisson process and the Brownian motion. They form the class of space-time homogeneous Markov processes and they are the prototypes of semi-martingales.

Definition 2.1. A $\mathbb{R}^d$-valued or d-dimensional *Lévy process* is a random variable X in a complete probability space $(\Omega, \mathcal{F}, P)$ with values in the canonical $D([0,\infty), \mathbb{R}^d)$ such that

(1) for any $n \geq 1$ and $0 \leq t_0 < t_1 < \cdots < t_n$ the $\mathbb{R}^d$-valued random variables $X(t_0)$, $X(t_1) - X(t_2)$,...,$X(t_n) - X(t_{n-1})$ are independent (i.e., independent increments),

(2) for any $s > 0$ the $\mathbb{R}^d$-valued random variables $X(t) - X(0)$ and $X(t+s) - X(s)$ have the same distributions (i.e., stationary increments),

(3) for any $s \geq 0$ and $\varepsilon > 0$ we have $P(|X(t) - X(s)| \geq \varepsilon) \to 0$ as $t \to s$ (i.e., stochastically continuous) and

(4) $P(X(0) = 0) = 1$.

An *additive process* is defined by means of the same properties except that condition (2) on stationary increments is removed. □

Usually the fact that the paths of a Lévy process are almost surely cad-lag is deduced from conditions (1),...,(4) after a modification of the given process. However, we prefer to impose a priori the cad-lag regularity. It is clear that under conditions (2) (stationary increments) and (4) we may replace condition (3) (on stochastically continuous paths) by condition $P(|X(t)| \geq \varepsilon) \to 0$ as $t \to 0$, for every $\varepsilon > 0$.

2.2.1 Generalities of LP

Recall that a classic tool to analyze distributions in $\mathbb{R}^d$ is *characteristic functions* (or Fourier transform). Thus, for a given distribution μ of a random variable ξ in $\mathbb{R}^d$, the characteristic function $\hat{\mu}\colon \mathbb{R}^d \to \mathbb{C}$ is defined by

$$\hat{\mu}(y) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\, x\cdot y}\, \mu(\mathrm{d}x) = \mathbb{E}\{\mathrm{e}^{\mathrm{i}\, y\cdot \xi}\}.$$

If μ_1 and μ_2 are the distributions of two $\mathbb{R}^d$-valued independent random variables ξ_1 and ξ_2 then the *convolution* $\mu_1 \star \mu_2$ defined by

$$(\mu_1 \star \mu_2)(B) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_B(x+y)\, \mu_1(\mathrm{d}x)\, \mu_2(\mathrm{d}y), \quad \forall B \in \mathcal{B}(\mathbb{R}^d)$$

is the distribution of the sum $\xi_1 + \xi_2$. We have $\widehat{\mu_1 \star \mu_2} = \hat{\mu}_1\, \hat{\mu}_2$, and therefore, the characteristic functions of independence of random variables is product of characteristic function of each variable.

If X is a Lévy process then we may consider the characteristic function of the $\mathbb{R}^d$-valued random variable $X(1)$, i.e.,

$$\hat{\mu}(y) = \mathbb{E}\{\mathrm{e}^{\mathtt{i}\, y\cdot X(1)}\}.$$

Since $X(1) = X(1/n) + [X(2/n) - X(1/n)] + \cdots + [X(1) - X(1-1/n)]$, the random variable $X(1)$ can be expressed as the sum of n independent identically distributed random variables, the distribution μ is the n-fold convolution of some distribution μ_n, i.e., $\mu = \mu_n^{n\star}$, μ_n is the distribution of $X(1/n)$. A distribution μ with the above property is called *infinitely divisible*. For instance, Gaussian, Cauchy and Dirac-δ distributions on $\mathbb{R}^d$, as well as Poisson, exponential and Γ distributions on $\mathbb{R}$, are infinitely divisible, for instance see Stroock [168, Section 3.2, pp. 139–153].

Any infinitely divisible distribution μ has a never vanishing characteristic function $\hat{\mu}$ which can be expressed as an exponential function, i.e.,

$$\hat{\mu}(y) = \exp[-\phi(y)], \quad \forall y \in \mathbb{R}^d,$$

where ϕ is uniquely determined as a complex-valued continuous function in $\mathbb{R}^d$ with $\phi(0) = 1$, which is called *characteristic exponent* or the Lévy symbol. Thus, we have $\mathbb{E}\{\mathrm{e}^{\mathtt{i}\, y\cdot X(t)}\} = \exp[-t\phi(y)]$ for t rational and by continuity for any $t \geq 0$. Since the Fourier transform is one-to-one, the expression

$$\widehat{\mu^{\star t}}(y) = \exp[-t\phi(y)], \quad \forall y \in \mathbb{R}^d,\ t > 0,$$

define the $^\star t$-convolution. Moreover, $\mu^{\star t}$ is also an infinitely divisible distribution.

A key result is Lévy-Khintchine formula states that a complex-valued function ϕ is the characteristic exponent of an infinitely divisible distributions μ if and only if

$$\phi(y) = \mathtt{i}\, g\cdot y + \frac{1}{2}\, Qy\cdot y + \int_{\mathbb{R}^d_*} \big[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x} + \mathtt{i}\, y\cdot x \mathbb{1}_{|x|<1}\big]\, m(\mathrm{d}x),$$

for every y in $\mathbb{R}^d$, where g belongs to $\mathbb{R}^d$, Q is a non-negative semi-definite $d\times d$-matrix and m is a Radon measure on $\mathbb{R}^d_* = \mathbb{R}^d \smallsetminus \{0\}$ which integrates the function $x \mapsto |x|^2 \wedge 1$. The representation of ϕ by (g, Q, m) is unique. However, the cut-off function $\mathbb{1}_{|x|<1}$ may be replaced by a bounded smooth function which is equal to 1 at the origin, e.g. $(1+|x|^2)^{-1}$. In this case, the parameter g changes and we have for every y in $\mathbb{R}^d$,

$$\phi(y) = \mathtt{i}\, f\cdot y + \frac{1}{2}\, Qy\cdot y + \int_{\mathbb{R}^d_*} \big[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x} + \mathtt{i}\frac{y\cdot x}{1+|x|^2}\big]\, m(\mathrm{d}x),$$

$$f = g + \int_{\mathbb{R}^d} x\big[\frac{1}{1+|x|^2} - \mathbb{1}_{|x|<1}\big]\, m(\mathrm{d}x).$$

We may also use $\sin x$ as in Krylov [103, Section 5.2, pp. 137–144], for the one-dimensional case.

2.2.2 Compound Poisson Processes

An important class of Lévy processes are the so-called (compound) Poisson processes. A Lévy process X is called a *Poisson process* with parameter $c > 0$, if $X(t)$ has a Poisson distribution with mean ct, for every $t \geq 0$, in other words, X is a cad-lag process with independent increments, $X(0) = 0$, and

$$P\big\{X(t) - X(s) = k\big\} = \frac{\mathrm{e}^{-c(t-s)}(c(t-s)^k)}{k!}, \quad \forall k = 0, 1, \dots, \ t \geq s \geq 0.$$

Similarly, a Lévy process X is called a *compound Poisson* process with parameters (c, γ), where $c > 0$ and γ is a distribution in $\mathbb{R}^d$ with $\gamma(\{0\}) = 0$ (i.e., γ is a distribution in $\mathbb{R}^d_*$), if $\mathbb{E}\{\mathrm{e}^{\mathrm{i}\, y\cdot X(t)}\} = \exp[-t\, c(\hat\gamma(y) - 1)]$, for any $t \geq 0$ and y in $\mathbb{R}^d$, with $\hat\gamma$ the characteristic function of the distribution γ. The parameters (c, γ) are uniquely determined by X and a simple construction is given as follows. If $\{\zeta_n : n = 1, 2, \dots\}$ is a sequence of independent identically distributed (with distribution law γ) random variables, and $\{\tau_n : n = 1, 2, \dots\}$ is another sequence of independent exponentially distributed (with parameter c) random variables, with $\{\zeta_n : n = 1, 2, \dots\}$ independent of $\{\tau_n : n = 1, 2, \dots\}$, then for $\theta_n = \tau_1 + \tau_2 + \cdots + \tau_n$ (which has a Gamma distribution with parameters c and n), the expressions

$$\begin{aligned}
&X(t) = \sum_{n=1}^{\infty} \zeta_n \mathbb{1}_{t \geq \theta_n}, \quad \text{with} \quad \delta X(t) = X(t) - X(t-) \\
&\delta X(\theta_n) = \zeta_n, \quad \text{and} \quad \delta X(t) = 0 \ \text{ if } t \neq \theta_n, \ \forall n, \ \text{or equivalently} \\
&X(t) = \zeta_1 + \zeta_2 + \cdots + \zeta_n \ \text{ if } \ \sum_{i=1}^{n} \tau_i = \theta_n \leq t < \theta_{n+1} = \sum_{i=1}^{n+1} \tau_i,
\end{aligned}$$

are realizations of a compound Poisson process and its associate point (or jump) process. Indeed, for any integer k, any $0 \leq t_0 < t_1 < \cdots < t_k$ and any Borel subsets $B_0, B_1, \dots, B_k$ of $\mathbb{R}^d$ we can calculate the finite-dimensional distributions of X by the formula

$$\begin{aligned}
P(X(t_0) \in B_0,\, X(t_1) - X(t_0) \in B_1, \dots, X(t_k) - X(t_{k-1}) \in B_k) = \\
= P\big(X(t_0) \in B_0\big)\, P\big(X(t_1) - X(t_0) \in B_1\big) \dots \\
\dots P\big(X(t_k) - X(t_{k-1}) \in B_k\big).
\end{aligned}$$

This yields the expression

$$\mathbb{E}\{\mathrm{e}^{\mathrm{i}\, y\cdot X(t)}\} = \exp[-t\, c\,(1 - \hat\gamma(y))], \quad \forall y \in \mathbb{R}^d, t \geq 0,$$

which is continuous in t. Then, all conditions in Definition 2.1, including the stochastic continuity of path (3), are satisfied. Note that for a pairwise disjoint family of Borel sets of the form $]s_i, t_i] \times B_i$, with $0 \leq s_i < t_i$, B_i in $\mathcal{B}(\mathbb{R}^d)$, $i = 1, 2, \dots, k$ the integer-valued random variables

$$\nu(]s_i, t_i] \times B_i) = \sum_{n=1}^{\infty} \mathbb{1}_{s_i < \theta_n \leq t_i}\, \mathbb{1}_{\zeta_n \in B_i}, \quad \forall i = 1, 2, \dots, k$$

are independent identically Poisson distributed, with parameter (or mean) $c\,(t_i - s_i)\gamma(B_i)$.

An interesting point is the fact that a compound Poisson process in $\mathbb{R}^+$, with parameters (c, σ) such that $c > 0$ and σ is a distribution in $(0, \infty)$, is increasing in t and its Laplace transform is given by

$$\mathbb{E}\{\mathrm{e}^{-\xi X(t)}\} = \exp\Big[- t\, c \int_{(0,\infty)} (\mathrm{e}^{-\xi x} - 1)\, \sigma(\mathrm{d}x)\Big], \quad \forall \xi \in \mathbb{R}, t \geq 0.$$

These processes are called *subordinator* and are used to model random *time changes*, possible discontinuous. Moreover, the Lévy measure $\mathtt{m}$ of any Lévy process with increasing path satisfies

$$\int_{\mathbb{R}^1_*} |x|\, \mathtt{m}(\mathrm{d}x) = \int_0^\infty x\, \mathtt{m}(\mathrm{d}x) < \infty,$$

e.g., see books Applebaum [1, Section 2.3, pp. 99-112], Bertoin [7, Chapter III, pp. 71-102], Itô [77, Section 1.11] and Sato [157, Chapter 6, pp. 197-236].

The interested reader, may consult the book by Applebaum [1], which discuss Lévy process at a very accessible level.

2.2.3 Wiener Processes

The next typical class Lévy processes is the Wiener processes or Brownian motions. A Lévy process X is called a *Brownian motion* or *Wiener* process in $\mathbb{R}^d$, with (vector) drift b in $\mathbb{R}^d$ and (matrix) co-variance σ^2, a nonnegative-definite $d \times d$ matrix, if $\mathbb{E}\{\mathrm{e}^{y \cdot X(t)}\} = \exp[-t(|\sigma y|^2/2 - i\, b)]$, for any $t \geq 0$ and y in $\mathbb{R}^d$, i.e., if $X(t)$ has a Gaussian distribution with (vector) mean $\mathbb{E}\{X(t)\} = bt$ and (matrix) co-variance $\mathbb{E}\{(X(t) - bt)^*(X(t) - bt)\} = t\sigma^2$. A *standard* Wiener process is when $b = 0$ and $\sigma^2 = 1$, the identity matrix. The construction of a Wiener process is a somehow technical and usually details are given for the standard Wiener process with t in a bounded interval. The general case is an appropriate transformation of this special case. First, let $\{\xi_n : n = 1, 2, \dots\}$ be a sequence of independent identically normally distributed (i.e., Gaussian with zero-mean and co-variance 1) random variables in $\mathbb{R}^d$ and let $\{e_n : n = 1, 2, \dots\}$ be a complete orthonormal sequence in $L^2(]0, \pi[)$, e.g., $e_n(t) = \sqrt{2/\pi}\cos(nt)$. Define

$$X(t) = \sum_{n=1}^{\infty} \xi_n \int_0^t e_n(s)\mathrm{d}s, \quad t \in [0, \pi].$$

It is not hard to show that X satisfies all conditions of a Wiener process, except for the stochastic continuity and the cad-lag sample property of paths. Next, essentially based on the (analytic) estimate: for any constants $\alpha, \beta > 0$ there exists a positive constant $C = C(\alpha, \beta)$ such that

$$|X(t) - X(s)|^\alpha \leq C\, |t - s|^\beta \int_0^\pi \mathrm{d}t \int_0^\pi |X(t) - X(s)|^\alpha\, |t - s|^{-\beta - 2}\mathrm{d}s,$$

for every t, s in $[0,\pi]$, we may establish that that series defining the process X converges uniformly in $[0,\pi]$ almost surely. Indeed, if X_k denotes the k partial sum defining the process X then an explicit calculations show that

$$\mathbb{E}\{|X_k(t)-X_\ell(s)|^4\}=\mathbb{E}\Big\{\Big|\sum_{n=\ell+1}^{k}\xi_n\int_s^t e_n(r)\mathrm{d}r\Big|^4\Big\}\le 3|t-s|^2,$$

for every $t\ge s\ge 0$ and $k>\ell\ge 1$. After using the previous estimate with $\alpha=4$ and $1<\beta<2$ we get

$$\mathbb{E}\{\sup_{|t-s|\le\delta}|X_k(t)-X_\ell(s)|^4\}\le C\,\delta^\beta,\quad \forall\delta>0,\ k>\ell\ge 1,$$

for a some constant $C>0$. This proves that X is a Wiener process with continuous paths. Next, the transformation $t\,X(1/t)$ (or patching k independent copies, i.e., $X_k(t)$ if $(k-1)\pi\le t<k\pi$, for $k\ge 1$.) produces a standard Wiener process in $[0,\infty)$ and the process $b\,t+\sigma\,X(t)$ yields a Wiener process with parameters b and σ.

The above estimate is valid even when t is multidimensional and a proof can be found in Da Prato and Zabczyk [29, Theorem B.1.5, pp. 311–316]. For more details on the construct arguments, see, e.g., Friedman [55] or Krylov [102].

For future reference, we state the general existence result without any proof.

Theorem 2.2 (construction)**.** *Let m be a Radon measure on $\mathbb{R}^d_*$ such that*

$$\int_{\mathbb{R}^d_*}|x|^2\wedge 1\,m(\mathrm{d}x)<\infty,$$

Q be a nonnegative-definite $d\times d$ matrix and g be a vector in $\mathbb{R}^d$. Then there exists a unique probability measure P on the canonical probability space $\Omega=D([0,\infty),\mathbb{R}^d)$ such that the canonical process $(X(t)=\omega(t):t\ge 0)$ is a Lévy process with characteristic *(g,Q,m), i.e.,*

$$\mathbb{E}\{\mathrm{e}^{\mathtt{i}\,y\cdot X(t)}\}=\exp[-t\,\phi(y)],\quad \forall y\in\mathbb{R}^d,\ t\ge 0,\qquad \textit{with}$$
$$\phi(y)=\mathtt{i}\,g\cdot y+\frac{1}{2}\,Qy\cdot y+\int_{\mathbb{R}^d_*}\big[1-\mathrm{e}^{\mathtt{i}\,y\cdot x}+\mathtt{i}\,y\cdot x\mathbb{1}_{|x|<1}\big]\,m(\mathrm{d}x).$$

Conversely, given a Lévy process X the characteristic (g,Q,m) is uniquely determined through the above formula.

Recall that any infinitely divisible probability measure on $\mathbb{R}^d$ can be viewed as the distribution of a Lévy process evaluated at time 1, and, an important point to remark is that the construction shows that any Lévy process is a Wiener process plus the limit of a sequence of compound Poisson processes. However, the structure of a typical graph of Levy process seems to elude us. For instance, almost surely, the jumping times $J=\{t:X(t,\omega)\ne X(t-,\omega)\}$ are countable, and (a) if the Levy measure satisfies $\mathtt{m}(\mathbb{R}^d)=\infty$ then J is dense in $[0,\infty)$ while (b) if $\mathtt{m}(\mathbb{R}^d)<\infty$ then J can be written as an increasing sequence $\{\tau_k:k\ge 1\}$, $\tau_k\le\tau_{k+1}$, of independent random variables having exponential distributions with mean $1/\mathtt{m}(\mathbb{R}^d)$, see Sato [157, Theorem 21.3, pp. 136–137].

2.2.4 Path-regularity for LP

To end this section, let us take a look at the path-regularity of the Lévy processes. If we drop the cad-lag condition in the Definition 2.1 then we use the previous expressions (for either Lévy or additive processes *in law*) to show that there exits a cad-lag version, see Sato [157, Theorem 11.5, p. 65], which is actually indistinguishable of the initial Lévy or additive process was a separable process.

Proposition 2.3. *Let y be an additive process in law on a (non-necessarily completed) probability space $(\Omega, \mathcal{F}, P)$, and let $\mathcal{F}_t^0(y)$ denote the σ-algebra generated by the random variables $\{y(s) : 0 \le s \le t\}$. Define $\mathcal{F}_t(y) = \mathcal{F}_t^0(y) \vee \mathcal{N}$, the minimal σ-algebra containing both $\mathcal{F}_t^0(y)$ and $\mathcal{N}$, where $\mathcal{N} = \{N \in \mathcal{F} : P(N) = 0\}$. Then $\mathcal{F}_t(y) = \bigcap_{s>t} \mathcal{F}_s(y)$, for any $t \ge 0$.*

Proof. Set $\mathcal{F}_t^+(y) = \bigcap_{s>t} \mathcal{F}_s(y)$ and $\mathcal{F}_\infty^0(y) = \vee_{t\ge 0}\mathcal{F}_t^0(y)$. Since both σ-algebras contain all null sets in $\mathcal{F}$, we should prove that $\mathbb{E}(Z \mid \mathcal{F}_t^+(y)) = \mathbb{E}(Z \mid \mathcal{F}_t(y))$ for any $\mathcal{F}_\infty^0(y)$-measurable bounded random variable Z, to get the right-continuity of the filtration. Actually, it suffices to establish that

$$\mathbb{E}\{e^{\mathtt{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_t^+(y)\} = \mathbb{E}\{e^{\mathtt{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_t(y)\}$$

for any choice of $0 \le s_1 \le s_2 \le \ldots \le s_n$, $(r_1, r_2, \ldots, r_n)$, and n. Moreover, only the case $s_1 > t$ need to be considered. To this purpose, we use the characteristic function $f_t(r) = \mathbb{E}\{e^{\mathtt{i}ry(t)}\}$ which satisfies $f_{t+s}(r) = f_t(r)f_s(r)$, and the martingale property of $M_t(r) = e^{\mathtt{i}ry(t)}/f_t(r)$ with respect to $\mathcal{F}_t(y)$.

Now, let $s_1 > t' \ge t$ and consider

$$\begin{aligned}
\mathbb{E}\{e^{\mathtt{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_{t'}(y)\} &= f_{s_n}(r_n)\mathbb{E}\{e^{\mathtt{i}\sum_{j=1}^{n-1} r_j y(s_j)} M_{s_n}(r_n) \mid \mathcal{F}_{t'}(y)\} = \\
&= f_{s_n}(r_n)\mathbb{E}\{e^{\mathtt{i}\sum_{j=1}^{n-1} r_j y(s_j)} M_{s_{n-1}}(r_n) \mid \mathcal{F}_{t'}(y)\} = \\
&= f_{s_n - s_{n-1}}(r_n) f_{s_{n-1}}(r_{n-1} + r_n) \times \\
&\qquad \times \mathbb{E}\{e^{\mathtt{i}\sum_{j=1}^{n-2} r_j y(s_j)} M_{s_{n-1}}(r_{n-1} + r_n) \mid \mathcal{F}_{t'}(y)\}) = \\
&= \ldots = f_{s_n - s_{n-1}}(r_n) f_{s_{n-1} - s_{n-2}}(r_{n-1} + r_n) \times \\
&\qquad \times f_{s_{n-2} - s_{n-3}}(r_{n-2} + r_{n-1} + r_n) \times \\
&\qquad \times \ldots \times f_{s_2 - s_1}(r_2 + \cdots + r_{n-2} + r_{n-1} + r_n) e^{\mathtt{i}r_1 y(s_1)},
\end{aligned}$$

i.e., we have

$$\mathbb{E}\{e^{\mathtt{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_{t+\varepsilon}(y)\} = \mathbb{E}\{e^{\mathtt{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_t(y)\}, \quad \forall \varepsilon > 0.$$

and the proof is finished by passing to the limit as $\varepsilon \to 0$. □

• *Remark* 2.4. Sometimes, an adapted process y (not necessarily cad-lag) is called *additive* with respect to a filtration $\mathbb{F}$ (non necessarily right-continuous or complete) if the random variable $y(s) - y(t)$ is independent of $\mathcal{F}(t)$, for any $s > t \ge 0$. Because y is adapted and $\mathcal{F}(t)$ increasing, this is equivalent to a

stronger condition, namely, the σ-algebra $\mathcal{G}(t)$ generated by $\{y(s_2) - y(s_1) : s_2 > s_1 \geq t\}$ is independent of $\mathcal{F}(t)$ for any $t \geq 0$. Now, let $\mathcal{N}$ be the σ-algebra of all null sets in $\mathcal{F}$ and set $\mathcal{F}(t+) = \bigcap_{\varepsilon>0} \mathcal{F}(t+\varepsilon)$. If y is right-continuous in probability then we want show that $\mathbb{E}\{\cdot \,|\, \mathcal{F}(t+)\} = \mathbb{E}\{\cdot \,|\, \mathcal{F}(t)\}$. Indeed, for any t there is a sequence $\{t_n\}$, $t_n > t$ convergent to t and a set of measure null such that $y(t_n, \omega) \to y(t, \omega)$, for every ω in $\Omega \setminus N$. Since $y(s) - y(t_n)$, $s > t$, is independent of $\mathcal{F}(t_n) \supset \mathcal{F}(t+)$, we have

$$\mathbb{E}\big\{f\big(y(s) - y(t_n)\big)\mathbb{1}_F\big\} = \mathbb{E}\big\{f\big(y(s) - y(t_n)\big)\big\}\,\mathbb{E}\{1_F\}, \quad \forall F \in \mathcal{F}(t+),$$

for every continuous function f. Hence, $y(s) - y(t)$, $s > t$ is independent of $\mathcal{F}(t+)$, i.e., $\mathcal{G}(t) \vee \mathcal{N}$ is independent of $\mathcal{F}(t+)$, for every $t \geq 0$. Now, if A is in $\mathcal{F}(t)$ and B in $\mathcal{G}(t) \vee \mathcal{N}$ then the $\mathcal{F}(t)$-measurable random variable $\mathbb{1}_A\, P(B)$ is a version of the conditional expectation $\mathbb{E}\{\mathbb{1}_A \mathbb{1}_B \,|\, \mathcal{F}(t+)\}$, and a class monotone argument shows that for any bounded and $\mathcal{F}(t) \vee \mathcal{G}(t) \vee \mathcal{N}$-measurable random variable h we have a $\mathcal{F}(t)$-measurable version of the $\mathbb{E}\{h \,|\, \mathcal{F}(t+)\}$. This proves that $\mathcal{F}(t+) = \mathcal{F}(t) \vee \mathcal{N}$, i.e., another way of proving the previous Proposition 2.3. □

The reader is referred to the books by Bremaud [19], Elliott [43], Protter [149]), and the comprehensive works by Bertoin [7, Chapters O and I, pp. 1–42] and Sato [157, Chapters 1 and 2, pp. 1–68].

2.3 Martingales in Continuous Time

Martingales plays a key role in stochastic analysis, and in all what follows a martingale is a cad-lag process X with the following property relative to the conditional expectation

$$\mathbb{E}\{X(t) \,|\, X(r),\, 0 \leq r \leq s\} = X(s), \quad \forall t \geq s > 0, \tag{2.1}$$

and when the $=$ sign replaced by the $\geq$ sign in the above property, the process X is called a sub-martingale, and similarly a super-martingale with the $\leq$ sign. The conditional expectation requires an integrable process, i.e., $\mathbb{E}\{|X(t)|\} < \infty$ for every $t \geq 0$ (for sub-martingale $\mathbb{E}\{[X(t)]^+\} < \infty$ and for super-martingale $\mathbb{E}\{[X(t)]^-\} < \infty$ are sufficient). Moreover, only a version of the process X is characterized by this property, so that a condition on the paths is also required. A minimal condition is to have a separable process X, but this theory becomes very useful when working with cad-lag process X. We adopted this point of view, so in this context, a *martingale is always a cad-lag integrable process*. Most of the time we replace the conditional expectation property with a more general statement, namely

$$\mathbb{E}\{X(t) \,|\, \mathcal{F}(s)\} = X(s), \quad \forall t \geq s > 0,$$

where now X is a cad-lag integrable process adapted to the filtration $(\mathcal{F}(t) : t \geq 0)$, which is always assumed right-continuous and even completed when necessary. However, the concept of martingale is independent of the filtration

as soon as some regularity on the paths is given. Actually, the conditional expectation property is equivalent to the condition

$$\mathbb{E}\Big\{X(t)\prod_{i=1}^{n}h_i\big(X(s_i)\big)\Big\}=\mathbb{E}\Big\{X(s)\prod_{i=1}^{n}h_i\big(X(s_i)\big)\Big\},$$

for every $0\le s_1<s_2\cdots\le s_n\le s<t$, any (real-valued) Borel and bounded functions h_i, $i=1,\dots,n$, any integer n. Nevertheless, to weaker the condition on integrability, a technical *localization* procedure is used, and a *local-martingale* is a cad-lag process X such that $X_k:t\mapsto X(t\wedge\tau_k)-X(0)$ is a martingale for some increasing sequence of stopping times τ_k satisfying $\tau_k\to\infty$. This forces the use of a filtration.

Note the contrast of the previous property and the Markov property valid for a Markov process X: for any $n=1,2\dots$, any bounded measurable (actually continuous suffices) functions $f_1,\dots,f_n$, $g_1,\dots,g_n$, h, and times $s_1\le\cdots\le s_n\le t\le t_1\le\cdots\le t_n$ we have

$$\mathbb{E}\big\{h(X_t)\big(\prod_{i=1}^{n}f(X_{s_i})\big)\big(\prod_{i=1}^{n}g(X_{t_i})\big)\big\}=$$
$$=\mathbb{E}\big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n}f(X_{s_i})\,|\,X_t\}\big\}\,\mathbb{E}\{\prod_{i=1}^{n}g(X_{t_i})\,|\,X_t\}\big\},$$

where $\mathbb{E}\{\prod_{i=1}^{n}f(X_{s_i}\,|\,X_t)\}$ and $\mathbb{E}\{\prod_{i=1}^{n}g(X_{t_i})\,|\,X_t\}$ are X_t-measurable functions satisfying

$$\mathbb{E}\Big\{h(X_t)\prod_{i=1}^{n}f(X_{s_i})\Big\}=\mathbb{E}\Big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n}f(X_{s_i})\,|\,X_t)\}\Big\},$$
$$\mathbb{E}\Big\{h(X_t)\prod_{i=1}^{n}g(X_{t_i})\Big\}=\mathbb{E}\Big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n}g(X_{t_i})\,|\,X_t)\}\Big\},$$

i.e., they are the conditional expectations with respect to the σ-algebra generated by the random variable X_t. This is briefly expressed by saying that the *past* and the *future* are independent given the *present*. Clearly, this condition involves only the finite-dimensional distributions of the process, and no condition on integrability for X is necessary for the above Markov property.

Also note that for a random process $X=\{X(t),t\ge0\}$ with independent increments, i.e., for any $n\ge1$ and $0\le t_0<t_1<\cdots<t_n$ the $\mathbb{R}^d$-valued random variables $X(t_0)$, $X(t_1)-X(t_2)$,..., $X(t_n)-X(t_{n-1})$ are independent, we have the following assertions: (a) if $\mathbb{E}\{|X(t)|\}<\infty$ for every $t\ge0$ then the random process $t\mapsto X(t)-\mathbb{E}\{X(t)\}$ satisfies the martingale inequality (2.1), and (b) if $\mathbb{E}\{X(t)\}=0$ and $\mathbb{E}\{|X(t)|^2\}<\infty$ for every $t\ge0$ then the random process $t\mapsto(X(t))^2-\mathbb{E}\{(X(t))^2\}$ also satisfies the martingale inequality (2.1).

For instance, the reader is referred to the books Chung and Williams [25], Bichteler [11], Dudley [37, Chapter 12, pp. 439–486], Durrett [40], Elliott [43], Kuo [107], Medvegyev [120], Protter [149], among others, for various presentations on stochastic analysis.

2.3.1 Dirichlet Class

We rephrase the above martingale concept

Definition 2.5 (martingale)**.** A martingale (process) relative to a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ is a random variable M (P-equivalence class) with values into the canonical space $D([0,\infty), \mathbb{R}^d)$ satisfying the *martingale property*

$$\mathbb{E}\{|M(t)|\} < \infty, \ \forall t, \qquad \mathbb{E}\{M(t) \,|\, \mathcal{F}(s)\} = M(s), \ \text{ a.s. } \ \forall t > s,$$

where $M(t) = M(\omega)(t)$. If the filtration $\mathbb{F} = \{\mathcal{F}(t) : t \geq 0\}$ is not mentioned, then it is assumed that $\{\mathcal{F}(t) : t \geq 0\}$ is the smallest filtration satisfying the usual condition, which renders the process $\{M(t) : t \geq 0\}$ adapted. Moreover, the martingale is called *continuous* if M take values into the canonical space $C([0,\infty), \mathbb{R}^d)$ almost surely, and it is called *uniformly integrable* if the family of random variables $\{M(t),\, t \geq 0\}$ is uniformly integrable, i.e., for any $\varepsilon > 0$ there is a $r > 0$ sufficiently large such that $P\{|M(t)| \geq r\} \leq \varepsilon$, for any t in $[0,\infty)$. When $d = 1$, i.e., with values in $\mathbb{R}$, we may define also *super-* or *sub-*martingale by replacing the equal sign by either $\leq$ or $\geq$ in the above condition. Sometimes, martingales are considered in a bounded (or unbounded) time interval instead of the semi-line $[0,\infty)$. □

First, note the role of uniformly integrability by mentioning Doob's martingale convergence and optional-sampling results

Theorem 2.6. *If M is martingale bounded in L^1, i.e., $\sup_t \mathbb{E}\{|M(t)|\} < \infty$, the limit $M(\infty) = \lim_{t\to\infty} M(t)$ exists almost surely and the convergence of $M(t)$ to $M(\infty)$ is in L^1 if and only if the martingale is uniformly integrable. On the other hand, if M is an uniformly integrable martingale then* (a) *the family of $\mathbb{R}^d$-valued random variable $\{M(\tau) : \tau$ is a stopping time$\}$ is uniformly integrable, and* (b) *for any stopping times $\tau \leq \theta$ the equality $\mathbb{E}\{M(\theta) \,|\, \mathcal{F}(\tau)\} = M(\tau)$ holds almost surely.* □

As in the discrete case, the proof is mainly based on the Doob's upcrossing estimate. A (super-/sub-) martingale M satisfying the property (a) of the above theorem is called of *class (D)* (Dirichlet class). Note that an uniformly integrable super(or sub)-martingale need not to be of class (D). However, for any nonnegative sub-martingale X we have

$$r\, P\big(\sup_{s\leq t} X(s) \geq r\big) \leq \mathbb{E}\{X(t)\, \mathbb{1}_{\sup_{s\leq t} X(s)\geq r}\} \leq \mathbb{E}\{X(t)\}, \tag{2.2}$$

and therefore

$$\|\sup_{s\leq t} X(s)\|_p \leq p'\, \|X(t)\|_p, \qquad \text{with } \ 1/p + 1/p' = 1, \tag{2.3}$$

actually, valid even if t is replaced by a stopping time τ. Here $\|\cdot\|_p$ denotes the norm in $L^p(\Omega, P, \mathcal{F})$. However, $p' = \infty$ for $p = 1$, this inequality becomes

$$\|\sup_{s\leq t} X(s)\|_1 \leq \frac{\mathrm{e}}{\mathrm{e}-1}\, \|X(t) \ln^+ X(t)\|_1, \tag{2.4}$$

where $\ln^+(\cdot)$ is the positive part of $\ln(\cdot)$, but this is rarely used.

2.3.2 Doob-Meyer Decomposition

The following decomposition is extremely useful to extend the previous result to sub-martingales.

Theorem 2.7 (Doob-Meyer)**.** *If X is a (continuous) sub-martingale of class (D) then there exists a uniformly integrable martingale M and an integrable predictable (continuous) monotone increasing process A, both null at time zero such that $X = X(0) + M + A$. Moreover, this decomposition is unique.* □

Note that (1) a martingale M is uniformly integrable if and only if for any $\varepsilon > 0$ there exists $\eta > 0$ such that $\mathbb{E}\{\mathbb{1}_{|M(t)>\eta|} M(t)\} \leq \varepsilon$ for every $t \geq 0$, while (2) an integrable monotone increasing process A means a process with cad-lag monotone paths such that $\mathbb{E}\{|A(t)|\} < \infty$, for every $t \geq 0$.

For instance, a comprehensive proof of this fundamental results can be found Rogers and Williams [153, Section VI.6, pp. 367–382]. In particular, if X is an adapted (cad-lag) increasing process satisfying $\mathbb{E}\{\sup_t |X(t)|\} < \infty$ then X is a sub-martingale of class (D) and the above decomposition yields the so-called predictable *compensator*. Certainly, this can be extended to integrable bounded variation processes, by using the positive and negative variation.

In view of Doob-Meyer Theorem 2.7, the previous convergence Theorem 2.6 can be extended to super-/sub-martingales of class (D) and the process $A = A_X$ is called the (predictable) *compensator* of the sub-martingale X, and because M is a martingale, the equality

$$\mathbb{E}\{A(\theta) - A(\tau)\} = \mathbb{E}\{X(\theta) - X(\tau)\},$$

holds true, for any stopping times $\tau \leq \theta$.

Let us denote by $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ the space of square-integrable martingales M null at time zero, i.e., besides the martingale conditions in Definition 2.5 we impose $M(0) = 0$ and $\sup_{t\geq 0} \mathbb{E}\{|M(t)|^2\} < \infty$. A square-integrable martingale M is uniformly integrable and the convergence theorem applies to produce a $\mathcal{F}(\infty)$-measurable random variable $M_\infty = M(\infty)$ with values in $\mathbb{R}$ (or $\mathbb{R}^d$) and square-integrable such that $M(t) = \mathbb{E}\{M(\infty) \,|\, \mathcal{F}(t)\}$. Hence, the space $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ can be identified with the closed subspace of the Hilbert space $L^2(\Omega, P, \mathcal{F}_\infty)$, $\mathcal{F}_\infty = \mathcal{F}(\infty)$, satisfying $\mathbb{E}\{M(\infty) \,|\, \mathcal{F}(0)\} = 0$. Note that if M^* denotes the sup-process defined by $M^*(t) = \sup_{s\leq t} |M(s)|$ and its limit $M^*(\infty) = \sup_{t\geq 0} |M(t)|$ then we have

$$\mathbb{E}\{|M^*(\infty)|^2\} \leq 4 \sup_{t\geq 0} \mathbb{E}\{|M(t)|^2\} = 4\,\mathbb{E}\{|M(\infty)|^2\},$$

after using Doob's estimate (2.3) with $p = 2$. Thus, $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ can be regarded as a Banach space with the norm $\|M^*(\infty)\|_p$, with $p = 2$, for any element M, without changing the topology. Moreover, the space of *continuous* square-integrable martingale processes, denoted by $M^2_c(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ is a closed subspace of the Hilbert space $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$.

Thus, we may consider its orthogonal complement referred to as *purely discontinuous* square-integrable martingale processes null at time zero and denoted by $M_d^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$, of all square-integrable martingale processes Y null at time zero satisfying $\mathbb{E}\{M(\infty)\,Y(\infty)\} = 0$ for all elements M in $M_c^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$, actually, M and Y are what is called *strongly orthogonal*, i.e., $(M(t)\,Y(t) : t \geq 0)$ is an uniformly integrable martingale. The concept of strongly orthogonal is actually stronger than the concept of *orthogonal* in M^2 and weaker than imposing $M(t) - M(s)$ and $Y(t) - Y(s)$ independent of $\mathcal{F}(s)$ for every $t > s$.

Let M be a (continuous) square-integrable martingale process null at time zero, in a given filtered space $(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$. Based on the above argument M^2 is a sub-martingale of class (D) and Doob-Meyer decomposition Theorem 2.7 applies to get a unique predictable (continuous) increasing process $\langle M \rangle$, referred to as the *predictable quadratic variation* process, such that $t \mapsto M^2(t) - \langle M \rangle(t)$ is a martingale. Thus, for a given element M in $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$, we have a unique pair M_c in $M_c^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ and M_d in $M_d^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ such that $M = M_c + M_d$. Applying Doob-Meyer decomposition to the sub-martingales M_c and M_d we may define (uniquely) the so-called *quadratic variation* (or optional quadratic variation) process by the formula

$$[M](t) = \langle M_c \rangle(t) + \sum_{s \leq t} (M_d(s) - M_d(s-))^2, \quad \forall t > 0. \tag{2.5}$$

Note that $[M_c] = \langle M_c \rangle$ and $M_d(t) - M_d(t-) = M(t) - M(t-)$, for any $t > 0$. We re-state these facts for a further reference

Theorem 2.8 (quadratic variations)**.** *Let M be a (continuous) square-integrable martingale process null at time zero, in a given filtered space $(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$. Then* (1) *there exists a unique predictable (continuous) integrable monotone increasing process $\langle M \rangle$ null at time zero such that $M^2 - \langle M \rangle$ is a (continuous) uniformly integrable martingale, and* (2) *there exists a unique optional (continuous) integrable monotone increasing process $[M]$ null at time zero such that $[M](t) - [M](t-) = (M(t) - M(t-))^2$, for any $t > 0$, and $M^2 - [M]$ is a (continuous) uniformly integrable martingale. Moreover $M = 0$ if and only if either $[M] = 0$ or $\langle M \rangle = 0$.* □

With all this in mind, for any two square-integrable martingale process null at time zero M and N we define the predictable and optional quadratic covariation processes by

$$\begin{aligned} \langle M, N \rangle &= \big(\langle M + N \rangle - \langle M - N \rangle\big)/4, \\ [M, N] &= \big([M + N] - [M - N]\big)/4, \end{aligned} \tag{2.6}$$

which are processes of integrable bounded variations.

Most of proofs and comments given in this section are standard and can be found in several classic references, e.g., the reader may check the books by Dellacherie and Meyer [32, Chapters V–VIII], Jacod and Shiryaev [84], Karatzas and Shreve [91], Neveu [137], Revuz and Yor [151], among others.

2.3.3 Local-Martingales

Starting from a (super-/sub-) martingale $(M(t) : t \geq 0)$ relative to a filtration $(\mathcal{F}(t) : t \geq 0)$ and a stopping time τ, we may stop M at time τ and preserve the martingale property, i.e., define a new (super-/sub-) martingale $(M(t \wedge \tau) : t \geq 0)$ relative to the (stopped) filtration $(\mathcal{F}(t \wedge \tau) : t \geq 0)$. Thus, the martingale property is stable under the above stopping time operation and give rise to the following concept.

Definition 2.9 (localization)**.** Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a given filtered space. The term *locally* or *local* is applied to a property relative to a stochastic processes $\{X(t) : t \geq 0\}$ with the understanding that there exists a sequence of stopping times τ_n, with $\tau_n \to \infty$, such that the stopped process $\{X(t \wedge \tau_n) : t \geq 0\}$ satisfies the required property for any n, e.g., we say that $\{X(t) : t \geq 0\}$ is a local-martingale or locally integrable or locally bounded if for any n the process $\{X(t \wedge \tau_n) : t \geq 0\}$ is respectively a martingale or integrable or bounded. The sequence $\{\tau_n : n = 1, 2, \ldots\}$ is called a *reducing sequence* for the process $\{X(t) : t \geq 0\}$. □

For any local sub-martingale we may define a reducing sequence as follows $\tau_n = \inf\{t \in [0, n] : |X(t)| \geq n\}$. Thus, a local sub-martingale is locally of class (D) and Theorem 2.7 applies to the stopped process. Thus the uniqueness yields the following local version of Doob-Meyer decomposition: A local sub-martingale X can be expressed as $X = X(0) + M + A$, where M is a local-martingale and A is a predictable locally integrable monotone increasing process, both null at time zero. The case where the (local) predictable compensator A is continuous is very import. As mentioned above, these are quasi-left continuous processes, which are characterized by the condition either $\mathcal{F}(\tau) = \mathcal{F}(\tau-)$ or $P\{X(\tau) = X(\tau-)\} = 1$ valid for any predictable stopping time τ.

Also remark that not all local-martingales are locally square-integrable martingale. For instance a local-martingale X with locally square-integrable jump process $\delta X = (X(t) - X(t-) : t > 0)$ is actually a locally square-integrable martingale, so that continuous local martingales are locally square-integrable martingale. Hence, for a given local-martingale M the predictable quadratic variation process $\langle M_c \rangle$ is defined as the unique predictable locally integrable monotone increasing process null at time zero such that $M_c^2 - \langle M_c \rangle$ is a (continuous) local-martingale. Next, the (optional) quadratic variation process $[M]$ is defined as

$$[M](t) = \langle M_c \rangle(t) + \sum_{s \leq t} [M(s) - M(s-)]^2, \quad \forall t \geq 0, \tag{2.7}$$

where the second term in the right-hand side is an optional monotone increasing process null at time zero, not necessarily locally integrable (in sense of the localization in Ω defined above).

Nevertheless, if the local-martingale M is also locally square-integrable then the predictable quadratic variation process $\langle M \rangle$ is defined as the unique predictable locally integrable monotone increasing process null at time zero such

that $M^2 - \langle M\rangle$ is a local-martingale. In this case $\langle M\rangle$ is the predictable compensator of $[M]$. Hence, via the predictable compensator we may define the angle-bracket $\langle M\rangle$ when M is only a local-martingale, but this is not actually used. An interesting case is when the predictable compensator process $\langle M\rangle$ is continuous, and therefore $[M] = \langle M\rangle$, which is the case when the initial local-martingale is a quasi-left continuous process. Finally, the optional and predictable quadratic variation processes are defined by coordinates for local-martingale with values in $\mathbb{R}^d$ and even the *co-variation* processes $\langle M, N\rangle$ and $[M, N]$ are defined by orthogonality as in (2.6) for any two local martingales M and N. For instance we refer to Rogers and Williams [153, Theorem 37.8, Section VI.7, pp. 389–391]) where it is proved that $[M, N]$ defined as above (for two local martingales M and N) is the unique optimal process such that $MN - [M, N]$ is a local-martingale where the jumps satisfy $\delta[M, N] = \delta M\, \delta N$.

It is of particular important to estimate the moments of a martingale in term of its quadratic variation. For instance, if M is a square-integrable martingale with $M(0) = 0$ then $\mathbb{E}\{|M(t)|^2\} = E\{[M](t)\} = \mathbb{E}\{\langle M\rangle(t)\}$. If M is only locally square-integrable martingale then

$$\mathbb{E}\{|M(t)|^2\} \le E\{[M](t)\} = \mathbb{E}\{\langle M\rangle(t)\}.$$

In any case, by means of the Doob's maximal inequality (2.3), we deduce

$$\mathbb{E}\{\sup_{0\le t\le T} |M(t)|^2\} \le 4\,\mathbb{E}\{\langle M\rangle(T)\},$$

for any positive constant T, even a stopping time. This can be generalized to the following estimate: for any constant p in $(0, 2]$ there exists a constant C_p depending only on p (in particular, $C_2 = 4$ and $C_1 = 3$) such that for any local-martingale M with $M(0) = 0$ and predictable quadratic variation $\langle M\rangle$ we have the estimate

$$\mathbb{E}\{\sup_{0\le t\le T} |M(t)|^p\} \le C_p\, \mathbb{E}\{\big(\langle M\rangle(T)\big)^{p/2}\}, \tag{2.8}$$

for every stopping time T. If $\langle M\rangle$ is continuous (i.e., M is quasi-left continuous), we can proceed as follows. For a given $r > 0$ and a local-martingale M we set $\tau_r = \inf\{t \ge 0 : \langle M\rangle(t) \ge r^2\}$, with $\tau_r = 0$ if $\langle M\rangle(t) < r^2$ for every $t \ge 0$. Since $\langle M\rangle$ is continuous we have $\langle M\rangle(\tau_r) \le r^2$ and $(M(t \wedge \tau_r) : t \ge 0)$ is a bounded martingale. Thus, for any $c > 0$ we have

$$P(\sup_{t\le T\wedge\tau_r} M^2(t) > c^2) \le \frac{1}{c^2}\mathbb{E}\{M^2(T \wedge \tau_r)\} = \\ = \frac{1}{c^2}\mathbb{E}\{\langle M\rangle(T \wedge \tau_r)\} \le \frac{1}{c^2}\mathbb{E}\{r^2 \wedge \langle M\rangle(T)\}.$$

Hence, for $r = c$ we obtain

$$P(\sup_{t\le T} M^2(t) > c^2) \le P(\tau_c < T) + P(\sup_{t\le T\wedge\tau_c} M^2(t) > c^2) \le \\ \le P(\langle M\rangle(t) > c^2) + \frac{1}{c^2}\mathbb{E}\{c^2 \wedge \langle M\rangle(T)\}.$$

Now, setting $c = r^{1/p}$, integrating in r and using Fubini's theorem we deduce

$$\mathbb{E}\{\sup_{t\le T}|M(t)|^p\} = \int_0^\infty P(\sup_{t\le T} M^2(t) > r^{2/p})\,dr \le$$
$$\le \int_0^\infty \big[P(\sup_{t\le T}\langle M\rangle(t) r^{2/p}) + \\ + \frac{1}{r^{2/p}}\mathbb{E}\{r^{2/p}\wedge\langle M\rangle(T)\}\big]\,dr = \frac{4-p}{2-p}\,\mathbb{E}\big\{\big[\langle M\rangle(T)\big]^{p/2}\big\},$$

so that we can take $C_p = (4-p)/(2-p)$, for $0 < p < 2$. If $\langle M\rangle$ is not continuous, then it takes longer to establish the initial bound in c and r, but the estimate (2.8) follows. This involves Lenglart–Robolledo inequality, see Liptser and Shiryayev [111, Section 1.2, pp. 66–68].

A very useful estimate is the so-called Davis-Burkhölder-Gundy inequality for local-martingales vanishing at the initial time, namely

$$c_p\,\mathbb{E}\{\big([M](T)\big)^{p/2}\} \le \mathbb{E}\{\sup_{t\le T}|M(t)|^p\} \le C_p\,\mathbb{E}\{\big([M](T)\big)^{p/2}\}, \tag{2.9}$$

valid for any $T \ge 0$ and $p \ge 1$ and some universal constants $C_p > c_p > 0$ independent of the filtered space, T and the local martingale M. In particular, we can take $C_1 = C_2 = 4$ and $c_1 = 1/6$. Moreover, a stopping time τ can be used in lieu of the time T and the above inequality holds true.

Remark that a Markov chain (in continuous time) $\{X_t : t \ge 0\}$ (i.e., cad-lag process having values in $\mathbb{N} = \{1, 2, \dots\}$ and transition function $p_{ij}(t-s) = P\{X_t = j \mid X_s = i\}$) with infinitesimal generator $\{q_{i,j} : i, j \in \mathbb{N}\}$ provides a typical example of martingale, i.e., if h is a nonnegative function from $\mathbb{N}\times\mathbb{N}$ into $[0,\infty[$ then the Lévy formula

$$\mathbb{E}\Big\{\sum_{s<r\le t} h(X_{r-}, X_r) \mid X_s\Big\} = \mathbb{E}\Big\{\int_s^t \sum_{j\ne X_r} q_{X_r,j}h(X_r, j)\mathrm{d}r \mid X_s\Big\},$$

with

$$\mathbb{E}\Big\{\int_0^t \sum_{j\ne X_r} q_{X_r,j}h(X_r, j)\mathrm{d}r \mid X_s\Big\} < \infty,$$

implies that the process

$$\sum_{0<r\le t} h(X_{r-}, X_r) - \int_0^t \sum_{j\ne X_r} q_{X_r,j}h(X_r, j)\mathrm{d}r, \quad t \ge 0$$

is a martingale.

An other interesting example is the following: If P and $\tilde P$ are two probabilities defined on the same measurable space $(\Omega, \mathcal{F})$ with a filtration $\mathbb{F} = \{\mathcal{F}_t : t \ge 0\}$, and $\tilde P$ is absolutely continuous with respect to P, then the restrictions P_t

and $\tilde{P}_t$ to $\mathcal{F}_t$ are such that $\tilde{P}$ may be used to define the process $M_t = \mathrm{d}\tilde{P}/\mathrm{d}P$, as the Radon-Nikodym derivative, which results a $(P, \mathcal{F}_t)$-martingale.

Note that when the martingale M is continuous the optional quadratic variation $[M]$ may be replaced with the predictable quadratic variation angle-brackets $\langle M \rangle$. Furthermore, the p-moment estimate (2.8) and (2.9) hold for any $p > 0$ as long as M is a continuous martingale. All these facts play an important role in the continuous time case. By means of this inequality we show that any local-martingale M such that $\mathbb{E}\{|M(0)| + \big(\sup_{t>0}[M](t)\big)^{1/2}\} < \infty$ is indeed a uniformly integrable martingale. For instance, we refer to Kallenberg [88, Theorem 26.12, pp. 524–526], Liptser and Shiryayev [111, Sections 1.5–1.6, pp. 70–84] or Dellacherie and Meyer [32, Sections VII.3.90–94, pp. 303–306], for a proof of the above Davis-Burkhölder-Gundy inequality for (non-necessary continuous) local-martingale and $p \geq 1$, and to Revuz and Yor [151, Section IV.4, pp. 160–171] for continuous local-martingales.

2.3.4 Semi-Martingales

Recall that an adapted (optional or predictable) monotone increasing locally integrable processes A means an adapted (optional or predictable) process such that $A(t) \geq A(s)$ for every $t \geq s$ and there exists a sequence $\{\tau_n\}$ of stopping times satisfying $P(\tau_n < \infty) \to 0$ as $n \to \infty$ and $\mathbb{E}\{A(t \wedge \tau_n)\} < \infty$, for every $n \geq 1$. Similarly, the difference of two such processes is called an adapted (optional or predictable) process with locally integrable bounded variation (or locally finite variation). Now, combining bounded variation processes with martingales processes and localization arguments, we are led to the following definition.

Definition 2.10 (semi-martingale)**.** Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a given filtered space. A *semi-martingale* is a random variable X (P-equivalence class) with values into the canonical space $D([0,\infty), \mathbb{R}^d)$ which can be expressed as $X = X(0) + A^+ - A^- + M$, where $X(0)$ is a $\mathbb{R}^d$-valued $\mathcal{F}(0)$-measurable random variable, A^+, A^-, are adapted monotone increasing locally integrable processes and M is a local-martingale, satisfying $A^+(0) = A^-(0) = M(0) = 0$. Thus, $A = A^+ - A^-$ is a process with locally integrable bounded variation. □

Based on the uniqueness of Doob-Meyer decomposition, a local martingale null at time zero with locally bounded variation is identically zero if it is predictable (in particular if it is continuous or deterministic). Since there are non-constant martingales with locally bounded variation paths (e.g., purely discontinuous local-martingales), the decomposition in the definition of semi-martingale is not necessarily unique. Usually, the above definition of semi-martingale is known as *special semi-martingale*, but this is sufficiently general for our study. These (special) semi-martingales include a natural condition of local integrability (local first moment) on the bounded variation part (the adapted process A). The equivalent of this local integrability property, applied to the martingale part (the process M), is actually a necessary condition for martingale. Unless explicitly mentioned, we drop the adjective *special* in using of the name semi-martingale but we may call *general* or *non-special* semi-martingale

when the process A in the above definition may not be locally integrable. Note that the only reason why the process A may not be integrable is because of the *large jumps*. It is clear then that a (special) semi-martingale is the difference of two local sub-martingales. Moreover, a local sub-martingale zero at the origin can be written in a unique manner as the sum of a local martingale and an increasing predictable process, both zero at the origin. Thus, the concept of special semi-martingales is equivalent to that of quasi-martingales, e.g. see Kallenberg [88], Protter [149].

Theorem 2.11. *Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a filtered space. Then every semi-martingale $X = (X(t) : t \geq 0)$ admits the unique canonical decomposition $X = X(0) + A + M$, where A is a predictable process with locally integrable variation and M is a local-martingale, both satisfying $A(0) = M(0) = 0$. Moreover, the quadratic variation $[M]$ defined by (2.7) is the unique optional monotone increasing process such that $M^2 - [M]$ is a local-martingale and the jumps $\delta[M] = \delta M\, \delta M$, where $\delta M(t) = M(t) - M(t-)$. Furthermore, the processes $\sqrt{[M]}$ (by coordinates) and $\sup\{|X(s) - X(0)| : 0 \leq s \leq t\}$ are locally integrable. If the semi-martingale X is quasi-left continuous (i.e., either $P\{X(\tau-) = X(\tau)\} = 1$ or $\mathcal{F}(\tau-) = \mathcal{F}(\tau)$ for every predictable stopping time τ), then the process A in the semi-martingale decomposition is continuous.* □

Note that the local-martingale appearing in the above expression has a unique representation $M = M_c + M_d$, where M_c (respectively M_d) is the continuous (purely discontinuous) part. Also, if M is a local-martingale with $M(0) = 0$ and $[M]$ denotes its (optional) quadratic variation (or characteristic) then for any $t > 0$ and any sequence of partitions $(\pi_k : k = 1, 2, \ldots)$, with π_k of the form $(0 = t_0 < t_1 < \cdots < t_n = t)$ and the mesh (or norm) of π_k going to zero we have $\mathrm{var}_2(M, \pi_k) \to [M](t)$ in probability as $k \to 0$, see Liptser and Shiryayev [111, Theorem 1.4, pp. 55–59].

Semi-martingales are stable under several operations, for instance under stopping times operations and localization, see Jacod and Shiryaev [84, Theorem I.4.24, pp. 44-45].

Observe that a process X with independent increments (i.e., which satisfies for any sequence $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n$ the random variables $\{X(t_0), X(t_1) - X(t_0), \ldots, X(t_n) - X(t_{n-1})\}$ are independent) is not necessarily a semi-martingale, e.g., deterministic cad-lag process null at time zero is a process with independent increments, but it is not a general semi-martingale (not necessarily special!) unless it has finite variation.

The only reason that a semi-martingale may not be special is essentially the non-integrability of large jumps. If X is a semi-martingale satisfying $|X(t) - X(t-)| \leq c$ for any $t > 0$ and for some positive (deterministic) constant $c > 0$, then X is special. Indeed, if we define $\tau_n = \inf\{t \geq 0 : |X(t) - X(0)| > n\}$ then $\tau_n \to \infty$ as $n \to \infty$ and $\sup_{0 \leq s \leq \tau_n} |X(s) - X(0)| \leq n + c$. Thus X is a special semi-martingale and its canonical decomposition $X = X(0) + A + M$ satisfies $|A(t) - A(t-)| \leq c$ and $|M(t) - M(t-)| \leq 2c$, for any $t > 0$.

Similar to (2.9), another very useful estimate is the Lenglart's inequality: If X and A are two cad-lag adapted processes such that A is monotone increas-

ing and $\mathbb{E}\{|X_\tau|\} \le \mathbb{E}\{A_\tau\}$, for every bounded stopping time τ, then for every stopping time τ and constants $\varepsilon, \eta > 0$ we have

$$P\Big\{\sup_{t\le\tau}|X_t| \ge \varepsilon\Big\} \le \frac{1}{\varepsilon}\Big[\eta + \mathbb{E}\Big\{\sup_{t\le\tau}(A_t - A_{t-})\Big\}\Big] + P\big\{A_\tau \ge \eta\big\}, \tag{2.10}$$

and if A is also predictable then the term with the jump $(A_t - A_{t-})$ is removed from the above estimate. A simple way to prove this inequality is first to reduce to the case where the stopping time τ is bounded. Then, defining $\theta = \inf\{s \ge 0 : |X_s| > \varepsilon\}$ and $\varrho = \inf\{s \ge 0 : A_s > \eta\}$, since A is not necessarily continuous, we have $A_{\varrho-} \le \eta$ and

$$\begin{aligned} &A_{\theta\wedge\tau\wedge\varrho} \le \eta + \sup_{t\le\tau}(A_t - A_{t-}),\\ &\Big\{\sup_{t\le\tau}|X_t| > \varepsilon\Big\} \subset \big\{\theta \le \tau < \varrho\big\} \cup \big\{A_\tau \ge \eta\big\}. \end{aligned}$$

Hence, by means of the inequality

$$P\big\{\theta \le \tau < \varrho\big\} \le P\big\{|X_{\theta\wedge\tau\wedge\varrho}| \ge \varepsilon\big\} \le \frac{1}{\varepsilon}\,\mathbb{E}\big\{A_{\theta\wedge\tau\wedge\varrho}\big\},$$

we obtain (2.10). However, if A is predictable then ϱ is a predictable time, and there is a sequence of stopping times $(\varrho_k : k \ge 1)$ converging to ϱ such that $\varrho_k < \varrho$ if $\varrho > 0$. Thus $A_{\theta\wedge\tau\wedge\varrho} \le A_{\varrho-}$ almost surely, which completes the argument. Given a local-martingale M, a good use of (2.10) is when the predictable compensator process $\langle M\rangle$ is continuous, and therefore $[M] = \langle M\rangle$, so that

$$P\Big\{\sup_{t\le\tau}|M_t| \ge \varepsilon\Big\} \le \frac{\eta}{\varepsilon^2} + P\big\{\langle M\rangle_\tau \ge \eta\big\}, \quad \forall \varepsilon, \eta > 0, \tag{2.11}$$

for any stopping time τ. Actually, this is the case of a quasi-left continuous local-martingale M.

In short, cad-lag (quasi-continuous) local-martingales could be expressed as the sum of (1) a local-martingales with continuous paths, which are referred to as continuous martingales, (2) a purely discontinuous local-martingale. The semi-martingales add an optional process with locally integrable bounded variation, which is necessarily predictable for quasi-continuous semi-martingales, and quasi-continuous means that the cad-lag semi-martingale is also continuous in probability, i.e., there is no deterministic jumps.

For a comprehensive treatment with proofs and comments, the reader is referred to the books by Dellacherie and Meyer [32, Chapters V–VIII], Liptser and Shiryayev [111, Chapters 2–4, pp. 85–360], Rogers and Williams [153, Section II.5, pp. 163–200], among others. For instances, a treatment of semi-martingale directly related with stochastic integral can be found in He et al. [68] and Protter [149], among others.

2.4 Gaussian Noises

The idea of a *noise* is the extension of a sequence of independent identically distributed random variables to the *continuous* context, where the two typical cases are Gaussian and Poisson noises. First, let us recall that we can build a (complete) probability space $(\Omega, \mathcal{F}, P)$, e.g., P is the Lebesgue measure on $(0,1)$, with the property that for any countable family of distributions $\{F_i\}$ on $\mathbb{R}^d$ there exists a family of independent random variables $\{\xi_i\}$ such that ξ_i is distributed accordingly to F_i, e.g., see Kallenberg [88, Theorem 3.19, pp. 55–57]. In particular, there exist two independent countable families of normally and exponentially distributed random variables, with parameters prescribed a priori, in some probability space $(\Omega, \mathcal{F}, P)$.

However, the situation is complicate in the case of a white noise $\dot{w} = \{\xi_n : n \geq 1\}$, where the independent random variables ξ_n are standard normally distributed.

2.4.1 The White Noise

The simplest construction of a Wiener process with the L^2 orthogonal theory begins with an orthogonal basis $\{\varphi_n : n \geq 0\}$ in $L^2(]0,T[)$, and a sequence $\{\xi_n : n \geq 0\}$ of independent standard normally distributed random variables, which forms also an orthonormal system in $L^2 = L^2(\Omega, \mathcal{F}, P)$. Each function φ in $L^2(]0,T[)$ can be written as a converging orthogonal series

$$\varphi(s) = \sum_n (\varphi, \varphi_n)\varphi_n(s), \qquad \text{a.e.} \quad s \in]0,T[,$$

where $(\cdot,\cdot)$ denotes the scalar product in $L^2(]0,T[)$, and $(\varphi,\varphi) = \sum_n |(\varphi,\varphi_n)|^2$. Thus the mapping $\varphi \mapsto w(\varphi) = \sum_n (\varphi,\varphi_n)\xi_n$ is an isometry from $L^2(]0,T[)$ into L^2 such that $w(\varphi)$ is a Gaussian random variable with $\mathbb{E}\{w(\varphi)\} = 0$ and $\mathbb{E}\{w(\varphi)w(\varphi')\} = (\varphi,\varphi')$, for every φ and φ' in $L^2(]0,T[)$. Hence, $\mathbb{1}_{]a,b]} \mapsto w(\mathbb{1}]a,b])$ could be regarded as a L^2-valued measure and $w(\varphi)$ is the integral. In particular, the orthogonal series

$$w(t) = w(\mathbb{1}_{]0,t[}) = \sum_n \xi_n \int_0^t \varphi_n(s)\mathrm{d}t, \qquad \forall t \geq 0$$

is converging in L^2, and

$$\mathbb{E}\{|w(t)|^2\} = \sum_n \Big| \int_0^t \varphi_n(s)\mathrm{d}s \Big|^2 = \int_0^T \mathbb{1}_{]0,t[}(s)\mathrm{d}s = t, \qquad \forall t \geq 0,$$

i.e., the above series yields a Gaussian process $t \mapsto w(t)$, which is continuous in L^2 and satisfies $\mathbb{E}\{w(t)\} = 0$ and $\mathbb{E}\{w(t)w(s)\} = t \wedge s$, for every t, s in $[0,T]$. Conversely, if a Wiener process $\{w(t) : t \geq 0\}$ is given then we can reconstruct the sequence $\{\xi_n : n \geq 0\}$ by means of the square-wave orthogonal

basis $\{\varphi_n : n \geq 0\}$, where the integral $w(\varphi_n)$ reduces to a finite sum, namely,

$$\xi_n = w(\varphi_n) = \sum_{i=1}^{2^n} (-1)^{i-1} T^{-1/2} \big[w(t_{i,n}) - w(t_{i-1,n}) \big],$$

with $t_{i,n} = i2^{-n}T$, $i = 0, \ldots, 2^n$, $n \geq 0$. Finally, the almost surely continuity of the path requires either taking a particular version or using some martingale inequality. It is clear that in this construction, there is not a precise way to handle which or how random variables are involves in $w(s)$ when s belongs to $[0, t]$. However, a small change along this previous argument makes the trick, as we see in what follows.

The closed linear subspace H of $L^2(\Omega, \mathcal{F}, P)$ generated by the orthonormal sequence $\{\xi_n : n \geq 0\}$ is called a white noise (or Gaussian) space. If $\mathcal{F}_\xi$ is the σ-algebra generated by the random variables $\{\xi_n : n \geq 0\}$ and the null sets in $\mathcal{F}$, then any random variable x in $L^2(\Omega, \mathcal{F}_\xi, P)$ which is independent of H is actually a constant, i.e., $x = \mathbb{E}\{x\}$. It is also clear that the Hilbert space H can be identified with the $L^2(]0, T[)$ via the above isometry. As discussed later on (see Section 2.6.1), based on the Hermit polynomials $h_n(x)$ we can construct an orthonormal basis

$$\Xi_n = \prod_i h_{n_i}(\xi_i)\sqrt{n_i!}, \quad n = (n_i), \text{ only a finite number nonzero,} \tag{2.12}$$

for the space $L^2(\Omega, \mathcal{F}_\xi, P)$, which can be written as an infinite orthogonal sum of subspaces. It is clear that $\mathcal{F}_\xi$ is equal to $\mathcal{F}_w$, the σ-algebra generated by the random variables $\{w_t : t > 0\}$, it seems not obvious how to use the above construction to get an orthonormal basis corresponding to the σ-algebra $\mathcal{F}(t)$ generated by the random variables $\{w_s : 0 < s \leq t\}$.

Sometimes, another Hilbert space H is preferred instead of $L^2(]0, T[)$, i.e., we may begin with an orthogonal basis $\{e_n : n \geq 1\}$ in H and a sequence $\{\xi_n : n \geq 1\}$ of independent standard normally distributed random variables (after some adaptation, perhaps, with values in $\mathbb{R}^d$ or in some infinite dimension Banach space), which forms also an orthonormal system in $L^2 = L^2(\Omega, \mathcal{F}, P)$. Each function h in H can be written as a converging orthogonal series $h = \sum_n (h, e_n)_H e_n$, and $(h, h)_H = \sum_n |(h, e_n)_H|^2$. Thus the mapping $h \mapsto w(h) = \sum_n (h, e_n)_H \xi_n$ is an isometry from H into L^2 such that $w(h)$ is a Gaussian random variable with $\mathbb{E}\{w(h)\} = 0$ and $\mathbb{E}\{w(h)w(h')\} = (h, h')_H$, for every h and h' in H. Of particular interest in the case where $H = L^2(X, \mathcal{X}, \mu)$ for a σ-finite measure space $(X, \mathcal{X}, \mu)$. Choose a semi-ring $\mathcal{K}$ of $\mathcal{X}$ with finite measure, i.e., $\mu(K) < \infty$, for every K in $\mathcal{K}$ to consider the map $\mathbb{1}_K \mapsto w(\mathbb{1}_K)$ as a L^2-valued measure and $w(h)$ is the integral. Moreover, we may single-out a time variable, i.e., replace X and μ with $X \times]0, T[$ and $\mu \times \mathrm{d}t$, and get an orthonormal system of the form $e_i \varphi_n$. Hence, by relabeling the sequence $\{\xi_{i,n} : i, n \geq 1\}$ the orthogonal series

$$w_i(t) = w(e_i \mathbb{1}_{]0,t[}) = \sum_n \xi_{i,n} \int_0^t \varphi_n(s)\mathrm{d}t, \qquad \forall t \geq 0,\ i = 1, 2, \ldots,$$

is converging in L^2, and

$$\mathbb{E}\{|w_i(t)|^2\} = \sum_n \Big| \int_0^t \varphi_n(s)\mathrm{d}s \Big|^2 = \int_0^T \mathbb{1}_{]0,t[}(s)\mathrm{d}s = t, \qquad \forall t \geq 0,$$

i.e., the above series yields Gaussian processes $t \mapsto w_i(t)$, which are continuous in L^2 and satisfy $\mathbb{E}\{w_i(t)\} = 0$, $\mathbb{E}\{w_i(t)w_i(s)\} = t \wedge s$, for every t, s in $[0, T]$, and the process $(w_i(t) : t \geq 0)$ is independent of $(w_j(t) : t \geq 0)$ for every $i \neq j$. This construction is referred to as a general Wiener noise or white noise (random) measure.

2.4.2 The White Noise (details)

Formally, assume that the Hilbert space $L^2 = L^2(\Omega, \mathcal{F}, P)$ contains a sequence $\{e_{i,n} : i = 1, 2, \ldots, 4^n,\ n \geq 1\}$ of independent standard normally distributed random variables, and set $\{e_r^n = e_{i,n} : r \in R_n\}$, indexed in r belonging to the dyadic numbers $R = \bigcup_n R_n$, with $R_n = \{r = i2^{-n} : i = 1, 2, \ldots, 4^n\}$. Because each $e_{i,n}$ has zero mean and are independent of each other, the sequence is orthogonal in L^2, actually, it is an orthonormal system since all variances are equal to 1. To simplify notation, assume that $\mathcal{F}$ is the sub σ-algebra generated by the sequence of random variables $\dot{w} = \{e_r^n : r \in R_n,\ n \geq 1\}$ (and all null sets). The closed linear subspace H of $L^2(\Omega, \mathcal{F}, P)$ generated by the elements in $\dot{w}$ is called a white noise (or Gaussian) space. The system $\dot{w}$ is the ideal expression of the *white noise*, which is the formal derivative of the Wiener process w.

To given details of a simple construction a Wiener process $\{w_t : t > 0\}$ as the integral of (the function $s \mapsto \mathbb{1}_{s \leq t}$ with respect to) the system $\dot{w}$, we make use of the dyadic property $t = \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t}$[1] to define the random variable

$$w_t = \sum_n 2^{-n} \sum_{i=1}^{4^n} e_{i,n} \mathbb{1}_{i2^{-n} \leq t}, \tag{2.13}$$

as a convergence series in $L^2(\Omega, \mathcal{F}, P)$, for every $t > 0$. Indeed, regard the expression as an orthogonal series expansion, and set $w_0 = 0$, for any $t \geq s \geq 0$, to have

$$\begin{aligned} \mathbb{E}\{|w_t - w_s|^2\} &= \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{E}\{|e_{i,n}|^2\} \mathbb{1}_{s < i2^{-n} \leq t} = \\ &= \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{s < i2^{-n} \leq t} = (t - s). \end{aligned}$$

Thus, $t \mapsto w_t$ provides a L^2-norm continuous random process satisfying (a) w_t is a Gaussian random variable with $\mathbb{E}\{w_t\} = 0$ and $\mathbb{E}\{|w_t|^2\} = t$, and (b) w_s is

[1] if $t = k2^{-m} = (k2^{n-m})2^{-n}$, $1 \leq k \leq 4^m$ then $k2^{n-m} \leq 4^n$, $\mathbb{1}_{i2^{-n} \leq t} = 1$ if and only if $i = 1, \ldots, k2^{n-m}$, which yields $\sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t} = k2^{n-m} = t2^n$ if $k2^{n-m} = t2^n \geq 1$.

independent of $w_t - w_s$ for every $t > s$. If a parameter (variance) $a > 0$ is included then the Gaussian random variables $\{\xi_n : n \geq 0\}$ and $\{e_r^n : r \in R_n,\ n \geq 1\}$ have variance a and $\mathbb{E}\{|w_t|^2\} = ta$. Moreover

$$P\{w_t \in \mathrm{d}x\} = \mathrm{e}^{-|x|^2/(2ta)}\mathrm{d}x \quad \text{and} \quad \mathbb{E}\{\mathrm{e}^{\mathrm{i}\xi w_t}\} = \mathrm{e}^{-ta|\xi|^2/2}$$

are the transition density and the characteristic function.

Next, to check that the process $\{w_t : t \geq 0\}$ has a continuous version, we recall that $w_t - w_s$ is a Gaussian variable with zero mean and variance $|t - s|$, so that we deduce $\mathbb{E}\{|w_t - w_s|^4\} = 3|t - s|^2$, and therefore, we are allowed to select a continuous version.

The concept of *stopping time* relative to a white noise $\dot{w}$ can be expressed as preserving orthogonality, i.e., a $[0, \infty]$-valued random variable τ is called $\dot{w}$-stopping time if $\{e_{i,n}\mathbb{1}_{i2^{-n} \leq \tau}\}$ (or equivalently $\{e_r^n \mathbb{1}_{r \leq \tau}\}$) remains an orthogonal system, for every $t > 0$. For instance, if τ is a $\dot{w}$-stopping time then the formula (2.13) shows that $\mathbb{E}\{|w(t \wedge \tau)|^2\} = \mathbb{E}\{t \wedge \tau\}$ as expected.

Note that if x belongs to H then

$$\mathbb{E}\{xw_t\} = \sum_n 2^{-n} \sum_{i=1}^{4^n} \mathbb{E}\{xe_{i,n}\}\mathbb{1}_{i2^{-n} \leq t},$$

and by taking $r = k2^{-m}$ with k some odd integer number between 1 and 4^m, we deduce $\mathbb{E}\{x(w_r - w_{r'})\} \to 2^{-m}\mathbb{E}\{xe_{k,m}\}$ as $r' \uparrow r$. This proves that any x in H which is orthogonal to any element in $\{w_t : t \geq 0\}$ is also orthogonal to any element in $\{e_{i,n} : i = 1, \ldots, 4^n,\ n \geq 1\}$, i.e, the white noise subspace H is indeed the closed linear span of $\{w_t : t \geq 0\}$. Therefore the projection operator

$$\mathbb{E}\{x \mid w_s,\ s \leq t\} = \sum_n \sum_{i=1}^{4^n} \mathbb{E}\{x\, e_{i,n}\}\, e_{i,n}\mathbb{1}_{i2^{-n} \leq t}, \tag{2.14}$$

is valid for every x in H. By means of the Hermit polynomials and $\{e_{i,n} : i2^{-n} = r \in R,\ r \leq t\}$ we can construct an orthonormal basis for $L^2(\Omega, \mathcal{F}(t), P)$ as in (2.12), which yields an explicit expression for the conditional expectation with respect to $\mathcal{F}(t)$, for any square-integrable random variable x. In this context, remark that we have decomposed the Hilbert space H into an orthogonal series $(n \geq 1)$ of finite dimensional subspaces generated by the orthonormal systems $\dot{w}^n = \{e_r^n : r \in R_n\}$.

2.4.3 The White Noise (converse)

Conversely, if a Wiener process $\{w_t : t \geq 0\}$ is given then the random variables $\bar{e}_{i,n} = 2^{n/2}\big[w_{i2^{-n}} - w_{(i-1)2^{-n}}\big]$, are identically standard normally distributed, and the system $\{\bar{e}_{i,n} : i = 1, \ldots, 4^n\}$ is independent, but $\{\bar{e}_{i,n} : i \geq 1,\ n \geq 1\}$ is not fully independent, i.e., $r = i2^{-n} = 2i2^{-n-1}$ yields $\sqrt{2}\bar{e}_{i,n} = \bar{e}_{2i,n+1} + \bar{e}_{2i-1,n+1}$, which produces correlations. In this case, the representation (2.13)

takes the form

$$w_t = \lim_n \Big[2^{-n/2} \sum_{i=1}^{4^n} \bar{e}_{i,n} \mathbb{1}_{i2^{-n} \le t} \Big],$$

or equivalently $w_t = \lim_n w_{k_n(t)2^{-n}}$, where $k_n(t)2^{-n} \le t < (k_n(t)+1)2^{-n}$, $1 \le k_n(t) \le 4^n$. Moreover, the projection operator becomes

$$\mathbb{E}\{x \mid w_s,\, s \le t\} = \lim_n \sum_{i=1}^{4^n} \mathbb{E}\{x\, \bar{e}_{i,n}\}\, \bar{e}_{i,n} \mathbb{1}_{i2^{-n} \le t},$$

which can be proved to be convergent (as a particular case of a stochastic integral considered later on) in L^2.

To recover a white noise $\dot{w} = \{e_r^n : r \in R_n,\, n \ge 1\}$ as a dyadic sequence of independent standard normally distributed random variables from a given Wiener process $w = \{w_t : t \ge 0\}$; we may use the square wave procedure, namely, for $i = 1, 2, \ldots$ consider the Haar-type functions $f_i(s) = \mathbb{1}_{2i-1<2s\le 2i} - \mathbb{1}_{2(i-1)<2s\le 2i-1}$ and $f_{i,n}(s) = 2^{-n/2} f_i(s2^n)$, for $n \ge 0$. By construction, if $n \ge m$ then $f_{i,n} f_{j,m} = 0$ except for i within $(j-1)2^{n-m}+1$ and $j2^{n-m}$, and moreover, $\{f_{i,n}\}$ is an orthonormal system in $L^2(]0,\infty[)$. Therefore

$$e^n_{i2^{-n}} = w(f_{i,n}) = 2^{-n/2} \big[w_{(i-1)2^{-n}} - 2w_{(2i-1)2^{-n-1}} + w_{i2^{-n}} \big], \tag{2.15}$$

for $i = 1, \ldots, 4^n$, $n \ge 1$, define a white noise which produces another Wiener process via (2.13), also given by the stochastic integral

$$\bar{w}_t = \sum_{n=1}^{\infty} 2^{-n} \sum_{i=1}^{4^n} w(f_{i,n}) \mathbb{1}_{i2^{-n} \le t} = \int_0^t f_T(s) \mathrm{d}w_s \quad \forall T \ge t > 0,$$

where the real-valued function

$$f_t = \sum_{n=1}^{\infty} 2^{-n} \sum_{i=1}^{4^n} f_{i,n} \mathbb{1}_{i2^{-n} \le t}, \qquad \int_0^{\infty} |f_t(s)|^2 \mathrm{d}s = t, \quad \forall t > 0,$$

is defined as an orthogonal series expansion in $L^2(]0,\infty[)$. Remark that $f_t(s) = f_T(s)$ a.e. s in $(0,t)$ for every $t \le T$, and $f_t(s) = 0$ a.e. for s in (t,∞). Actually, for the probability measure $\mathrm{d}t/T$ defined on Borel σ-algebra on $]0,T[$, the family of random variables $\{\sqrt{T} f_t : t \in [0,T]\}$ is a Wiener process.

Furthermore, if a factor 2^{k-1} is added to the orthogonal series (2.13) then we may begin the sum with $n = k$ instead of $n = 1$. Comparing with the initial isometry given via orthonormal sequences, we note that the orthonormal system $\{f_{i,n}\}$ can be completed to be a basis by adding the functions $\tilde{f}_i(s) = \tilde{f}_{i,0}(s) = \mathbb{1}_{(i-1)<s\le i}$, for $i = 1, 2 \ldots$. Indeed, it suffices to check that $1/2\{\tilde{f}_{i,0}\} \pm 1/2\{f_{i,0}\}$ yields $\{\tilde{f}_{i,1}(s) = \mathbb{1}_{i-1<2s\le i}\}$, and $1/2\{\tilde{f}_{i,1}\} \pm 1/2\{f_{i,1}\}$ yields $\{\tilde{f}_{i,2}(s) = \mathbb{1}_{i-1<4s\le i-1}\}$ and so on. Thus, the isometry $w(f_{i,n}) = e_{i,n}$ and $w(\tilde{f}_i) = \tilde{e}_i$

mapping the basis $\{f_{i,n} : i = 1, \dots, 4^n, n \geq 0\} \cup \{\tilde{f}_i : i \geq 1\}$ in $L^2(]0,\infty[)$ into an orthornormal system $\{e_{i,n} : i = 1, \dots, 4^n, n \geq 0\} \cup \{\tilde{e}_i : i \geq 1\}$ in L^2 produces an expression very similar to (2.13), namely,

$$\begin{aligned} \tilde{w}_t &= \sum_{i=1}^{\infty} \tilde{c}_i(t)\tilde{e}_i + \sum_{n=0}^{\infty}\sum_{i=1}^{4^n} c_{i,n}(t)e_{i,n}, \\ \tilde{c}_i(t) = \int_0^t \tilde{f}_i(s)\mathrm{d}s, &\qquad c_{i,n}(t) = \int_0^t f_{i,n}(s)\mathrm{d}s, \end{aligned} \tag{2.16}$$

where the first series in i is a finite sum for each fixed $t > 0$, and the series in n reduces to a finite sum if $t = j2^{-m}$ for some $m \geq 0$ and $j \geq 1$. Essentially based on Borel-Cantelli Lemma and the estimates

$$q_n = \max_{t\geq 0} \sum_{i=1}^{4^n} |c_{i,n}(t)e_{i,n}| = 2^{-n/2} \max_{i=1,\dots,4^n} |e_{i,n}|,$$

$$P\big(|e_{i,n}| > a\big) \leq \frac{2}{\sqrt{\pi}}\mathrm{e}^{-a^2/2}, \qquad P\Big(\max_{i=1,\dots,4^n} |e_{i,n}| > a\Big) \leq 4^n \frac{2}{\sqrt{\pi}}\mathrm{e}^{-a^2/2},$$

$$P\big(q_n > \theta(2^{-n}\ln 8^n)^{1/2}\big) \leq \frac{2}{\sqrt{\pi}} 4^{n(1-\theta^2)}, \qquad \theta > 1,$$

a more careful analysis shows the uniform convergence on any bounded time interval, almost surely. Actually, this is almost Ciesielski-Levy's construction as described in McKean [119, Section 1.2, pp. 5–8] or Karatzas and Shreve [91, Section 2.3, pp. 56–59]. Remark that with the expression (2.16), we cannot easily deduce a neat series expansion like (2.14) for the projection operator, i.e., since the functions $\{c_{i,n}\}$ have disjoint support only as i changes, for a fixed $t > 0$, the orthogonal systems $\{\tilde{c}_i(s)\tilde{e}_i,\, c_{i,n}(s)e_{i,n} : s \leq t,\ i,n\}$ and $\{\tilde{c}_i(s)\tilde{e}_i,\, c_{i,n}(s)e_{i,n} : s > t,\ i,n\}$ are not orthogonal to each other, as in the case of the orthogonal series expansion (2.13). In the context of the orthogonal series expansion (2.16), the series

$$\langle \dot{\tilde{w}}, \phi\rangle = \sum_{i=1}^{\infty} \tilde{e}_i \langle \tilde{f}_i, \phi\rangle + \sum_{n=0}^{\infty}\sum_{i=1}^{4^n} e_{i,n}\langle f_{i,n}, \phi\rangle, \qquad \forall \phi \in \mathcal{S}(]0,\infty[),$$

could be referred to as white noise, the derivative in the sense of Schwartz distribution of a Wiener process, meaningful only as a generalized process.

On the other hand, note that we cannot take a fractional derivative to recover a white noise, i.e., the limit $(t-r)^{-1/2}\big[w_t - w_r\big] \to e_r$ as $t \downarrow r$ for a particular sequence of $\{t\}$. Indeed, if $r < t < s$ then $w_t - w_r$ and $w_s - w_t$ are independent, and hence

$$\mathbb{E}\Big\{\Big|\frac{w_s - w_r}{\sqrt{s-r}} - \frac{w_t - w_r}{\sqrt{t-r}}\Big|^2\Big\} = 2 - 2\mathbb{E}\Big\{\Big(\frac{w_s - w_r}{\sqrt{s-r}}\Big)\Big(\frac{w_t - w_r}{\sqrt{t-r}}\Big)\Big\} = 2\Big(1 - \frac{\sqrt{t-r}}{\sqrt{s-r}}\Big).$$

Thus, if $\alpha_n \to 0$ then $(1 - \sqrt{\alpha_{n+k}}/\sqrt{\alpha_n})$ does not converges to 0 as $n, k \to \infty$, which implies that the sequence $(w_{r+\alpha_n} - w_r)/\sqrt{\alpha_n}$ cannot be a Cauchy sequence in L^2. Therefore, we may have a subsequence such that $(w_{r+\alpha_n} - w_r)/\sqrt{\alpha_n} \to e_r$ weakly in L^2, but $E\{|e_r|^2\} \neq 1$, since otherwise, the sequence would converge in the L^2 norm.

2.4.4 The White Noise (another)

With the previous observation in mind, consider a countable family $\{e_r^n\}$ of standard normally distributed random variables, indexed for r in the diadic numbers $R = \bigcup_n R_n = \{r = i2^{-n} : i = 1, \dots, 4^n\}$ as early; but, we assume only that the finite family $\{e_r^n : r \in R_n\}$ is independent, for each fixed $n \geq 1$. Based on the dyadic property $2^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t} = \max\{r : r \in R_n,\, r \leq t\} \to t$, define the sequence of normally distributed random variables $w_0^n = 0$ and

$$w_t^n = 2^{-n/2} \sum_{i=1}^{4^n} e_{i2^{-n}}^n \mathbb{1}_{i2^{-n} \leq t}, \quad \forall t > 0. \tag{2.17}$$

Note that $\mathbb{E}\{w_t^n\} = 0$ and $\mathbb{E}\{|w_r^n|^2\} = r$, for every r in R. Thus, the classic Central Theorem shows that $\{w_r^n : n \geq 1\}$ converges in law and $\lim_n \mathbb{E}\{|w_t^n - w_r^n|^2\} = t - r$, for any $t > r > 0$. Since, for Gaussian variables with zero-mean we have the equality

$$\mathbb{E}\{|w_r^n - w_s^n|^4\} = 3\big(\mathbb{E}\{|w_r^n - w_s^n|^2\}\big)^2 = 3|r - s|^2, \quad \forall r, s \in R,$$

this construction yields a Wiener measure W, i.e., a probability measure on $\Omega = C([0, \infty[)$ such that the coordinate functions $\omega : \Omega \mapsto \omega(t) = w_t(\omega)$ define a Wiener process.

Contrary to the previous series (2.13), the convergence in L^2 of the whole sequence $\{w_t^n : n \geq 1\}$ is not automatically insured, we need to assume that the system $\{e_r^n : r \in R_n\}$ is compatible with the diadic numbers in the sense that without ambiguity we can remove the super-index n in e_r^n and use the notation $\{e_r : r \in R_n\}$. Indeed, e.g., by compactness, we can extract a subsequence $\{n_k\}$ such that $w_r^{n_k} \to w_r$ in L^2, for every r in R (i.e., only the random variables $\{e_r^{n_k} : r \in R_{n_k},\, k \geq 1\}$ were used), but another convergent subsequence may have another limit (which uses another subset of random variables $\{e_r^n\}$). This previous argument can not used if we impose the condition $e_r^n = e_r^m$, for every n, m and r, i.e., compatibility with the dyadic numbers. Moreover, under this assumption, we can single out all terms in the sum defining w_r^n using $\{e_r : r \in R_n\} \setminus \{e_r : r \in R_{n-1}\}$ to obtain the relation $\sqrt{2} w_r^n = w_r^{n-1} + v_r^n$, for r in R_{n-1}, with v_r^n being a normally distributed random variable independent of w_r^n satisfying $\mathbb{E}\{v_r^n\} = 0$ and $\mathbb{E}\{|v_r^n|^2\} = r$. By induction, we deduce

$$w_r^n = 2^{(n(r)-n-1)/2} w_r^{n(r)-1} + \sum_{k=n(r)}^{n} 2^{(k-n-1)/2} v_r^k, \tag{2.18}$$

where $n(r) = \inf\{n \geq 1 : r \in R_n\}$, $w_r^0 = 0$ and $\{w_r^{n(r)-1}, v_r^{n(r)}, \dots, v_r^n\}$ is an orthogonal system. This implies that the whole sequence $\{w_r^n : n \geq 1\}$ converges in L^2, i.e., the limit

$$w_t = \lim_n \Big[2^{-n/2} \sum_{i=1}^{4^n} e_{i2^{-n}} \mathbb{1}_{i2^{-n} \leq t}\Big], \quad \forall t > 0 \tag{2.19}$$

exits in L^2, almost as an orthogonal series expansion if r belongs to R. Anyway, only the random variables $\{e_r^n : r \in R_n, n \geq 1, r \leq t\}$ intervene in getting $\{w_s : s \leq t\}$, and the projector operator has the form

$$\mathbb{E}\{x \mid w_s,\, s \leq t\} = \lim_n \sum_{i=1}^{4^n} \mathbb{E}\{x\, e_{i2^{-n}}\}\, e_{i2^{-n}} \mathbb{1}_{i2^{-n} \leq t}, \tag{2.20}$$

as a limit in L^2 of almost an orthogonal series expansion, valid for any x in the closed linear span of $\{w_t : t \geq 0\}$. Hermit polynomials are needed to get a series expansion for any x in $L^2(\Omega, \mathcal{F}, P)$.

Let us summarize the main points proved above:

Proposition 2.12. *Let $\{e_{i,n} : i = 1, \dots, 4^n, n \geq 1\}$ be a countable family of identically distributed random variables with $\mathbb{E}\{e_{i,n}\} = 0$ and $\mathbb{E}\{|e_{i,n}|^2\} = 1$, and such that each $\dot{w}^n = \{e_{i,n} : i = 1, \dots, 4^n\}$ is a sub-family of independent random variables. For convenience we may take all $e_{i,n}$ normally distributed, but this is not necessary.*

(a) *If $\dot{w}^1, \dots, \dot{w}^n, \dots$ are independent then the orthogonal series expansion (2.13) yields a Wiener process, and the conditional expectation operator (2.14).*

(b) *If $e_r^n = e_{i,n}$, with $r = i2^{-n}$, $i = 1, \dots, 4^n$, then the limit of the expression (2.17) exists in law and defines a Wiener measure.*

(c) *If, besides the conditions of (b), also we suppose the dyadic compatibility assumption, i.e., $\{e_r^n = e_r : r = i2^{-n}, i = 1, \dots, 4^n, n \geq 1\}$ is an independent family, then the limit (2.19) exits in L^2, almost as an orthogonal series expansion, and the conditional expectation operator takes the form (2.20).* □

• *Remark* 2.13. If $\dot{w} = \{e_r^n : r \in R_n, n \geq 1\}$ is a diadic family of identically distributed random variables with $\mathbb{E}\{e_r^n\} = 0$ and $\mathbb{E}\{|e_r^n|^2\} = 1$ then the diadic compatibility assumption as in (c) of Proposition 2.12 is satisfied for the diadic family $\dot{w}' = \{e_r : r \in R_n, n \geq 1\}$ defined by $e_r = \sum_{k=1}^{\infty} 2^{k/2} e_r^{k+n(r)-1}$, with $n(r) = \min\{n \geq 1 : r \in R_n\}$. In this case, note that the orthogonal series expansion (2.13) for the white noise $\dot{w}$ is very similar to the expression (2.19) for the white noise $\dot{w}'$. It is clear that there are infinite many choices to obtain a white noise $\dot{w}'$ from the initial $\dot{w}$, e.g., any sequence $\{k_n : n \geq 1\}$ with $k_n \geq n(r)$ will produce a suitable $\dot{w}'$, where $e_r = e_r^{k(n)}$, for r in $R_n \setminus R_{n-1}$, with $R_0 = \emptyset$. □

• *Remark* 2.14. Under the compatibility assumption as in (c) of Proposition 2.12,

we may use the equality (2.18) to obtain

$$\sum_{n=n(r)}^{\infty} 2^{-n/2} w_r^n = 2^{(n(r)-1)/2} w_r^{n(r)-1} + \sum_{n=n(r)}^{\infty} 2^{-n} \sum_{k=n(r)}^{n} 2^{(k-1)/2} v_r^k,$$

and exchanging the double sum

$$\sum_{n=n(r)}^{\infty} 2^{-n} \sum_{k=n(r)}^{n} 2^{(k-1)/2} v_r^k = \sum_{k=n(r)}^{\infty} 2^{(k-1)/2} v_r^k \sum_{n=k}^{\infty} 2^{-n} =$$

$$= \sum_{k=n(r)}^{\infty} 2^{-(k-1)/2} v_r^k.$$

This shows that the series (2.13), with $e_{i,n} = e_{i2^{-n}}$, converges in L^2, as an orthogonal series expansion relative to $\{w_r^{n(r)-1}, v_r^{n(r)}, \dots, v_r^n, \dots\}$, with $t = r$ in R. For a non-diadic t, we have an almost orthogonal series expansion. □

• *Remark* 2.15. The above arguments can be used to construct the integral of a function φ belonging to $L^2(]0,\infty[)$. For instant, if $\varphi_n(s) = \varphi(i2^{-n})$ for s in $](i-1)2^{-n}, i2^{-n}]$, $i = 1, \dots, 4^n$, then

$$2^{-n} \sum_{i=1}^{4^n} |\varphi(i2^{-n})|^2 = \int_0^{4^n} |\varphi_n(s)|^2 \mathrm{d}s.$$

Therefore, e.g., we may replace (2.13) and (2.19) with

$$w(\varphi) = \sum_n 2^{-n} \sum_{i=1}^{4^n} \varphi(i2^{-n}) e_{i,n} \quad \text{and} \quad 2^{-n/2} \sum_{i=1}^{4^n} \varphi(i2^{-n}) e_{i2^{-n}} \to w(\varphi),$$

to obtain the integral of φ with respect to the Wiener noise $\dot{w}$. Essentially, this is like using the diadic system of functions $\phi_{i,n} = (-1)^{i-1} \mathbb{1}_{](i-1)2^{-n}, i2^{-n}]}$, $i = 1, \dots, 4^n$, and $\{(-1)^{i-1} e_{i,n}\}$ to establish a mapping similar (after a proper scaling) to the initial isometry in the beginning of this subsection. Note that $\{\phi_{i,n}\}$ is not a fully orthogonal system, but $\phi_n = \sum_{i=1}^{4^n} \phi_{i,n}$ is a finite sum of functions with disjoint supports and ϕ_n is orthogonal to ϕ_m if $n \neq m$. It is clear that behind is the orthonormal system obtained from (2.15). □

2.5 Poisson Noises

A simple Poisson noise with parameter $\lambda > 0$ can be regarded as a sequence $\dot{p} = \{\tau_n : n \geq 1\}$ of independent exponentially (with parameter $1/\lambda$) distributed random variables. Since $P(\lim_n \sum_{i=1}^n \tau_i = \infty) = 1$, the counting process $p_t = \sum_{n=1}^{\infty} \mathbb{1}_{\tau_1 + \dots + \tau_n \leq t}$, i.e.

$$p_t = \begin{cases} 0 & \text{if } t < \tau_1, \\ n & \text{if } \sum_{i=1}^n \tau_i \leq t < \sum_{i=1}^{n+1} \tau_i, \end{cases}$$

is defined almost surely and called a Poisson process, i.e., $p_0 = 0$, $p_t - p_s$ is Poisson distributed with mean $\lambda(t-s)$ and independent of p_s, for any $t > s \geq 0$. The paths are piecewise constant with jumps equal to 1. Moreover, if δ_n denotes the Dirac measure concentrated at n then

$$P\{p_t \in \mathrm{d}x\} = \mathrm{e}^{-\lambda t} \sum_{n=0}^{\infty} \delta_n(\mathrm{d}x) \frac{(\lambda t)^n}{n!} \quad \text{and} \quad \mathbb{E}\{\mathrm{e}^{\mathrm{i}\xi p_t}\} = \exp\left[t\lambda(\mathrm{e}^{\mathrm{i}\xi} - 1)\right]$$

are the transition density and the characteristic function. It is also clear that for $q_t = p_t - t\lambda$,

$$P\{q_t \in \mathrm{d}x\} = \mathrm{e}^{-\lambda t}\Big[\delta_0(\mathrm{d}x) + \sum_{n=1}^{\infty} \big(\delta_n(\mathrm{d}x) - 1\big)\frac{(\lambda t)^n}{n!}\Big]$$

is the transition function.

Properties of Poisson processes are well known, for instance, the fact that the waiting times $\{\tau_n\}$ have an exponential distribution produces the so-called 'memoryless' effect, which yields the Markov character of the process. For instance, if the waiting times is an identical distributed sequence of random variables (with a distribution other than exponential) then the corresponding counting process is a semi-Markov, i.e., one more piece of information (the time elapsed since the last jump) is necessary to produce a Markov process.

2.5.1 The Poisson Measure

The construction of Poisson (random) measure and some of its properties are necessary to discuss general Poisson noises. One way is to follow the construction of the general Wiener noise or white noise (random) measure, but using Poisson (random) variables instead of Gaussian (random) variables.

If $\{\tau_{i,n} : i \geq 1\}$ is a sequence of independent exponentially (with parameter 1) distributed random variables then random variables $\zeta_n(\lambda) = \sum_k \mathbb{1}_{\theta_{k,n} \leq \lambda}$, with $\theta_{k,n} = \tau_{1,n} + \cdots + \tau_{n,n}$, is a sequence of independent identically distributed random variables having a Poisson distribution with parameter λ. Hence, $\tilde{\zeta}_n(\lambda) = \zeta_n(\lambda) - \lambda$ has mean zero and variance $\mathbb{E}\{|\tilde{\zeta}_n(\lambda)|^2\} = \lambda$. If $\{h_n : n \geq 1\}$ is a complete orthogonal system in a Hilbert space H with $(h_n, h_n)_H = 1/k_n$, then any function h in H can be written as a converging orthogonal series $h = \sum_n (h, h_n)_H k_n h_n$, and $(h,h)_H = \sum_n |(h, h_n)_H|^2 k_n$. Thus the mapping $h \mapsto q(h) = \sum_n (h, h_n)_H \tilde{\xi}_n(k_n)$ is a linear isometry from H into $L^2 = L^2(\Omega, \mathcal{F}, P)$, and if $(h, h_n)_H = 1$ for any n in a finite subset of indexes N_h and $(h, h_n)_H = 0$ otherwise then $p(h) = \sum_n (h, h_n)_H \xi_n(k_n)$ is Poisson random variable with parameter $\sum_{n \in N_h} k_n$. In any case, if the series $m(|h|) = \sum_n |(h, h_n)_H| k_n < \infty$ then $p(h) = \sum_n (h, h_n)_H \xi_n(k_n)$ is convergent, and $p(h) = q(h) + m(h)$, with $\sum_n (h, h_n)_H k_n$.

Another construction is developed for a more specific Hilbert space, namely, $H = L^2(Y, \mathcal{Y}, \mu)$ with a σ-finite measure space $(Y, \mathcal{Y}, \mu)$, where the Poisson character is imposed on the image of $\mathbb{1}_K$ for any K in $\mathcal{Y}$ with $\mu(K) < \infty$.

Two steps are needed, first assume $\mu(Y) < \infty$ and choose a sequence $\{\zeta_n : n \geq 1\}$ of independent identically distributed following the probability law given by $\mu/\mu(Y)$ and also choose an independent Poisson distributed variable η with parameter $\lambda = \mu(Y)$. Define $p(A) = 0$ when $\eta = 0$ and $p(A) = \sum_{n=1}^{\eta} \mathbb{1}_{\zeta_n \in A}$ otherwise, for every A in $\mathcal{Y}$. The random variable $p(A)$ takes only nonnegative integer values, $p(Y) = \eta$, and if $A_1, \ldots, A_k$ is a finite partition of Y, i.e., $Y = \sum_i A_i$, and $n_1 + \cdots + n_k = n$ then

$$P\big(p(A_1) = n_1, \ldots, p(A_k) = n_k\big) = \\ = P\big(p(A_1) = n_1, \ldots, p(A_k) = n_k \,:\, p(Y) = n\big) P\big(p(Y) = n\big),$$

which are multinomial and Poisson distribution, and so

$$P\big(p(A_1) = n_1, \ldots, p(A_k) = n_k\big) = \\ = n! \frac{\big(\mu(A_1)\big)^{n_1}}{\big(\mu(Y)\big)^{n_1} n_1!} \cdots \frac{\big(\mu(A_k)\big)^{n_k}}{\big(\mu(Y)\big)^{n_k} n_k!} \, \mathrm{e}^{-\mu(Y)} \frac{\big(\mu(Y)\big)^n}{n!},$$

and summing over $n_1, \ldots, n_k$ except in n_i , we obtain

$$P\big(p(A_i) = n_i\big) = \mathrm{e}^{-m(A_i)} \frac{\big(\mu(A_i)\big)^{n_i}}{n_i!}.$$

Thus the mapping $A \mapsto p(A)$ satisfies:

(1) for every ω, $A \mapsto p(A, \omega)$ is measure on Y;

(2) for every measurable set A, the random variable $p(A)$ has a Poisson distribution with parameter (or mean) $\mu(A)$;

(3) if $A_1, \ldots, A_k$ are disjoint then $p(A_1), \ldots, p(A_k)$ are independent.

In the previous statements, note that if $\mu(A) = 0$ then the random variable $p(A) = 0$, which is (by convention) also referred to as having a Poisson distribution with parameter (or intensity) zero.

For the second step, because μ is σ-finite, there exists a countable partition $\{Y_k : k \geq 1\}$ of Y with finite measure, i.e., $Y = \sum_k Y_k$ and $\mu(Y_k) < \infty$. Now, for each k with construct p_k (as above) corresponding to the finite measure μ_k, with $\mu_k(A) = \mu(A \cap Y_k)$, in a way that the random variable involved $\zeta_{k,n}$ and η_k are all independent in k. Hence the mapping $A \mapsto p_k(A)$ satisfies (1), (2) and (3) above, and also:

(4) for every choice $k_1, \ldots, k_n$ (all different of each other) and $A_1, \ldots, A_n$ in $\mathcal{A}$, the random variables $p_{k_1}(A_1), \ldots, p_{k_n}(A_n)$ are independent.

Since a sum of independent Poisson (random) variables is again a Poisson variable, the series $p(A) = \sum_k p_k(A)$ defines a Poisson (random) variable with parameter (or mean) $\mu(A)$ whenever $\mu(A) < \infty$. If $\mu(A) = \infty$ then

$$\sum_k P\big(p_k(A) \geq 1\big) = \sum_n \big(1 - \mathrm{e}^{-\mu(A \cap Y_k)}\big) = \infty,$$

since $\lim_n(1-\mathrm{e}^{-\mu(A\cap Y_k)})=1$, i.e., if $\mu(A)=\infty$ then $p(A)=\infty$ almost surely. Hence, the mapping $A\mapsto p(A)$ satisfies (1), (2) and (3), as long as a random variable which is equal to infinite (or to zero) is considered a Poisson variable with parameter $\lambda=\infty$ (or $\lambda=0$). In this case, a Poisson variable with $\lambda=\infty$ (or $\lambda=0$) means a (degenerate) random variable which is equal to $+\infty$ (or to 0) almost surely. Remark that contrary to the Wiener process, to define the Poisson measure, the previous construction uses *independence* instead of orthogonality.

Recall that for convenience, if μ is a measure and f a function then $\mu(f)$ denotes the integral of f with respect to μ. In particular, $\mu(A)=\mu(\mathbb{1}_A)$ and $p(A)=p(\mathbb{1}_A)$.

Proposition 2.16. *If $(Y,\mathcal{Y},\mu)$ is a σ-finite measure space then the previous construction yields a Poisson random measure p with intensity measure μ, i.e., (1), (2), (3) above are satisfied. Moreover, if $\mu(Y)<\infty$ and $\varphi: Y\to\mathbb{R}^m$ is a measurable function then*

$$p(\varphi)=\int_Y \varphi(y)\,p(\mathrm{d}y)$$

defines a random variable on $\mathbb{R}^m$ with compound Poisson distribution, namely,

$$\mathbb{E}\big\{\mathrm{e}^{\mathrm{i}r\cdot p(\varphi)}\big\}=\exp\Big[\int_Y\big(\mathrm{e}^{\mathrm{i}r\cdot\varphi(y)}-1\big)\mu(\mathrm{d}y)\Big],\quad \forall r\in\mathbb{R}^m.$$

Moreover, if $\mu(|\varphi|^2)<\infty$ then $\mathbb{E}\{p(\varphi)\}=\mu(\varphi)$ and $\mathbb{E}\{|q(\varphi)|^2\}=\mu(|\varphi|^2)$, where $q=p-\mu$. Furthermore, if $A_1,\dots,A_n$ are disjoint measurable sets then the random variables $p(\varphi\mathbb{1}_{A_1}),\dots,p(\varphi\mathbb{1}_{A_n})$ are independent.

Proof. From the construction we check that for every ω, the measure $A\mapsto p(A,\omega)$is supported in a finite number of points, namely, $\zeta_i(\omega)$ for $i=1,\dots,\eta(\omega)$. Thus, the expression of the random variable $p(A)$ is finite. Using a diadic approximation of φ, i.e., we partition $\mathbb{R}^m$ into diadic cubes of the form $C_{j,n}=\,]\!](j_1-1)2^{-n},j_1 2^{-n}]\times\cdots\times](j_m-1)2^{-n},j_m 2^{-n}]$, with $j=(j_1,\dots,j_m)$ and set $\varphi_n(x)=j2^{-n}$ for every x in $C_{j,n}$, we have $|\varphi(x)-\varphi_n|\le 2^{-n}\sqrt{m}$. Since

$$p(\varphi_n,\omega)=\sum_j (j2^{-n})p(\varphi^{-1}(C_{j,n},\omega),$$

from the definition of the Poisson measure p we deduce

$$\mathbb{E}\big\{\mathrm{e}^{\mathrm{i}r\cdot p(\varphi_n)}\big\}=\prod_j\mathbb{E}\big\{\mathrm{e}^{\mathrm{i}r\cdot j2^{-n}p(\varphi^{-1}(C_{j,n}))}\big\}=$$

$$=\prod_j\exp\big[(\mathrm{e}^{\mathrm{i}r\cdot j2^{-n}}-1)\mu(\varphi^{-1}(C_{j,n}))\big]=\exp\Big[\int_Y(\mathrm{e}^{\mathrm{i}r\cdot\varphi_n(y)}-1)\mu(\mathrm{d}y)\Big]$$

and the first part follows as $n\to\infty$.

Once the expression of the characteristic function have been proved, the mean and the variance are calculated by taking derivative with respect to the parameter r, and the last part, regarding the independence, is deduced by the convergence of $p(\varphi_n)$ to $p(\varphi)$ and the property (3) of Poisson measure discussed above. □

Remark that as it was defined, for each ω fixed, the Poisson random measure $p(\cdot, \omega)$ is a finite sum of Dirac measures. Hence, p is also called Poisson point measure. The companion measure $q = p - \mu$ is referred to as a centered (or martingale) Poisson (random or point) measure.

Sometimes, we may single-out a time variable, i.e., replace Y and μ with $Y \times]0, \infty[$ and $\mu \times \mathrm{d}t$. In this case, the variable η can be specifically constructed as a Poisson process with parameter $\lambda = \mu(Y) < \infty$, i.e.,

$$\eta(t) = \sum_n \mathbb{1}_{t \geq \theta_n}, \quad \forall t > 0,$$

where $\theta_n = \tau_1 + \cdots + \tau_n$ and $\{\tau_n : n \geq 1\}$ is a sequence of independent exponentially distributed (with parameter λ) random variable. In this case

$$p(A \times]a, b]) = \sum_{n=1}^{\eta(b)} \mathbb{1}_{\zeta_n \in A} - \sum_{n=1}^{\eta(a)} \mathbb{1}_{\zeta_n \in A} = \sum_n \mathbb{1}_{\zeta_n \in A} \mathbb{1}_{a < \theta_n \leq b}, \quad \forall a \leq b.$$

If $\mu(Y) = \infty$ then express the space Y as countable number of disjoint sets with finite measure (i.e., $Y = \sum_k Y_k$ with $\mu(Y_k) < \infty$), and find sequences of independent variables $\zeta_{n,k}$ with distribution $\mu(\cdot \cap Y_k)/\mu(Y_k)$ and $\tau_{n,k}$ exponentially distributed with parameter $\mu(Y_k)$, for any $n, k \geq 1$. The Poisson measure is given by

$$p(A \times]a, b]) = \sum_{n,k} \mathbb{1}_{\zeta_{n,k} \in A} \mathbb{1}_{a < \theta_{n,k} \leq b}, \quad \forall a \leq b,$$

where $\theta_{k,n} = \tau_{1,k} + \cdots + \tau_{n,k}$. Our interest is the case where $Y = \mathbb{R}^d_*$ and $\zeta_{n,k}$ is interpreted as the jumps of a Lévy process.

2.5.2 The Poisson Noise I

Another type of complications appear in the case of the compound Poisson noise, i.e., like a Poisson process with jumps following some prescribed distribution, so that the paths remain piecewise constant.

Consider $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$ and $\mathcal{B}_* = \mathcal{B}(\mathbb{R}^d_*)$, the Borel σ-algebra, which is generated by a countable semi-ring $\mathcal{K}$. (e.g., the family of d-intervals $]a, b]$ with closure in $\mathbb{R}^d_*$ and with rational end points). Now, beginning with a given (non-zero) finite measure m in $(\mathbb{R}^d_*, \mathcal{B}_*)$, we construct a sequence $\dot{q} = \{(z_n, \tau_n) : n \geq 1\}$ of independent random variables such that each τ_n is exponentially distributed with parameter $m(\mathbb{R}^d_*)$ and z_n has the distribution law $A \mapsto m(A)/m(\mathbb{R}^d_*)$, thus, the random variables $\theta_n = \tau_1 + \cdots + \tau_n$ have $\Gamma(m(\mathbb{R}^d_*), n)$ distribution. The series

$\eta_t = \sum_n \mathbb{1}_{t\geq\theta_n}$ is almost surely a finite sum and defines a Poisson process with parameter $m(\mathbb{R}^d_*)$, satisfying $\mathbb{E}\{\eta_t\} = tm(\mathbb{R}^d_*)$ and $\mathbb{E}\{|\eta_t - tm(\mathbb{R}^d_*)|^2\} = tm(\mathbb{R}^d_*)$. Moreover, we may just suppose given a $\mathbb{R}^d$-valued compound Poisson process $\{N_t : t \geq 0\}$ with parameter $\lambda = \mathtt{m}(\mathbb{R}^d_*)$ and $\mathtt{m}/\lambda$, or simply $\mathtt{m}$, i.e., with the following characteristic function

$$\mathbb{E}\{\mathrm{e}^{\mathrm{i}\zeta\cdot N_t}\} = \exp\Big\{t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathrm{i}\zeta\cdot z} - 1\big]m(\mathrm{d}z)\Big\}, \qquad \forall\xi\in\mathbb{R}^d,$$

as a Lévy process, with $N_t = \sum_n z_n \mathbb{1}_{t\geq\theta_n}$.

In any case, the counting measure either

$$p_t(K) = \sum_n \mathbb{1}_{z_n\in K}\mathbb{1}_{t\geq\theta_n}, \quad \forall K\in\mathcal{K},\ t\geq 0,$$

or equivalently

$$p_t(K) = \sum_{n=1}^{\eta(t)} \mathbb{1}_{z_n\in K}, \qquad \eta(t) = \sum_n \mathbb{1}_{t\geq\theta_n}, \quad \forall K\in\mathcal{K},\ t\geq 0,$$

is a Poisson process with parameter $m(K)$, $\eta(t)$ is also a Poisson process with parameter $tm(\mathbb{R}^d_*)$. Moreover, if $K_1,\dots,K_k$ are any disjoint sets in $\mathcal{K}$ then $p_t(K_1),\dots,p_t(K_k)$ are independent processes. Indeed, if $n = n_1+\cdots+n_k$ and $\mathbb{R}^d_* = K_1\cup\cdots\cup K_k$ then

$$\begin{aligned}
&P\big(p_t(K_1) = n_1,\dots p_t(K_k) = n_k\big) = \\
&\quad = P\big(p_t(K_1) = n_1,\dots p_t(K_k) = n_k \mid p_t(K) = n\big)P\big(p_t(K) = n\big) = \\
&\quad = P\big(\sum_{i=1}^n \mathbb{1}_{z_i\in K_1} = n_1,\dots \sum_{i=1}^n \mathbb{1}_{z_i\in K_k} = n_k \mid p_t(\mathbb{R}^d_*) = n\big)P\big(\eta(t) = n\big),
\end{aligned}$$

which are multinomial and Poisson distribution, and so

$$\begin{aligned}
&P\big(p_t(K_1) = n_1,\dots p_t(K_k) = n_k\big) = \\
&\qquad = \frac{\big(m(K_1)\big)^{n_1}}{\big(m(\mathbb{R}^d_*)\big)^{n_1}n_1!}\cdots\frac{\big(m(K_k)\big)^{n_k}}{\big(m(\mathbb{R}^d_*)\big)^{n_k}n_k!}\,\mathrm{e}^{-m(\mathbb{R}^d_*)}\frac{\big(m(\mathbb{R}^d_*)\big)^n}{n!},
\end{aligned}$$

and summing over $n_1,\dots,n_k$ except in n_j , we obtain

$$P\big(p_t(K_j) = n_j\big) = \mathrm{e}^{-m(K_j)}\frac{\big(m(K_j)\big)^{n_j}}{n_j!},$$

which proves that $p_t(K_j)$ are independent Poisson processes. This implies that

$$\mathbb{E}\{p_t(K)\} = tm(K), \qquad \mathbb{E}\big\{|p_t(K) - tm(K)|^2\big\} = tm(K),$$

for every K in $\mathcal{K}$ and $t \geq 0$. Hence, the (martingale or centered) measure

$$q_t(K) = \sum_n \mathbb{1}_{z_n \in K} \mathbb{1}_{t \geq \theta_n} - tm(K), \qquad \mathbb{E}\{q_t(K)\} = 0, \quad \forall K \in \mathcal{K}$$

satisfies $\mathbb{E}\{q_t^2(K)\} = tm(K)$, and if $K \cap K' = \emptyset$ then $q_t(K)$ and $q_t(K')$ are orthogonal and, in general, $\mathbb{E}\{q_t(K)q_t(K')\} = tm(K \cap K')$.

• *Remark* 2.17. Sometime it is more convenient not to distinguish the time t in the Poisson measure, i.e., to consider p as a random integer measure on $\mathbb{R}_*^d \times (0, \infty)$. In this case, either two steps are necessary or only a construction on $\mathbb{R}_*^d \times (0, b]$ $(b < \infty)$ is achieved. For instance, given a bounded measure Π on $\mathbb{R}_*^d \times (0, b]$ proceed as follows: (1) find a sequence $\{z_n : n \geq 1\}$ of independent random variables with identical distribution Π/c, $c = \Pi(\mathbb{R}_*^d \times (0, b])$, and (2) find an independent Poisson distributed (with parameter c) random variable η, and then define $p(B) = \sum_{n=1^\eta} \mathbb{1}_{z_n \in B}$. By using independent copies of p, we can patch the definition of p from $\mathbb{R}_*^d \times (0, b]$ into $\mathbb{R}_*^d \times (b, 2b]$ and so on, to get p defined on the whole $\mathbb{R}_*^d \times (0, \infty)$, and clearly $\Pi(\mathrm{d}z, \mathrm{d}t) = m(\mathrm{d}z)\mathrm{d}t$. In this construction, the origin $\{0\}$ plays not particular role, so that the intensity Π needs only to be a σ-finite Borel measure on some Polish space. Later, to integrate the function z to reproduce the jumps, the Lévy measure condition appears. □

Now, a (real-valued) simple function relative to the semi-ring $\mathcal{K}$ is a finite sum of terms (with disjoint K's) of the form $\alpha \mathbb{1}_K(z)$ (which is equal to α when z belongs to K and 0 otherwise). Each term integrates with respect to $p_t(\mathrm{d}z)$ and $q_t(\mathrm{d}z)$ as follows

$$\int_{\mathbb{R}_*^d} \alpha \mathbb{1}_K(z)\, q_t(\mathrm{d}z) = \alpha\, q_t(K), \qquad \mathbb{E}\Big\{ \Big| \int_{\mathbb{R}_*^d} \alpha \mathbb{1}_K(z)\, q_t(\mathrm{d}z) \Big|^2 \Big\} = \alpha^2 tm(K).$$

This definition is extended by linearity (uniquely) to any simple function, ψ and because each $\{q_t(K)\}$ are independent when the K's are disjoint, we preserve the relation

$$\mathbb{E}\{|q_t(\psi)|^2\} = \mathbb{E}\Big\{ \Big| \int_{\mathbb{R}_*^d} \psi(z)\, q_t(\mathrm{d}z) \Big|^2 \Big\} = t \int_{\mathbb{R}_*^d} |\psi(z)|^2 m(\mathrm{d}z) = tm(|\psi|^2).$$

Remark that to simplify the notation, we write $q_t(\psi)$ and $m(\psi)$ to symbolize the integral of a function ψ, e.g., $m(K) = m(\mathbb{1}_K) = m(|\mathbb{1}_K|^2)$. Moreover, because m is a finite measure, if $m(|\psi|^2) < \infty$ then $m(|\psi|) < \infty$.

Again, this integral $\psi \mapsto q_t(\psi)$ is extended as a linear isometry between Hilbert spaces, from $L^2(m) = L^2(\mathbb{R}_*^d, \mathcal{B}_*, tm)$ into $L^2(\Omega, \mathcal{F}, P)$, and

$$q_t(\psi) = \sum_n \psi(z_n) \mathbb{1}_{t \geq \theta_n} - tm(\psi), \qquad \text{with} \quad \mathbb{E}\{q_t(\psi)\} = 0, \tag{2.21}$$

reduces to a finite sum almost surely. This is the same argument as the case of random orthogonal measures, but in this case, this is also a pathwise argument. Indeed, we could use the almost surely finite sum (2.21) as definition.

A priori, the above expression of $q_t(\psi)$ seems to depend on the pointwise definition of ψ, however, if $\psi = \psi'$ m-almost everywhere then $q_t(\psi) = q_t(\psi')$ almost surely. Moreover, $\mathbb{E}\{q_t(\psi)q_s(\psi')\} = (t \wedge s)m(\psi\psi')$ and the process $t \mapsto q_t(\psi)$ is continuous in the L^2-norm.

As mentioned early, $N_t = \sum_n z_n \mathbb{1}_{t \geq \theta_n}$ is a $\mathbb{R}^d$-valued compound Poisson process, and therefore, the expression

$$t \mapsto p_t(\psi) = \sum_n \psi(z_n)\mathbb{1}_{t \geq \theta_n}, \qquad \forall \psi \in L^2(\mathbb{R}^d_*, \mathcal{B}_*, m)$$

defines a real-valued compound Poisson process with characteristic function

$$\mathbb{E}\{\mathrm{e}^{\mathtt{i}p_t(\psi)}\} = \exp\Big\{t \int_{\mathbb{R}^d_*} \big[\mathrm{e}^{\mathtt{i}\psi(z)} - 1\big] m(\mathrm{d}z)\Big\}.$$

This yields

$$\mathbb{E}\{\mathrm{e}^{\mathtt{i}q_t(\psi)}\} = \exp\Big\{t \int_{\mathbb{R}^d_*} \big[\mathrm{e}^{\mathtt{i}\psi(z)} - 1 - \mathtt{i}\psi(z)\big] m(\mathrm{d}z)\Big\}, \tag{2.22}$$

for every ψ in $L^2(\mathbb{R}^d_*, \mathcal{B}_*, m)$.

If $m(|z|) < \infty$ then $\mathbb{E}\{|z_n|\} < \infty$ and $\mathbb{E}\{|N_t|\} = tm(|z|)$. Moreover, we can define the $\mathbb{R}^d$-valued Lévy process $q_t(z) = N_t - tm(z)$ with characteristic $(0, 0, m)$, i.e.,

$$\mathbb{E}\{\mathrm{e}^{\mathtt{i}\zeta \cdot q_t(z)}\} = \exp\Big\{t \int_{\mathbb{R}^d_*} \big[\mathrm{e}^{\mathtt{i}\zeta \cdot z} - 1 - \mathtt{i}\zeta \cdot z\big] m(\mathrm{d}z)\Big\} \tag{2.23}$$

and transition density

$$\begin{aligned} P(q_t(z) \in \mathrm{d}x) &= \mathrm{e}^{-m(\mathbb{R}^d_*)t}\Big[\delta_0(\mathrm{d}x) + \sum_{n=1}^{\infty} \big(m^{\star n}(\mathrm{d}x) - m^{\star n}(\mathbb{R}^d_*)\big)\frac{t^n}{n!}\Big], \\ m^{\star(n+1)}(B) &= (m^{\star n} \star m)(B) = \int_{\mathbb{R}^d_* \times \mathbb{R}^d_*} \mathbb{1}_B(x+y)\, m^{\star n}(\mathrm{d}x)\, m(\mathrm{d}y), \end{aligned} \tag{2.24}$$

where $m^{\star 1} = m$ and $m^{\star n}(\mathbb{R}^d_*) = (m(\mathbb{R}^d_*))^n = \lambda^n$. Next, remarking that $t \mapsto q_t(z)$ is continuous except for $t = \theta_n$ that $q_t(z) - q_{t-}(z) = N_t - N_{t-} = z_n$, the expression

$$q_t(K) = \sum_{s \leq t} \mathbb{1}_{q_t(z) - q_{t-}(z) \in K} - tm(K) \tag{2.25}$$

is a finite sum almost surely, and can be used to reconstruct the counting measure $\{q_t(K) : K \in \mathcal{K}\}$ from the $\{q_t(z) : t \geq 0\}$. Indeed, just the knowledge that the paths $t \mapsto q_t(z)$ are cad-lag, implies that the series (2.25) reduces to a finite sum almost surely.

The terms $\psi(z_n)\mathbb{1}_{t\geq\theta_n}$ in the series (2.21) are not independent, but setting $\lambda = m(\mathbb{R}^d_*)$ and $m' = m/\lambda$ we compute

$$\mathbb{E}\big\{|\psi(z_n)\mathbb{1}_{t\geq\theta_n}|^2\big\} = m'(|\psi|^2)\, r_n(t),$$
$$\mathbb{E}\big\{\psi(z_n)\mathbb{1}_{t\geq\theta_n}\psi(z_k)\mathbb{1}_{t\geq\theta_k}\big\} = |m'(\psi)|^2 r_n(t), \quad \forall k > n \geq 1,$$

where

$$\mathbb{E}\{\mathbb{1}_{t\geq\theta_n}\} = \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!}\mathrm{e}^{-\lambda s}\mathrm{d}s = r_n(t)$$

with $\sum_n r_n(t) = tm(\mathbb{R}^d_*)$. Thus, the Gram-Schmidt orthogonalization procedure can be used to define $e_0(\psi,t) = -tm(\psi)$, $e_1(\psi,t) = \psi(z_1)\mathbb{1}_{t\geq\theta_1} - m'(\psi)r_1(t)$, $\mathbb{E}\{|e_1(\psi,t)|^2\} = \big(m'(|\psi|^2) - |m'(\psi)|^2 r_1(t)\big)r_1(t)$, and

$$e_2(\psi,t) = \psi(z_2)\mathbb{1}_{t\geq\theta_2} - m'(\psi)r_2(t) - \frac{|m'(\psi)|^2}{m'(|\psi|^2) - |m'(\psi)|^2 r_1(t)}\, e_1(\psi,t),$$

and a more complicate expression for $n \geq 2$. Actually, this is equivalent to

$$e_n(\psi,t) = \psi(z_n)\mathbb{1}_{t\geq\theta_n} - \mathbb{E}\{\psi(z_n)\mathbb{1}_{t\geq\theta_n} \,|\, \psi(z_i)\mathbb{1}_{t\geq\theta_i},\ i = 1,\dots,n-1\}$$

the conditional expectation, and $q_t(\psi) = \sum_{n\geq 1} e_n(\psi,t)$.

Alternatively, if $e'_n(\psi,t) = \big(\psi(z_n) - m'(\psi)\big)\mathbb{1}_{t\geq\theta_n}$ then $\mathbb{E}\{e'_n(\psi,t)\} = 0$, $\mathbb{E}\{|e'_n(\psi,t)|^2\} = \big(m'(|\psi|^2) - |m'(\psi)|^2\big)r_n(t)$, and for $k > n \geq 1$,

$$\mathbb{E}\{e'_n(\psi,t)e'_k(\psi,t)\} = \mathbb{E}\{(\psi(z_n) - m'(\psi))(\psi(z_k) - m'(\psi))\mathbb{1}_{t\geq\theta_n}\} = 0.$$

Also, define $e''_n(\psi,t) = m'(\psi)\big(\mathbb{1}_{t\geq\theta_n} - r_n(t)\big)$, which satisfies $\mathbb{E}\{e''_n(\psi,t)\} = 0$, $\mathbb{E}\{|e''_n(\psi,t)|^2\} = |m'(\psi)|^2 r_n(t)\big(1 - r_n(t)\big)$, $\mathbb{E}\{e''_n(\psi,t)e''_k(\psi,t)\} = 0$ if $n \neq k$, and $\mathbb{E}\{e''_n(\psi,t)e'_k(\psi,t)\} = 0$ for any n,k. Therefore $\{e'_n(\psi,t), e''_k(\psi,t) : n,k \geq 1\}$ is an orthogonal system such that $q_t(\psi) = \sum_n e'_n(\psi,t) + \sum_k e''_n(\psi,t)$ or $q_t(\psi) = \sum_n e'_n(\psi,t) + m'(\psi)N_t$, where $N_t = \sum_n \mathbb{1}_{t\geq\theta_n}$ is a Poisson process with parameter $m(\mathbb{R}^d_*)$. Comparing with the white noise, the orthogonality is not necessary since the series defining (2.21) is finite almost surely.

If $\mathcal{F}_\psi$ is σ-algebra generated by $\{q_s(\psi) : s \leq t\}$ (or equivalently by the countable family $\{e_r(\psi) : r \leq t,\ r \in R\}$), then the closure linear subspace H_ψ of $L^2(\Omega,\mathcal{F}_\psi,P)$ spanned by $\{q_t(\psi) : t \geq 0\}$ could be called the Poisson noise relative to any nonzero ψ in $L^2(\mathbb{R}^d_*,\mathcal{B},m)$. If we normalize the orthogonal system then the projection operator

$$\mathbb{E}\{x \,|\, q_s(\psi),\ s \leq t\} = \sum_n \frac{\mathbb{E}\{x\, e_n(\psi,t)\}}{\mathbb{E}\{|e_n(\psi,t)|^2\}}\, e_n(\psi,t), \tag{2.26}$$

valid only for x in H_ψ. Contrary to the white noise, there is not an equivalent to the Hermit polynomials (in general), and we do not have an easy construction of an orthonormal basis for $L^2(\Omega,\mathcal{F}_\psi,P)$.

• *Remark* 2.18. The above argument used to construct $q_t(\psi)$ for every ψ in $L^2(m)$ can be adapted to define $q_t(\Psi) = q(\Psi \mathbb{1}_{]0,t]})$ as the double integral of functions $\Psi = \Psi(t,z)$ belonging to $L^2(]0,\infty[\times\mathbb{R}^d_*, \mathrm{d}t \times \mathrm{d}m)$, where

$$\mathbb{E}\{|q(\Psi)|^2\} = \int_0^\infty \mathrm{d}t \int_{\mathbb{R}^d_*} |\Psi(t,z)|^2 m(\mathrm{d}z),$$

and $\mathbb{E}\{q(\Psi)\} = 0$. Even $\mathbb{R}^n$-valued functions Ψ can be handled with the same argument. □

For instance, the reader may be interested in checking the so-called marked processes as described in the books Bremaud [19, 20] and Jacobsen [81], among others.

2.5.3 The Poisson Noise II

Even more complicate is the case of the general Poisson noise, which is regarded as Poisson point process or Poisson measure, i.e., the paths are cad-lag functions, non necessary piecewise constant.

Let m be a σ-finite measure in $(\mathbb{R}^d_*, \mathcal{B}_*)$, with the Borel σ-algebra being generated by a countable semi-ring $\mathcal{K}$. We partition the space $\mathbb{R}^d_*$ is a disjoint union $\mathbb{R}^d_* = \sum_k \mathtt{R}_k$ with $0 < m(\mathtt{R}_k) < \infty$ to apply the previous construction for the finite measures $m_k = m(\cdot \cap \mathtt{R}_k)$ in such a way that the sequences $\dot{q}_k = \{(z_{n,k}, \tau_{n,k}) : n \geq 1\}$ are independent for $k \geq 1$. Therefore, the sequence of counting measures $\{q_{t,k}(K) : k \geq 1\}$ is orthogonal, with $\mathbb{E}\{|q_{t,k}(K)|^2\} = tm(K \cap \mathtt{R}_k)$, and the series $q_t(K) = \sum_k q_{t,k}(K)$ is now defined as a limit in $L^2(\Omega, \mathcal{F}, P)$ satisfying $\mathbb{E}\{q_t(K)\} = 0$ and $\mathbb{E}\{|q_t(K)|^2\} = tm(K)$. Remark that if we assume given a sequence $\{N_{t,k} : k \geq 1\}$ of independent $\mathbb{R}^d$-valued compound Poisson processes with parameter m_k, the series $\sum_k N_{t,k}$ may not be convergent.

Next, the same argument applies for the integrals, i.e., $q_t(\psi) = \sum_k q_t(\psi)$ makes sense (as a limit in the L^2-norm) for every ψ in $L^2(\mathbb{R}^d_*, \mathcal{B}_*, m)$, and $\mathbb{E}\{q_t(\psi)\} = 0$, $\mathbb{E}\{|q_t(\psi)|^2\} = tm(|\psi|^2)$. However the (double) series

$$q_t(\psi) = \sum_k \Big[\sum_n \psi(z_{n,k}) \mathbb{1}_{t \geq \theta_{n,k}} - t m_k(\psi) \Big], \quad \forall \psi \in L^2(\mathbb{R}^d_*, \mathcal{B}_*, m), \quad (2.27)$$

does not necessarily reduces to a finite sum almost surely, $m(|\psi|)$ may not be finite and the pathwise analysis can not be used anymore.

Nevertheless, if we add the condition that any K in $\mathcal{K}$ is contained in a finite union of $\mathtt{R}_k$, then $q_t(K) = \sum_k q_{t,k}(K)$ does reduce to a finite sum almost surely, and we can construct the integral almost as in the case of the composed Poisson noise. This is to say that, for any K in $\mathcal{K}$, the path $t \mapsto q_t(K)$ is a piecewise constant function almost surely. Similarly, if ψ vanishes outside of a finite number of $\mathtt{R}_k$ then the series (2.27) reduces to a finite sum almost surely.

The martingale estimate[2]

$$\mathbb{E}\{\sup_{t\le T}|q_t(\psi)|^2\}\le 3m(|\psi|^2)\,T,\qquad \forall T>0,$$

shows that the limit defining $q_t(\psi)$ converges uniformly on any bounded time interval $[0,T]$, and so, it is a cad-lag process. Another way is to make use of the estimate $\mathbb{E}\{|q_t(\psi)-q_s(\psi)|^2\}=m(\psi)|t-s|$ (and the property of independent increments) to select a cad-lag version.

Therefore, the (double) integral $q_t(\psi)$ is defined above as a L^2-continuous random process by means of a L^2 converging limit as in (2.27).

Actually, the random measure $\{q_t(\mathrm{d}z):t\ge 0,\ z\in\mathbb{R}^d_*\}$ is a *centered Poisson measure* Lévy measure m, namely, for every ψ in $L^2(\mathbb{R}^d_*,\mathcal{B}_*,m)$, the integral $\{q_t(\psi):t\ge 0\}$ is a Lévy process with characteristic $(0,0,m_\psi)$, where m_ψ is pre-image measure of m, i.e., $m_\psi(B)=m(\psi^{-1}(B))$, for every Borel set B in $\mathbb{R}$, and the expression (2.22) of the characteristic function of $q_t(\psi)$ is valid.

Since the measure m is not necessarily finite, only if $m(\psi)<\infty$ we can add the counting process to define the integral $p_t(\psi)$ as in the case of a compound Poisson process, i.e., the (double) series

$$p_t(\psi)=\sum_{n,k}\varphi(z_{n,k})\,\mathbb{1}_{t\le\theta_{n,k}}$$

converges in $L^1(\Omega,\mathcal{F},P)$, but does not necessarily reduces to a finite sum almost surely. Any way, we have the equality $\mathbb{E}\{q_t(\psi)q_s(\psi')\}=(t\wedge s)m(\psi\psi')$, for every ψ and ψ' in $L^2(\mathbb{R}^d_*,\mathcal{B}_*,m)$ and any $t,s>0$.

Thus, if $m(|z|)<\infty$ then the series $\sum_k\sum_n|z_n|\mathbb{1}_{t\ge\theta_{n,k}}=\sum_k m_k(|z|)=m(|z|)$ converges, and therefore, the $\mathbb{R}^d$-valued Lévy process

$$N_t=\sum_k\sum_n z_{n,k}\mathbb{1}_{t\ge\theta_{n,k}}$$

is meaningful and $N_t=q_t(z)+tm(z)$. In general, if only $m(|z|^2\wedge|z|)<\infty$ then the $\mathbb{R}^d$-valued Lévy process $\{q_t(z):t\ge0\}$ with characteristic function (2.23) remains meaningful, and the expression (2.25) allows us to reconstruct the counting measure $\{q_t(K):K\in\mathcal{K}\}$ from the $\{q_t(z):t\ge0\}$. However, the expression of the transition density is not so immediately, for each finite measure $m_k=m(\cdot\cap\mathtt{R}_k)$ we have an explicit series but the limit in k is not so clear. Any way, for a bounded set B with $m(B)<\infty$, the transition density of $\{q_t(z\mathbb{1}_B):t\ge0\}$ is given by a series similar to (2.24).

Observe that if the measure m integrates the function $z\mapsto|z|^2$ then

$$q_t(z)=\sum_k\Big[\sum_n z_{n,k}\mathbb{1}_{t\ge\theta_{n,k}}-tm_k(z)\Big]$$

[2]Note that $\{q_t(\psi):t\ge0\}$ is a separable martingale, so that Doob's inequality or regularization suffices to get a cad-lag version

converges in L^2, and because $P(\lim_{n,k}\theta_{n,k}=\infty)=1$ and $m(\mathbb{1}_{|z|\geq\varepsilon})<\infty$, $\varepsilon>0$, the series $\sum_k\sum_n \mathbb{1}_{|z_{n,k}|\geq\varepsilon}\mathbb{1}_{t\geq\theta_{n,k}}$ is a finite sum almost surely, for every $\varepsilon>0$. Therefore, a convenient semi-ring $\mathcal{K}$ is the countable class of d-intervals $]a,b]$ with closure in $\mathbb{R}^d_*$ and with rational end points, in this way, if $m(|z|^2\wedge 1)<\infty$ then $q_t(K)$, given by either (2.27) or (2.25), reduces to a finite sum almost surely, for every K in $\mathcal{K}$. Usually, an *intensity measure* m (not necessarily in $\mathbb{R}^d_*$) is associated with $\{q_t(\mathrm{d}z)\}$ (regarded as a Poisson martingale measure), whist a *Lévy measure* m (on $\mathbb{R}^d_*$), which necessary satisfies $m(|z|^2\wedge 1)<\infty$, is associated with $\{q_t(z)\}$ (regarded as a $\mathbb{R}^d$-valued centered Poisson process). However, we prefer to assume $m(|z|^2)<\infty$ to obtain a $\mathbb{R}^d$-valued Lévy process $\{q_t(z):t\geq 0\}$ with finite second-order moments.

If $\mathcal{K}$ is a countable semi-ring (with each K separated from $\{0\}$) generating the Borel σ-algebra in $\mathbb{R}^d_*$ then, perhaps, the system $\dot{q}_t=\{e_{n,k}(K,t):n,k\geq 1,\ K\in\mathcal{K}\}$, with

$$e_{n,k}(K,t)=\big[\mathbb{1}_{z_{n,k}\in K}\mathbb{1}_{\theta_{n,k}\leq t}-\mathbb{E}\{\mathbb{1}_{z_{n,k}\in K}\}\mathbb{E}\{\mathbb{1}_{\theta_{n,k}\leq t}\}\big],$$

is the ideal expression of a Poisson noise with Lévy measure m. Similarly, if ψ in $L^2(\mathbb{R}^d_*,\mathcal{B}_*,m)$ then for every $n\geq 1$, $\{e_{n,k}(\psi,t):k\geq 1\}$, with

$$e_{n,k}(\psi,t)=\big[\psi(z_{n,k})\mathbb{1}_{\theta_{n,k}\leq t}-\mathbb{E}\{\psi(z_{n,k})\}\mathbb{E}\{\mathbb{1}_{\theta_{n,k}\leq t}\}\big],$$

is an orthogonal system in $L^2(\Omega,\mathcal{F},P)$, with $\mathbb{E}\{e_{n,k}(\psi,t)\}=0$ and repeating the orthogonalization of the case with finite measure m, an orthogonal system $\{\tilde{e}_{n,k}(\psi,t):n,k\geq 1\}$ can be constructed. Hence, the projection operator has a form similar to (2.26). It is also clear that we can extend Remark 2.18 to this general Poisson noise.

To conclude this long introduction (of Wiener and Poisson noises or processes), let us mention that the previous arguments could be used to define a Wiener process $\{w_t:t\geq 0\}$ and a $\mathbb{R}^d$-valued (centered) Poisson process $\{q_t(z):t\geq 0\}$ or martingale Poisson measure $\{q_t(\mathrm{d}z):t\geq 0\}$ with Lévy measure m on $\mathbb{R}^d_*$, independent one of each other. Essentially, the arguments go as follows: a convergent orthogonal (better, independent identically distributed random variables with zero mean) series is used for the Wiener process, and a two procedure is needed for the Poisson measure, namely, an almost surely finite series (when the Lévy measure is finite) and next a convergent orthogonal series. As mentioned early, the above constructions can handle real-valued functions in $L^2(]0,\infty[)$ or $L^2(m(\mathrm{d}z)\times\mathrm{d}t)$ instead of just integrating functions constant in time (1 and ψ), and eventually random functions which are appropriate limits of a linear combination of terms like $\mathbb{1}_{]0,\tau]}$, with a bounded stopping time τ.

Summing-up, these constructions, specially the extension to random functions, are called *stochastic integrals*. The class of random functions that are *integrable* with respect to either a $\mathbb{R}^d$-valued Wiener process w or a Poisson martingale measure q with Lévy measure m in $\mathbb{R}^d_*$ are processes either $(f(t):t\geq 0)$ or $\{g(z,t):z\in\mathbb{R}^d_*,\ t\geq 0\}$ satisfying almost surely the integrability condition

$$\text{either}\quad \int_0^T|f(t)|^2\mathrm{d}t<\infty\quad\text{or}\quad\int_0^T\mathrm{d}t\int_{\mathbb{R}^d_*}|g(z,t)|^2\pi(\mathrm{d}z)<\infty,$$

and the non-anticipative assumption, i.e., for every $t \geq 0$, either $f(t)$ or $g(z,t)$ is independent of the increments, either $\{w(s) - w(t) : s > t\}$ or $\{q_s(K) - q_t(K) : s > t,\ K \in \mathcal{K}\}$, with $\mathcal{K}$ a countable semi-ring (each K separated from $\{0\}$) generating the Borel σ-algebra in $\mathbb{R}^d_*$. This non-anticipating property with respect to the previous constructions translates into an independent condition of either $f(t)$ or $g(z,t)$ with respect to the sequence od random variables

$$\begin{aligned} &\text{either} \quad \{e_{i,n} : i = 1, \dots, 4^n,\ n \geq 1,\ i2^{-n} > t\} \\ &\text{or} \quad \{\mathbb{1}_{z_{n,k} \in K} \mathbb{1}_{s \geq \theta_{n,k} > t} : n, k \geq 1,\ K \in \mathcal{K},\ s > t\}, \end{aligned}$$

with the notation (2.13) and (2.27). The main point of these constructions is to note that the stochastic integrals are intrinsically connected with the construction of Lévy processes. However, in what follows, the focus is on the integrands (i.e., processes that are integrable) with respect to a Lévy process.

2.6 Probability Measures and Processes

We are interested in the law of two particular type of Lévy processes, the Wiener and the Poisson processes in Hilbert spaces. There is a rich literature on Gaussian processes, but less is known in Poisson processes, actually, we mean *compensated* Poisson processes. For stochastic integration we also use the Poisson random measures and in general integer random measures.

Definition 2.19 (Lévy Space)**.** For any nonnegative symmetric square matrix a and any σ-finite measure π in $\mathbb{R}^d_* = \mathbb{R}^d \smallsetminus \{0\}$ satisfying

$$\int_{\mathbb{R}^d_*} \big(|y|^2 \wedge 1\big)\pi(\mathrm{d}y) < \infty,$$

there exists a unique probability measure $P_{a,\pi}$, called Lévy *noise* space, on the space $\mathcal{S}'(\mathbb{R}, \mathbb{R}^d)$ of Schwartz tempered distributions on $\mathbb{R}$ with values in $\mathbb{R}^d$ such that

$$\begin{aligned} \mathbb{E}\big\{\mathrm{e}^{\mathrm{i}\langle \cdot, \phi\rangle}\big\} = \exp\Big(-\frac{1}{2}\int_{\mathbb{R}} a\phi(t)\cdot\phi(t)\mathrm{d}t\Big)\times \\ \times \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{\mathbb{R}^d_*} \big[\mathrm{e}^{\mathrm{i}\phi(t)\cdot y} - 1 - \mathrm{i}\mathbb{1}_{\{|y|<1\}}\,\phi(t)\cdot y\big]\pi(\mathrm{d}y)\Big), \end{aligned}$$

for any test function ϕ in $\mathcal{S}(\mathbb{R}, \mathbb{R}^d)$. Therefore, a cad-lag version ℓ of the stochastic process $t \mapsto \langle \cdot, \mathbb{1}_{(0,t)}\rangle$ is well define and its law P on the canonical sample space $\mathbb{D} = D([0,\infty), \mathbb{R}^d)$ with the Skorokhod topology and its Borel σ-algebra $\mathcal{B}(\mathbb{D})$ is called the canonical *Lévy space* with parameters a and π, the diffusion covariance matrix a and the Lévy measure π. □

Clearly, ℓ is a Lévy process (see Section 2.2 in Chapter 2)

$$\langle \omega, \phi\rangle = \int_{\mathbb{R}} \omega(t)\cdot\phi(t)\,\mathrm{d}t, \quad \forall \omega \in \mathcal{S}'(\mathbb{R}, \mathbb{R}^d),\ \phi \in \mathcal{S}(\mathbb{R}, \mathbb{R}^d)$$

and $\cdot$ denotes the scalar product in the Euclidian space $\mathbb{R}^d$. To simplify notation and not to the use $\mathbb{1}_{\{|y|<1\}}$, we prefer to assume a stronger assumption on the Lévy measure, namely

$$\int_{\mathbb{R}^d_*} \big(|y|^2 \wedge |y|\big)\pi(\mathrm{d}y) < \infty,$$

and even to have a finite second moment, we assume

$$\int_{\mathbb{R}^d_*} |y|^2\pi(\mathrm{d}y) < \infty.$$

The existence of the probability $P_{a,\pi}$ was discussed in Section 1.4 of Chapter 1, and obtained via a Bochner's type theorem in the space of Schwartz tempered distributions (we may also use the Lebesgue space $L^2(]0,T[,\mathbb{R}^d)$, for $T>0$).

The expression of the characteristic function contains most of the properties of a Lévy space. For instance, we can be construct $P_{a,\pi}$ as the product $P_a \times P_\pi$ of two probabilities, one corresponding to the first exponential (called Wiener white noise, if a is the identity matrix)

$$\exp\Big(-\frac{1}{2}\int_{\mathbb{R}} ax(t)\cdot x(t)\mathrm{d}t\Big),$$

which has support in $C([0,\infty),\mathbb{R}^d)$, and another one corresponding to the second exponential (called compensated Poisson noise)

$$\exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}x(t)\cdot y}-1-\mathtt{i}\mathbb{1}_{\{|y|<1\}}\,x(t)\cdot y\big]\pi(\mathrm{d}y)\Big).$$

The canonical process corresponding to P_a and P_π, denoted by $w(t)$ and $\bar{p}(t)$, are independent. Moreover, they may be assumed to take valued in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. The topological space $\Omega = C([0,\infty),\mathbb{R}^n)\times D([0,\infty),\mathbb{R}^m)$ with the probability $P = P_w \times P_{\bar{p}}$ on the Borel σ-algebra $\mathcal{F}$ and the two canonical process w and $\bar{p}$ is called the canonical Wiener-Poisson space.

On the other hand, also the process

$$t \mapsto \frac{\exp\big[\mathtt{i}x\cdot\ell(t)\big]}{\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}x\cdot\ell(t)}\big\}}$$

is a complex-valued martingale, where

$$\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}x\cdot\ell(t)}\big\} = \exp\Big(-\frac{t}{2}ax\cdot x + t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}x\cdot y}-1-\mathtt{i}\mathbb{1}_{\{|y|<1\}}\,x\cdot y\big]\pi(\mathrm{d}y)\Big),$$

for any x in $\mathbb{R}^d$. The process ℓ is a $\mathbb{R}^d$-valued martingale itself, with $\ell(0)=0$, and $\ell = w+\bar{p}$, where w is a Wiener process (continuous martingale) and $\tilde{p}$ is a *compensated Poisson process* (purely discontinuous martingale).

A generalization of this to infinite-dimensional spaces involves Sazonov's Theorem 1.25 and Minlos' Theorem 1.26, and the concept of nuclear operators,

see Section 1.4 in Chapter 1. For instance, a Wiener random measure and a (compensated) Poisson random measure are constructed as follows, replacing $\mathbb{R}^d$ by and L^2 space. Given a (nonnegative) Radon measure m on $\mathbb{R}^d$, we get a probability measure P_m on $L^2_m = L^2(\mathbb{R}\times\mathbb{R}^d, \mathrm{d}t\times \mathrm{d}m)$ such that

$$\mathbb{E}\big\{\mathrm{e}^{\mathrm{i}(\cdot,h)}\big\} = \exp\Big(-\frac{1}{2}\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}|h(t,x)|^2 m(\mathrm{d}x)\Big),\quad \forall h\in L^2_m,$$

where $(\cdot,\cdot)$ denotes the scalar product in L^2_m. Then we choose a continuous version $w(t,B)$ of the stochastic process $(t,B)\mapsto(\cdot,\mathbb{1}_{(0,t)}\mathbb{1}_B)$, $t\geq 0$, B in $\mathcal{B}(\mathbb{R}^d)$ and bounded. Thus, $t\mapsto w(t,B)/m(B)$ is a standard Wiener process, and $B\mapsto w(t,B)$ is a (random) measure. Moreover, if $B_1,\ldots,B_n$ are disjoint sets then $w(t,B_1),\ldots,w(t,B_n)$ are independent processes. Similarly, given a σ-finite measure π in $\mathbb{R}^d_*$ as in Definition 2.19, we get a probability measure P_π on $L^2_\pi(\mathbb{R}\times\mathbb{R}^d_*)$ with the product measure $\mathrm{d}t\times\pi(\mathrm{d}y)$ such that

$$\mathbb{E}\big\{\mathrm{e}^{\mathrm{i}(\cdot,\phi)}\big\} = \exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathrm{i}\phi(t,y)}-1-\mathrm{i}\phi(t,y)\big]\pi(\mathrm{d}y)\Big),$$

for any function ϕ in $L^2_\pi(\mathbb{R}\times\mathbb{R}^d_*)$, where now $(\cdot,\cdot)$ denotes the scalar product in $L^2_\pi(\mathbb{R}\times\mathbb{R}^d_*)$. Therefore, we can justify the use of $\phi(t,y)=\mathbb{1}_{(a,b)}(t)\mathbb{1}_B(y)$, and then we choose a cad-lag version $\tilde p(t,B)$ of the stochastic process $(t,B)\mapsto(\cdot,\mathbb{1}_{(0,t)}\,\mathbb{1}_B)$, $t\geq 0$, B in $\mathcal{B}(\mathbb{R}^d_*)$, with $\bar B\cap\{0\}=\emptyset$, $\bar B$ is the closure. Moreover, $B\mapsto\tilde p(t,B)$ is a (random) measure, and if $B_1,\ldots,B_n$ are disjoint sets then $\tilde p(t,B_1),\ldots,\tilde p(t,B_n)$ are independent processes. Actually, $p(t,B)=\tilde p(t,B)+t\pi(B)$ is a (Poisson) integer-valued measure because

$$\mathbb{E}\big\{\mathrm{e}^{\mathrm{i}rp(t,B)}\big\} = \exp\Big(t\pi(B)\big[\mathrm{e}^{\mathrm{i}r}-1\big]\Big),\quad \forall r\in\mathbb{R},\tag{2.28}$$

for any B in $\mathcal{B}(\mathbb{R}^d_*)$, with $\bar B\cap\{0\}=\emptyset$, and any $t\geq 0$.

• *Remark* 2.20. First recall the separability of the σ-algebra $\mathcal{B}(\mathbb{R}^d)$ or $\mathcal{B}(\mathbb{R}^d_*)$, i.e., there is an increasing sequence of finite σ-algebras $\{\mathcal{B}_k\}$ such that $\mathcal{B}=\vee_k\mathcal{B}_k$, e.g. see Malliavin [115, Section 6.1, pp. 219–220]. It is clear now that we are able to show that for any t in a countable set and for each ω outside of a null set, the function $B\mapsto w(t,B,\omega)$ (or $B\mapsto\tilde p(t,B,\omega)$) is a (positive) measure on any $\mathcal{B}_k$, $k\geq 1$. Hence, we can take a *version* so that for any B in $\mathcal{B}$ the process $t\mapsto w(t,B)$ (or $t\mapsto\tilde p(t,B)$) is continuous or cad-lag, and for any $t\geq 0$ the set function $B\mapsto w(t,B)$ (or $B\mapsto\tilde p(t,B)$) is a measure on $\mathcal{B}(\mathbb{R}^d)$ (or $\mathcal{B}(\mathbb{R}^d_*)$, respectively). Actually, w and $\tilde p$ are random measures in both variables, i.e., in $\mathbb{R}\times\mathbb{R}^d$. Note that sometimes it is convenient to use the notation $p(B,t)$, $\tilde p(B,t)$ and $\bar p(B,t)$, i.e., we may exchange the order of the variable t and B as long no confusion is made. □

As discussed later to study the jumps, we may use the construction of the $\mathbb{R}^d$-valued compensated Poisson process $\bar p(t)$ or the compensated Poisson "point" process if the emphasis is on the jumps $\delta p(s)=p(s)-p(s-)$. We define the

$\mathbb{R}^d$-valued Poisson measure

$$\bar{p}(t,B) = \sum_{0<s\le t} [\bar{p}(s) - \bar{p}(s-)] \mathbb{1}_{\{\bar{p}(s)-\bar{p}(s-)\in B\}}, \qquad \forall B \in \mathcal{B}_*,$$

where the sum has a finite number of terms and $\mathcal{B}_*$ denotes the ring of Borel sets B in $\mathcal{B}(\mathbb{R}^d_*)$ satisfying $\bar{B} \cap \{0\} = \emptyset$, $\bar{B}$ is the closure. We have

$$\mathbb{E}\{e^{\mathtt{i}x\cdot\bar{p}(t,B)}\} = \exp\Big(t \int_B [e^{\mathtt{i}x\cdot y} - 1]\pi(\mathrm{d}y)\Big), \quad \forall x \in \mathbb{R}^d,\ B \in \mathcal{B}_*,$$

which implies

$$\mathbb{E}\{x \cdot \bar{p}(t, B \cap \{|y| < 1\})\} = t \int_B x \cdot y\, \mathbb{1}_{\{|y|<1\}} \pi(\mathrm{d}y), \quad \forall x \in \mathbb{R}^d,\ B \in \mathcal{B}_*,$$

for any $t \ge 0$.

Sometimes, instead of using the (Poisson) point processes $\bar{p}(t)$ or (Poisson) vector-valued measure $\bar{p}(t,B)$, we prefer to use the (Poisson) counting (integer) measure

$$p(t,B) = p(]0,t] \times B) = \sum_{0<s\le t} \mathbb{1}_{\{\bar{p}(s)-\bar{p}(s-)\in B\}}, \quad \forall B \in \mathcal{B}_*,$$

which is a Poisson process with parameter $\pi(B)$, i.e., (2.28) holds for any B in $\mathcal{B}_*$, or equivalently

$$P\{p(t,B) = n\} = \frac{\big(t\pi(B)\big)^n}{n!} e^{-t\pi(B)}, \quad \forall B \in \mathcal{B}_*,\ n = 0, 1, \ldots,$$

for any $t > 0$. Moreover, because there are a finite number of jumps within B, the integral

$$\bar{p}(t,B) = \int_{]0,t]\times B} z p(\mathrm{d}t, \mathrm{d}z), \quad \forall B \in \mathcal{B}_*,\ t > 0$$

is finite and reproduces the $\mathbb{R}^d$-valued Poisson measure initially defined. To reproduce $\bar{p}(t)$ on this context, we should make sense to the limit

$$\bar{p}(t) = \bar{p}(t, \{|y| \ge 1\}) + \lim_{\varepsilon\to 0} \big[\bar{p}(t, \{\varepsilon \le |y| < 1\}) - t\pi(\{\varepsilon \le |y| < 1\})\big],$$

by means of the stochastic integral. All theses facts are particular cases of the theory of random measures, martingale theory and stochastic integral.

2.6.1 Gaussian Processes

A $\mathbb{R}^d$-valued random variable ξ is *Gaussian distributed* (also called *normally distributed*) with parameters (c, C) if its (complex-valued) characteristic function has the following form

$$\mathbb{E}\{\exp(\mathtt{i}\,\lambda\cdot\xi)\} = \exp(\mathtt{i}\,\lambda\cdot c - \lambda\cdot C\lambda/2), \quad \forall \lambda \in \mathbb{R}^d,$$

or equivalently if its distribution has the form

$$P(\xi \in B) = \int_B \left[(2\pi)^n \sqrt{\det(C)}\right]^{-1/2} \exp\big(-\frac{C^{-1}(x-c)\cdot(x-c)}{2}\big)\mathrm{d}x,$$

for every Borel subset of $\mathbb{R}^d$, where c is the (vector) mean $\mathbb{E}\{\xi\}$ and C is the (matrix) covariance $\mathbb{E}\{(\xi-c)^2\}$. When $c=0$ the random variable ξ is called *centered* or symmetric. Notice that the expression with the characteristic function make sense even if C is only a symmetric nonnegative definite matrix, which is preferred as the definition of Gaussian variable. It is clear that a $\mathbb{R}^d$-valued Gaussian variable has moments of all orders and that a family of centered $\mathbb{R}^d$-valued Gaussian variables is independent if and only if the family is orthogonal in $L^2(\Omega,\mathcal{F},P)$. Next, an infinite sequence $(\xi_1,\xi_2,\dots)$ of real-valued (or $\mathbb{R}^d$-valued) random variables is called *Gaussian* if any (finite) linear combination $c_1\xi_1+\cdots+c_n\xi_n$ is a Gaussian variable. Finally, a probability measure μ on the Borel σ-algebra $\mathcal{B}$ of a separable Banach space B is called a *(centered) Gaussian measure* if any continuous linear functional h is (centered) Gaussian real-valued random variable when considered on the probability space $(B,\mathcal{B},\mu)$. If $B{=}H$ a separable Hilbert space then the mean c value and covariance C operator are well defined, namely,

$$(c,h) = \int_H (h,x)\mu(\mathrm{d}x), \quad \forall h\in H,$$
$$(Ch,k) = \int_H (h,x)(k,x)\mu(\mathrm{d}x) - (c,h)(c,k), \quad \forall h,k\in H,$$

where $(\cdot,\cdot)$ is the inner product in H. Moreover, the covariance C operator is a trace-class operator, i.e., for any (or some) orthonormal basis $\{e_n : n\ge n\}$ in H the series $\mathrm{Tr}(C)=\sum_n (Ce_n,e_n)$ converges.

A fundamental result is the following Fernique's bound

$$\int_B \mathrm{e}^{\lambda\,\|x\|^2}\,\mu(\mathrm{d}x) \le \mathrm{e}^{16\,\lambda\,r^2} + \frac{\mathrm{e}^2}{\mathrm{e}^2-1}, \tag{2.29}$$

valid for any centered Gaussian measure μ on the separable Banach space $(B,\mathcal{B})$ and any λ, $r>0$ such that

$$\ln\big(1-\mu(\{x:\|x\|\le r\})\big) + 32\,\lambda\,r \le \ln\big(\mu(\{x:\|x\|\le r\})\big) - 1,$$

where $\|\cdot\|$ is the norm in B.

In particular, any continuous linear functional φ on B has a finite second moment. Thus, the dual space B' of B can be identified with a subspace of $L^2(B,\mathcal{B},\mu)$ and call $\bar{B}'$ the Hilbert space obtained as the closure of this subspace. Recalling that any φ in B' is a centered Gaussian variable with covariance $|\varphi|^2_{L^2}$, we define the mapping J by setting

$$J : L^2(B,\mathcal{B},\mu) \to B,$$
$$J(\varphi) = \int_B x\,\varphi(x)\,\mu(\mathrm{d}x), \quad \forall \varphi\in L^2(B,\mathcal{B},\mu),$$

but we consider J only acting from $\bar{B}' \subset L^2(B, \mathcal{B}, \mu)$ into B. Since the linearity and continuity of φ and Hölder inequality yield

$$\varphi(J(\varphi)) = \int_B |\varphi(x)|^2 \, \mu(\mathrm{d}x), \quad \text{and} \quad \|J(\varphi)\|^2 \le |\varphi|_{L^2}^2 \int_B \|x\|^2 \, \mu(\mathrm{d}x),$$

so that the mapping J is one-to-one, continuous and linear. The image $H = J(\bar{B}')$ is continuously embedded in B as a Hilbert space with the inner product

$$(f, g)_H = \int_B J^{-1}(f)(x) \, J^{-1}(g)(x) \, \mu(\mathrm{d}x), \quad \forall f, g \in H.$$

Moreover, any φ in the dual space B' is a centered Gaussian random variable on $(B, \mathcal{B}, \mu)$ with covariance $|\varphi|_H^2$, where the dual norm is given by $|\varphi|_H^2 = \sup\{|\varphi(x)| : |x|_H \le 1\}$. The space $H = H_\mu$ is called a *reproducing kernel space* for the centered Gaussian measure $(B, \mathcal{B}, \mu)$. Now, denote by H_μ^0 the image of B' by J, i.e., $H_\mu^0 = J(B')$, which is dense in H_μ.

Let $\{e_1, e_2, \ldots\}$ be a orthonormal basis in H_μ with elements in H_μ^0, and let $\{\xi_1, \xi_2, \ldots\}$ be a sequence of independent real-valued random variables with standard normal distribution (i.e., Gaussian with parameters $0, 1$) relative to a (complete) probability space $(\Omega, \mathcal{F}, P)$. Then, it can be proved that the sequence of partial sums $\{\sum_{k=1}^n \xi_k \, e_k : n = 1, 2, \ldots\}$ converges almost surely in B to a random variable ξ with law μ. Notice that the above series does not converges almost surely in H_μ, but the map $h \mapsto X(h) = \sum_k \xi_k \, (h, e_k)$ is well defined for any h in H_μ, and called *white noise*, see Da Prato and Zabczyk [28, Theorems 1.2.6–12, pp. 37–48].

This procedure can be done backward, i.e., starting from the Hilbert space H. With respect to the previous construction, now H is the dual space of H_μ. Pick an orthonormal basis $\{e_1, e_2, \ldots\}$ in H and a sequence $\{\xi_1, \xi_2, \ldots\}$ of independent real-valued random variables with standard normal distribution in a (complete) probability space $(\Omega, \mathcal{F}, P)$. Since

$$\mathbb{E}\{[\sum_{k=\ell}^n \xi_k \, (h, e_k)]^2\} = \sum_{k=\ell}^n [(h, e_k)]^2 \mathbb{E}\{|\xi_k|^2\} = \sum_{k=\ell}^n [(h, e_k)]^2,$$

for every $n > \ell \ge 1$, we may define $X(h) = \sum_k \xi_k \, (h, e_k)$, for any h in H, as a convergent series in $L^2(\Omega, \mathcal{F}, P)$. The map $h \mapsto X(h)$ is linear, $X(h)$ is a centered Gaussian random variable with covariance $\mathbb{E}\{[X(h)]^2\} = |h|^2$, for any h in H. Actually, the series also converges almost surely and $X(h)$ is indeed an equivalence class. The space $\{X(h) : h \in H\}$ is a Gaussian subspace of $L^2(\Omega, \mathcal{F}, P)$, which is isomorphic to H. In particular $\mathbb{E}\{X(f) \, X(g)\} = (f, g)$, for any f and g in H. This show that $X(f)$ is independent of $X(g)$ if and only if f and g are orthogonal (because independence and orthogonality are equivalent in a Gaussian space). The family $\{X(h) : h \in H\}$ is called an *isonormal Gaussian* stochastic process. If $H = L^2(A, \mathcal{A}, \mu)$, where $(A, \mathcal{A}, \mu)$ is a σ-finite measure space, the mapping X is called a *Gaussian measure* or *white noise* with intensity μ on $(A, \mathcal{A})$. When F belongs to $\mathcal{A}$ we write $X(F) = X(\mathbb{1}_F)$. Thus, if F and G

are sets with $\mu(F) < \infty$ and $\mu(G) < \infty$ then $\mathbb{E}\{X(F)\,X(G)\} = \mu(F \cap G)$, and so that $X(F)$ and $X(G)$ are independent when F and G are disjoint. Notice that if $\{F_k : k = 1, 2, \ldots\}$ is a pairwise disjoint sequence of subset in $\mathcal{A}$, $F = \cup_k F_k$ with $\mu(F) < \infty$ then $X(F) = \sum_k X(F_k)$ *almost surely* so that some regularity (as in the case of regular conditional probability) is need to ensure the existence of a good selection, in order that $F \mapsto X(F, \omega)$ is a measure for ω outside of a set of probability zero.

Sometimes, the initial point is a family of centered Gaussian random variables $X = \{X(h) : h \in H\}$ in a complete probability space $(\Omega, \mathcal{F}, P)$, where the index H is a separable Hilbert space, the σ-algebra $\mathcal{F}$ is the smallest complete σ-algebra such that $X(h)$ is measurable for any h in H and $\mathbb{E}\{X(f)\,X(g)\} = (f, g)_H$, for any f and g in H. This is called a Gaussian process on H. Notice that mapping $h \mapsto X(h)$ has to be linear and provides an isometry from H onto a closed subspace of $L^2(\Omega, \mathcal{F}, P)$, where all elements are zero-mean Gaussian random variables.

Consider the Hermite polynomials, which are defined by

$$h_0(x) = 1, \quad h_n(x) = \frac{(-1)^n}{n!} \mathrm{e}^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \mathrm{e}^{-x^2/2}, \; n = 1, 2, \ldots,$$

which satisfies several properties, e.g.,

$$\exp\big[\frac{x^2}{2} - \frac{(x-t)^2}{2}\big] = \sum_{n=0}^{\infty} t^n\, h_n(x), \quad \forall t, x,$$

$h_n' = h_{n-1}$, $(n+1)h_{n+1}(x) = x\,h_n(x) - h_{n-1}(x)$, $h_n(-x) = (-1)^n\,h_n(x)$, $h_n(0) = 0$ if n is odd and $h_{2n}(0) = (-1)^n/(2^n\,n!)$. It is not hard to show that for any two random variables ξ and ζ with joint standard normal distribution we have $\mathbb{E}\{h_n(\xi)\,h_m(\zeta)\} = (\mathbb{E}\{\xi\,\zeta\})/n!$ if $n = m$ and $\mathbb{E}\{h_n(\xi)\,h_m(\zeta)\} = 0$ otherwise. Essentially based on the one-to-one relation between signed measures and their Laplace transforms, we deduce that only the null element ξ in $L^2(\Omega, \mathcal{F}, P)$ (recall that $\mathcal{F}$ is generated by $\{X(h) : h \in H\}$) satisfies $\mathbb{E}\{\xi\,\exp(X(h))\} = 0$, for any h in H. Hence, the space H can be decomposed into an infinite orthogonal sum of subspaces, i.e.,

$$L^2(\Omega, \mathcal{F}, P) = \oplus_{n=0}^{\infty} H_n,$$

where H_n is defined as the subspace of $L^2(\Omega, \mathcal{F}, P)$ generated by the family random variables $\{h_n(X(h)) : h \in H, |h|_H = 1\}$. Thus, H_0 is the subspace of constants and H_1 the subspace generated by $\{X(h) : h \in H\}$. This analysis continues with several applications, the interest reader is referred to Hida et al. [72], Kallianpur and Karandikar [89], Kuo [106], among others.

Going back to our main interest, we take $H = L^2(\mathbb{R}^+)$ with the Lebesgue measure, initially the Borel σ-algebra, and we construct the family of equivalence classes of centered Gaussian random variables $\{X(h) : h \in H\}$ as above. Thus we can pick a random variable $b(t)$ within the equivalence class $X([0, t]) = X(\mathbb{1}_{[0,t]})$. This stochastic process $b = (b(t) : t \geq 0)$ has the following properties:

(1) The process b has *independent increments*, i.e. for any sequence $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n$ the random variables $\{b(t_0),\, b(t_1) - b(t_0), \ldots,\, b(t_n) - b(t_{n-1})\}$ are independent. Indeed, they are independent since $b(t_k) - b(t_{k-1})$ is in the equivalence class $X(]t_{k-1}, t_k])$ which are independent because the interval $]t_{k-1}, t_k]$ are pairwise disjoint.

(2) The process b is a Gaussian process, i.e., for any sequence $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n$ the $\mathbb{R}^{n+1}$-valued random variable $(b(t_0), b(t_1), \ldots, b(t_n))$ is a Gaussian random variables. Indeed, this follows from the fact that $\{b(t_0),\, b(t_1) - b(t_0), \ldots,\, b(t_n) - b(t_{n-1})\}$ is a family of independent real-valued Gaussian random variable.

(3) For each $t > 0$ we have $\mathbb{E}\{b^2(t)\} = t$ and $b(0) = 0$ almost surely. Moreover, using the independence of increments we find that the covariance $\mathbb{E}\{b(t)\, b(s)\} = t \wedge s$.

(4) Given a function f in $L^2(\mathbb{R}^+)$ (i.e., in H) we may pick an element in the equivalence class $X(f\, \mathbb{1}_{[0,t]})$ and define the integral with respect to b, i.e., $X(f\, \mathbb{1}_{[0,t]})$.

(5) The hard part in to show that we may choose the random variables $b(t)$ in the equivalence class $X([0,t])$ in a way that the path $t \mapsto b(t,\omega)$ is continuous (or at least cad-lag) almost surely. A similar question arises when we try to show that $F \mapsto X(\mathbb{1}_F)$ is a measure almost surely. Because $b(t) - b(s)$ is Gaussian, a direct calculation show that $\mathbb{E}\{|b(t) - b(s)|^4\} = 3|t-s|^2$. Thus, Kolmogorov's continuity criterium (i.e., $\mathbb{E}\{|b(t) - b(s)|^\alpha\} \leq C|t-s|^{1+\beta}$ for some positive constants α, β and C) is satisfied. This show the existence of a continuous stochastic process B as above, which is called *standard Brownian motion* or *standard Wiener process*. The same principle can be used with the integral $\langle f, b\rangle(t) = X(f\, \mathbb{1}_{[0,t]})$, as long as f belongs to $L^\infty(\mathbb{R}^+)$. This continuity holds true also for any f in $L^2(\mathbb{R}^+)$, by means of theory of stochastic integral as seen later.

It is clear that we may have several independent copies of a real-valued standard Brownian motion and then define a $\mathbb{R}^d$-valued standard Brownian motion. Moreover, if for instance, the space $L^2(\mathbb{R}, \mathcal{X})$, for some Hilbert $\mathcal{X}$ (or even co-nuclear) space, is used instead of $L^2(\mathbb{R})$ then we obtain the so called cylindrical Brownian motions or space-time Wiener processes, which is not considered here. We may look at B as a random variable with values in the canonical sample space $C = C([0,\infty), \mathbb{R}^d)$, of continuous functions with the locally uniform convergence (a separable metric space), and its Borel σ-algebra $\mathcal{B} = \mathcal{B}(C)$. The law of B in the canonical sample space C define a *unique* probability measure W such that the coordinate process $X(t) = \omega(t)$ is a standard Brownian motion, which is called the *Wiener measure*. Thus $(C, \mathcal{B}, W)$ is referred to as a *Wiener space*.

Generally, a *standard Wiener process* is defined as a real-valued continuous stochastic process $w = (w(t) : t \geq 0)$ such that (1) it has independent increments and (2) its increments $w(t) - w(s)$, $t > s \geq 0$, $k+1, 2, \ldots, d$ are normally distributed with zero-mean and variance $t - s$. This definition is extended to

a d-dimensional process by coordinates, i.e., $\mathbb{R}^d$-valued where each coordinate w_k is a real-valued standard Wiener process and $\{w_1, w_2, \dots, w_n\}$ is a family of independent processes. For any f in $L^\infty(\mathbb{R}^+)$, the integral with respect to the standard Wiener process $w = (w_1, \dots, w_d)$ is defined as a $\mathbb{R}^d$-valued continuous centered Gaussian process with independent increments and independent coordinates such that for any $k = 1, 2, \dots, d$

$$\int_0^t f(s)\,\mathrm{d}w_k(s) = X_k(f\,\mathbb{1}_{[0,t]}),$$
$$\mathbb{E}\Big\{\Big(\int_0^t f(s)\,\mathrm{d}w_k(s)\Big)^2\Big\} = \int_0^t f^2(s)\mathrm{d}s,$$

for any $t \geq 0$. Notice that the second equality specifies the covariance of the process.

Similarly, we can define the Gaussian-measure process $w(t, \cdot)$, by using the Hilbert space $L^2(\mathbb{R}^+ \times \mathbb{R}^d)$ with the product measure $\mathrm{d}t \times m(\mathrm{d}x)$, where $m(\mathrm{d}x)$ is a Radon measure on $\mathbb{R}^d$ (i.e., finite on compact subsets). In this case $w(t, K)$ is a Wiener process with diffusion $m(K)$ (and mean zero) and $w(t, K_1), \dots, w(t, K_n)$ are independent if $K_1, \dots, K_n$ are disjoint. Clearly, this is related with the so-called *white noise measure* (e.g., see Bichteler [11, Section 3.10, pp. 171–186]) and *Brownian sheet* or *space-time* Brownian motion. The reader is referred to Kallianpur and Xiong [90, Chapters 3 and 4, pp. 85–148] for the infinite dimensional case driven by a space-time Wiener process and a Poisson random measure. This requires the study of martingales with values in Hilbert, Banach and co-nuclear spaces, see also Métivier [127].

The following Lévy's characterization of a Wiener process is a fundamental results, for instance see Revuz and Yor [151, Theorem IV.3.6, pp. 150].

Theorem 2.21 (Lévy)**.** *Let X be an adapted $\mathbb{R}^d$-valued continuous stochastic process in a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$. Then X is a Wiener if and only if X is a (continuous) local-martingale and one of the two following conditions is satisfied:*

(1) *$X_i\,X_j$ and $X_i^2 - t$ are local-martingales for any $i, j = 1, \dots, d$, $i \neq j$,*

(2) *for any $f_1, f_2, \dots, f_d$ functions in $L^\infty(\mathbb{R}^+)$ the (exponential) process*

$$Y_f(t) = \exp\Big\{\mathtt{i}\sum_k \int_0^t f_k(s)\,dX_k(s) + \frac{1}{2}\sum_k \int_0^t f_k^2(s)\mathrm{d}s\Big\},$$

defined for every $t \geq 0$, is a (bounded) complex-valued martingale. □

Clearly, condition (1) means that the (matrix-valued) predictable quadratic variation process $\langle X\rangle$ associated with X is such that $\langle X_i, X_i\rangle(t) = t$ and $\langle X_i, X_j\rangle(t) = 0$ for any $i, j = 1, \dots, d$, $i \neq j$. In condition (2) we may also take f_k in $L^2(\mathbb{R}^+)$ and even adapted processes. The assumption on *continuity* is essential to the above Lévy's theorem.

It can be proved that a Gaussian semi-martingale X is continuous if and only if it is stochastically continuous, i.e., $P(|X(t) - X(s)| > \varepsilon)$ goes to zero

as $t \to s$, for any $\varepsilon > 0$. Moreover, a centered Gaussian local-martingale X with $X(0) = 0$ and independent increments, is actually a locally square integrable and its predictable quadratic variation (non-necessarily continuous) satisfies $\langle X\rangle(t) \wedge \langle X\rangle(s) = \mathbb{E}\{X(t)\,X(s)\}$, for any $t \geq s \geq 0$. It is also clear that for a centered Gaussian martingale X with $X(0) = 0$, the covariance matrix $c(t) = (\mathbb{E}\{X_i(t)\,X_j(t)\} : t \geq 0,\, i,j = 1,2,\ldots,d)$ satisfies

$$\mathbb{E}\{\exp[\mathtt{i}\,\lambda \cdot (X(t) - X(s))]\} = \exp[-\lambda \cdot (c(t) - c(s))\lambda/2],$$

for every λ in $\mathbb{R}^d$ and $t \geq s \geq 0$. This property completely characterizes the finite distributions of X, see Liptser and Shiryayev [111, Section 4.9, pp. 270–306].

The Ornstein-Uhlenbeck process is another typical example of Gaussian process that is given by

$$X(t) = \exp(-\alpha t)X_0 + \int_0^t \exp[-\alpha(t-s)]\sigma \mathrm{d}w(s), \quad \forall t \geq 0,$$

where α and σ are matrices, α has positive eigenvalues, X_0 is an initial random variable normally distributed and w is an standard Wiener process. Even more general, if $\Phi(t,s)$ denotes the fundamental (matrix) solution of a linear ordinary differential equation with matrix $\alpha(t)$, i.e.,

$$\dot{\Phi}(t,s) = -\alpha(t)\Phi(t,s), \quad \forall t \neq s, \quad \text{and} \quad \Phi(s,s) = 1, \quad \forall s,$$

then

$$X(t) = \Phi(t,0)X_0 + \int_0^t \Phi(t,s)\sigma(s)\mathrm{d}w(s), \quad \forall t \geq 0,$$

is a Gaussian process with mean $m_i(t) = \mathbb{E}\{X_i(t)\}$ and covariance matrix $v_{ij}(s,t) = \mathbb{E}\{[X_i(s) - m_i(s)][X_j(t) - m_j(t)]\}$, which can be explicitly calculated. For instance, in the one-dimensional case with constant α and σ we have

$$\begin{aligned} m(t) &= \mathbb{E}\{X(t)\} = \mathrm{e}^{-\alpha t}m(0), \\ v(s,t) &= \mathbb{E}\{[X(s) - m(s)][X(t) - m(t)]\} = \\ &= \Big\{\frac{\sigma^2}{2\alpha}\big[\mathrm{e}^{2\alpha(s\wedge t)} - 1\big] + v(0)\Big\}\mathrm{e}^{-\alpha(s+t)}. \end{aligned}$$

Therefore, if the initial random variable has mean zero and the variance is equal to $v_0 = \sigma^2/(2\alpha)$, then X is a stationary, zero-mean Gaussian process with covariance function $\rho(s,t) = v_0 \exp(-\alpha|t-s|)$.

2.6.2 Compensated Poisson Processes

A $\mathbb{R}^d$-valued random variable ξ has a *compensated Poisson distributed* (also called *centered Poisson distributed*) with parameter π if its (complex-valued) characteristic function has the following form

$$\mathbb{E}\{\exp(\mathtt{i}\,\lambda \cdot \xi)\} = \exp\Big(\int_{\mathbb{R}^d_*} \big[\mathrm{e}^{\mathtt{i}\,\lambda\cdot x} - 1 - \mathtt{i}\,\lambda \cdot x\big]\pi(\mathrm{d}x)\Big), \quad \forall \lambda \in \mathbb{R}^d,$$

where π is a Radon measure on $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$ satisfying

$$\int_{\mathbb{R}^d_*} |x|^2 \pi(\mathrm{d}x) < \infty.$$

Usually, the arguments begin with a compound Poisson variable p in $\mathbb{R}^d$ (mainly, d=1) with a finite measure π as parameter, i.e.,

$$\mathbb{E}\{\exp(\mathtt{i}\,\lambda \cdot p)\} = \exp\Big(\int_{\mathbb{R}^d_*} \big[\mathrm{e}^{\mathtt{i}\,\lambda\cdot x} - 1\big]\pi(\mathrm{d}x)\Big), \quad \forall \lambda \in \mathbb{R}^d,$$

Then define $\xi = p - \mathbb{E}\{p(t)\}$ as a centered Poisson distribution random variable, i.e.,

$$\mathbb{E}\{\exp(\mathtt{i}\,\lambda \cdot \xi)\} = \exp\Big(\int_{\mathbb{R}^d_*} \big[\mathrm{e}^{\mathtt{i}\,\lambda\cdot x} - 1 - \lambda\cdot x\big]\pi(\mathrm{d}x)\Big), \quad \forall \lambda \in \mathbb{R}^d.$$

Next, the construction and properties of the compensated Poisson (or centered Poisson) random variable ξ are extended for characteristic measures π as above.

It is called *symmetric* if π satisfies

$$\int_{\mathbb{R}^d_*} \big[\mathrm{e}^{\mathtt{i}\,\lambda\cdot x} - 1 - \mathtt{i}\,\lambda\cdot x\big]\pi(\mathrm{d}x) = \int_{\mathbb{R}^d_*} \big[\mathrm{e}^{-\mathtt{i}\,\lambda\cdot x} - 1 + \mathtt{i}\,\lambda\cdot x\big]\pi(\mathrm{d}x),$$

for every λ in $\mathbb{R}^d$. It is clear that a $\mathbb{R}^d$-valued compensated Poisson variable ξ has finite first and second moments, i.e.,

$$\mathbb{E}\{|\xi|^2\} = \int_{\mathbb{R}^d_*} |x|^2 \pi(\mathrm{d}x),$$

and if we add the condition

$$\int_{\mathbb{R}^d_*} |x|^p \pi(\mathrm{d}x) < \infty, \quad \forall p \geq 2.$$

then all moments are finite.

An infinite sequence $(\xi_1, \xi_2, \ldots)$ of real-valued (or $\mathbb{R}^d$-valued) random variables is called *compensated Poisson* process if any (finite) sum $\xi_1 + \cdots + \xi_n$ is a compensated Poisson variable (the sequence is necessarily independent). Next, given a (nuclear) countably Hilbertian space $\Phi = \bigcap_{n\geq 0} \Phi_n$, its dual space $\Phi' = \bigcup_{n\geq 0} \Phi_{-n}$ (see Section 1.4 in Chapter 1), a probability measure μ on the Borel σ-algebra $\mathcal{B}(\Phi')$ is called a *compensated Poisson measure* if $\langle \cdot, \varphi\rangle$ is a compensated Poisson real-valued random variable, for any φ in Φ, when considered on the probability space $(\Phi', \mathcal{B}(\Phi'), \mu)$, i.e., there exists a σ-finite measure π on $\Phi'_* = \Phi' \setminus \{0\}$ such that

$$\mathbb{E}_\mu\big\{\mathrm{e}^{\mathtt{i}\langle\cdot,\varphi\rangle}\big\} = \exp\Big(\int_{\Phi'_*} \big[\mathrm{e}^{\mathtt{i}\langle\cdot,\varphi\rangle} - 1 - \mathtt{i}\langle\cdot,\varphi\rangle\big]\mathrm{d}\pi\Big), \quad \forall \varphi \in \Phi.$$

Similarly to the finite-dimensional case, besides the condition

$$\int_{\Phi'_*} |\langle \cdot, \varphi \rangle|^2 \mathrm{d}\pi < \infty, \quad \forall \varphi \in \Phi,$$

we assume that

$$\int_{\Phi'_*} |\langle \cdot, \varphi \rangle|^2 \mathrm{d}\pi \le C_0 \|\varphi\|_n^2, \quad \forall \varphi \in \Phi, \tag{2.30}$$

for some $n \ge 0$ and some constant $C_0 > 0$.

• *Remark* 2.22. Minlos' Theorem 1.26 ensures the existence of a probability measure μ for any given σ-finite measure such that

$$\varphi \mapsto \int_{\Phi'_*} \left[\mathrm{e}^{\mathtt{i}\langle \cdot, \varphi \rangle} - 1 - \mathtt{i}\langle \cdot, \varphi \rangle \right] \mathrm{d}\pi$$

is continuous, in particular if (2.30) holds. Note that (2.30) is equivalent to the condition that

$$\varphi \mapsto \int_{\Phi'_*} |\langle \cdot, \varphi \rangle|^2 \mathrm{d}\pi$$

is continuous. However, if we wish to replace the space Φ by a Banach space B some difficulties appears and we cannot guarantee the existence of a probability measure μ, e.g., see Rudiger [155]. □

Under the assumption (2.30), there is a separable Hilbert space $\Phi \subset H \subset \Phi_0$, with continuous and dense inclusion, and a nonnegative symmetric trace-class operator R in $L_1(H)$ (i.e., $R^{1/2}$ is a Hilbert-Schmidt operator), such that the support of π is included in $R(H) \subset H \subset \Phi_0$, i.e.,

$$\pi\big(\{\chi \in \Phi' : \langle \chi, \varphi \rangle \le r\}\big) = \pi\big(\{h \in R(H) : \langle h, \varphi \rangle \le r\}\big), \quad \forall \varphi \in \Phi, r \in \mathbb{R},$$

and defining

$$\pi_0(B) = \pi(\{h \in H_* : R^{-1}h \in B\}), \quad \forall B \in \mathcal{B}(H_*)$$

or equivalently $\pi_0 = R\pi$, with $H_* = H \setminus \{0\}$, we have

$$\int_{\Phi'} \mathrm{e}^{\mathtt{i}\langle \cdot, \varphi \rangle} \mathrm{d}\mu = \exp\Big(\int_{H_*} \left[\mathrm{e}^{\mathtt{i}\langle R\cdot, \varphi \rangle} - 1 - \mathtt{i}\langle R\cdot, \varphi \rangle \right] \mathrm{d}\pi_0 \Big), \quad \forall \varphi \in \Phi.$$

The integrability condition becomes

$$\int_{H_*} |(R\cdot, \varphi)|^2 \mathrm{d}\pi_0 \le C_0 \|\varphi\|_H^2, \quad \forall \varphi \in H,$$

for some constant $C_0 > 0$, which yields

$$\int_{H_*} (Rh, h) \pi_0(\mathrm{d}h) \le C_0 \operatorname{Tr}(R) < \infty.$$

Hence, Sazonov's Theorem 1.25 shows that μ is actually supported in H, i.e., μ is a *compensated Poisson measure* with parameter $\pi = R^{-1}\pi_0$ satisfying

$$\int_H \mathrm{e}^{\mathtt{i}(h,k)}\mu(\mathrm{d}k) = \exp\Big(\int_{H_*}\big[\mathrm{e}^{\mathtt{i}(Rh,k)} - 1 - \mathtt{i}(Rh,k)\big]\pi_0(\mathrm{d}k)\Big), \quad \forall h \in H.$$

Thus, by working on a nuclear countably Hilbertian space we are reduced to the case of a Hilbert space. Now, we can justify

$$\mathbb{E}_\mu\big\{\langle\cdot,\varphi\rangle\big\} = 0 \quad \text{and} \quad \mathbb{E}_\mu\big\{|\langle\cdot,\varphi\rangle|^2\big\} = \int_{H_*}|\langle Rh,\varphi\rangle|^2\pi_0(\mathrm{d}h), \quad \forall\varphi\in\Phi,$$

actually, we may take φ in H, replace the duality $\langle\cdot,\cdot\rangle$ by $(\cdot,\cdot)$, and assume $H = \Phi_0$.

Hence, the map $\varphi \mapsto \langle\cdot,\varphi\rangle$ allows us to identify the space H with a subspace of $L^2(\Phi',\mathcal{B}(\Phi'),\mu) = L^2(H,\mathcal{B}(H),\mu)$ and then to call $\bar H$ the Hilbert space obtained as the closure of this subspace. Recalling that any φ in H the random variable $\langle\cdot,\varphi\rangle$ is a compensated Poisson variable with with parameter π, we define the mapping J by setting

$$J : L^2(H,\mathcal{B}(H),\mu) \to H, \qquad J(\varphi) = \int_H h\,\varphi(h)\,\mu(\mathrm{d}h),$$

but we consider J only as being from $\bar H \subset L^2(H,\mathcal{B}(H),\mu)$ into H. Since the linearity and continuity of φ and Hölder inequality yield

$$\begin{aligned}
&\varphi(J(\varphi)) = \int_H |\varphi(x)|^2\,\mu(\mathrm{d}x),\\
&\|J(\varphi)\|_H^2 \le \Big(\int_H |\varphi(x)|^2\,\mu(\mathrm{d}x)\Big)\Big(\int_H \|h\|_H^2\,\mu(\mathrm{d}h)\Big),
\end{aligned}$$

the mapping J is one-to-one, continuous and linear. The image $H_\mu = J(\bar H)$ is continuously embedded in H as a Hilbert space with the inner product

$$(f,g)_{_\mu} = \int_H J^{-1}(f)(h)\,J^{-1}(g)(h)\,\mu(\mathrm{d}h), \quad \forall f,g\in H_\mu.$$

Now, set $H_\mu^0 = J(H)$, which is dense in $H_\mu \subset H \subset \bar H$. Clearly, if f and g belong to H_μ^0 then $(f,g)_{_\mu} = (J^{-1}f, J^{-1}g) = \langle J^{-1}f, J^{-1}g\rangle$.

Let $\{e_1,e_2,\ldots\}$ be an orthonormal basis in H_μ with elements in H_μ^0, and for every h in H, consider the expression $X = \sum_j\langle h, J^{-1}e_j\rangle$, which is a sum of independent real-valued random variables $\xi_j(\omega) = \langle\omega, J^{-1}e_j\rangle$, with joint compensated Poisson distribution

$$\mathbb{E}_\mu\big\{e^{\mathtt{i}\sum_{j=1}^n c_j\xi_j}\big\} = \int_{\mathbb{R}_*^n}\big[\mathrm{e}^{\mathtt{i}\sum_{j=1}^n c_j s_j} - 1 - \mathtt{i}\sum_{j=1}^n c_j s_j\big]\bar\pi_n(\mathrm{d}s), \quad \forall c\in\mathbb{R}^n,$$

where $\bar\pi_n$ is the projection on $\mathbb{R}_*^n$ of π, i.e., with $h_j = (h, J^{-1}e_j)$,

$$\bar\pi_n(B) = \pi\big(\{h\in H_* : (h_1,\cdots,h_n)\in B,\, h_j = 0,\, \forall j > n\}\big), \quad \forall B\in\mathcal{B}(\mathbb{R}_*^n).$$

Thus

$$\mathbb{E}_\mu\{\sum_{j=1}^n |\xi_j|^2\} = \int_{\mathbb{R}_*^n} \Big(\sum_{j=1}^n s_j^2\Big)\bar\pi_n(\mathrm{d}s) = \int_{H_*} \langle Rh^n, h^n\rangle \pi_0(\mathrm{d}h),$$

where $h^n = \sum_{j=1}^n \langle h, J^{-1}e_j\rangle e_j$. Hence, the series $X = \sum_{j=1}^\infty \xi_j(\omega)e_j$ converges in $\bar H \subset L^2(H, \mathcal{B}(H), \mu)$, i.e., it can be considered as a $\bar H$-valued random variable on the probability space $(\Omega, \mathcal{F}, P) = (H, \mathcal{B}(H), \mu)$. Because $\{e_1, e_2, \ldots\}$ is an orthonormal basis in H_μ, the mapping

$$X(h) = \langle X, h\rangle = \sum_{j=1}^n \xi_j \, \langle J^{-1}e_j, h\rangle = \sum_{j=1}^n \xi_j \, (e_j, Jh)_\mu$$

is a H_μ-valued random variable (almost surely) well defined for any $h = J^{-1}Jh$ in H_μ, and called *Poisson white noise*.

Let $\{\xi_1, \xi_2, \ldots\}$ be a sequence of independent real-valued compensated Poisson random variables with parameters $\{\pi_1, \pi_2, \ldots\}$ in (complete) probability space $(\Omega, \mathcal{F}, P)$, i.e.,

$$\mathbb{E}\big\{e^{\mathtt{i}r\xi_j}\big\} = \exp\Big(\int_{\mathbb{R}_*} \big[\mathrm{e}^{\mathtt{i}rs} - 1 - \mathtt{i}rs\big]\pi_j(\mathrm{d}s)\Big), \quad \forall r \in \mathbb{R},\, j \ge 1,$$

with π_j satisfying

$$\int_{\mathbb{R}_*} s^2\pi_j(\mathrm{d}s) \le C_0, \quad \forall j \ge 1, \tag{2.31}$$

for some constant $C_0 > 0$. Now, for any given sequence of nonnegative real numbers $r = \{r_1, r_2, \ldots\}$, define the measures $\bar\pi_{r,n}$ and π_{j,r_j} on $\mathbb{R}^n$ as

$$\int_{\mathbb{R}^n} f(s)\bar\pi_{r,n}(\mathrm{d}s) = \sum_{j=1}^n \int_{\mathbb{R}} f_j(\sqrt{r_j}s_j)\pi_j(\mathrm{d}s_j) = \sum_{j=1}^n \int_{\mathbb{R}^n} f(s)\pi_{j,r_j}(\mathrm{d}s),$$

for any $n \ge 1$ and for every positive Borel function f in $\mathbb{R}^n$ satisfying $f(0) = 0$, where $s = (s_1, \ldots, s_n)$ and $f_1(s_1) = f(s_1, 0, \ldots, 0)$, $f_2(s_2) = f(0, s_2, \ldots, 0)$, $\ldots$, $f_n(s_n) = f(0, 0, \ldots, s_n)$, i.e.,

$$\pi_{j,r_j}(\mathrm{d}s) = \delta_0(\mathrm{d}s_1)\ldots\delta_0(\mathrm{d}s_{j-1})\pi_j(r_j^{-1/2}\mathrm{d}s_j)\delta_0(\mathrm{d}s_{j+1})\ldots\delta_0(\mathrm{d}s_n),$$

where δ_0 is the Dirac measure at 0 and $\pi_j(r_j^{-1/2}\mathrm{d}s_j) = 0$ if $r_j = 0$. We can check that $\bar\xi_{r,n} = (\sqrt{r_1}\xi_1, \ldots, \sqrt{r_n}\xi_n)$ has a compensated Poisson distribution with parameter $\bar\pi_{r,n}$, i.e.,

$$\mathbb{E}\big\{e^{\mathtt{i}c\cdot\bar\xi_{r,n}}\big\} = \int_{\mathbb{R}_*^n} \big[\mathrm{e}^{\mathtt{i}c\cdot s} - 1 - \mathtt{i}c\cdot s\big]\bar\pi_{r,n}(\mathrm{d}s), \quad \forall c \in \mathbb{R}^n,$$

where the dot "$\cdot$" denotes the scalar product in $\mathbb{R}^n$. Clearly, (2.31) implies

$$\sum_{j=1}^{n}\int_{\mathbb{R}^n_*}|s_j|^2\bar\pi_{r,n}(\mathrm{d}s)\le C_0\sum_{j=1}^{n}r_j,\quad\forall n\ge 1,$$

with the same constant $C_0>0$.

Moreover, we may regard the measures $\bar\pi_{r,n}$ and π_{j,r_j} as being defined either on $\mathbb{R}^n$ or directly on the infinite product $\mathbb{R}^\infty$ (the space of all sequences), namely,

$$\int_{\mathbb{R}^\infty}f(s)\bar\pi_{r,n}(\mathrm{d}s)=\int_{\mathbb{R}^n}f(s_1,\dots,s_n,0,0,\dots)\bar\pi_{r,n}(\mathrm{d}s)$$

or equivalently,

$$\pi_{j,r_j}(\mathrm{d}s)=\delta_0(\mathrm{d}s_1)\dots\delta_0(\mathrm{d}s_{j-1})\pi_j(r_j^{-1/2}\mathrm{d}s_j)\delta_0(\mathrm{d}s_{j+1})\delta_0(\mathrm{d}s_{j+2})\dots,$$

and $\bar\pi_{r,n}=\sum_{j=1}^n\pi_{j,r_j}$. Note the projection type property

$$\bar\pi_{r,n}(B)=\bar\pi_{r,n+k}\big(\{s\in\mathbb{R}^\infty:(s_1,\dots,s_n)\in B,\,s_j=0,\,j>n\}\big),$$

for any B in $\mathcal{B}(\mathbb{R}^n)$. Therefore, the series $\bar\pi_r=\sum_{j=1}^\infty\pi_{j,r_j}$ defines a measure on $\mathbb{R}^\infty$. Hence, if the series $\sum_{j=1}^\infty r_j$ is convergent then

$$\int_{\mathbb{R}^\infty}|s|^2\bar\pi_r(\mathrm{d}s)=\sum_{j=1}^{\infty}\int_{\mathbb{R}^n_*}|s_j|^2\bar\pi_{r,n}(\mathrm{d}s)\le C_0\sum_{j=1}^{\infty}r_j<\infty,\tag{2.32}$$

i.e., $\bar\pi_r$ becomes a σ-finite measure on $\ell^2_*=\ell^2\setminus\{0\}$, where ℓ^2 is the Hilbert space of square-convergent sequences. Also, we have

$$\int_{\ell^2_*}f(s)\bar\pi_r(\mathrm{d}s)=\lim_n\int_{\ell^2_*}f(s)\bar\pi_{r,n}(\mathrm{d}s)=\sum_{j=1}^{\infty}\int_{\ell^2_*}f(s)\pi_{j,r_j}(\mathrm{d}s),$$

for any continuous function f such that $|f(s)|\le|s|^2$, for any s in ℓ^2_*. Moreover, since $r_j=0$ implies $\pi_{j,r_j}=0$ on ℓ^2_*, we also have $\pi_{j,r_j}(R^{-1}\{0\})=0$ for any j, where R is the nonnegative symmetric trace-class operator $s\mapsto(r_1s_1,r_2s_2,\dots)$. Hence $\bar\pi_r(R^{-1}\{0\})=0$. This means that support of $\bar\pi_r$ is contained in $R(\ell^2_*)$ and we could define a new pre-image measure by setting $\bar\pi_0(B)=\bar\pi_r(RB)$, for any B in $\mathcal{B}(\ell^2_*)$ with the property

$$\int_{\ell^2_*}f(s)\bar\pi(\mathrm{d}s)=\int_{\ell^2_*}f(Rs)\bar\pi_0(\mathrm{d}s),\quad\forall f\ge 0\ \text{ and measurable.}$$

It is clear that estimate (2.32) identifies the measures only on ℓ^2_* and so, we may (re)define all measures at $\{0\}$ by setting

$$\bar\pi_r(\{0\})=\bar\pi_{r,n}(\{0\})=\pi_{j,r_j}(\{0\})=\bar\pi_0(\{0\})=0.$$

Then we can consider the measures as σ-finite defined either on ℓ^2 or on ℓ^2_*.

Now, let H be a separable Hilbert space, R be a nonnegative symmetric (trace-class) operator in $L_1(H)$, and $\{e_1, e_2, \dots\}$ be an orthonormal basis of eigenvectors of R, i.e., $Re_j = r_j e_j$, $(e_j, e_k) = 0$ if $j \neq k$, $|e_j| = 1$, for every j, and $\operatorname{Tr}(R) = \sum_{j=1}^{\infty} r_j < \infty$, $r_j \geq 0$. Note that the kernel of R may be of infinite dimension, i.e., there infinite many $r_j = 0$. Consider the product measure π on $H_* = H \setminus \{0\}$, with support in $R(H)$, defined as

$$\pi\big(\{h \in H_* : (h, e_j) \in Be_j,\ \forall j\}\big) = \bar{\pi}_r(B), \quad \forall B \in \mathcal{B}(H_*)$$

or equivalently

$$\int_{H_*} f(h)\pi(\mathrm{d}h) = \int_{\ell_*} f(s_1 e_1 + \cdots + s_n e_n + \cdots)\bar{\pi}_r(\mathrm{d}s),$$

nonnegative Borel function f in H_*. In particular,

$$\int_H |h|^2 \pi(\mathrm{d}h) = \int_{H_*} \Big|\sum_{j=1}^{\infty} s_j e_j\Big|^2 \bar{\pi}_r(\mathrm{d}s) = \sum_{j=1}^{\infty} \int_{\ell_*} s_j^2 r_j \pi_j(\mathrm{d}s) \leq C_0 \operatorname{Tr}(R)$$

and if $\pi_0 = R\pi$, i.e., $\pi_0(B) = \pi(RB)$, for every B in $\mathcal{B}(H_*)$, then

$$\int_{H_*} f(h)\pi(\mathrm{d}h) = \int_{H_*} f(Rh)\pi_0(\mathrm{d}h),$$

for any nonnegative Borel measurable function f on H_*.

• *Remark* 2.23. Recall the following result, e.g., see Federer [46, Section 2.2.13, pp. 69]. Let X be a complete separable metric space, Y be a Hausdorff space, $f\colon X \to Y$ be a continuous function, and μ be a measure Y such that every closed subset of Y is μ measurable. Then the f image of every Borel subset of X is μ measurable. This result is classic in the general study of Borel σ-algebras, analytic sets and universally measurable sets, i.e., the fact that a projection of a Borel measurable set is analytic and so, it is universally measurable, e.g., see Dellacherie and Meyer [32, Section III.75–85, pp. 243–254], Dudley [37, Section 13.2, pp 493–499] or Parthasarathy [141, Section I.3]. It is now clear that the above measure π_0 can be defined in term of π, provided that π has support contained in $R(H_*)$. Note that for any orthonormal basis $\{e_j\}$ in H and any measure m on H with $m(\{0\}) = 0$ we have $m(B) = \sum_j m(e_j B)$, for any B in $\mathcal{B}(H)$, where $e_j B$ is the (orthogonal) projection of B in the e_j direction, i.e., $e_j B = \{(b, e_j)e_j : b \in B\}$. Thus, for any integrable function f with $f(0) = 0$ we have

$$\int_H f(h)m(\mathrm{d}h) = \sum_j \int_H f(h)m(e_j \mathrm{d}h) = \sum_j \int_H f(e_j h)m(\mathrm{d}h),$$

where $f(e_j h) = f\big((h, e_j)e_j\big)$ and $m(e_j \mathrm{d}h)$ is the measure $B \mapsto m(e_j B)$. □

Therefore, the H-valued random variable

$$X = \sum_{j=1}^{\infty} \sqrt{r_j}\, \xi_j\, e_j$$

satisfies

$$\mathbb{E}\{|X|^2\} = \sum_{j=1}^{\infty} r_j\, \mathbb{E}\{|\xi_j|^2\} = \sum_{j=1}^{\infty} r_j \int_{\mathbb{R}_*} s^2 \pi_j(\mathrm{d}s),$$

and

$$\mathbb{E}\big\{e^{\mathtt{i}(h,X)}\big\} = \prod_{j=1}^{\infty} \mathbb{E}\big\{e^{\mathtt{i}\sqrt{r_j}(h,e_j)\xi_j}\big\} =$$

$$= \exp\Big(\sum_{j=1}^{\infty} \int_{\mathbb{R}_*} \big[\mathrm{e}^{\mathtt{i}\sqrt{r_j}(h,e_j)s_j} - 1 - \mathtt{i}\sqrt{r_j}(h,e_j)s_j\big]\pi_j(\mathrm{d}s_j)\Big) =$$

$$= \exp\Big(\sum_{j=1}^{\infty} \int_{\ell^2_*} \big[\mathrm{e}^{\mathtt{i}(h,e_j)s} - 1 - \mathtt{i}(h,e_j)s\big]\bar{\pi}_r(\mathrm{d}s)\Big),$$

i.e.,

$$\mathbb{E}\big\{e^{\mathtt{i}(h,X)}\big\} = \exp\Big(\int_{H_*} \big[\mathrm{e}^{\mathtt{i}(h,k)} - 1 - \mathtt{i}(h,k)\big]\pi(\mathrm{d}k)\Big) =$$

$$= \exp\Big(\int_{H_*} \big[\mathrm{e}^{\mathtt{i}(Rh,k)} - 1 - \mathtt{i}(Rh,k)\big]\pi_0(\mathrm{d}k)\Big),$$

for every h in H. Thus, X is a compensated Poisson random variable with values in H and Lévy measure π in H_*. Next, the mapping

$$h \mapsto X(h) = \sum_{j=1}^{\infty} \sqrt{r_j}\, \xi_j\, (h, e_j)$$

from H into $L^2(\Omega, \mathcal{F}, P)$ is linear, $X(h)$ is a (real-valued) compensated Poisson random variable with covariance $\mathbb{E}\{[X(h)]^2\} = |h|^2$, for any h in H. Thus the space $\{X(h) : h \in H\}$ is a Poisson subspace of $L^2(\Omega, \mathcal{F}, P)$, which is isomorphic to H. In particular $\mathbb{E}\{X(f)\, X(g)\} = (f, g)$, for any f and g in H, and $X(f)$ is a compensated Poisson variable independent of $X(g)$ if $(f, g) = 0$, i.e., if f and g are orthogonal. The family $\{X(h) : h \in H\}$ is called an *compensated Poisson* stochastic process. If $H = L^2(A, \mathcal{A}, \mu)$, where $(A, \mathcal{A}, \mu)$ is a σ-finite measure space, the mapping X is called a *Poisson measure* or *Poisson white noise* with intensity μ on $(A, \mathcal{A})$. When F belongs to $\mathcal{A}$ we write $X(F) = X(\mathbb{1}_F)$. Thus, if F and G are sets with $\mu(F) < \infty$ and $\mu(G) < \infty$ then $\mathbb{E}\{X(F)\, X(G)\} = \mu(F \cap G)$, and so that $X(F)$ and $X(G)$ are independent when F and G are disjoint. Notice that if $\{F_k : k = 1, 2, \ldots\}$ is a pairwise disjoint sequence of

subset in $\mathcal{A}$, $F = \cup_k F_k$ with $\mu(F) < \infty$ then $X(F) = \sum_k X(F_k)$ *almost surely* so that some regularity (as in the case of regular conditional probability) is need to ensure the existence of a good selection, in order that $F \mapsto X(F, \omega)$ is a measure for ω outside of a set of probability zero.

Sometimes, the initial point is a family of compensated Poisson random variables $X = \{X(h) : h \in H\}$ in a complete probability space $(\Omega, \mathcal{F}, P)$, where the index H is a separable Hilbert space, the σ-algebra $\mathcal{F}$ is the smallest complete σ-algebra such that $X(h)$ is measurable for any h in H and $\mathbb{E}\{X(f)\,X(g)\} = (f,g)_H$, for any f and g in H. This is called a compensated Poisson process on H. For the particular case of a standard Poisson process (and some similar one, like symmetric jumps) we have the so-called Charlier polynomials $c_{n,\lambda}(x)$, an orthogonal basis in $L^2(\mathbb{R}^+)$ with the weight $\alpha(x) = \sum_{n=1}^{\infty} \mathbb{1}_{\{x \geq n\}} \mathrm{e}^{-\lambda} \lambda^n / n!$, $\lambda \neq 0$, which are the equivalent of Hermit polynomials in the case of a Wiener process. Charlier polynomials are defined by the generating function

$$t \mapsto \mathrm{e}^{-\lambda t}(1+t)^x = \sum_{n=0}^{\infty} c_{n,\lambda}(x) \frac{t^n}{n!},$$

or explicitly by the expression

$$c_{n,\lambda}(x) = \sum_{k=0}^{n} \binom{n}{k} \binom{x}{k} k!\,(-\lambda)^{n-k}$$

and they satisfy the orthogonal relations

$$\int_0^{\infty} c_{m,\lambda}(x)\, c_{n,\lambda}(x)\, \mathrm{d}\alpha(x) = \sum_{k=1}^{\infty} c_{m,\lambda}(k)\, c_{n,\lambda}(k)\, \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} = 0, \quad \text{if } m \neq n$$

and

$$\int_0^{\infty} c_{n,\lambda}(x)\, c_{n,\lambda}(x)\, \mathrm{d}\alpha(x) = \sum_{k=1}^{\infty} c_{n,\lambda}(k)\, c_{n,\lambda}(k)\, \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} = \lambda^n n!.$$

Also the three-terms recurrence formula

$$c_{\lambda,n+1}(x) = (x - n - \lambda) c_{\lambda,n}(x) - n\lambda c_{\lambda,n-1}(x),$$

and several other properties, e.g., see Chihara [22] or Szegö[170].

The previous analysis applied to the *particular case* when a Lévy measure π is given on a separable Hilbert space H. The measure π is constructed from a nonnegative symmetric (trace-class) operator R in $L_1(H)$ with eigenvalues and eigenvectors $\{r_j, e_j,\, j = 1, 2, \ldots\}$, where $\{e_j\}$ is a orthonormal basis in H and $\sum_j r_j < \infty$, and a sequence $\{\pi_j\}$ of Lévy measures on $\mathbb{R}_*$ satisfying (2.31). Thus, we define the Lévy measures $\bar{\pi}_{r,n}$ on $\mathbb{R}^n$ satisfying (2.32), which induces the Lévy measures $\bar{\pi}_r$ and $\bar{\pi}$ on ℓ^2_* and π and π_0 on H_*, satisfying

$$\int_{H_*} |h|_H^2 \pi(\mathrm{d}h) = \int_{H_*} |R^{1/2} h|_H^2 \pi_0(\mathrm{d}h) < \infty. \tag{2.33}$$

By means of Sazonov's Theorem 1.25, there is a probability measure P on $(\Omega, \mathcal{F})$, with $\Omega = L^2(\mathbb{R}, H_*)$ and $\mathcal{F} = \mathcal{B}(\Omega)$, such that for any ϕ in $L^2(\mathbb{R}, H_*)$ we have

$$\mathbb{E}\{e^{\mathtt{i}\langle\cdot,\phi\rangle}\} = \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_*} \big[\mathrm{e}^{\mathtt{i}(\phi(t),h)} - 1 - \mathtt{i}(\phi(t),h)\big]\pi(\mathrm{d}h)\Big) = \\ = \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_*} \big[\mathrm{e}^{\mathtt{i}(R\phi(t),h)} - 1 - \mathtt{i}(R\phi(t),h)\big]\pi_0(\mathrm{d}h)\Big),$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in $L^2(\mathbb{R}, H_*)$ and $(\cdot,\cdot)$ is the inner product in H. Hence, we can pick a H_*-valued random variable $\bar{p}(t)$ in $\omega \mapsto (\omega, \cdot\, \mathbb{1}_{(0,t)})$ such that $t \mapsto \bar{p}(t)$ is a cad-lag stochastic process, called a (H_*-valued) compensated Poisson point process with Lévy measure π.

On the other hand, consider the space $\Omega = L^2_\pi(\mathbb{R}\times H_*)$ with the σ-finite product measure $\mathrm{d}t \times \pi(\mathrm{d}h)$ on $\mathbb{R} \times H_*$. Again, by means of Sazonov's Theorem 1.25 (remark that the condition (2.33) is not being used), there is a probability measure P on $(\Omega, \mathcal{F})$, with $\mathcal{F} = \mathcal{B}(\Omega)$, such that

$$\mathbb{E}\{e^{\mathtt{i}\langle\cdot,\varphi\rangle}\} = \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_*} \big[\mathrm{e}^{\mathtt{i}\varphi(t,h)} - 1 - \mathtt{i}\varphi(t,h)\big]\pi(\mathrm{d}h)\Big),$$

for any φ in $L^2_\pi(\mathbb{R} \times H_*)$, where now $\langle\cdot,\cdot\rangle$ denotes the inner product in $L^2_\pi(\mathbb{R} \times H_*)$. Note that if $\{(t,y) : \varphi_1(t,y) \neq 0\}$ and $\{(t,y) : \varphi_2(t) \neq 0\}$ are disjoint in $\mathbb{R} \times H_*$ (except for a set of $\mathrm{d}t \times \pi(\mathrm{d}y)$ measure zero), then the random variables (ω, φ_1) and (ω, φ_2) are independent. Now, in particular, if $\varphi = \mathbb{1}_{(0,t)}\mathbb{1}_B$, $t > 0$ and B in $\mathcal{B}(H_*)$, with $\pi(B) < \infty$, we can pick a real-valued random variable $p(t, B)$ in

$$\omega \mapsto \int_{\mathbb{R}} \Big(\int_B \omega(t,h)\pi(\mathrm{d}h) + \pi(B)\Big)\varphi(t)\mathrm{d}t,$$

such that $t \mapsto p(t, B)$ is a cad-lag stochastic process and $B \mapsto p(t, B)$ is a (random) measure, called a Poisson (integer) measure. Actually, p is a measure in both variables. These stochastic process has the following properties:

(1) For any B in $\mathcal{B}(H_*)$, with $\pi(B) < \infty$, the real-valued process $p(\cdot, B)$ or the H-valued process $\bar{p}$ has independent increments.

(2) For any sequence of disjoint sets $B_1, \dots, B_n$ in $\mathcal{B}(H_*)$ the stochastic processes $p(t, B_1), \dots, p(t, B_n)$ are independent.

(3) The process $p(t, B)$ is a Poisson process with parameter $\pi(B)$ and $\bar{p}(t)$ is a compensated Poisson point process, i.e., for any sequences of disjoint sets $B_1, \dots, B_n$ in $\mathcal{B}(H_*)$ with $\pi(B_i) < \infty$, and $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n$ we have

$$\mathbb{E}\Big\{e^{\mathtt{i}\sum_{j=1}^n r_j(p(t_j,B_j)-p(t_{j-1},B_j))}\Big\} = \exp\Big(\sum_{j=1}^n (t_j - t_{j-1})\pi(B_j)\big[\mathrm{e}^{\mathtt{i}r_j} - 1\big]\Big),$$

for any sequence $r_1, \dots, r_n$ in $\mathbb{R}$, whilst for the H-valued process $\bar{p}(t)$ we obtain

$$\begin{aligned}\mathbb{E}\Big\{ e^{\mathtt{i}\sum_{j=1}^n (\bar{p}(t_j)-\bar{p}(t_{j-1}),h_j)} \Big\} = \\ = \exp\Big(\sum_{j=1}^n (t_j - t_{j-1}) \int_{H_*} \big[\mathrm{e}^{\mathtt{i}(h_j,h)} - 1 - \mathtt{i}(h_j,h) \big] \pi(\mathrm{d}h) \Big) = \\ = \exp\Big(\sum_{j=1}^n (t_j - t_{j-1}) \int_{H_*} \big[\mathrm{e}^{\mathtt{i}(Rh_j,h)} - 1 - \mathtt{i}(Rh_j,h) \big] \pi_0(\mathrm{d}h) \Big),\end{aligned}$$

for any sequence $h_1, \dots, h_n$ in H.

(4) For each $s > t \geq 0$, we have $\mathbb{E}\{\bar{p}(t)\} = 0$,

$$\mathbb{E}\big\{ |\bar{p}(s) - \bar{p}(t)|^2 \big\} = (s-t) \int_{H_*} |h|^2 \pi(\mathrm{d}h) = (s-t) \int_{H_*} |R^{1/2} h|^2 \pi_0(\mathrm{d}h),$$

and $p(0, B) = 0$ almost surely. Moreover, using the independence of increments we find that

$$\mathbb{E}\big\{ |\bar{p}(r) - \bar{p}(s)|^2 |\bar{p}(s) - \bar{p}(t)|^2 \big\} = (s-t)(r-s) \Big(\int_{H_*} |h|^2 \pi(\mathrm{d}h) \Big)^2,$$

for any $r > s > t \geq 0$.

(5) For any deterministic function φ in $L^2_\pi(\mathbb{R} \times H)$ and ϕ in $L^2(\mathbb{R}, H)$, we can define the (stochastic) integrals

$$\begin{aligned}\int_{\mathbb{R}\times H_*} \varphi(t,h) \tilde{p}(\mathrm{d}t, \mathrm{d}h) &= \langle \cdot, \varphi \rangle_{L^2_\pi(\mathbb{R}\times H)} = \int_{\mathbb{R}} \mathrm{d}t \int_{H_*} \omega(t,h) \varphi(t,h) \pi(\mathrm{d}h), \\ \int_{\mathbb{R}} \big(\phi(t), \bar{p}(\mathrm{d}t) \big)_H &= \langle \cdot, \phi \rangle_{L^2(\mathbb{R},H)} = \int_{\mathbb{R}} \big(\omega(t), \phi(t) \big)_H \mathrm{d}t,\end{aligned}$$

where $\tilde{p}(t, B) = p(t, B) - t\pi(B)$. In particular, if we assume (2.33) then π integrates $h \mapsto |h|^2$, and we can define the stochastic integral

$$\omega \mapsto \int_{H_*} h p(t, \mathrm{d}h) = \int_{(0,t]} \mathrm{d}t \int_{H_*} \omega(t,h) h \pi(\mathrm{d}h),$$

which has the same distribution as the compensated Poisson point process $\bar{p}(t)$ obtained before.

The law of the process $\bar{p}$ on the canonical space either $D([0,\infty), H)$ or $D([0,\infty), H_*)$ is called a (H-valued) compensated Poisson measure with Lévy measure π. The reader may want to check other sources regarding the so-called Malliavin calculus, e.g., Bichteler [12], Ishikawa [76], Nualart [138], Sanz-Sole [156], among others.

2.7 Integer Random Measures

In the same way that a measure (or distribution) extends the idea of a function, random measures generalize the notion of stochastic processes. In terms of random noise, the model represents a noise distribution in time and some other *auxiliary space variable*, generalizing the model of noise distribution only in the time variable. Loosely speaking, we allow the index to be a measure. The particular class where the values of the measure are only positive integers is of particular interest to study the jumps of a random process.

2.7.1 Integrable Finite Variation

Before looking at random measure, first consider processes with paths having bounded variation. Usually, no specific difference is made in a pathwise discussion regarding paths with bounded variation within any bounded time-interval and within the half (or whole) real line, i.e., bounded variation paths (without any other qualification) refers to any bounded time-interval, and so the limit $A(+\infty)$ for a monotone paths could be infinite. Moreover, no condition on integrability (with respect to the probability measure) was assumed, and as seen later, this integrability condition (even locally) is related to the concept of martingales.

Now, we mention that an important role is played by the so-called *integrable increasing processes* in $[0,\infty)$, i.e., processes A with (monotone) increasing path such that

$$\mathbb{E}\{\sup_t A(t)\} = \mathbb{E}\{\lim_{t\to\infty} A(t)\} = \mathbb{E}\{A(\infty)\} < \infty,$$

and processes with *integrable bounded variation* or *integrable finite variation* on $[0,\infty)$, i.e., processes A where the variation process $\{\mathrm{var}(A,[0,t]) : t \geq 0\}$ satisfies

$$\mathbb{E}\{\sup_t \mathrm{var}(A,[0,t])\} = \mathbb{E}\{\mathrm{var}(A,[0,\infty[)\} < \infty,$$

or equivalently, $A = A^+ - A^-$ where A^+ and A^- are integrable increasing processes in $[0,\infty)$. These two concepts are localized as soon as a filtration is given, e.g., if there exists a (increasing) sequence of stopping times $(\tau_n : n \geq 1)$ satisfying $P(\lim_n \tau_n = \infty) = 1$ such that the stopped process $A_n(t) = A(t \wedge \tau_n)$ is an integrable increasing process in $[0,\infty)$ for any n then A is a *locally integrable increasing process* in $[0,\infty)$. Note that processes with *locally integrable bounded variation* or *locally integrable finite variation* on $[0,\infty)$, could be misinterpreted as processes such that their variations $\{\mathrm{var}(A,[0,t]) : t \geq 0\}$ satisfy $\mathbb{E}\{\mathrm{var}(A,[0,t])\} < \infty$, for any $t > 0$. It is worth to remark that any predictable process of bounded (or finite) variation (i.e., its variation process is finite) is indeed of locally integrable finite variation, e.g., see Jacod and Shiryaev [84, Lemma I.3.10]. Moreover, as mentioned early, the qualifiers increasing or bounded (finite) variation implicitly include a cad-lag assumption,

also, the qualifier locally implicitly includes an adapted condition. In the rare situation where an adapted assumption is not used, the tern *raw* will be explicitly used.

Going back to the relation of locally bounded variation process X with a Borel (positive or signed) measure on $[0,\infty)$

$$\mu(\{0\}) = X(0,\omega), \qquad \mu(]a,b]) = X(b,\omega) - X(a,\omega), \quad 0 < a < b$$

and abandon the pathwise analysis. Similar to the null sets in Ω, a key role is played by *evanescent* sets in $[0,\infty)\times\Omega$, which are defined as all sets N in the product σ-algebra $\mathcal{B}([0,\infty))\times\mathcal{F}$ such that $P(\{\cup_t N_t\}) = 0$, where N_t is the t section $\{\omega : (\omega,t)\in N\}$ of N. For a given process A of *integrable bounded variation*, i.e., such that

$$\mathbb{E}\{\sup_t \operatorname{var}(A,[0,t]\} < \infty,$$

we may define (bounded) signed measure μ_A (this time) on $[0,\infty)\times\Omega$ by the formula

$$\mu_A(]a,b]\times F) = \mathbb{E}\Big\{\mathbb{1}_F \int_{]a,b]} \mathrm{d}A(t)\Big\}, \quad \forall b > a \geq 0,\ F\in\mathcal{F}. \tag{2.34}$$

Since progressively, optional or predictable measurable sets are naturally identified except an evanescent set, the measure μ_A correctly represents a process A with integrable bounded variation. Conversely, a (bounded) signed measure μ on $[0,\infty)\times\Omega$ corresponds to some process A if and only if μ is a so-called signed P-measure, namely, if for any set N with vanishing sections (i.e., satisfying $P\{\omega : (\omega,t)\in N\} = 0$ for every t) we have $\mu(N) = 0$. A typical case is the point processes, i.e.,

$$A(t) = \sum_n a_n \mathbb{1}_{\tau_n\geq t},$$

where $\tau_{n-1}\leq\tau_n$ and $\tau_{n-1}<\tau_n$ if $\tau_n<\infty$ is a sequence of stopping times and a_n is $\mathcal{F}(\tau_n)$-measurable random variable with values in $\mathbb{R}_* = \mathbb{R}\setminus\{0\}$, for every n. Then, for each fixed ω the function $t\to A(t,\omega)$ is piecewise constant, but even if all the random variable a_n are bounded, the variation of the process A may not be integrable. The measure μ_A takes the form

$$\mu_A(X) = \sum_n E\{a_n\, X(\tau_n)\} = \mathbb{E}\Big\{\int_{[0,\infty)}\int_{\mathbb{R}_*} a\, X(t,\omega)\,\nu_A(\mathrm{d}t,\mathrm{d}a,\omega)\Big\},$$

$$\nu_A(B,\omega) = \#\{n : (\tau_n(\omega), a_n(\omega))\in B\},$$

for every B in $\mathcal{B}([0,\infty)\times\mathbb{R}_*)$, where # denotes the number of elements in a set and X is any bounded measurable process, in particular of the form $X(t,\omega) = \mathbb{1}_{]a,b]}(t)\,\mathbb{1}_F(\omega)$, for some set F in $\mathcal{F}$. It may seem more complicate to use the

random measure ν_A defined on $[0,\infty)\times\mathbb{R}_*$, but indeed this is characteristic to jumps processes. The reader is referred to the discussions in the books by Dellacherie and Meyer [32, Section VI.2, pp. 113–164], Jacod and Shiryaev [84, Section 1.3, pp. 27–32], Métivier and Pellaumail [128, Chapter 5, pp. 147–161], Rogers and Williams [153, Sections VI.19–21, pp. 347–352], and Elliott [43], Protter [149], among others, to complement the above remarks and following theorem–definition

Definition 2.24 (compensator)**.** Let $(\Omega,\mathcal{F},P,\mathcal{F}(t):t\geq 0)$ be a given filtered space. For any bounded (or integrable) measurable process X there exists a unique predictable process ${}^pX$, called *predictable projection*, such that for any predictable stopping time τ we have $\mathbb{E}\{{}^pX\mathbb{1}_{\tau<\infty}\}=\mathbb{E}\{X\mathbb{1}_{\tau<\infty}\}$. It is proved that a process A with integrable bounded variation is predictable if and only if $\mu_A(X)=\mu_A({}^pX)$ for any bounded measurable process X, see (2.34). Now, given a process A with integrable bounded variation with a corresponding signed P-measure μ_A on $[0,\infty)\times\Omega$, the *dual predictable projection* of μ_A is defined by duality as follows

$$\mu_A^p(X)=\mathbb{E}\Big\{\int_{[0,\infty)}{}^pX(t,\omega)\mathrm{d}A(t,\omega)\Big\},$$

for any bounded measurable process X. Since μ_A^p is a signed P-measure which commutes with the predictable projection, its corresponding process with integrable bounded variation, denoted by A^p, is predictable and satisfies

$$\mathbb{E}\{\int_{[0,\infty)}X(t,\omega)\mathrm{d}A^p(t,\omega)=\mathbb{E}\{\int_{[0,\infty)}{}^pX(t,\omega)\mathrm{d}A(t,\omega),$$

for any bounded measurable process X, and called the *compensator* of A. □

Similarly to above, we may define the *optional projection*, and *dual optional projection*, with the notations ${}^oX$, μ_A^o and A^o. Clearly, the above statements can be localized, i.e., the process X can only be assumed locally bounded or locally integrable, and the process A can only be supposed with locally integrable finite variation.

It will be stated later that the dual predictable projection μ_A^p corresponding to a signed P-measure μ_A of an adapted process A with integrable bounded variation is actually characterized by the fact that the (Stieltjes integral) process

$$\int_{[0,t]}X(t-,\omega)\mathrm{d}A(t,\omega)-\int_{[0,t]}X(t-,\omega)\mathrm{d}A^p(t,\omega).\quad t\geq 0$$

is a martingale for any bounded adapted process X. It is clear that $t\mapsto X(t-)$ is a predictable process and that in the above martingale condition it suffices to take processes of the form $X(t)=\mathbb{1}_{t\leq\tau}$ for some stopping time τ, i.e., the process $t\mapsto A(t\wedge\tau)-A^p(t\wedge\tau)$ is a martingale.

Related with the compensator definition is the (unique) decomposition of any positive increasing adapted right-continuous process A into the sum of a

continuous increasing adapted process ${}^cA$ with ${}^cA(0) = 0$ and a right-continuous increasing adapted process ${}^jA$ which can be expressed as follows:

$$ {}^jA(t) = \sum_n a_n \, \mathbb{1}_{t \geq \tau_n}, $$

where $\{\tau_n\}$ is a sequence of stopping times with bounded disjoint graphs and a_n is a bounded positive $\mathcal{F}(\tau_n)$-measurable function for every n. The proof of this fact is rather simple, first define inductively $\tau^{i,0} = 0$ and

$$ \tau^{i,j} = \inf\{t > \tau^{i,j} : A(t+) - A(t-) \geq 1/i\}, $$

and then $\tau_k^{i,j} = \tau^{i,j}$ if $A(t+) - A(t-) \leq k+1$ and $\tau^{i,j} \leq k$, and $\tau_k^{i,j} = \infty$ otherwise. Clearly $\{\tau_k^{i,j}\}$ is countable and can be rewritten as $\{\tau_n' : n = 1, 2, \dots\}$, which is a sequence of stopping times with bounded graphs. Again, defining $\tau_n = \tau_n'$ if $\tau_i \neq \tau_n$ for every $i = 1, \dots, n$ and $\tau_n = \infty$ otherwise, we get the desired sequence, with $a_n = A(\tau_n+) - A(\tau_n-)$.

Similarly, if A is as above and $\varphi : [0, \infty) \to [0, \infty)$ is a continuously differentiable function and for a given $r \geq 0$ we set

$$ \tau_r = \inf\{t \geq 0 : A(t) \geq r\} \quad \text{and} \quad \theta_r = \inf\{t \geq 0 : A(t) > r\}, $$

which are both stopping times (as seen later, τ_r is predictable), then for every bounded measurable process X we have

$$ \int_0^\infty X(s) \mathrm{d}\varphi(A(s)) = \int_0^\infty X(\tau_r) \, \varphi'(r) \, \mathbb{1}_{\tau_t < \infty} \mathrm{d}r = $$
$$ = \int_0^\infty X(\theta_r) \, \varphi'(r) \, \mathbb{1}_{\theta_t < \infty} \mathrm{d}r. $$

Details on the proof of these results can be found in Bichteler [11, Section 2.4, pp. 69–71].

2.7.2 Counting the Jumps

Returning to the sample space, we know that an element ω in $D([0, \infty), \mathbb{R}^d)$ has at most a countable number of jumps, with only a finite number of jumps of size greater than a positive quantity. For any Borel set B in $\mathcal{B}(\mathbb{R}_*^d)$ with $\mathbb{R}_*^d = \mathbb{R}^d \setminus \{0\}$ (so-called punctured d-space) the number of jumps before a time t and with values in B are finite if B is compact. Thus, for any (cad-lag) stochastic process with values in $\mathbb{R}^d$ or equivalently for any random variable X with values in $D([0, \infty), \mathbb{R}^d)$ we can define a measure ν_X with integer values, as the number of jumps in B within a bounded time interval, i.e.,

$$ \nu_X(B \times]a, b], \omega) = \#\{t : a < t \leq b, \; X(t, \omega) - X(t-, \omega) \in B\}, \tag{2.35} $$

for any $b > a \geq 0$, B in $\mathcal{B}(\mathbb{R}_*^d)$, and where $\#$ denotes the number of elements (which may be infinite) of a set. Sometime we use the notation $\nu_X(B,]a, b], \omega)$

and we may look at this operation as a *functional* on $D([0,\infty),\mathbb{R}^d)$, i.e., for every $b > a \geq 0$ and B in $\mathcal{B}(\mathbb{R}^d_*)$,

$$\nu(B,]a,b],\omega) = \sum_{a<t\leq b} \mathbb{1}_B\big(\omega(t)-\omega(t-)\big),$$

so that $\nu_X(B\times]a,b],\omega) = \nu(B,]a,b],X(\cdot,\omega))$. For each ω, this is Radon measure on $\mathbb{R}^d_*\times(0,\infty)$ with integer values. By setting $\nu(\mathbb{R}^d_*\times\{0\}) = 0$ we may consider ν as a measure on $\mathbb{R}^d_*\times[0,\infty)$.

This measure ν is used as a characterization of the jumps $\delta X = (\delta X(t) = X(t)-X(t-) : t>0)$, in the sense that ν vanishes if and only if the process X is continuous. Note that for any continuous function $f(t,x)$ which vanishes near $x=0$ we have

$$\int_{\mathbb{R}^d_*\times(a,b]} f(x,t)\nu(\mathrm{d}\zeta,\mathrm{d}t) = \sum_{\delta X(t)\neq 0} \mathbb{1}_{\{a<t\leq b\}} f\big(\delta X(t),t\big),$$

where the sum is finite. In this sense, the random measure ν contains all information about the jumps of the process X. Moreover, remark that ν is a sum of Dirac measures at $(\delta X(t),t)$, for $\delta X(t)\neq 0$. This sum is finite on any set separated from the origin, i.e., on any sets of the form

$$\big\{(x,t)\in\mathbb{R}^d_*\times[0,\infty) : t\in]a,b],\ |x|\geq\varepsilon\big\},$$

for every $b>a\geq 0$ and $\varepsilon>0$.

Recall that the Skorokhod's topology, given by the family of functions defined for ω in $D([0,\infty),\mathbb{R}^d)$ by the expression

$$w(\omega,\delta,]a,b]) = \inf_{\{t_i\}}\ \sup_i\ \sup\{|\omega(t)-\omega(s)|\ :\ t_{i-1}\leq s<t<t_i\}$$

where $\{t_i\}$ ranges over all partitions of the form $a=t_0<t_1<\cdots<t_{n-1}<b\leq t_n$, with $t_i-t_{i-1}\geq\delta$ and $n\geq 1$, makes $D([0,\infty),\mathbb{R}^d)$ a complete separable metric space. Again, note that

$$\nu(\{z\in\mathbb{R}^d : |z|\geq w(\omega,\delta,]a,b])\},]a,b],\omega) \leq \frac{b-a}{\delta},$$

for every $\delta>0$, and $b>a\geq 0$.

Another point is the following fact that for any set B in $\mathbb{R}^d_*$ with a positive distance to the origin, we can define the sequence of jump-times and jump-size as

$$\begin{aligned}
&\tau_0(B,]a,b],\omega) = a, \quad \zeta_0(B,]a,b],\omega) = 0,\\
&\tau_k(B,]a,b],\omega) = \inf\{t\in]\tau_{k-1},b] : \omega(t)-\omega(t-)\in B\}, \quad k\geq 1,\\
&\zeta_k(B,]a,b],\omega) = \omega(\tau_k)-\omega(\tau_k-), \quad k\geq 1,
\end{aligned}$$

for any $b\geq a\geq 0$ and $1\leq k\leq\nu(B,]a,b],\omega)$. Thus, if ω_n is a sequence converging to ω in $D([0,\infty),\mathbb{R}^d)$, and B is also an open set with boundary ∂B

satisfying $\nu(\partial B,]a,b], \omega) = 0$, and such that the first jump $\omega(a) - \omega(a-)$ and the last jump $\omega(b) - \omega(b-)$ have a positive distance to B, then

$$\nu(B,]a,b], \omega_n) \to \nu(B,]a,b], \omega),$$
$$\tau_k(B,]a,b], \omega_n) \to \tau_k(B,]a,b], \omega),$$
$$\zeta_k(B,]a,b], \omega_n) \to \zeta_k(B,]a,b], \omega),$$

for any $k = 0, 1, \dots, \nu(B,]a,b], \omega)$.

Definition 2.25 (integer measure)**.** Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a filtered space. A *random measure* on a Polish space E is a random variable ν with values in the space of σ-finite measures on the Borel σ-algebra $\mathcal{B}(E)$. In most of the cases, the Polish space E is locally compact and the random variable ν take values in the space of Radon (nonnegative) measures (finite on every compact sets) on $\mathcal{B}(E)$. If the time-variable is singled-out, e.g., $E = \mathbb{R}^m_* \times [0, \infty)$ then it is required that $\nu(\mathbb{R}^m_* \times \{0\}) = 0$. In this case a random measure on $\mathbb{R}^m_* \times [0, \infty)$ is called a *optional or predictable (respectively, locally integrable)* if for any stopping time $\tau < \infty$ and any compact subset K of $\mathbb{R}^m_*$ the stochastic process $t \mapsto \nu(K \times [0, t \wedge \tau])$ is optional or $t \mapsto \nu(K \times [0, t \wedge \tau[)$ is predictable (respectively, $\mathbb{E}\{\nu(K \times [0, t \wedge \tau])\} < \infty$ for every $t > 0$). Moreover, an optional locally integrable random measure ν is called *integer measure* or *integer-valued random measure* if it takes values in $\{0, 1, \dots, \infty\}$, $\nu(\mathbb{R}^m_* \times \{0\}) = 0$ and $\nu(\mathbb{R}^m_* \times \{t\}) = 0$ or $= 1$ for any $t > 0$. □

When referring to an integer-valued random measure, the above definition implies that we mean an optional locally integrable integer-valued random measure. Moreover, the local integrability ensures that the *product measure* $\nu(\mathrm{d}x \times \mathrm{d}t, \omega)\, P(\mathrm{d}\omega)$ is σ-finite. It is clear that we may replace $\mathbb{R}^m_*$ by a locally compact Polish E. An essential point is the use of the following two properties: (1) the σ-algebra $\mathcal{E}$ is generated by a countable algebra and (2) any $(E, \mathcal{E})$-valued random variable x on a probability space $(\Omega, \mathcal{F}, P)$ admits a regular conditional distribution relative to a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$. This *disintegration property* (2) can be restated as: for any positive and finite measure m on the product space $(E \times B, \mathcal{E} \times \mathcal{B})$ there exist a measurable kernel $k(\mathrm{d}x, b)$ such that $m(\mathrm{d}x, \mathrm{d}b) = k(\mathrm{d}x, b)\, m_B(\mathrm{d}b)$, where $m_B(\mathrm{d}b) = m(E, \mathrm{d}b)$ is the B-marginal distribution of m. Clearly, this is related to the conditional property, and this is used to define the *compensator*, a key instrument for the stochastic integral. These properties are satisfied by the so-called *Blackwell spaces*, see Dellacherie and Meyer [32]. Only the case of locally compact Polish spaces will be used here.

A typical example of optional (respectively, predictable) integer measure on $\mathbb{R}^m_*$ is the one constructed by (2.35) for an adapted (i.e., optional) (respectively, predictable) locally integrable stochastic process with values in $\mathbb{R}^m$. Notice that integrability at infinity is not an issue in the above definition of integer-valued measure, the key part is the integrability away of the origin, i.e., we may use $\mathbb{E}\{\nu(B \times [0,t])\} < \infty$, for any Borel subset B of $\mathbb{R}^m_*$ with a positive distance

to the origin. Certainly, this can be viewed as a *localization* (via a sequence of stopping times) of the integral condition

$$\mathbb{E}\Big\{ \int_{\mathbb{R}^m_* \times [0,t]} (|\zeta|^2 \wedge 1) \nu(\mathrm{d}\zeta, \mathrm{d}t) \Big\} < \infty,$$

for every $t \geq 0$, which is used later for Lévy measures.

Given an integer-valued random measure ν on $\mathbb{R}^m_*$, the set $\{t : \nu(\mathbb{R}^m_* \times \{t\}) = 1\}$ is countable for any ω and can be written as a sequence $(\tau_n(\omega) : n = 1, 2, \ldots)$. Moreover, because ν assumes only integers values, there is a sequence $(a_n(\omega) : n = 1, 2, \ldots)$ such that $\nu(\{(a_n, \tau_n)\}) = 1$ and $\nu(\mathbb{R}^m_* \times [0, \infty) \smallsetminus \{(a_n, \tau_n)\}) = 0$. Because ν is finite on compact subsets of $\mathbb{R}^d_*$, for each ε, $t > 0$ there exists only a finite number of (a_n, τ_n) such that $\varepsilon \leq |a_n| \leq 1/\varepsilon$ and $\tau_n \leq t$. Hence we may always rewrite ν as

$$\begin{aligned} \nu(B, \omega) &= \sum_n \mathbb{1}_{(a_n(\omega), \tau_n(\omega)) \in B}, \quad \forall B \in \mathcal{B}(\mathbb{R}^m_* \times [0, \infty)), \\ A^\varepsilon_\nu(t, \omega) &= \sum_n a_n(\omega)\, \mathbb{1}_{\varepsilon \leq |a_n| \leq 1/\varepsilon}\, \mathbb{1}_{\tau_n(\omega) \leq t}, \quad \forall t \geq 0, \end{aligned} \tag{2.36}$$

this determines an optional locally integrable jump process A^ε_ν on $\mathbb{R}^m_*$, and so the following expression for every F in $\mathcal{F}(s)$ and $t \geq s \geq 0$,

$$\mu^\varepsilon(]s, t] \times F) = \mathbb{E}\big\{ [A^\varepsilon_\nu(t) - A^\varepsilon_\nu(s)]\, \mathbb{1}_F \big\}, \tag{2.37}$$

defines a bounded ($\mathbb{R}^d$-valued) measure on $[0, \infty) \times \Omega$.

If the jump processes $\{A^\varepsilon_\nu : \varepsilon > 0\}$ have a uniformly locally integrable bounded variation, i.e., $\mathbb{E}\{\sum_n |a_n|\} < \infty$,, then $A_\nu = (\sum_n a_n\, \mathbb{1}_{\tau_n \leq t} : t \geq 0)$ has a locally integrable bounded variation (when $d = 1$ we have a signed measure μ^ε) and a measure μ (limit as $\varepsilon \to 0$, which is called Doléans measure), can be defined. To come back from this ($\mathbb{R}^d$-valued) measure μ^ε to the process A^ε (or to the integer-valued random measure ν), we need μ^ε to vanish for any evanescent set, i.e., $\mu^\varepsilon(N) = 0$ for any subset N of $[0, \infty) \times \Omega$ such that $P(\cup_t \{\omega : (t, \omega) \in N)\}) = 0$. The point is that the integer measure ν captures all the features of the family of processes A^ε, even when A can not be defined. In other words, if A^ε is a semi-martingale we will see that μ^ε may define a measure as ε vanishes.

2.7.3 Compensating the Jumps

Returning to the compensator, as in Definitions 2.24 (in Chapter 2) and 2.25, we have a unique dual predictable projection ν^p of any optional locally integrable random measure ν, characterized (almost surely) as being a predictable random measure such that $\mathbb{E}\{\nu(K \times [0, t \wedge \tau]) - \nu^p(K \times [0, t \wedge \tau])\} = 0$ for any stopping time $\tau < \infty$, any compact subset K of $\mathbb{R}^m_*$ and any $t > 0$, or equivalently the process $t \mapsto \nu(K \times]0, t]) - \nu^p(K \times]0, t])$ is a martingale. Hence, by an argument

of monotone class, we have

$$\mathbb{E}\Big\{ \int_{\mathbb{R}_*^m \times [0,\infty)} X(z,t)\, \nu(\mathrm{d}z, \mathrm{d}t) \Big\} = \mathbb{E}\Big\{ \int_{\mathbb{R}_*^m \times [0,\infty)} X(z,t)\, \nu^p(\mathrm{d}z, \mathrm{d}t) \Big\},$$

for any nonnegative function $(z,t,\omega) \mapsto X(z,t,\omega)$ measurable with respect to the product σ-algebra $\mathcal{B}(\mathbb{R}_*^m) \times \mathcal{O}$ (with $\mathcal{O}$ being the optional σ-algebra) where the product measure $\nu(\mathrm{d}z, \mathrm{d}t, \omega)\, P(\mathrm{d}\omega)$ is defined. Recall that we assume $\nu(\mathbb{R}_*^m \times \{0\}) = 0$, so that $\nu(K \times \{0\}) = \nu^p(K \times \{0\}) = 0$. Moreover, based on the disintegration property, the predictable compensator can be written as $\nu^p(\mathrm{d}z, \mathrm{d}t, \omega) = k(\mathrm{d}z, t, \omega)\mathrm{d}A(t, \omega)$, where A is a integrable predictable increasing process and $k(\mathrm{d}z, t, \omega)$ is a measurable kernel. We refer to Bichteler [11, Sections 3.10, 4.3, pp. 171–186, 221–232], He et al. [68], Jacod and Shiryaev [84, Section II.1, pp. 64–74], and Kallenberg [87] for a full discussion on random measures, only some results are reported here.

Theorem 2.26. *Let ν^p be compensator of an integer-valued random measure ν. Then the predictable random measure ν^p (which is not necessarily an integer-valued random measure) has the following properties. First* (a) *its predictable support, namely the set $\{(t,\omega) : 0 < \nu^p(\mathbb{R}_*^m \times \{t\}, \omega) \le 1\}$, can be written as a sequence of predictable stopping times, i.e., $\{(\tau_n^p(\omega), \omega) : n = 1, 2, \ldots\}$ with τ_n^p a predictable stopping time for any n, and $P(\{\omega : 0 < \nu^p(\mathbb{R}_*^m \times \{t\}, \omega) \le 1\}) = 1$, for any $t \ge 0$. Next* (b) *we have*

$$\nu^p(K \times \{\tau\}) = \mathbb{E}\Big\{ \sum_n \mathbb{1}_{a_n \in K} \,|\, \mathcal{F}(\tau-) \Big\},$$

on the predictable support, for any predictable stopping time $\tau < \infty$ and any compact subset K of $\mathbb{R}_*^m$. Moreover, if ν is defined as the number of jumps (2.35) of a (special) semi-martingale X then the predictable processes in $t > 0$,

$$\sqrt{\sum_{0<s\le t} \nu^p(\mathbb{R}_*^m \times \{s\})} \quad \text{and} \quad \sqrt{\int_{\mathbb{R}_*^m \times]0,t]} (|z|^2 \wedge |z|)\, \nu^p(\mathrm{d}z, \mathrm{d}t)},$$

are locally integrable. They also are integrable or (locally) square integrable if the semi-martingale X has the same property. Furthermore, X is quasi-left continuous if and only if its predictable support is an empty set, i.e., $\nu^p(\mathbb{R}_*^m \times \{t\}) = 0$, for any $t \ge 0$. □

Note if $\nu(\mathrm{d}z, \mathrm{d}t, \omega)$ is a quasi-left continuous integer random measure then its predictable compensator can be written as $\nu^p(\mathrm{d}z, \mathrm{d}t, \omega) = k(\mathrm{d}z, t, \omega)\mathrm{d}A(t, \omega)$, where k is a measurable (predictable) kernel and A is a continuous increasing process.

To check the point regarding the quasi-left continuity for a square integrable martingale X, let $\tau < \theta < \infty$ be given two stopping times. Since, for any compact subset K of $\mathbb{R}_*^d$ the quantity

$$\mathbb{E}\Big\{ \sum_{\tau < t \le \theta} \mathbb{1}_{\delta X(t) \in K} |\delta X(t)|^2 \Big\} = \mathbb{E}\Big\{ \int_{K \times]\tau,\theta]} |z|^2\, \nu(\mathrm{d}z, \mathrm{d}t) \Big\}$$

is a finite, the number of jumps is finite for each ω and ν can be replaced by ν^p in the above equality, we deduce

$$\varepsilon^2\,\mathbb{E}\{\nu(K\times]\tau,\theta])\,|\,\mathcal{F}(\tau)\} \le \mathbb{E}\Big\{\int_{K\times]\tau,\theta]} |z|^2\,\nu(\mathrm{d}z,\mathrm{d}t)\,|\,\mathcal{F}(\tau)\Big\} \le \\ \le \mathbb{E}\{|X(\theta)|^2 - |X(\tau)|^2\,|\,\mathcal{F}(\tau)\},$$

where $\{|z| < \varepsilon\} \cap K = \emptyset$, $\varepsilon > 0$. Hence, $\nu(K \times [0,t])$ and $\nu^p(K \times [0,t])$ are quasi-left continuous if and only if X is quasi-left continuous.

Note that the previous theorem selects a particular representation (or realization) of the compensator of an integer-valued random measure suitable for the stochastic integration theory. Thus, we always refer to the compensator satisfying the properties in Theorem 2.26. Moreover, given an integer-valued random measure ν the process $\nu_{qc}(]0,t\wedge\tau]\times K)$ given by the expression

$$\nu_{qc}(K\times]0,t\wedge\tau]) = \nu(K\times]0,t\wedge\tau]) - \sum_{0<s\le t\wedge\tau} \nu^p(K\times\{s\}),$$

is quasi-left continuous, and its compensator is the continuous part of the compensator ν^p, denoted by ν^p_c. Hence, for any stopping time $\tau < \infty$ and any compact subset K of $\mathbb{R}^m_*$ the stochastic process $t \mapsto \tilde\nu_{qc}(K\times]0,t\wedge\tau])$, with $\tilde\nu_{qc} = \nu_{qc} - \nu^p_c$ is a local (purely discontinuous) martingale, whose predictable quadratic variation process obtained via Doob-Meyer decomposition is actually the process $\nu^p_c(K\times]0,t\wedge\tau])$, i.e.,

$$\langle\tilde\nu_{qc}(K\times[0,\cdot\wedge\tau])\rangle(t) = \nu^p_c(K\times]0,t\wedge\tau]), \quad \forall\, t\ge 0.$$

Thus, the optional locally integrable random measure $\tilde\nu = \nu - \nu^p = \tilde\nu_{qc}$ is called the (local) *martingale random measure* associated with ν or with the cad-lag process X.

• *Remark* 2.27. Two (or more) random measures ν_1 and ν_2 are called independent if for any stopping time $\tau < \infty$ and any compact subset K of $\mathbb{R}^m_*$ the stochastic process $t \mapsto \nu_1(K\times[0,t\wedge\tau])$ and $t \mapsto \nu_2(K\times[0,t\wedge\tau])$ are independent. Therefore, if ν_1 and ν_2 are defined as the number of jumps (2.35) of two (or more) (special) semi-martingale X_1 and X_2 then the (purely) jumps processes δX_1 and δX_2 are independent if and only the random measures ν_1 and ν_2 (and therefore ν^p_1 and ν^p_1) are so. However, the random measure associated via (2.35) with the jumps $(\delta X_1, \delta X_2)$ considered in $\mathbb{R}^{2m}_*$, and the almost *product* measure ν in $\mathbb{R}^{2m}_*$ defined by $\nu(K_1\times K_2\times]0,t]) = \nu_1(K_1\times]0,t])\,\nu_1(K_2\times]0,t])$, for every K_1, K_2 compact subset of $\mathbb{R}^m_*$ and $t > 0$ may not agree. Certainly, they are the same if the process X_1 and X_2 do not jump simultaneously. In particular, if X_1 and X_2 are Poisson processes with respect to the same filtration then they are independent if and only if they never jump simultaneously. □

A fundamental example of jump process is the simple point process $(N(t) : t \ge 0)$ which is defined as a increasing adapted cad-lag process with nonnegative integer values and jumps equal to 1, i.e., $\delta N(t) = 0$ or $\delta N(t) = 1$ for every $t \ge 0$,

and $N(t)$ represents the number of *events* occurring in the interval $(0,t]$ (and so more then one event cannot occur exactly a the same time). Given $(N(t) : t \geq 0)$ we can define a sequence $\{T_n : n \geq 0\}$ of stopping times $T_n = \{t \geq 0 : N(t) = n\}$. Notice that $T_0 = 0$, $T_n < T_{n+1}$ on the set $\{T_{n+1} < \infty\}$, and $T_n \to \infty$. Since

$$N(t) = \sum_{n=0}^{\infty} \mathbb{1}_{T_n \leq t}, \quad \forall t \geq 0,$$

the sequence of stopping times completely characterizes the process, and because $N(T_n) \leq n$, any point process is locally bounded. An *extended Poisson process* N is an adapted point process on the filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ satisfying:

(1) $\mathbb{E}\{N(t)\} < \infty$, for every $t \geq 0$,

(2) $N(t) - N(s)$ is independent of $\mathcal{F}(s)$, for every $t \geq 0$,

The function $a(t) = \mathbb{E}\{N(t)\}$ is called intensity (of N). It can be proved that if the function a is continuous then N is a *Poisson process* and if $a(t) = t$ for every $t \geq 0$ then N is a *standard Poisson process*. In this example, the compensator can be calculated, it can be proved (e.g., Jacod and Shiryaev [84, Proposition I.3.27, pp. 34–35]) that the compensator of an extended Poisson process is equal to its intensity, i.e., $N^p(t) = \mathbb{E}\{N(t)\}$ and that N is quasi-left continuous if and only if it is a Poisson process (i.e., its intensity is continuous). In general, even though the jumps are always countable they can not be ordered as in a point process. This yields the notion of integer-valued random measures.

Our main interest is on integer-valued random measure ν_X associated with a quasi-left continuous semi-martingale X, so that $t \mapsto \nu_X^p(K\times]0, t \wedge \tau])$ is continuous and for $\tilde{\nu}_X = \nu_X - \nu_X^p$ we have the following expressions for the optional and predictable quadratic variation processes

$$[\tilde{\nu}_X(K\times]0, \cdot \wedge \tau])](t) = \langle \tilde{\nu}_X(K\times]0, \cdot \wedge \tau])\rangle(t) = \nu_X^p(K\times]0, t \wedge \tau]), \quad (2.38)$$

for any $t > 0$, any stopping time $\tau < \infty$ and any compact subset K of R_*^m. Ignoring the local character of the semi-martingale X, this yields the compensated jumps equality

$$\mathbb{E}\Big\{\Big|\int_{K\times]0,t\wedge\tau]} \varphi(z,s)\,\tilde{\nu}_X(\mathrm{d}z,\mathrm{d}s)\Big|^2\Big\} = \\ = \mathbb{E}\Big\{\int_{K\times]0,t\wedge\tau]} |\varphi(z,s)|^2\,\nu_X^p(\mathrm{d}z,\mathrm{d}s)\Big\}$$

and estimate

$$\mathbb{E}\Big\{\sup_{0\leq t\leq T}\Big|\int_{K\times]0,t\wedge\tau]} \varphi(z,s)\,\tilde{\nu}_X(\mathrm{d}z,\mathrm{d}s)\Big|^2\Big\} \leq \\ \leq 4\,\mathbb{E}\Big\{\int_{K\times]0,T\wedge\tau]} |\varphi(z,s)|^2\,\nu_X^p(\mathrm{d}z,\mathrm{d}s)\Big\},$$

for any Borel measurable function $\varphi(z,s)$ such that the right-hand side is finite. Thus, we can define the integral of φ with respect to $\tilde{\nu}_X$

$$\tilde{\nu}_X(\varphi \mathbb{1}_{]0,t\wedge\tau]}) = \lim_{\varepsilon\to 0} \int_{\{|x|\geq\varepsilon\}\times]0,t\wedge\tau]} \varphi(z,s)[\nu_X(\mathrm{d}z,\mathrm{d}s) - \nu_X^p(\mathrm{d}z,\mathrm{d}s)], \quad (2.39)$$

where φ vanishes for $|z|$ large and for $|z|$ small. All this is developed with the stochastic integral, valid for any predictable process instead of $\varphi\mathbb{1}_{]0,t\wedge\tau]}$. The point here is that the integral

$$\int_{\{|x|<1\}\times]0,t\wedge\tau]} z\,\tilde{\nu}_X(\mathrm{d}z,\mathrm{d}s)$$

is meaningful as a limit in L^2 for every φ square integrable with respect to ν_X^p, and the compensated jumps estimate holds.

In this way, the stochastic process X and the filtered space $(\Omega,\mathcal{F},P,\mathcal{F}(t) : t\geq 0)$ determine the predictable compensator ν_X^p. Starting from a given integer-valued random measure ν and by means of the previous Theorem 2.26, we can define its *compensated martingale random measure* $\tilde{\nu} = \nu - \nu^p$, where ν^p is the compensator. The Doléans measure on $\mathbb{R}^m_* \times [0,\infty)\times\Omega$ relative to the integer measure ν is defined as the *product measure* $\mu = \nu(\mathrm{d}z,\mathrm{d}s,\omega)\,P(\mathrm{d}\omega)$, i.e., associated with the jumps process Z_K induced by ν, namely, for every compact subset K of $\mathbb{R}^m_*$

$$Z_K(t,\omega) = \int_{K\times]0,t]} z\,\nu(\mathrm{d}z,\mathrm{d}s), \quad \forall t\geq 0.$$

Therefore whenever ν integrate the function $z\mapsto|z|$ we can consider the process $Z_{\mathbb{R}^m_*}$ as in (2.37). Conversely, if a given (m-valued) Doléans measure μ vanishes on any evanescent set, i.e., $\mu(K\times N)=0$ for every compact K of $\mathbb{R}^m_*$ and for any subset N of $[0,\infty)\times\Omega$ such that $P(\cup_t\{\omega : (t,\omega)\in N)\})=0$, then there is an optional ($\mathbb{R}^m$-valued) jump process A with integrable bounded variation associated with μ. This argument can be localized as long as we assume $\nu^p(]0,t\wedge\tau)\times K)<\infty$, for any compact K in $\mathbb{R}^m$ (not only in $\mathbb{R}^m_*$) to get a jump process A with locally integrable bounded variation path associated to ν. Now, for this jump process A we can defined an integer-valued measure ν with the same initial predictable compensator ν^p.

The following *canonical representation* of (special) semi-martingale holds. Let ν_X be the (random) integer measure associated with the semi-martingale X, namely, $\nu_X(B\times]a,b])$ is the number of jumps on the time interval $(a,b]$ of the process X with a value δX belonging to the set B, i.e. for every $b>a\geq 0$ and B in $\mathcal{B}(\mathbb{R}^d_*)$,

$$\nu_X(B\times]a,b]) = \#\{t : a<t\leq b,\ X(t)-X(t-)\in B\},$$

and let ν_X^p be its (dual predictable) compensator (satisfying the properties given in Theorem 2.26), so that $\tilde{\nu}_X = \nu_X - \nu_X^p$ is a local-martingale measure, then

$$X(t) = X(0) + A(t) + X^c(t) + \int_{\mathbb{R}^d_*\times]0,t]} z\tilde{\nu}_X(\mathrm{d}z,\mathrm{d}s), \quad \forall t\geq 0, \quad (2.40)$$

where A is a predictable process with locally integrable variation and X^c is a continuous local-martingale, both satisfying $A(0) = X^c(0) = 0$ and X^c is uniquely determined. Clearly, the integer measure ν depends only on the jump process δX, i.e., only the discontinuous part of X determines ν_X. If the semi-martingale X is quasi-left continuous (i.e., either $\mathcal{F}(\tau-) = \mathcal{F}(\tau)$ for every predictable stopping time τ or equivalently the predictable compensator ν^p_X satisfies $\nu^p_X(\mathbb{R}^d_* \times \{t\}) = 0$ almost surely), then the process A in (2.40) is continuous and uniquely determined.

Note the characteristic elements of a semi-martingale X, which are (1) the predictable process A with locally integrable variation (which is uniquely determined only when the semi-martingale is quasi-left continuous), (2) the predictable quadratic variation $\langle X^c \rangle$ and (3) the (dual predictable) compensator measure ν^p_X. If $X = M$ is a quasi-left continuous local-martingale then $A = 0$ and there are only two characteristic elements to consider: (a) the predictable quadratic variation $\langle M^c \rangle$ (or the optional quadratic variation $[M]$) and (b) the predictable compensator ν^p (or the integer-valued measure ν). If the *special* character of the semi-martingale is removed, then the jumps may be not locally integrable and then the predictable compensator ν^p may be not integrable at infinity, i.e., only the function $z \mapsto |z|^2 \wedge 1$ in ν^p-integrable, so that the predictable process

$$t \mapsto \int_{\mathbb{R}^d_*} \int_{]0,t]} (|z|^2 \wedge 1)\, \nu^p(\mathrm{d}z, \mathrm{d}s)$$

is locally integrable. Thus the representation (2.40) becomes

$$X(t) = X(0) + A(t) + X^c(t) + \int_{\{|z|<1\}\times]0,t]} z\, \tilde{\nu}_X(\mathrm{d}z, \mathrm{d}s), \quad \forall t \geq 0, \tag{2.41}$$

where A contains a term of the form

$$\int_{|z|\geq 1} \int_{]0,t]} z\, \nu_X(\mathrm{d}z, \mathrm{d}t),$$

and $h(z) = z\mathbb{1}_{|z|<1}$ is used as the truncation function. However, our main interest is on processes with finite moments of all order, so that ν^p should integrate $z \mapsto |z|^n$ for all $n \geq 2$. The reader may consult He et al. [68, Section XI.2, pp. 305–311], after the stochastic integral is covered.

2.7.4 Poisson Measures

A fundamental example is the Poisson measures. We have

Definition 2.28 (Poisson-measure)**.** Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a filtered space. An integer-valued random measure ν on $\mathbb{R}^m_* \times [0,\infty)$ is called *Poisson measure* if

(a) the (nonnegative) measure $\Pi(B) = \mathbb{E}\{\nu(B)\}$ is a Radon measure on $\mathbb{R}^m_* \times [0,\infty)$, i.e., $\mathbb{E}\{\nu(K \times [0,t])\} < \infty$ for any compact subset K of $\mathbb{R}^m_*$ and for any

$t \geq 0$,

(b) for any Borel measurable subset B of $\mathbb{R}^m_* \times (t, \infty)$ with $\Pi(B) < \infty$ the random variable $\nu(B)$ is independent of the σ-algebra $\mathcal{F}(t)$,

(c) Π satisfies $\Pi(\mathbb{R}^m_* \times \{t\}) = 0$ for every $t \geq 0$.

The measure Π is called *intensity measure* relative to the Poisson measure ν. If Π has the form $\Pi(\mathrm{d}z, \mathrm{d}t) = \pi(\mathrm{d}z) \times \mathrm{d}t$ for a (nonnegative) Radon measure π on $\mathbb{R}^m_*$ then ν is called a *homogeneous (or standard) Poisson measure*. If the condition (c) is not satisfied then ν is called *extended Poisson measure*. □

A standard Poisson measure ν on a Polish space $\mathcal{O} \times [0, \infty)$ (e.g., $\mathbb{R}^m_* \times [0, \infty)$ or even a non-locally compact separable metric space) relative to a σ-finite measure $\pi \times \mathrm{d}t$ on $\mathcal{B}\big(\mathcal{O} \times [0, \infty)\big)$ (called intensity) can be also defined as a random measure satisfying (a) for any Borel subset B of $\mathcal{O}$ with $\pi(B) < \infty$ and $t \geq 0$ the random variable $\nu(B \times]0, t]) = \nu(B, t)$ has a Poisson distribution with parameter $t\pi(B)$ and (b) for any $n \geq 1$ and any disjoint Borel sets $B_1, B_2, \ldots, B_n$ and $0 \leq t_0 < t_1 < \cdots t_n$ the random variables $\nu(B_1, t_1) - \nu(B_1, t_0), \nu(B_2, t_2) - \nu(B_2, t_1), \ldots, \nu(B_n, t_n) - \nu(B_n, t_{n-1})$ are independent. Given a σ-finite measure π on $\mathcal{B}(\mathcal{O})$, a standard Poisson measure ν can be constructed as follows. First, if π is a finite measure then we can find a sequence $\{\tau_1, \tau_2, \ldots\}$ of independent exponentially distributed random variables, with parameter $c = \pi(\mathcal{O})$ and a sequence $(\xi_1, \xi_2, \ldots)$ of $\mathcal{O}$-valued independent identically distributed random variables, with distribution $\pi/\pi(\mathcal{O})$ and independent of $\{\tau_1, \tau_2, \ldots\}$, in some (complete) probability space $(\Omega, \mathcal{F}, P)$. Thus the random integer measure on $\mathcal{O}$ defined by

$$\nu(B, t) = \sum_{i=1}^{\infty} \mathbb{1}_{\xi_i \in B} \mathbb{1}_{\tau_i \leq t}, \quad \forall B \in \mathcal{B}(\mathcal{O}),\ \forall t > 0$$

is the desired standard Poisson measure satisfying $\mathbb{E}\{\nu(B, t)\} = t\pi(B)$. Next, if ν is merely σ-finite then there exists a Borel partition of the whole space, $\mathcal{O} = \bigcup_n \mathcal{O}_n$, with $\pi(\mathcal{O}_n) < \infty$ and $\mathcal{O}_n \cap \mathcal{O}_k = \emptyset$ for $n \neq k$. For each $\mathcal{O}_n$ we can find a Poisson measure ν_n as above, and make the sequence of integer-valued random measure $(\nu_1, \nu_2, \ldots)$ independent. Hence $\nu = \sum_n \nu_n$ provides a standard Poisson measure with intensity π. Remark that ν_n is a finite standard Poisson measure on $\mathcal{O}_n \times [0, \infty)$ considered on the whole $\mathcal{O} \times [0, \infty)$ with intensity π_n, $\pi_n(B) = \pi(B \cap \mathcal{O}_n)$.

Moreover, if $\mathcal{O} = \mathbb{R}^d_*$ then the jump random process corresponds to the measure π restricted to $\mathcal{O}_n$

$$\mathtt{j}_n(t, \omega) = \sum_{i=1}^{\infty} \xi_i^n \mathbb{1}_{\tau_i^n \leq t}, \quad \forall t > 0$$

is properly defined, and if π integrates the function $z \mapsto |z|$ the jumps $\mathtt{j} = \sum_n \mathtt{j}_n$ (associated with ν_n) are defined almost surely. However, if π integrates only the function $z \mapsto |z|^2 \wedge 1$ then the stochastic integral is used to define the compensated jumps, formally $\mathtt{j} - \mathbb{E}\{\mathtt{j}\}$.

The same arguments apply to Poisson measures, if we start with an intensity measure defined on $\mathcal{O}\times[0,\infty)$. In this case, the (compensated) jumps is defined as a stochastic process, by integrating on $\mathcal{O}\times[0,t]$.

If the variable t is not explicitly differentiated, the construction of a Poisson (random is implicitly understood) measures ν on a Polish space Z, relative to a σ-finite measure Π can be simplified as follows: First, if Π is a finite measure then we can find a Poisson random variable η with parameter $c=\Pi(Z)$ and a sequence $(\zeta_1,\zeta_2,\ldots)$ of Z-valued independent identically distributed random variables, with distribution Π/c and independent of η in some (complete) probability space $(\Omega,\mathcal{F},P)$. Then $\nu(B)=\sum_{k=1}^{\eta}\mathbb{1}_{\zeta_k\in B}$, for any B in $\mathcal{B}(Z)$, satisfies

$$\mathbb{E}\{\nu(B)\}=\sum_n\mathbb{E}\Big\{\sum_{k=1}^{n}\mathbb{1}_{\zeta_k\in B}\,|\,\eta=n\Big\}=\sum_n\frac{n\Pi(B)}{c}P(\eta=n)=\Pi(B).$$

In particular, if $Z=\mathcal{O}\times[0,\infty)$ and $\Pi=\pi\times\mathrm{d}t$ then $\mathbb{E}\{\nu(B\times]0,t])\}=t\pi(B)$, for every B in $\mathcal{B}(\mathcal{O})$ and $t\geq 0$.

Thus, if Π is only σ-finite then partition the space $Z=\sum_n Z_n$ into sets with finite measure $\Pi(Z_n)<\infty$, and redo the construction with independent sequences $\{\eta^n\}$ and $\{\zeta_i^n\}$ to define $\nu(B)=\sum_n\sum_k\mathbb{1}_{k\leq\eta^n}\mathbb{1}_{\zeta_k^n\in B}$.

As in Çınlar [26, Theorems 3.2 and 3.19, Chapter 6, pp. 264-270], we can now consider

Proposition 2.29. *Let $Z=\sum_n Z_n$ and $X=\sum_n X_n$ be partitions of the Polish spaces Z and X, and let $\mathtt{m}_n(z,\mathrm{d}y)$ be a transition kernel from Z_n into X_n, i.e., (a) for every B in $\mathcal{B}(X)$ the mapping $z\mapsto\mathtt{m}_n(z,B)$ is $\mathcal{B}(Z_n)$-measurable and (b) for every z in Z_n the set function $B\mapsto\mathtt{m}_n(z,B)$ is a probability on X_n. Suppose that $\{\xi_1^n,\xi_2^n,\ldots\}$ are X_n-valued random variables conditionally independent given $\{\eta^n,\zeta_i^n:i\geq 1\}$, for each $n\geq 1$, such that ξ_i^n has distribution $\mathtt{m}(\zeta_i^n,\cdot)$. Then $\mu(B)=\sum_n\sum_{k=1}^{\eta_n}\mathbb{1}_{\xi_k^n\in B}$, for any B in $\mathcal{B}(X)$, is a Poisson measure with (marginal) intensity $\sum_n\int_{Z_n}\mathtt{m}_n(z,\cdot)\Pi(\mathrm{d}z)$ in X, and $\lambda(B)=\sum_n\sum_{k=1}^{\eta_n}\mathbb{1}_{(\zeta_k^n,\xi_k^n)\in B}$, for any B in $\mathcal{B}(Z\times X)$, is a Poisson measure with (product) intensity $\sum_n\Pi_n\times\mathtt{m}_n=\sum_n\mathtt{m}_n(z,\mathrm{d}x)\mathbb{1}_{Z_n}\Pi(\mathrm{d}z)$ in $Z\times X$.*

Proof. Since the random variable $\{(\zeta_i^n,\xi_i^n):i\geq 1\}$ is a sequence of independent identically distributed random variables with (product) distribution

$$P\big\{(\zeta_i^n,\xi_i^n)\in B\big\}=\int_{Z_n}\Pi(\mathrm{d}z)\int_{X_n}\mathbb{1}_B\mathtt{m}_n(z,\mathrm{d}x),$$

based on the above construction, we deduce that λ is a Poisson measure with (product) intensity $\sum_n\Pi_n\times\mathtt{m}_n$. Moreover, conditioning with respect to $\mathcal{B}(Z)$, we obtain the first assertion. Note that the marginal distribution is indeed

$$B\mapsto\mathtt{m}_n(\cdot,B)\Pi_n=\int_{Z_n}\mathtt{m}_n(z,B\cap X_n)\Pi(\mathrm{d}z),$$

for every B in $\mathcal{B}(X)$. □

If φ is a random transformation from Z into X, i.e., $(\omega, z) \mapsto \varphi(\omega, z)$ is a $\mathcal{F} \times \mathcal{B}(Z)$-measurable function from $\Omega \times Z$ into X. Then the marginal distributions

$$\mathtt{m}(z, B) = P\{\varphi(\omega, z) \in B\}, \quad \forall z \in Z, \quad \forall B \in \mathcal{B}(X)$$

defined a transition kernel as in Proposition 2.29. If ν is a Poisson measure with intensity Π on Z then

$$\mu(B) = \int_Z \mathbb{1}_{\{\varphi(\cdot,z)\in B\}}\, \nu(\mathrm{d}z) = \sum_n \sum_{k=1}^{\eta_n} \mathbb{1}_{\varphi(\zeta_k^n,\cdot)\in B}, \quad \forall B \in \mathcal{B}(X)$$

and

$$\lambda(B) = \int_Z \mathbb{1}_{\{(z,\varphi(\cdot,z))\in B\}}\, \nu(\mathrm{d}z) = \sum_n \sum_{k=1}^{\eta_n} \mathbb{1}_{(\zeta_k^n,\varphi(\xi_k^n,\cdot))\in B},$$

for every $B \in \mathcal{B}(Z \times X)$, are Poisson measures with intensities $\int_Z \mathtt{m}(z, \cdot)\Pi(\mathrm{d}z)$ on X and $\Pi \times \mathtt{m} = \mathtt{m}(z, \mathrm{d}x)\Pi(\mathrm{d}z)$ on $Z \times X$.

It is clear that $Z = \mathbb{R}_*^m$ and $X = \mathbb{R}_*^d \times [0, \infty)$ are special cases. The (nonnegative) intensity measure can be written as sum of its continuous and discontinuous parts, i.e.,

$$\Pi = \Pi^c + \Pi^d, \quad \Pi^d(\mathrm{d}z, \mathrm{d}t) = \mathbb{1}_{\{t\,:\,\Pi(\mathbb{R}_*^m \times \{t\})>0\}}\, \Pi(\mathrm{d}z, \mathrm{d}t).$$

There is a characterization of Poisson measures as follows

Theorem 2.30. *An integer-valued random measure ν is a Poisson measure if and only if its compensator ν^p is deterministic and continuous, i.e., $\nu^p = \Pi$ and $\Pi(\mathbb{R}_*^m \times \{t\}) = 0$ for every $t \geq 0$. Moreover, for any Poisson measure ν and any pairwise disjoint measurable sets $(B_1, B_2, \ldots, B_n)$ with finite Π-measure, the set $\{\nu(B_1), \nu(B_2), \ldots, \nu(B_1)\}$ is a family of independent random variables and $\nu(B_i)$ has a Poisson distribution with mean $\Pi(B_i)$, for any i.* □

In view of the above characterization, $\nu^p = \Pi$ for a Poisson measure and because of the previous Theorem 2.26 we deduce that ν should integrate the function $|z|^2 \wedge 1$ when the jumps process A associated with the Poisson measure ν is a general semi-martingale. For an (special) semi-martingale the intensity Π should integrate $|z|^2 \wedge |z|$. Thus, we are only interested in Lévy measures ν which necessarily integrate the function $|z|^2 \wedge 1$.

It is clear that homogeneous (or standard) Poisson measures are associated with the jumps of Lévy processes via (2.35), and with Remark 2.27 in mind, the integer measures ν_i associated with each component of X_i in $\mathbb{R}_*$ may not reconstruct the measure ν associated with the X in $\mathbb{R}_*^m$, even if each component is independent of the others.

For a proof (including extended Poisson measure) we refer to Jacod and Shiryaev [84, Theorem II.4.8, pp. 104–106]. The reader may consult, for instance, Bremaud [19], where jump processes are discussed as point processes in the framework of the queue theory.

Chapter 3

Stochastic Calculus I

This is the first chapter dedicated to the stochastic integral. Certainly, there are many excellent books on stochastic integrals, e.g., Bichteler [11], Chung and Williams [25], Da Prato [27], Kuo [107], Medvegyev [120], Métivier and Pellaumail [128], Protter [149], each with a particular objective. Our interest is stochastic integrals with respect to a Wiener process and a Poisson measure. Indeed, in the first section, a more analytic approach is used to introduce the concept of random orthogonal measures. Section 2 develops the stochastic integrals, first relative to a Wiener process, second relative to a Poisson measure, and then in general relative to a semi-martingale and ending with the vector valued stochastic integrals. The third and last Section is mainly concerned with the stochastic differential or Itô formula, first for Wiener-type integrals and then for Poisson-type integrals. The last two subsections deal with the previous construction as its dependency with respect to the filtration. First, non-anticipative processes are discussed and then quick analysis on functional representation is given.

3.1 Random Orthogonal Measures

Before going further, we take a look at the $\mathcal{L}^p$ and L^p spaces, for $1 \leq p < \infty$. Let μ be a complete σ-finite measure on the measurable space $(S, \mathcal{B})$ and Π be a total π-system of finite measure sets, i.e., (a) if F and G belong to Π then $F \cap G$ also belongs to Π, (b) if F is in Π then $m(F) < \infty$, and (c) there exists a sequence $\{S_k : k \geq 1\} \subset \Pi$ such that $S_k \subset S_{k+1}$ and $S = \bigcup_k S_k$. For any measurable function f with values in the extended real line $[-\infty, +\infty]$ we may define the quantity

$$\|f\|_p = \Big(\int |f|^p \, \mathrm{d}\mu \Big)^{1/p},$$

which may be infinite. The set of step or elementary functions $\mathcal{E}(\Pi, \mu)$ is defined as all functions of the form

$$e = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i},$$

where c_i are real numbers and A_i belongs to Π for every $i = 1, \dots, n$, i.e., the function e assumes a finite number of non-zero real values on sets in Π. Denote by $\mathcal{L}^p(\Pi, \mu)$ the sets of $\mathcal{B}$-measurable functions f with values in $[-\infty, +\infty]$ for which there exists a sequence $(e_1, e_2, \dots)$ of step functions $\mathcal{E}(\Pi, \mu)$ such that $\|f - e_n\|_p \to 0$ as $n \to \infty$. Since

$$|f|^p \le 2^{p-1}|e_n|^p + 2^{p-1}|f - e_n|^p,$$

all functions in $\mathcal{L}^p(\Pi, \mu)$ satisfies $\|f\|_p < \infty$, and in view of the triangle inequality

$$\|f + g\|_p \le \|f\|_p + \|g\|_p, \quad \forall f, g \in \mathcal{L}^p(\Pi, \mu),$$

the map $f \mapsto \|\cdot\|_p$ is a semi-norm. For $p = 2$, we may use the bilinear form

$$(f, g) = \int f\, g \, \mathrm{d}\mu, \quad \forall f, g \in \mathcal{L}^2(\Pi, \mu)$$

as a semi-inner product, which yields the semi-norm $\|\cdot\|_2$.

If f, f_n belong to $\mathcal{L}^p(\Pi, \mu)$ and $\|f - f_n\|_p \to 0$ as $n \to \infty$ we say that f_n converges to f in $\mathcal{L}^p(\Pi, \mu)$. Also if f_m, f_n belong to $\mathcal{L}^p(\Pi, \mu)$ and $\|f_m - f_n\|_p \to 0$ as $m, n \to \infty$ we say that $\{f_n\}$ is a Cauchy sequence in $\mathcal{L}^p(\Pi, \mu)$. It is clear that any Cauchy sequence in $\mathcal{L}^p(\Pi, \mu)$ has a almost everywhere convergent sub-sequence. Next, essentially based on the triangular inequality and Fatou's Lemma, we deduce that $\mathcal{L}^p(\Pi, \mu)$ is a complete vector space, i.e., (1) for any f, g in $\mathcal{L}^p(\Pi, \mu)$ and any a, b in $\mathbb{R}$ the function $af + bg$ is in $\mathcal{L}^p(\Pi, \mu)$ and (2) any Cauchy sequence in $\mathcal{L}^p(\Pi, \mu)$ converges to a function in $\mathcal{L}^p(\Pi, \mu)$. Thus, if $\sigma_\mu(\Pi)$ is the smaller sub-σ-algebra of $\mathcal{B}$ containing Π and all μ-null sets then $\mathcal{L}^p(\Pi, \mu) = \mathcal{L}^p(\sigma_\mu(\Pi), \mu)$, after using a monotone class argument.

If we should identify functions which are equals almost everywhere, i.e., use classes of equivalence $f \sim g$ if and only if $f = g$ almost everywhere, then the quotient space $L^p(\Pi, \mu) = \mathcal{L}^p(\Pi, \mu)/_\sim$ is a Banach (Hilbert for $p = 2$) space.

Definition 3.1 (random orthogonal measure)**.** A family of real-valued random variables $\{\zeta(A) : A \in \Pi\}$ on a complete probability space $(\Omega, \mathcal{F}, P)$ is called a *random orthogonal measure with structural measure* μ if

(a) $\mathbb{E}\{|\zeta(A)|^2\} < \infty$ for any A in Π,

(b) $\mathbb{E}\{\zeta(A)\,\zeta(B)\} = \mu(A \cap B)$ for any A, B in Π.

Note that the random variables $\zeta(A)$ are almost surely defined, i.e., they are elements in $L^2(\Omega, \mathcal{F}, P)$, and the measure μ and the π-system Π are as above. □

Clearly, the above condition (b) translates the *orthogonal* condition, whist the word *measure* can be justified as follows: if A is a disjoint union of sets in Π, i.e., $A = \cup_i A_i$, $A_i \cap A_j = \emptyset$ if $i \neq j$, then

$$\Big(\mathbb{1}_A - \sum_i \mathbb{1}_{A_i}\Big)^2 = \mathbb{1}_{A \smallsetminus \cup_i A_i} = \mathbb{1}_A - \sum_i \mathbb{1}_{A_i},$$

which yields

$$\mathbb{E}\Big\{\Big(\mathbb{1}_A - \sum_i \mathbb{1}_{A_i}\Big)^2\Big\} = \mu\Big(\mathbb{1}_{A \smallsetminus \cup_i A_i}\Big),$$

i.e., for each sequence $\{A_i\}$ as above, there exists a set Ω_0 in $\mathcal{F}$ with $P(\Omega_0) = 1$ such that $\zeta(A, \omega) = \sum_i \zeta(A_i, \omega)$ for every ω in Ω_0. This is not to say that a *regular* selection exists, i.e., to show that (except for set of probability zero) the mapping $A \mapsto \zeta(A)$ can be extended to a measure in $\sigma(\Pi)$, which involves a countable generated π-system Π and some topology on ω, as in the case of regular conditional probability measures.

Let as define the operator $e \mapsto I(e)$ from the set of elementary (or step) functions $\mathcal{E}(\Pi, \mu)$ into the Hilbert space $L^2(\Omega, \mathcal{F}, P) = L^2(\mathcal{F}, P)$ by the formula

$$\text{if} \quad e = \sum_{i=1}^n c_i \mathbb{1}_{A_i} \quad \text{then} \quad I(e) = \int e \mathrm{d}\zeta = \sum_{i=1}^n c_i\, \zeta(A_i), \tag{3.1}$$

which is clearly independent of the particular representation of the given elementary function. Thus, we have

$$\begin{aligned} &(I(e), \zeta(A))_{\mathcal{F}} = (e, \mathbb{1}_A)_\Pi, \quad \forall A \in \Pi, \\ &\|I(e)\|_{2,\mathcal{F}} = \|e\|_{2,\Pi}, \end{aligned}$$

where $(\cdot,\cdot)_{\mathcal{F}}$ and $(\cdot,\cdot)_\Pi$ denote the inner or scalar products in the Hilbert spaces $L^2(\Omega, \mathcal{F}, P)$ and $L^2(\Pi, \mu) = L^2(\sigma_\mu(\Pi), \mu)$, respectively. Next, by linearity the above definition is extended to the vector space generated by $\mathcal{E}(\Pi, \mu)$, and by continuity to the whole Hilbert space $L^2(\Pi, \mu)$. Hence, this procedure constructs a linear isometry map between the Hilbert spaces $L^2(\Pi, \mu)$ and $L^2(\mathcal{F}, P)$ satisfying

$$\begin{aligned} &I : f \mapsto \int f \mathrm{d}\zeta, \quad \forall f \in L^2(\Pi, \mu), \\ &(I(f), I(g))_{\mathcal{F}} = (f, g)_\Pi, \quad \forall f, g \in L^2(\Pi, \mu). \end{aligned} \tag{3.2}$$

Certainly, there is only some obvious changes if we allow integrand functions with complex values, and if the spaces $L^p(\Pi, \mu)$ are defined with complex valued functions, and so, the inner product in L^2 need to use the complex-conjugation operation.

The above construction does not give a preferential role to the *time* variable as in the case of stochastic processes, and as mentioned in the book by

Krylov [102, Section III.1, pp. 77-84], this procedure is used in several opportunities, not only for the stochastic integral. The interested reader may consult Gikhman and Skorokhod [61, Section V.2] for a detailed analysis on (vector valued) orthogonal measures.

3.1.1 Orthogonal or Uncorrelated Increments

Random orthogonal measures is a generalization of stochastic processes with orthogonal (or uncorrelated) increments, the reader is referred to the classic book Doob [33, Chapter IX, pp. 425–451] for more details. A $\mathbb{R}^d$-valued (for complex valued use conjugate) x is said to have *uncorrelated increments* if the increments are square-integrable and uncorrelated, i.e., if (a) $\mathbb{E}\{|x(t)-x(s)|^2\} < \infty$, for every $t > s \geq 0$ and (b) $\mathbb{E}\{(x(t_1) - x(s_1))\,(x(t_2) - x(s_2))\} = \mathbb{E}\{x(t_1) - x(s_1)\}\,\mathbb{E}\{x(t_2)-x(s_2)\}$ for any $0 \leq s_1 < t_1 \leq s_2 < t_2$. Similarly, x has *orthogonal increments* if $\mathbb{E}\{(x(t_1) - x(s_1))\,(x(t_2) - x(s_2))\} = 0$. It is clear that a process with independent increments is also a process with uncorrelated increments and that we may convert a process x with uncorrelated increments into a process with orthogonal (and uncorrelated) increments y by subtraction its means, i.e., $y(t) = x(t) - \mathbb{E}\{x(t)\}$. Thus, we will discuss only orthogonal increments.

If y is a process with orthogonal increments then we can define the (deterministic) monotone increasing function $F_y(t) = \mathbb{E}\{|y(t) - y(0)|^2\}$, for any $t \geq 0$, with the property that $\mathbb{E}\{|y(t) - y(s)|^2\} = F_y(t) - F_y(s)$, for every $t \geq s \geq 0$. Because the function F_y has a countable number of discontinuities, the mean-square left $y(t-)$ and right $y(t+)$ limit of $y(t)$ exist at any $t \geq 0$ and, except for a countable number of times $y(t-) = y(t) = y(t+)$. Therefore, we can define real-valued random variables $\{\zeta(A) : A \in \Pi_+\}$, where Π_+ is the π-system of semi-open intervals $(a, b]$, $b \geq a \geq 0$ and

$$\zeta(A) = y(b+) - y(a+), \quad A = (a, b],$$

which is a random orthogonal measure with structural measure μ, the Lebesgue-Stieltjes measure generated by F_y, i.e., $\mu(A) = F_y(b+) - F_y(a+)$, for any $A = (a, b]$. Certainly, we may use the π-system Π_- of semi-open intervals $[a, b)$, $b \geq a \geq 0$ and $\zeta(A) = y(b-) - y(a-)$, with $A = [a, b)$, or the combination of the above π-system, and we get the same structural measure (and same extension of the orthogonal measure ζ). Moreover, we may even use only the π-system of interval of the form $[0, b)$ (or $(0, b]$) to initially define the random orthogonal measure.

Now, applying the previous we can define the stochastic integral for any (deterministic) function in $L^2(\sigma_\mu(\Pi), \mu)$

$$\int_{\mathbb{R}} f(t)\mathrm{d}y(t) = \int f\mathrm{d}\zeta$$

as an equivalence class of square-integrable random variables, even if we actually think of a particular member. Moreover, the way how this is defined (via limit

of elementary or step functions) allows us to that the stochastic integral process

$$\Phi(s) = \int_{\mathbb{R}} \varphi(s,t)\mathrm{d}y(t)$$

can be chosen measurable if φ is a measurable function with respect to the Lebesgue-Stieltjes measure $\mathrm{d}s\,\mathrm{d}F(t)$ satisfying

$$\int_{\mathbb{R}} \varphi(s,t)\mathrm{d}F(t),$$

all s except in a set of zero Lebesgue measure. Clearly, the stochastic integral over a Borel (even μ-measurable) set of time A can be define as

$$\int_A f(t)\mathrm{d}y(t) = \int f\mathbb{1}_A \mathrm{d}\zeta.$$

A Fubini type theorem holds, for the double integral, and in particular, if h is an absolutely continuous function and $\mathbb{1}_{\{s\leq t\}}$ denotes the function equal to 1 when $s \leq t$ and equal to 0 otherwise, then exchanging the order of integration we deduce

$$\int_a^b \mathrm{d}s \int_{(a,b]} h'(s)\mathbb{1}_{\{s\leq t\}}\mathrm{d}y(s) = \int_{(a,b]} [h(t)-h(a)]\mathrm{d}y(t) =$$
$$= [h(b)-h(a)][y(b+)-y(a+)] - \int_a^b [y(t)-y(a+)]\mathrm{d}t,$$

for any $b > a \geq 0$.

3.1.2 Typical Examples

There two typical constructions of random orthogonal measures, based on the Poisson and the Gaussian distributions, or equivalent on the Poisson process and the Wiener process, both are processes with independent increments.

Perhaps a simple (constructed) example of a random orthogonal measure begins with a given (structural) finite measure m on $S = \mathbb{R}^d_* = \mathbb{R}^d \smallsetminus \{0\}$, where the π-system Π plays almost not role. Let $\{\tau_n, z_n : n \geq 1\}$ be a sequence of independent random variables in a probability space $(\Omega, \mathcal{F}, P)$, such that each τ_n is exponentially distributed with parameter $m(\mathbb{R}^d_*)$ and z_n has the distribution law $A \mapsto m(A)/m(\mathbb{R}^d_*)$. Define the compound Poisson process $p_t = \sum_n z_n\,\mathbb{1}_{t\geq\theta_n}$, with $\theta_n = \tau_1 + \cdots + \tau_n$. This can be written as $p_t = \sum_{n=1}^{N_t} z_n$, where $N_t = \sum_n \mathbb{1}_{t\geq\theta_n}$ is the Poisson process counting the jumps, which has a Poisson distribution with intensity $\lambda = m(\mathbb{R}^d_*)$, i.e., $P\{N_t = n\} = \mathrm{e}^{-\lambda t}(\lambda t)^n/n!$, $n = 0, 1, 2, \ldots$, and thus $E\{N_t\} = \lambda t$ and $E\{(N_t - \lambda t)^2\} = \lambda t$.

If the emphasis is only on the jumps then the series defining the Poisson process p_t is regarded as the sum-of-jumps of the sequence of jumps $\{z_n, \theta_n : n \geq 1\}$, which is referred to as a Poisson point process, where z_n is the size of

the jump at the time θ_n. Note that if initially the measure m is given on $\mathbb{R}^d$ and $m(\{0\}) \neq 0$ then the above expression of N_t does not count the actual jumps of the compound Poisson process p_t, i.e., the random process $q_t = \sum_n \mathbb{1}_{z_n=0}\mathbb{1}_{t\geq\theta_n}$ intervenes.

The independence of the random variables $\{z_n\}$ and $\{\theta_n\}$ and the fact all random variables z_n have the same distribution, imply that

$$\mathbb{E}\{p_t\} = \mathbb{E}\{z\}\sum_n \mathbb{E}\{\mathbb{1}_{t\geq\theta_n}\} = m(z)t,$$

where $m(z)$ means the integral of the function $z \mapsto z$ with respect to the measure m, i.e., $m(z) = \mathbb{E}\{z_1\}m(\mathbb{R}^d_*)$. Similarly, if $m(|z|^2) = \mathbb{E}\{|z_1|^2\}m(\mathbb{R}^d_*)$ then more calculations show that the variance $\mathbb{E}\{|p_t - m(z)t|^2\} = m(|z|^2)t$, and also

$$\mathbb{E}\{\mathrm{e}^{irp_t}\} = \exp\big[m(\mathbb{R}^d_*)t\big(\mathbb{E}\{\mathrm{e}^{irz_1}\} - 1\big)\big] = \exp\big[tm(\mathrm{e}^{irz} - 1)\big]$$

is its characteristic function. Moreover, these distributions also imply that

$$\mathbb{E}\{\mathbb{1}_{z_n\in A}\} = \frac{m(A)}{m(\mathbb{R}^d_*)} \qquad \text{and} \qquad \sum_k \mathbb{E}\{\mathbb{1}_{\theta_{n+k}\leq t}\} = m(\mathbb{R}^d_*)t,$$

for every $t \geq 0$ and A in Π. Therefore, this yields the Poisson orthogonal measure

$$\zeta_t(A) = \sum_n \big[\mathbb{1}_{z_n\in A}\,\mathbb{1}_{t\geq\theta_n} - \mathbb{E}\{\mathbb{1}_{z_n\in A}\,\mathbb{1}_{t\geq\theta_n}\}\big], \quad \forall A \in \Pi.$$

Indeed, by construction $\mathbb{E}\{\zeta_t(A)\} = 0$, $\sum_n \mathbb{E}\{\mathbb{1}_{z_n\in A}\,\mathbb{1}_{t\geq\theta_n}\} = m(A)t$, and

$$\mathbb{E}\{\mathbb{1}_{z_n\in A}\,\mathbb{1}_{t\geq\theta_n}\,\mathbb{1}_{z_k\in B}\,\mathbb{1}_{t\geq\theta_k}\} = \frac{m(A)\,m(B)}{m(\mathbb{R}^d_*)}\,\mathbb{E}\{\mathbb{1}_{t\geq\theta_{n\vee k}}\},$$
$$\mathbb{E}\{\mathbb{1}_{z_n\in A}\,\mathbb{1}_{t\geq\theta_n}\,\mathbb{1}_{z_n\in B}\,\mathbb{1}_{t\geq\theta_n}\} = \frac{m(A\cap B)}{m(\mathbb{R}^d_*)}\,\mathbb{E}\{\mathbb{1}_{t\geq\theta_n}\}, \quad \forall n,k,$$

and because $\sum_{n,k} = \sum_n + 2\sum_n\sum_{k=n+1}$ we have

$$\sum_{n,k}\mathbb{E}\{\mathbb{1}_{z_n\in A}\,\mathbb{1}_{t\geq\theta_n}\,\mathbb{1}_{z_k\in B}\,\mathbb{1}_{t\geq\theta_k}\} = m(A\cap B)t + 2m(A)m(B)t$$

which yields

$$\mathbb{E}\{\zeta_t(A)\zeta_t(B)\} = \sum_n \mathbb{E}\{\mathbb{1}_{z_n\in A\cap B}\,\mathbb{1}_{t\geq\theta_n}\} = t\,m(A\cap B),$$

as desired. Recall that the mapping $A \mapsto \zeta(A,\omega)$ is regarded as defined for any A in Π and taking values in $L^2(\Omega,\mathcal{F},P)$, i.e., properly saying the symbol $\zeta(A)$ is a class of equivalence of square-integrable random variables.

In general, if m is measure in $\mathbb{R}^d_* = \mathbb{R}^d \smallsetminus \{0\}$ that integrates the function $x \mapsto 1 \wedge |x|^2$ and $\{R_k : k \geq 1\}$ is a countable partition of finite m-measure, i.e.,

$\mathbb{R}^d_* = \bigcup_k R_k$ with $m(R_k) < \infty$ and $R_k \cap R_n = \emptyset$, for $k \neq n$, then we repeat the previous procedure with the finite measure $A \mapsto m(A \cap R_k)$ to construct an independent sequence of compound Poisson processes $\{p_t(R_k) : k \geq 1\}$, which yields the independent sequence Poisson orthogonal measures $\{\zeta_t(R_k) : k \geq 1\}$. Since $\mathbb{E}\{\zeta_t(R_k)\} = 0$, the sequence of Poisson orthogonal measures is an orthogonal system in $L^2(\Omega, \mathcal{F}, P)$, and so the series $\zeta_t(A) = \sum_k \zeta_k(A)$, for every A in Π, defines a Poisson orthogonal measure with structural measure $A \mapsto t\, m(A)$. Summing-up, if for a fixed $k = 1, 2, \ldots$, $(N^k_t, t \geq 0)$ is the Poisson process and $\{z^k_n : n \geq 1\}$ is the iid sequence with distribution $m(\cdot \cap R_k)$ then the compound Poisson processes $p_t(R_1), p_t(R_2), \ldots$ are independent and the series of jumps $\sum_k \sum_{n=1}^{N^k_t} z^k_n$ defines a Poisson point process with Lévy measure m, which yields the same Poisson orthogonal measure, namely,

$$\zeta_t(A) = \sum_k \Big[\sum_{n=1}^{N^k_t} \mathbb{1}_{z^k_n \in A} - \mathbb{E}\Big\{ \sum_{n=1}^{N^k_t} \mathbb{1}_{z^k_n \in A} \Big\} \Big], \quad \forall A \in \Pi,\ t \geq 0, \tag{3.3}$$

where the series (in the variable k) converges in the L^2-norm, i.e., for each k the series in n reduces to a finite sum for each ω, but the series in k defines $\zeta_t(A)$ as an element in $L^2(\Omega, \mathcal{F}, P)$. Note that in this construction, the variable t is considered fixed, and that $A \mapsto \mu(A) = tm(A)$ is the structural measure associated with the Poisson orthogonal measure $A \mapsto \zeta_t(A)$. Therefore, any square-integrable (deterministic) function f, i.e., any element in $L^2(\Pi, \mu) = L^2(\sigma_\mu(\Pi), \mu)$.

As seen in the previous section, any process with orthogonal increments yields a random orthogonal measure, in particular, a one-dimensional standard Wiener process $(w(t), t \geq 0)$ (i.e., $w(t)$ is a standard normal variables, $t \mapsto w(t)$ is almost surely continuous, and $\mathbb{E}\{w(t) \wedge w(s)\} = t \wedge s$) has independent increments and thus the expression $\zeta(]a, b]) = w(b) - w(a)$ defines a random orthogonal measure on the π-system of semi-open intervals $\Pi_+ = \{]a, b] : a, b \in \mathbb{R}\}$ with the Lebesgue measure as its structural measure, i.e., $\mathbb{E}\{\zeta(]a, b])\} = b - a$.

Similarly, the Poisson orthogonal measure $\zeta_t(A)$ defined previously can be regarded as a random orthogonal measure on π-system Π (which is composed by all subsets of $S = \mathbb{R}^d_* \times (0, \infty)$ having the form $K \times (0, t]$ for a compact set and a real number $t \geq 0$) with structural measure $\mu = m \times \mathrm{d}t$, where $\mathrm{d}t$ is the Lebesgue measure.

With this argument, we are able to define the stochastic integral of an (deterministic) integrand function $L^2(\sigma_\mu(\Pi), \mu)$ with respect to a random orthogonal measure constructed form either a Poisson point process with Lévy measure m or a (standard) Wiener process, which are denoted by either

$$\zeta(K \times (0, t]) = \tilde{p}(A \times (0, t]), \quad \text{and} \quad \int_{\mathbb{R}^d_* \times]0,T} f(t) \tilde{p}(\mathrm{d}z, \mathrm{d}t),$$

or

$$\zeta(]a, b]) = w(]a, b]), \quad \text{and} \quad \int_a^b f(t) \mathrm{d}w(t).$$

Note that this is not a pathwise integral, e.g., the paths of the Wiener process are almost surely of unbounded variation on any bounded time-interval and something similar holds true for the Poisson point process depending on the Lévy measure.

Perhaps a simple construction of a Wiener process begins with a sequence of independent standard normally distributed random variables $\{e_{i,n} : i = 1, 2, \dots, 4^n,\ n \geq 1\}$. Since each $e_{i,n}$ has zero mean and are independent of each other, the sequence is orthogonal in $L^2 = L^2(\Omega, \mathcal{F}, P)$, actually, it is an orthonormal system since all variances are equal to 1. Recalling the dyadic expressions that if $t = k2^{-m} = (k2^{n-m})2^{-n}$, $1 \leq k \leq 4^m$ then $k2^{n-m} \leq 4^n$, $\mathbb{1}_{i2^{-n} \leq t} = 1$ if and only if $i = 1, \dots, k2^{n-m}$, which yields $\sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t} = k2^{n-m} = t2^n$ if $k2^{n-m} = t2^n \geq 1$, we deduce $t = \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t}$, so that the random variable

$$w(t) = \sum_n 2^{-n} \sum_{i=1}^{4^n} e_{i,n} \mathbb{1}_{i2^{-n} \leq t}, \tag{3.4}$$

is defined as a convergent series in $L^2(\Omega, \mathcal{F}, P)$, for every $t > 0$. Indeed, regard the expression as an orthogonal series expansion, and set $w(0) = 0$, for any $t \geq s \geq 0$, to have

$$\mathbb{E}\{|w(t) - w(s)|^2\} = \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{E}\{|e_{i,n}|^2\} \mathbb{1}_{s < i2^{-n} \leq t} =$$

$$= \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{s < i2^{-n} \leq t} = (t - s).$$

Thus, $t \mapsto w(t)$ provides a L^2-norm continuous random process satisfying (a) $w(t)$ is a Gaussian random variable with mean $\mathbb{E}\{w(t)\} = 0$ and variance $\mathbb{E}\{|w(t)|^2\} = t$, and (b) $w(s)$ is independent of $w(t) - w(s)$ for every $t > s$. The fact that there is a continuous version of the limiting process $(w(t) : t \geq 0)$, which is called a Wiener process plays not an important role in this analysis. Indeed, the expressions (3.4) of a Wiener process and (3.3) of a centered Poisson point process are cad-lag and therefore, the corresponding random orthogonal measures are measures, for every fixed ω almost surely.

Certainly, for dimension d higher than 1 we should use the covariance matrix, i.e., $\mathbb{E}\{w(t)w^*(t)\} = tI_d$, with I_d the identity matrix. In this case, this could take us to discuss vector-valued random orthogonal measure, or simply consider a sum of independent Wiener processes and their corresponding orthogonal measures.

However, with little effort, an index $j = 1, \dots, d$ could be added to the iid sequence $\{e^j_{i,n}\}$, so that d-intervals $(0, t] = (0, t_1] \times \cdots \times (0, t_d]$ on $S = (0, +\infty)^d$ could be used to define

$$w(t_1, \dots, t_d) = \sum_n 2^{-n} \sum_{i=1}^{4^n} \sum_{j=1}^{d} e^j_{i,n} \mathbb{1}_{i2^{-n} \leq t_1, \dots, i2^{-n} \leq t_d}, \tag{3.5}$$

as a convergent series in $L^2(\Omega, \mathcal{F}, P)$. Besides being a Gaussian random variable with mean zero, note that

$$\mathbb{1}_{i2^{-n}\le t_1} \cdots \mathbb{1}_{i2^{-n}\le t_d} = \mathbb{1}_{i2^{-n}\le t_1,\,\ldots,\,i2^{-n}\le t_d}$$

implies

$$\mathbb{E}\{|w(t_1,\ldots,t_d)|^2\} = \sum_n 4^{-n} \sum_{i=1}^{4^n} \sum_{i=1}^{d} \mathbb{E}\{|e^j_{i,n}|^2\}\, \mathbb{1}_{i2^{-n}\le t_1,\,\ldots,\,i2^{-n}\le t_d} = \\ = \prod_{j=1}^{d} \Big[\sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n}\le t_j} \Big] = \prod_{j=1}^{d} t_j,$$

which yields the (random) Gaussian orthogonal measure $\zeta(]0,t]) = w(t_1,\ldots,t_d)$ in $\mathbb{R}^d$, with the Lebesgue measure on $(0,\infty)^d$.

Clearly, this last example is related with the so-called *white noise measure*, and *Brownian sheet* or *space-time* Brownian motion, e.g., see Kallianpur and Xiong [90, Section 3.2, pp. 93–109].

3.1.3 Filtration and Martingales

At this point, only deterministic integrand can be taken when the integrator is a standard Wiener process or a Poisson point process with Lévy measure m. To allow stochastic integrand a deeper analysis is needed to modify the π-system. Indeed, the two typical examples of either the Poisson or the Gaussian orthogonal measure suggests a π-system of the form either $\Pi = \{K\times]0,\tau] \subset \Omega\times\mathbb{R}^d_*\times(0,\infty)\}$, with the structural measure $\mu(K\times]0,\tau]) = \mathbb{E}\{\tau\}m(K)$ for the underlying product measure $P\times m\times \mathrm{d}t$, or $\Pi = \{]0,\tau] \subset \Omega\times(0,\infty)\}$, with the structural measure $\mu(]0,\tau]) = \mathbb{E}\{\tau\}$ for the underlying product measure $P\times \mathrm{d}t$, for a compact set K of $\mathbb{R}^d_*$ and a *bounded stopping time* τ. This means that there is defined a filtration $\mathbb{F}$ in the probability space $(\Omega,\mathcal{F},P)$, i.e., a family of sub σ-algebras $\mathcal{F}_t \subset \mathcal{F}$ such that (a) $\mathcal{F}_t \subset \mathcal{F}_s$ if $s > t \ge 0$, (b) $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ if $t \ge 0$, and (c) N belongs to $\mathcal{F}_0$ if N is in $\mathcal{F}$ and $P(N) = 0$. Therefore, of relevant interest is to provide some more details on the square-integrable functions that can be approximated by a sequence of Π-step functions, i.e., the Hilbert space $L^2(\Pi,\mu)$ or better $\mathcal{L}^2(\Pi,\mu)$.

This filtration $\mathbb{F}$ should be such that either $t \mapsto \zeta(K\times]0,t])$ or $t \mapsto \zeta(]0,t])$ is a $\mathbb{F}$-martingale. Because, both expressions (3.4) of a Wiener process and (3.3) of a Poisson point process have zero-mean with independent increments, the martingale condition reduces to either $\zeta(K\times]0,t])$ or $\zeta(]0,t])$ being $\mathcal{F}_t$-measurable, i.e., adapted to the filtration $\mathbb{F}$.

Under this $\mathbb{F}$-martingale condition, either the Poisson or the Gaussian orthogonal measure can be considered as defined on the above Π with structural (product) measure either $P\times m\times \mathrm{d}t$ or $P\times \mathrm{d}t$, i.e., just replacing a deterministic time t with a bounded stopping time τ. All this requires some work. In particular, a key role is played by the so-called predictable σ-algebra $\mathcal{P}$ in

either $\Omega \times \mathbb{R}^d_* \times (0,\infty)$ or $\Omega \times (0,\infty)$, which is the σ-algebra generated by the π-system Π, and eventually completed with respect to the structural measure μ. For instance, in this setting, a real-valued process $(f(t), t \geq 0)$ is an integrand (i.e., it is an element in $L^2(\Pi,\mu)$) if and only if (a) it is square-integrable, (i.e., it belongs to $L^2(\mathcal{F}\times\mathcal{B},\mu)$, $\mathcal{B}$ is the Borel σ-algebra either in $\mathbb{R}^d_*\times]0,\infty[$ or in $]0,\infty[$), and (b) its μ-equivalence contains a predictable representative. In other words, square-integrable predictable process are the good integrand, and therefore its corresponding class of μ-equivalence. Sometimes, stochastic intervals are denoted by $]\!]a,b]\!]$ (or $[\![a,b[\![$) to stress the randomness involved. Certainly, this argument also applies to the multi-dimensional Gaussian orthogonal measures (or Brownian sheet). On the other hand, the martingale technique is used to define the stochastic integral with respect to a martingale (non-necessarily with orthogonal), and various definitions are proposed. In any way, the stochastic integral becomes very useful due to the stochastic calculus that follows.

Among other sources, regarding random orthogonal measures and processes, the reader may consult the books by Krylov [102, Section III.1, pp. 77-84], Doob [33, Section IX.5, pp. 436–451], Gikhman and Skorokhod [61, Section V.2] and references therein for a deeper analysis.

3.2 Stochastic Integrals

Let us fix a filtered space $(\Omega,\mathcal{F},P,\mathcal{F}_t : t \geq 0)$. A *simple predictable process* (or piecewise constant over stochastic intervals) is a stochastic process of the form $Y(t) = Y(\tau_{i-1})$ if $\tau_{i-1} < t \leq \tau_i$ with some $i = 1,\dots,n$, where $0 = \tau_0 \leq \tau_1 \leq \cdots \leq \tau_n$ are stopping times and $Y(\tau_{i-1})$ is a $\mathcal{F}(\tau_{i-1})$ measurable random variable for any i, and $Y(t) = 0$ otherwise. It is called *bounded* if all $Y(\tau_{i-1})$ are bounded random variables. Note that any simple predictable process Y is left continuous with right-hand limits, so that $t \mapsto Y(t+)$ is a cad-lag process.

If X is an optional cad-lag process then we define the expression

$$Z(t) = \int_{(0,t]} Y(s)\mathrm{d}X(s) = \sum_{i=1}^{n} Y(\tau_{i-1})[X(t\wedge\tau_i) - X(t\wedge\tau_{i-1})], \tag{3.6}$$

as the integral of the simple predictable process (integrand) Y with respect to the optional cad-lag process (integrator) X. This integral process Z is cad-lag and optional, which is also continuous if X is so. On the other hand, the integration-by-part formula

$$X(b)Y(b) - X(a)Y(a) = \int_{(a,b]} X(t-)\mathrm{d}Y(t) + \int_{(a,b]} Y(t-)\mathrm{d}X(t) + \\ + \sum_{a<t\leq b} [X(t) - X(t-)]\,[Y(t+) - Y(t)], \tag{3.7}$$

yields the expression

$$Z(t) = \int_{(0,t]} Y(s)\mathrm{d}X(s) = \\ = X(t)Y(t) - \sum_{i=1}^{n} X(\tau_i)[Y(t \wedge \tau_i) - Y(t \wedge \tau_{i-1})], \quad (3.8)$$

which can be used to define the same integral process.

If $t \mapsto X(t,\omega)$ has also locally bounded variation for almost every ω then the measure theory can be used on (3.6) to extend the definition of the integral to a class of predictable processes, including all continuous adapted processes. On the other hand, we can use (3.8) to extend the definition of the integral to a class of predictable processes, including all continuous adapted processes with locally bounded variation. In either case, with this *pathwise analysis*, we are unable to see how two continuous processes of unbounded variation can be integrated, which is the case of a Wiener process as integrand and as integrator. In contrast with what follows, the fact that we use adapted processes is irrelevant in pathwise analysis. The reader may want to consult the classic reference McKean [119] for a comprehensive treatment. Also remark that recent development (e.g., see Dudley and Norvaisa [36], Lyons and Qian [113], Lyons et al. [114]) allows to view the stochastic integral as a pathwise Young integral with process of infinite variation (by means of the p-variation norm and without using martingales), referred to as processes with rough paths.

3.2.1 Relative to Wiener Processes

Let $(w(t) : t \geq 0)$ be a real-valued standard Wiener process in a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, i.e., $w(t)$ and $w^2(t) - t$ are continuous martingales relative to the filtration $(\mathcal{F}_t : t \geq 0)$, with $w(0) = 0$. Denote by $\mathcal{E}$ the vector space of all processes of the form $f(t,\omega) = f_{i-1}(\omega)$ if $t_{i-1} < t \leq t_i$ with some $i = 1, \ldots, n$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers and f_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $f(t,\omega) = 0$ otherwise. Elements in $\mathcal{E}$ are called elementary predictable processes. it is clear what the integral should be for any integrand in $\mathcal{E}$, namely

$$\int f(s)\mathrm{d}w(s) = \sum_{i=1}^{n} f_{i-1}[w(t_i) - w(t_{i-1})], \quad (3.9)$$

and

$$\int_{(0,t]} f(s)\mathrm{d}w(s) = \sum_{i=1}^{n} f_{i-1}[w(t \wedge t_i) - w(t \wedge t_{i-1})], \quad \forall t \geq 0,$$

$$\int_{(a,b]} f(s)\mathrm{d}w(s) = \int_{(0,b]} f(s)\mathrm{d}w(s) - \int_{(0,a]} f(s)\mathrm{d}w(s), \quad \forall b > a \geq 0.$$

Note that

$$\int_{(a,b]} f(s)\mathrm{d}w(s) = \int f(s)\,\mathbb{1}_{(a,b]}(s)\mathrm{d}w(s),$$

for every $b > a \geq 0$. This definition (3.9) is independent of the particular representation used and because f_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable we obtain

$$\mathbb{E}\Big\{\Big|\int f(s)\mathrm{d}w(s)\Big|^2\Big\} = \sum_{i=1}^{n} \mathbb{E}\{|f_{i-1}|^2(t_i - t_{i-1})\} = \mathbb{E}\{\int |f(s)|^2\mathrm{d}s\}, \tag{3.10}$$

for every f in $\mathcal{E}$. Moreover the processes

$$\int_{(0,t]} f(s)\mathrm{d}w(s) \quad \text{and} \quad \Big|\int_{(0,t]} f(s)\mathrm{d}w(s)\Big|^2 - \int_0^t |f(s)|^2\mathrm{d}s, \tag{3.11}$$

for every $\forall t \geq 0$, are continuous martingales, and

$$\mathbb{E}\Big\{\Big[\int f(s)\mathrm{d}w(s)\Big]\,\Big[\int g(s)\mathrm{d}w(s)\Big]\Big\} = \mathbb{E}\Big\{\int f(s)\,g(s)\mathrm{d}s\Big\}, \tag{3.12}$$

for any two stochastic processes f and g in $\mathcal{E}$. Denote by $\bar{\mathcal{E}}$ the L^2-closure of $\mathcal{E}$, i.e., the Hilbert space of all processes f for which there exists a sequence $(f_1, f_2, \ldots)$ of processes in $\mathcal{E}$ such that

$$\lim_{n\to\infty} \mathbb{E}\Big\{\int |f_n(s) - f(s)|^2\mathrm{d}s\Big\} = 0.$$

Based on the martingale inequality

$$\mathbb{E}\Big\{\sup_{0\leq t\leq T}\Big|\int_{(0,t]} f(s)\mathrm{d}w(s)\Big|^2\Big\} \leq 4\,\mathbb{E}\Big\{\int_0^T |f(s)|^2\mathrm{d}s\Big\}, \tag{3.13}$$

for every $T \geq 0$, and the isometry identity (3.10), this linear operation can be extended to the closure $\bar{\mathcal{E}}$, preserving linearity and the properties (3.10), (3.11), (3.12). This is called Itô integral or generally *stochastic integral*. Besides a density argument, the estimate (3.13) is used to show that the stochastic integral on $(0,t]$ is a continuous process as a function of $t \geq 0$, for any f in $\bar{\mathcal{E}}$.

If τ and θ are stopping times with $\theta \leq \tau \leq T$ (with T a constant) then the process

$$\mathbb{1}_{]\!]\theta,\tau]\!]} : (\omega, t) \mapsto \mathbb{1}_{\theta(\omega)<t\leq\tau(\omega)}$$

is elementary predictable process, indeed, for any partition $0 = t_0 < t_1 < \cdots < t_n$, with $t_n \geq T$ we have

$$\mathbb{1}_{]\!]\theta,\tau]\!]} = \sum_{i=1}^{n} \mathbb{1}_{[\theta\leq t_{i-1}]\setminus[\tau\leq t_i]}\ \mathbb{1}_{]t_{i-1},t_i]},$$

so that

$$\int \mathbb{1}_{]\!]\theta,\tau]\!]}(s)\mathrm{d}w(s) = \sum_{i=1}^{n} \mathbb{1}_{[\theta\le t_{i-1}]}\ \mathbb{1}_{[\tau\le t_i]}\ [w(t_i) - w(t_{i-1}] = \\ = \sum_{0\le i<j\le n} \mathbb{1}_{[\theta=t_i]}\ \mathbb{1}_{[\tau=t_j]}\ [w(\tau) - w(\theta)] = w(\tau) - w(\theta),$$

Even more general, we have the equality

$$\int_{(\theta,\tau]} c\, f(s)\mathrm{d}w(s) = c\int_{(\theta,\tau]} f(s)\mathrm{d}w(s), \tag{3.14}$$

for every bounded random variable c which is $\mathcal{F}_\theta$-measurable and any f in $\bar{\mathcal{E}}$. A way of proving (3.14) is to approximate the stopping times by finite-valued stopping times, which also show that in (3.9) we may replace the deterministic times t_i by stopping times τ_i, i.e.,

$$\int_{(0,t]} f(s)\mathrm{d}w(s) = \sum_{i=1}^{n} f_{i-1}[w(t\wedge\tau_i) - w(t\wedge\tau_{i-1})], \tag{3.15}$$

for every $t \ge 0$ and any processes of the form $f(t,\omega) = f_{i-1}(\omega)$ if $\tau_{i-1} < t \le \tau_i$ with some $i = 1,\dots,n$, where $0 = \tau_0 < \tau_1 < \cdots < \tau_n \le T$, with T a real number, and f_i are $\mathcal{F}(\tau_i)$ measurable bounded random variable for any i, and $f(t,\omega) = 0$ otherwise.

Now, we may extend this stochastic integral by localizing the integrand, i.e., denote by $\bar{\mathcal{E}}_{\mathrm{loc}}$ the space of all processes f for which there is a sequence $(\tau_1 \le \tau_2 \le \cdots)$ of stopping times such that $P(\tau_n < \infty)$ converges to zero and the processes $f_k(t,\omega) = f(t,\omega)$ for $t \le \tau_k$ (with $f_k(t,\omega) = 0$ otherwise) belong to $\bar{\mathcal{E}}$. Since, almost surely we have

$$\int_{(0,t]} f_k(s)\mathrm{d}w(s) = \int_{(0,t]} f_n(s)\, dw(s), \quad \forall t \le \tau_k,\ k \le n,$$

and both processes are continuous, we can define

$$\int_{(0,t]} f(s)\mathrm{d}w(s) = \lim_n \int_{(0,t]} f_n(s)\, dw(s), \quad \forall t \ge 0,$$

in a unique way and independent of the *localizing* sequence $(\tau_1 \le \tau_2 \le \cdots)$ used. For processes in $\bar{\mathcal{E}}_{\mathrm{loc}}$ the equalities (3.10) and (3.12) are no longer meaningful, but the processes (3.11) become continuous local-martingales. A very useful estimate, similar to the martingale inequality (3.13) but adapted to the local case is the following inequality

$$P\Big\{\sup_{0\le t\le T}\Big|\int_{(0,t]} f(s)\mathrm{d}w(s)\Big| \ge \varepsilon\Big\} \le \frac{\delta}{\varepsilon^2} + P\Big\{\int_0^T |f(s)|^2\mathrm{d}s \ge \delta\Big\}, \tag{3.16}$$

for any positive numbers T, ε and δ. Note that the martingale estimate (3.13) could be obtained either from Doob's maximal inequality

$$\mathbb{E}\big\{\sup_{0\le t\le T}|M(t)|^2\big\}\le 4\,\mathbb{E}\big\{|M(T)|^2\big\},$$

which is applied to the sub-martingale

$$t\mapsto M(t)=\Big|\int_{(0,t]}f(s)\mathrm{d}w(s)\Big|,$$

and the isometric equality (3.12), or from Davis-Burkhölder-Gundy inequality after identifying the optional quadratic variation by means of the second martingale assertion in (3.11). Similarly, instead of the martingale estimate (3.16), which is usually refer to as Lenglart's inequality, we could apply Doob's maximal inequality

$$\varepsilon P\big\{\sup_{0\le t\le T}|M(t)|\ge\varepsilon\big\}\le\mathbb{E}\big\{|M(T)|\mathbb{1}_{\{\sup_{0\le t\le T}|M(t)|\ge\varepsilon\}}\big\}\le\mathbb{E}\big\{|M(T)|\big\},$$

the stopping times

$$\tau_r=\inf\big\{t\in(0,T]:\int_0^t|f(s)|^2\mathrm{d}s\ge r/2\big\},\quad r>0,$$

which satisfy $P\{\tau_r<T\}\to 0$ and

$$\int_0^t|f_n(s)|^2\mathrm{d}s\big)\le r,\quad\forall t\le\tau_r,$$

for any n sufficiently large (after using the triangular inequality for the L^2-norm), to deduce

$$P\big\{\sup_{0\le t\le T}\Big|\int_{(0,t]}f_n(s)\mathrm{d}w(s)-\int_{(0,t]}f(s)\mathrm{d}w(s)\Big|\ge\varepsilon\big\}\le$$
$$\le P\{\tau_r<T\}+\frac{1}{\varepsilon^2}\mathbb{E}\big\{r\wedge\int_0^T|f_n(s)-f(s)|^2\mathrm{d}s\big\}.$$

Hence, by letting firstly $n\to\infty$ and secondly $r\to\infty$, the stochastic integral is defined for processes in $\bar{\mathcal{E}}_{\rm loc}$, such that the two processes in (3.11) are continuous local-martingales.

It is important to remark that the stochastic integral is initially defined in a L^2 space, where an element is an equivalence class relative to the product measure $P\times\mathrm{d}\ell$, with $\mathrm{d}\ell$ the Lebesgue measure on the semi-line $[0,\infty)$. For the sake of simplicity, we write $\Omega\times[0,\infty)$ or $[0,\infty)\times\Omega$ indistinctly as long as no confusion may arrive, i.e., processes are written $f(t,\omega)$ or $f(\omega,t)$. Next, by means of martingale properties we can select a *good* version to make the processes (3.11) continuous (local) martingales. By a simple argument of monotone classes, we

deduce that $\bar{\mathcal{E}}$ contains the Hilbert space $L^2(\Omega\times[0,\infty),\mathcal{P},P\times\mathrm{d}\ell)$. On the other hand, it is also clear that any stochastic process in $\bar{\mathcal{E}}_{\rm loc}$ is measurable relative to the σ-algebra $\bar{\mathcal{P}}$, generated by $\mathcal{P}$ and all $P\times\mathrm{d}\ell$-null subsets of $\Omega\times[0,\infty)$. As mentioned above, all concepts (in particular the stochastic integral) are up to or except to an evanescent set. However, the stochastic integral is defined up to a $P\times\mathrm{d}\ell$-null subset of $\Omega\times[0,\infty)$, and then a good version is chosen. Thus, the next question is what processes are in $\bar{\mathcal{E}}$ or $\bar{\mathcal{E}}_{\rm loc}$ besides those that are predictable, i.e., what can be said about completion σ-algebra $\bar{\mathcal{P}}$ of the predictable σ-algebra $\mathcal{P}$.

Adapted, Predictable and Other Properties

Recall that in a filtered probability space $(\Omega,\mathbb{F},P)$, elementary predictable processes are denoted by $\mathcal{E}$, i.e., $\mathcal{E}$ is the vector space of all processes of the form $f(t,\omega)=f_{i-1}(\omega)$ if $t_{i-1}<t\le t_i$ with some $i=1,\dots,n$, where $0=t_0<t_1<\cdots<t_n$ are real numbers and f_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $f(t,\omega)=0$ otherwise.

Also $\bar{\mathcal{E}}$ denotes the L^2-closure of $\mathcal{E}$, i.e., the Hilbert space of all processes f for which there exists a sequence $(f_1,f_2,\dots)$ of processes in $\mathcal{E}$ such that

$$\lim_{n\to\infty}\mathbb{E}\Big\{\int|f_n(s)-f(s)|^2\mathrm{d}s\Big\}=0,$$

while $\bar{\mathcal{E}}_{\rm loc}$ denotes its localization, i.e., the space of all processes f for which there is a sequence $(\tau_1\le\tau_2\le\cdots)$ of stopping times such that $P(\tau_n<\infty)$ converges to zero and the processes $f_k(t,\omega)=f(t,\omega)$ for $t\le\tau_k$ (with $f_k(t,\omega)=0$ otherwise) belong to $\bar{\mathcal{E}}$.

Proposition 3.2. *Any adapted square integrable process $f(t,\omega)$ is in $\bar{\mathcal{E}}$.*

Proof. This follows Doob [33] arguments.

Step 1 First assume f is bounded and vanishes for t outside of a bounded interval. Then partition the real line $\mathbb{R}$ into intervals $(k\varepsilon,(k+1)\varepsilon]$ with $k=0,\pm1,\pm2,\dots,\varepsilon>0$, and define $f_{\varepsilon,s}(t,\omega)=f(\alpha_\varepsilon(t-s)+s,\omega)$, where $\alpha_\varepsilon(r)=k\varepsilon$ for any r in the subinterval $(k\varepsilon,(k+1)\varepsilon]$, where f has been extended for $t\le0$. The restriction to $[0,\infty)$ of the process $f_{\varepsilon,s}$ belongs to $\mathcal{E}$, for any $\varepsilon>0$ and s in $\mathbb{R}$. The claim is that there exist a sequence $(\varepsilon_1>\varepsilon_2>\cdots)$ and some s such that

$$\lim_{n\to\infty}\mathbb{E}\Big\{\int|f_{\varepsilon_n,s}(t,\omega)-f(t,\omega)|^2\mathrm{d}t\Big\}=0.$$

Indeed, the continuity of the translation in $\mathbb{R}$ with respect to the Lebesgue measure and the fact that $\alpha_\varepsilon(r)-r\to0$ as $\varepsilon\to0$ show that

$$\lim_{\varepsilon\to0}\int|f(\alpha_\varepsilon(t)+s,\omega)-f(t+s,\omega)|^2\mathrm{d}s=0,\quad\forall t,\omega.$$

Since all processes considered are bounded and vanish outside of a fixed finite interval, we have

$$\lim_{\varepsilon\to 0}\int \mathbb{E}\{\int \big|f(\alpha_\varepsilon(t)+s,\omega)-f(t+s,\omega)\big|^2\mathrm{d}s\}\mathrm{d}t=0.$$

Fubini's Theorem allows us to exchange the integration order of the variables s and t, proving the claim.

Step 2 For the general case, define $f_n(t,\omega)=f(t,\omega)$ if $0\le t\le n$ and $|f(t,\omega)|\le n$, and $f_n(t,\omega)=0$ otherwise. Applying the previous approximation to f_n the proof is completed. Sometimes, it may be convenient to redo this argument on the compact time-interval $[0,T]$ instead of the semi-line $[0,\infty)$. □

Proposition 3.3. *Any measurable adapted process $f(t,\omega)$ is measurable with respect to the $(P\times \mathrm{d}t)$-completion $\bar{\mathcal{P}}$ of the predictable σ-algebra $\mathcal{P}$. Moreover, if $f(t,\omega)$ is locally square integrable, i.e.,*

$$P\{\int_0^t |f(s,\omega)|^2\mathrm{d}s<\infty\}=1,\quad \forall t\ge 0 \tag{3.17}$$

then $f(t,\omega)$ belongs to $\bar{\mathcal{E}}_{loc}$.

Proof. It is clear that the first assertion follows from the previous Proposition 3.2. To establish the second assertion, first assume that $f(t,\omega)$ is also progressively measurable i.e., the restriction of $(t,\omega)\mapsto f(t,\omega)$ to $[0,T]\times\Omega$ is measurable with respect to $\mathcal{B}([0,T])\times\mathcal{F}$, for every $T>0$. In this case, the expression

$$\tau_n=\inf\{t\ge 0:\int_0^t |f(t,\omega)|^2\mathrm{d}s\ge n\}$$

define a localizing sequence of stopping times for the process f, which proves the claim.

However, when f is only a measurable adapted process, τ_n may not be a stopping time. In this case, we can always approximate f by truncation, i.e, $f_n(t,\omega)=f(t,\omega)$ if $|f(t,\omega)|\le n$ and $f_n(t,\omega)=0$ otherwise, so that

$$\lim_n P\{\int_0^T |f_n(t,\omega)-f(t,\omega)|^2\mathrm{d}s\ge\delta\}=0,\quad \forall T,\delta\ge 0.$$

Since f_n belongs to $\bar{\mathcal{E}}$, for every $n\ge 1$, the estimate (3.16) proves also that $\bar{\mathcal{E}}_{\mathrm{loc}}$ contains all measurable adapted processes satisfying (3.17). □

Now, going back to the stochastic integral relative the a Wiener process, if f is a cad-lag adapted process then $t\mapsto f(t-)$ and f are progressively measurable then

$$\int_{(0,t]} f(s)\mathrm{d}w(s)=\int_{(0,t]} f(s-)\mathrm{d}w(s),\quad \forall t>0.$$

and condition (3.17) is satisfied. Moreover, regarding a process regarded as an equivalent class with respect to the product measure $\mathrm{d}t\times P$ on $[0,\infty)\times\Omega$,

Proposition 3.4. *If the equivalence class of a process f contains an element in $\bar{\mathcal{E}}_{loc}$ then the stochastic integral is defined. Moreover, the stochastic integral can be written with respect to any predictable representative of its equivalence class, e.g., adapted stochastically left continuous process is suitable predictable representative of its equivalent class.*

Proof. It is clear that in the case of a cad-lag adapted process f, $t \mapsto f(t-)$, for any $t > 0$, is a suitable predictable representative of its equivalent class.

Since the stochastic integral is defined as an equivalent class of processes, only the situation when f is an adapted stochastically left continuous process needs some details. Indeed, if $0 = \tau_0^n \leq \tau_1^n < \cdots < \tau_k^n < \cdots$ is a sequence of stopping times such that $P\{\sup_k \tau_k^n < \infty\} \to 0$ and $P\{\sup_k(\tau_k^n - \tau_{k-1}^n > \delta\} \to 0$, for any $\delta > 0$, as $n \to \infty$, then define the sequence of simple predictable processes $f_{n,m}(t,\omega) = f(\tau_k^n, \omega)$ when $|f(\tau_k^n, \omega)| \leq m$ and $\tau_k^n < t \leq \tau_{k+1}^n$, $k = 0, 1, 2, \ldots$, and $f_{n,m}(t,\omega) = 0$ otherwise. A typical case is when $\tau_k^n = k2^{-n}$.

Thus, it is clear that

$$\lim_n P\{|f_{n,m}(t,\omega) - f_m(t,\omega)| \geq \delta\} = 0, \quad \forall t, \delta, m > 0,$$

where $f_m(t,\omega) = f(t,\omega)$ if $|f(t,\omega)| \leq m$ and $f_m(t,\omega) = 0$ otherwise. Since $|f_{m,n}|$ is bounded by m, the limit

$$\lim_n P\Big\{ \int_0^T |f_{n,m}(t,\omega) - f_m(t,\omega)|^2 \mathrm{d}t \geq \delta \Big\} = 0, \quad \forall T, \delta, m > 0,$$

follows. Hence, by means of (3.16)

$$\lim_n P\Big\{ \sup_{0 \leq t \leq T} \Big| \int_{]0,t]} [f_{n,m}(t) - f_m(t)] \mathrm{d}w(t) \Big| \geq \varepsilon \Big\} = 0,$$

for every $T, \varepsilon, m > 0$. Thus, for each $t, m > 0$, the expression

$$\int_{]0,t]} f_{n,m}(s) \mathrm{d}w(s) = \sum_{k=0}^{\infty} f_m(\tau_k^n, \omega) \, [w(t \wedge \tau_{k+1}^n, \omega) - w(t \wedge \tau_k^n, \omega)],$$

for every $t > 0$, is an approximation of the stochastic integral provided f satisfies (3.17). Recall that $f_m(t,\omega) = f(t,\omega)$ if $|f(t,\omega)| \leq m$, so that f_m converges to f almost surely in L^2. □

It can be proved, see Dellacherie and Meyer [32, Theorem VIII.1.23, pp. 346-346] that for any f in $\bar{\mathcal{E}}_{\mathrm{loc}}$ we have

$$\text{if } \; f(s,\omega) = 0, \; \forall (s,\omega) \in]a,b] \times F, \; F \in \mathcal{F} \qquad \text{then } \int_{(a,b]} f(s) \mathrm{d}w(s) = 0 \;\; \text{a.s. on } \; F. \tag{3.18}$$

This expresses the fact that even if the construction of the stochastic integral is not *pathwise*, it retains some local character in Ω.

From the definition it follows that if f is a cad-lag adapted process with locally bounded variation then

$$\int_{(0,t]} f(s)\mathrm{d}w(s) = \int_{(0,t]} f(s-)\mathrm{d}w(s) = f(t)w(t) - \int_{(0,t]} w(s)\mathrm{d}f(s),$$

where the last integral is in the Riemann-Stieltjes or Lebesgue-Stieltjes sense. However, the Wiener process w has unbounded local variation. Let $\varpi = (0 = t_0 < t_1 < \cdots < t_n = t)$ be a partition of $[0,t]$, with mesh $|\varpi| = \max_i(t_i - t_{i-1})$ and consider the Riemann sums

$$S_\varpi = \sum_{i=1}^n w(t_i^*)[w(t_i) - w(t_{i-1})], \quad \text{with} \quad t_{i-1} \le t_i^* \le t_i,$$

which can be rewritten as

$$S_\varpi = \frac{w^2(t)}{2} - \frac{1}{2}\sum_{i=1}^n [w(t_i) - w(t_{i-1})]^2 + \sum_{i=1}^n [w(t_i^*) - w(t_{i-1})]^2 + \\ + \sum_{i=1}^n [w(t_i) - w(t_i^*)][w(t_i^*) - w(t_{i-1})].$$

Since

$$\mathbb{E}\Big\{ \sum_{i=1}^n [w(t_i) - w(t_{i-1})]^2 \Big\} = \sum_{i=1}^n [t_i - t_{i-1}] = t,$$

$$\mathbb{E}\Big\{ \Big[\sum_{i=1}^n [w(t_i^*) - w(t_{i-1})]^2 - \sum_{i=1}^n (t_i^* - t_{i-1}) \Big]^2 \Big\} = \\ = \mathbb{E}\Big\{ \sum_{i=1}^n [w(t_i^*) - w(t_{i-1})]^4 \Big\} - \sum_{i=1}^n (t_i^* - t_{i-1})^2 \le 2t|\varpi|,$$

and

$$\mathbb{E}\Big\{ \Big[\sum_{i=1}^n [w(t_i) - w(t_i^*)][w(t_i^*) - w(t_{i-1})] \Big]^2 \Big\} = \\ = \sum_{i=1}^n (t_i - t_i^*)(t_i^* - t_{i-1}) \le t|\varpi|,$$

we deduce that

$$\lim_{\|\varpi\| \to 0} \Big[S_\varpi - \sum_{i=1}^n (t_i^* - t_{i-1}) \Big] = \frac{w^2(t)}{2} - \frac{t}{2},$$

in the L^2-sense. In the Itô integral, $t_i^* = t_{i-1}$ so that

$$\int_{(0,t]} w(s)\mathrm{d}w(s) = \frac{w^2(t)}{2} - \frac{t}{2}, \quad \forall t \ge 0.$$

However, any choice $t_i^* = (1-r)t_{i-1} + rt_i$, with $0 \le r \le 1$, could be possible. In particular Fisk-Stratonovich integral, where $r = 1/2$, $t_i^* = (t_{i-1} + t_i)/2$, yields a symmetric calculus, very useful in some physical and mechanical models. However, Itô integral, i.e., the choice $r = 1$, $t_i^* = t_{i-1}$, is more oriented to control models, where the adapted (or predictable) character (i.e., non-interaction with the future) is an essential property. Moreover, the martingale property is preserved.

Working by coordinates, this stochastic integral can be extended to a $\mathbb{R}^d$-valued Wiener process and $n \times d$ matrix-valued predictable processes.

3.2.2 Relative to Poisson Measures

Let $\{p(t) : t \ge 0\}$ be a real-valued compound Poisson process with parameters (c, ν), where $c > 0$ and γ is a distribution in $\mathbb{R}_*^d = \mathbb{R}^d \setminus \{0\}$, in a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \ge 0)$. This means that

$$p(t,\omega) = \begin{cases} 0 & \text{if } \ t < \theta_1(\omega), \\ Z_n(\omega) & \text{if } \ \theta_n(\omega) \le t < \theta_{n+1}(\omega), \end{cases}$$

where $\theta_n = \tau_1 + \tau_2 + \cdots + \tau_n$, $\{\tau_n : n = 1, 2, \ldots\}$ is a sequence of independent exponentially distributed (with parameter c) random variables, $Z_n = \zeta_1 + \zeta_2 + \cdots + \zeta_n$, $\{\zeta_n : n = 1, 2, \ldots\}$ is a sequence of independent identically distributed (with distribution law γ) random variables, independent of the sequence $\tau_1, \tau_2 \ldots$. In particular, if γ is δ_1, the Dirac measure at $z = 1$ then $Z_n = n$, the case of a standard Poisson process. Note that $p(t) - ct$ and $p^2(t) - ct$ are martingales relative to the filtration $(\mathcal{F}_t : t \ge 0)$, with $p(0) = 0$. Since the function $t \mapsto p(t, \omega)$ is cad-lag, piecewise constant and with bounded variation for any ω, the integral with respect to $p(t)$ is covered by the measure theory, i.e., a pathwise integration. For a bounded left-continuous process $f(t, \omega)$ we can define

$$\int_{(0,t]} f(s,\omega)\mathrm{d}p(s,\omega) = \sum_{n=1}^{\infty} f(\theta_n(\omega),\omega)\, \mathbb{1}_{\theta_n(\omega)\le t} = \\ = \sum_{n=1}^{N(t,\omega)} f(\theta_n(\omega),\omega), \quad (3.19)$$

for each ω, where $N(t, \omega) = n$ if $\theta_n(\omega) \le t < \theta_{n+1}(\omega)$, i.e., a standard Poisson process. Because $t \mapsto \mathbb{E}\{p(t)\}$ is continuous, we have

$$\int_{(0,t]} f(s,\omega)\mathrm{d}\mathbb{E}\{p(s,\cdot)\} = \int_{(0,t]} f(s+,\omega)\mathrm{d}\mathbb{E}\{p(s,\cdot)\},$$

but

$$\int_{(0,t]} p(s-,\omega)\mathrm{d}p(s,\omega) = \sum_{n=1}^{\infty} p(\theta_n(\omega)-,\omega)\,\mathbb{1}_{\theta_n(\omega)\le t} = \\ = \sum_{k=1}^{N(t,\omega)} Z_{k-1}(\omega)\,\zeta_k(\omega),$$

and

$$\int_{(0,t]} p(s,\omega)\mathrm{d}p(s,\omega) = \sum_{n=1}^{\infty} p(\theta_n(\omega),\omega)\,\mathbb{1}_{\theta_n(\omega)\le t} = \sum_{k=1}^{N(t,\omega)} Z_k(\omega)\,\zeta_k(\omega).$$

Thus, for a given compound Poisson process $p(t)$ as above and a left-continuous (or only predictable) process $f(t)$ (without begin locally integrable), we can use (3.19) to define the *stochastic integral*, which is just a *pathwise* sum (integral) in this case, with is a jump process similar to the compound Poisson process. Similar arguments apply to the centered compound Poisson process $t \mapsto \big(p(t) - \mathbb{E}\{p(t)\}\big)$, and the integral is the difference of random pathwise integral and a deterministic integral.

Next step is to consider a standard Poisson measure $\{p(\cdot,t) : t \ge 0\}$ with Lévy (intensity) measure $\pi(\cdot)$ in a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \ge 0)$, i.e., (a) $\pi(\cdot)$ is a Radon measure on $\mathbb{R}^m_* = \mathbb{R}^m \setminus \{0\}$, i.e., $\pi(K) < \infty$ for any compact subset K of $\mathbb{R}^m_*$; (b) $\{p(B,t) : t \ge 0\}$ is a Poisson process with parameter $\pi(B)$, for any Borel subset B in $\mathbb{R}^d_*$ with $\pi(B) < \infty$ (here $p(B,t) = 0$ if $\pi(B) = 0$); (c) the Poisson processes $p(\cdot,B_1), p(\cdot,B_2), \dots, p(\cdot,B_n)$ are independent if $B_1, B_2, \dots, B_n$ are disjoint Borel set in $\mathbb{R}^m_*$ with $\pi(B_i) < \infty$, $i = 1, \dots, n$.

Given a Radon measure π in $\mathbb{R}^m_*$ (which integrates the function $|z|^2 \wedge 1$, so that it can be called a Lévy measure), we write $\pi = \sum_k \pi_k$, where $\pi_k(B) = \pi(B \cap R_k)$, $\mathbb{R}^m_* = \bigcup_k R_k$, $\pi(R_k) < \infty$ and $R_k \cap R_\ell = \emptyset$ if $k \ne \ell$. To each π_k we may associate a compound Poisson process and a point process by the expressions

$$Y_k(t) = \sum_{n=1}^{\infty} \zeta_{n,k}\,\mathbb{1}_{t\ge\theta_{n,k}} \quad \text{and} \quad \delta Y_k(t) = Y_k(t) - Y_k(t-) = \zeta_{n,k}\,\mathbb{1}_{t=\theta_{n,k}},$$

for $t > 0$, where $\theta_{n,k} = \tau_{1,k} + \tau_{2,k} + \cdots + \tau_{n,k}$, $\{\tau_{n,k} : n = 1, 2, \dots\}$ is a sequence of independent exponentially distributed (with parameter $\pi(R_k) = c_k$) random variables, and $\{\zeta_{n,k} : n = 1, 2, \dots\}$ is another sequence of independent identically distributed (with distribution law π_k/c_k) random variables, and the two sequences $\{\tau_{n,h} : n, k \ge 1\}$, $\{\zeta_{n,k} : n, k \ge 1\}$ are independent. The jump process $\delta Y = \sum_k \delta Y_k$ is indeed a Poisson point process with characteristic measure π, i.e., with $Z_{n,k} = \zeta_{1,k} + \zeta_{2,k} + \cdots + \zeta_{n,k}$,

$$p(B\times]s,t]) = \sum_{n,k=1}^{\infty} \mathbb{1}_{s<\theta_{n,k}\le t}\,\mathbb{1}_{Z_{n,k}\in B}, \quad \forall t > s \ge 0,\ B \in \mathcal{B}(\mathbb{R}^m_*),$$

is a standard Poisson random measure with intensity measure

$$\mathbb{E}\{p(B\times]s,t])\} = (t-s)\pi(B).$$

In general, we cannot arrange the jumps in the increasing order like the case of a compound Poisson process, because there may occur accumulation of small jumps. With any of the notation $p(B,t)$ or $p(B\times]0,t])$ or $p(B,]0,t])$ the integer-valued random measure p (see Section 2.7) is also called a standard Poisson random measure. From the process viewpoint, $p(B,]s,t])$ is defined as the (finite) number of jumps (of a cad-lag process Y) belonging to B within the interval $]s,t]$. Note that the predictable compensator of the optional random measure $p(\cdot,t)$ is the deterministic process πt. Thus, for a predictable process of the form $F(z,t,\omega) = f(t,\omega)\,\mathbb{1}_{z\in B}$ the expression

$$\int_{\mathbb{R}_k\times]0,t]} F(z,s,\omega)\,p(\mathrm{d}z,\mathrm{d}s) = \sum_{n=1}^{\infty} f(\theta_{n,k}(\omega),\omega)\,\mathbb{1}_{0<\theta_{n,k}(\omega)\le t}\,\mathbb{1}_{Z_{n,k}(\omega)\in B}$$

is indeed a finite stochastic pathwise sum (as previously). However, the passage to the limit in k is far more delicate and requires more details.

With the above introduction, let ν be an integer-valued random measure, which is a Poisson measure as in Definition 2.28, with Lévy measure

$$\Pi(B\times]s,t]) = \mathbb{E}\{\nu(B\times]s,t])\},\quad \Pi(\mathbb{R}^m_*\times\{t\}) = 0,\quad \text{for every}\ \ t\ge 0,$$

and local-martingale measure $\tilde\nu = \nu - \Pi$, in a given filtered space $(\Omega,\mathcal{F},P,\mathcal{F}_t : t\ge 0)$. In particular, a standard Poisson measure $\{p(\cdot,t) : t\ge 0\}$ with Lévy (characteristic or intensity) measure $\pi(\cdot)$, and $\Pi(\mathrm{d}z,\mathrm{d}t) = \pi(\mathrm{d}z)\times\mathrm{d}t$. Note that we reserve the notation p for a standard Poisson measure. Denote by $\mathcal{E}$ the vector space of all processes of the form $f(z,t,\omega) = f_{i-1,j}(\omega)$ if $t_{i-1} < t \le t_i$ and z belongs to K_j with some $i = 1,\dots,n$, and $j = 1,\dots,m$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers, K_j are disjoint sets with compact closure in $\mathbb{R}^m_*$ and $f_{i-1,j}$ is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $f(t,\omega) = 0$ otherwise. Elements in $\mathcal{E}$ are called elementary predictable processes. It is clear what the integral should be for any integrand in $\mathcal{E}$, namely

$$\begin{aligned}
\int_{\mathbb{R}^m_*\times(0,\infty)} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) &= \sum_{i=1}^{n}\sum_{j=1}^{m} f_{i-1,j}\,\tilde\nu(K_j\times]t_{i-1},t_i]),\\
\int_{\mathbb{R}^m_*\times(a,b]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) &= \int f(z,s)\,\mathbb{1}_{(a,b]}(s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s),
\end{aligned}\tag{3.20}$$

for every $b > a \ge 0$. Note that

$$\begin{aligned}
\int_{\mathbb{R}^m_*\times(0,\infty)} f(z,s)\,\mathbb{1}_{(0,t]}(s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) &= \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} f_{i-1,j}\,\tilde\nu(K_j\times]t\wedge t_{i-1}, t\wedge t_i])
\end{aligned}$$

and

$$\int f(z,s)\,\mathbb{1}_{(a,b]}(s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) = \\ = \int_{\mathbb{R}^m_*\times(0,b]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) - \int_{\mathbb{R}^m_*\times(0,a]} f(z,s)\,\tilde\nu(\mathrm{d}s,\mathrm{d}z),$$

for every $t>0$.

If ν is a standard (or homogeneous) Poisson measure, i.e., $\mathbb{E}\{\nu(B\times(]s,t])\} = (t-s)\pi(B)$, then $p(K,t)=\nu(K\times]0,t])$ is a Poisson process with parameter $\pi(K)$, then for any left-continuous adapted process of the form $f(z,t,\omega) = f_j(t,\omega)$ when z belongs to K_j, we can calculate the stochastic integral, namely,

$$\int_{\mathbb{R}^m_*\times(0,t]} \sum_{j=1}^m f_i(s)\,\mathbb{1}_{K_j}(z)\,\nu(\mathrm{d}z,\mathrm{d}s) = \sum_{j=1}^m \sum_{k=1}^{p(t,K_j,\omega)} f_j(\theta_k(\omega,K_j),\omega),$$

for every $t\ge 0$, where $\theta_k(\omega,K_j)$ is the time of the k jumps of the Poisson process $t\mapsto p(K_j,t)$. In the case of a compound-Poisson process as above, we may forget about the K dependency, and make the previous pathwise definition, both concepts agree. In general, from $\nu=\tilde\nu+\Pi$, with $\Pi=\pi\times\mathrm{d}t$, we can define the stochastic integral relative to an integer-valued random measure ν.

This definition is independent of the particular representation used. Since for any K_1 disjoint of K_2 and any $t\ge 0$ the random variables $p(K_1,t)$ and $p(K_2,t)$ are orthogonal, and because f_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable we obtain

$$\mathbb{E}\big\{\big|\int_{\mathbb{R}^m_*\times(0,t]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s)\big|^2\big\} = \mathbb{E}\{\int_{\mathbb{R}^m_*\times(0,t]} |f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s)\}, \quad (3.21)$$

for every f in $\mathcal{E}$. Moreover the processes

$$\int_{\mathbb{R}^m_*\times(0,t]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) \quad \text{and} \\ \Big|\int_{\mathbb{R}^d_*\times(0,t]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s)\Big|^2 - \int_{\mathbb{R}^m_*\times(0,t]} |f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s), \quad (3.22)$$

with $t\ge 0$ are cad-lag (quasi-left continuous) martingales, and

$$\mathbb{E}\big\{\big[\int_{\mathbb{R}^m_*\times(0,\infty)} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s)\big]\big[\int_{\mathbb{R}^m_*\times(0,\infty)} g(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s)\big]\big\} = \\ = \mathbb{E}\big\{\int_{\mathbb{R}^m_*\times(0,\infty)} f(z,s)\,g(z,s)\,\Pi(\mathrm{d}z,\mathrm{d}s)\big\}, \quad (3.23)$$

for any two stochastic processes f and g in $\mathcal{E}$. Denote by $\bar{\mathcal{E}}_\Pi$ the L^2-closure of $\mathcal{E}$, i.e., the Hilbert space of all processes f for which there exists a sequence $(f_1,f_2,\dots)$ of processes in $\mathcal{E}$ such that

$$\lim_{n\to\infty} \mathbb{E}\big\{\int_{(0,\infty)\times\mathbb{R}^m_*} |f_n(z,s)-f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s)\big\} = 0.$$

As in the previous section, the martingale inequality

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} \Big| \int_{\mathbb{R}^d_*\times(0,t]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s)\Big|^2\Big\} \le \\ \le 4\,\mathbb{E}\Big\{ \int_{\mathbb{R}^d_*\times(0,T]} |f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s)\Big\}, \quad (3.24)$$

holds for every $T \ge 0$, and also the isometric identity (3.21). Hence, this linear operation can be extended to the closure $\bar{\mathcal{E}}_\Pi$, preserving linearity and the properties (3.21), (3.22), (3.23). This is called Itô integral or generally *stochastic integral*, with respect to a Poisson measure. Next, by localizing the integrand, this definition is extended to $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$, the space of all processes f for which there is a sequence $(\tau_1 \le \tau_2 \le \cdots)$ of stopping times such that $P(\tau_n < \infty)$ converges to zero and the processes $f_k(t,\omega) = f(t,\omega)$ for $t \le \tau_k$ (with $f_k(t,\omega) = 0$ otherwise) belong to $\bar{\mathcal{E}}_\Pi$. As in the case of the Wiener process, a key role is played by the following inequality

$$P\Big\{ \sup_{0\le t\le T} \Big| \int_{\mathbb{R}^m_*\times(0,t]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s)\Big| \ge \varepsilon\Big\} \le \frac{\delta}{\varepsilon^2} + \\ + P\Big\{ \int_{\mathbb{R}^m_*\times(0,T]} |f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s) \ge \delta\Big\}, \quad (3.25)$$

for any positive numbers T, ε and δ.

The class of processes that we can integrate are those in $\bar{\mathcal{E}}_\Pi$ or more general in $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$, but the stochastic integral is initially defined in a L^2 space, where an element is an equivalence class relative to the product measure $P\times\Pi$, with $\Pi = \Pi(\mathrm{d}z,\mathrm{d}s)$ the Lévy measure on $\mathbb{R}^m_*\times[0,\infty)$. Again, for the sake of simplicity, we write $\Omega\times\mathbb{R}^m_*\times[0,\infty)$ or $\mathbb{R}^m_*\times]0,\infty)\times\Omega$ or $]0,\infty)\times\mathbb{R}^m_*\times\Omega$ indistinctly as long as no confusion may arrive, i.e., processes are written $f(\omega,t,z)$ or $f(z,\omega,t)$ or $f(t,z,\omega)$. Next, by means of martingale properties we can select a *good* version to make the process (3.22) a cad-lag (local) martingale. By a simple argument of monotone classes, we deduce that (as in the case of the Wiener process) the closure $\bar{\mathcal{E}}_\Pi$ (of all elementary processes in $\mathbb{R}^d_*\times[0,\infty)$) contains the Hilbert space $L^2(\mathbb{R}^d_*\times[0,\infty)\times\Omega, \mathcal{B}\times\mathcal{P}, \Pi\times P)$, $\Pi = \Pi(\mathrm{d}z,\mathrm{d}s)$.

On the other hand, it is also clear that any stochastic process in $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$ is measurable relative to the σ-algebra $\overline{\mathcal{B}\times\mathcal{P}}$, generated by $\mathcal{B}\times\mathcal{P}$ and all $\Pi\times P$-null subsets of $\mathbb{R}^d_*\times[0,\infty)\times\Omega$. Again, we note that the value at time 0 is irrelevant. It is also clear that the arguments in Subsection 3.2.1 relative to *Adapted, Predictable and Other Properties* can be repeated any stochastic integral, not necessarily relative to a Wiener process.

Comments on Lévy and Point Processes

If the Lévy measure is absolutely continuous with respect to the Lebesgue measure $\mathrm{d}\ell$ on $[0,\infty)$, i.e., $\Pi(\mathrm{d}z,\mathrm{d}s) = \pi(\mathrm{d}z)\times\mathrm{d}\ell$, then (as in the case of the

Wiener process) any measurable adapted process $f(z,s,\omega)$ is equivalent to a $\mathcal{B}\times\mathcal{P}$-measurable process, so it belongs to $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$ whenever

$$P\Big\{\int_{\mathbb{R}^d_*\times(0,T]}|f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s)<\infty\Big\}=1,\quad\forall\,T>0 \tag{3.26}$$

is satisfied. This holds for standard Poisson measures.

Because the Lévy measure does not charge on $\mathbb{R}^d_*\times\{t\}$, for every $t\geq 0$, see Theorem 2.30, the stochastic integral is a cad-lag quasi-left continuous and the argument developed for Wiener processes applies proving that any progressively measurable process satisfying (3.26) belongs to $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$.

The above stochastic integral can be constructed also for an *extended Poisson measure* (see Jacod and Shirayaev [84, Definition 1.20, Chapter 2, p. 70]), where $\Pi(\mathbb{R}^d_*\times\{t\})$ may not vanish for some $t>0$. Actually, the stochastic integral can be constructed for any *orthogonal* measures, see Definition 3.1 in Chapter 3.

On the other hand, a (homogeneous) Poisson measure $p(\mathrm{d}z,\mathrm{d}s)$ with Lévy measure π always satisfies $p(\mathbb{R}^m_*,\{0\})=0$ and can be approximated by another Poisson measure $p_\varepsilon(\mathrm{d}z,\mathrm{d}s)$ with Lévy measure $\pi_\varepsilon=\mathbb{1}_{K_\varepsilon}\pi$, where the support $K_\varepsilon=\{0<\varepsilon\leq|z|\leq 1/\varepsilon\}$ of π_ε is a compact on $\mathbb{R}^m_*$, i.e., all jumps smaller than ε or larger than $1/\varepsilon$ have been eliminated. The integer measure p_ε is associated with a compound Poisson process and has a finite (random) number of jumps, i.e., for any $t>0$ there is an integer $N=N(t,\omega)$, points $z_i=z_i(t,\omega)$ in K_ε for $i=1,\dots,N$ and positive reals $\theta_i=\theta_i(t,\omega)$, $i=1,\dots,N$ such that $p(B,]a,b],\omega)=\sum_{n=1}^N\mathbb{1}_{z_i\in B}\mathbb{1}_{a<\theta_i\leq b}$, for every $B\in\mathcal{B}(\mathbb{R}^m_*)$, $0\leq a<b\leq t$. In this case, the forward stochastic integral can be written as

$$\int_{\mathbb{R}^m_*\times(0,t]}f(z,s)\,\tilde{p}(\mathrm{d}z,\mathrm{d}s)=\sum_{i=1}^N f(z_i,\theta_i-)-\int_0^t\mathrm{d}s\int_K f(z,s)\pi(\mathrm{d}z), \tag{3.27}$$

for any adapted cad-lag process $f(z,s)$, continuous in z.

Alternatively, we may regard the integer measure ν as a point process, i.e.,

$$\nu(B,]a,b])=\sum_{i=1}^\infty\mathbb{1}_{\{p_i\in B\}}\mathbb{1}_{\{a<\tau_i\leq b\}}$$

to consider the pathwise integrals

$$\int_{\mathbb{R}^m\times]0,t]}f(z,s)\,\nu(\mathrm{d}z,\mathrm{d}s)=\sum_{i=1}^\infty f(p_i,\tau_i)\mathbb{1}_{0<\tau_i\leq t}$$

and $\displaystyle\int_{\mathbb{R}^m\times]0,t]}f(z,s)\,\Pi(\mathrm{d}z,\mathrm{d}s)$,

defined for integrable (with respect to ν and Π) processes f. Later, a martingale argument allows the extension to square-integrable with respect Π, e.g., see Ikeda and Watanabe [75, Chapter 2]. Both approaches are equivalent and the

expression (3.27) remains valid for f integrable with respect to ν and square-integrable with respect to Π.

It should be clear that the starting point is an integer-valued random measure ν (see Definition 2.25) which yields a compensated local-martingale measure $\tilde{\nu} = \nu - \nu^p$, where ν^p is the (unique dual) predictable projection of ν (see Theorem 2.26 and Definition 3.2.24 of the previous chapter). Recall that a local-martingale M is called *purely discontinuous* if $M(0) = 0$ and the product $M\,N$ is a local martingale for any continuous local-martingale N. Stochastic integrals with respect to a compensated local-martingale measure $\tilde{\nu}$ are purely discontinuous local-martingales. Also, given an optional locally integrable process X with $X(0) = 0$ there exists a unique predictable projection ${}^pX$, i.e. a predictable locally integrable process such that $\mathbb{E}\{{}^pX\,\mathbb{1}_{\tau<\infty}\} = \mathbb{E}\{X\,\mathbb{1}_{\tau<\infty}\}$ for any predictable stopping time τ, such that $t \mapsto \mathbb{1}_{\tau\le t}$ is a predictable process. In particular (e.g., Jacod and Shirayaev [84, Theorem 2.28, Corallary 2.31, Chapter 1, p. 23–24]) for a local-martingale M we have ${}^pM(t) = M(t-)$ and $\delta M(t) = 0$ for every $t > 0$.

• *Remark* 3.5. Let $p(\mathrm{d}z, \mathrm{d}s)$ be a Poisson measure with Lévy measure given by $\Pi(\mathrm{d}z, \mathrm{d}s) = \pi(\mathrm{d}z, s)\mathrm{d}s$ in $\mathbb{R}^m_* \times [0,\infty)$ with $\Pi(\mathbb{R}^m_*, \{0\}) = 0$ and let γ be a Borel function from $\mathbb{R}^m_* \times [0,\infty)$ into $\mathbb{R}^d$ square-integrable with respect to Π on any set of the form $\mathbb{R}^m_* \times (0,T]$, for any constant $T > 0$, and cad-lag in $[0,\infty)$. The Poisson measure p can be viewed as a Poisson point process in $\mathbb{R}^m_*$, i.e.,

$$p(B,]a,b]) = \sum_{i=1}^{\infty} \mathbb{1}_{\{p_i\in B\}}\,\mathbb{1}_{\{a<\tau_i\le b\}},$$

where the masses $\{p_i\}$ are in $\mathbb{R}^m_*$ and $\{\tau_i\}$ are stopping times (non necessary non-decreasing in i). Then we may define the stochastic integral

$$I(t,\tilde{p}) = \int_{\mathbb{R}^m_*\times(0,t]} \gamma(z,s)\,\tilde{p}(\mathrm{d}z,\mathrm{d}s),$$

which has a jump only at $t = \tau_i$ if $\gamma(p_i, \tau_i-) \neq 0$ for some i. If $z \mapsto \gamma(z,\cdot)$ is integrable with respect to p and Π (e.g., bounded, continuous in z and vanishing near $z = 0$) then

$$I(t,\gamma,\tilde{p}) = \sum_{i=1}^{\infty} \gamma(p_i,\tau_i-)\,\mathbb{1}_{\{0<\tau_i\le t\}} - \int_0^t \mathrm{d}s \int_{\mathbb{R}^m_*} \gamma(z,s)\,\pi(\mathrm{d}z,s)\mathrm{d}s,$$

which is a pathwise integral. The integer measure p_γ associate with the martingale $t \mapsto I(t,\gamma,\tilde{p})$ satisfies

$$p_\gamma(B,]a,b]) = \sum_{i=1}^{\infty} \mathbb{1}_{\{\gamma(p_i,\tau_i-)\in B\}}\,\mathbb{1}_{\{a<\tau_i\le b\}},$$

which is a Poisson measure with

$$\pi_\gamma(B,s) = \pi\big(\{(z,s)\in\mathbb{R}^m_*\times[0,\infty) : \gamma(z,s)\in B\}, s\big)$$

and $\Pi_\gamma(\mathrm{d}z,\mathrm{d}s) = \pi_\gamma(\mathrm{d}z,s)\mathrm{d}s$ as its Lévy measure on $\mathbb{R}^d_0 \times [0,\infty)$. □

Recall that δ denotes the jumps operator $\delta X(t) = X(t) - X(t-)$, the jumps of a local-martingale have the following structure,

Theorem 3.6 (jump structure)**.** *Let X be an optional locally integrable process with $X(0) = 0$. Then there exists a (unique purely discontinuous) local-martingale M such that δM and X are indistinguishable (i.e., except on a set of measure zero we have $\delta M(t) = X(t)$, for every $t \geq 0$) if and only if the predictable projection ${}^pX = 0$ and the increasing process $t \mapsto \sqrt{\sum_{s\leq t} |X(s)|^2}$ is (locally) integrable. Moreover, M is a (locally) square integrable martingale if and only if $t \mapsto \sum_{s\leq t} |X(s)|^2$ is (locally) integrable and M is a local-martingale with (locally) bounded variation paths if and only if $t \mapsto \sum_{s\leq t} |X(s)|$ is (locally) integrable.*

Proof. One part of the argument goes as follows. (1) First, if X is locally square integrable predictable process with ${}^pX = 0$ then a local martingale M satisfying $\delta M(t) = X(t)$, for every $t \geq 0$, can be constructed, essentially the case of the stochastic integral. (2) Second, if X is locally integrable predictable process with ${}^pX = 0$ then $A(t) = \sum_{s\leq t} X(s)$ and $A - A^p$ have locally integrable bounded variation paths, where A^p is its compensator. Since $\delta(A^p) = {}^p(\delta A) = {}^pX = 0$, we can set $M = A - A^p$ to obtain $\delta M = X$, which is a local-martingale with locally integral bounded variation paths. Finally, the general case is a superposition of the above two arguments. Indeed, let X be an optional process with ${}^pX = 0$ and $\sqrt{A}$ locally integrable, where $A = \sum_{s\leq t} |X(s)|^2$. Set $Y = X\mathbb{1}_{|X|>1}$, $X'' = Y - {}^pY$ and $X' = X - X''$, so ${}^pX' = {}^pX'' = 0$. The increasing process $B(t) = \sum_{s\leq t} |Y(s)|$ satisfies $|\delta B| \leq \sqrt{|\delta A|}$ so that B is locally integrable. Because ${}^p(\delta B) = \delta(B^p)$ we have $\sum_{s\leq t} |{}^pY(s)| \leq B^p(t)$, so that $\alpha(t) = \sum_{s\leq t} |X''(s)|$ is also locally integrable. In view of the previous argument (2), there is a local martingale M'' with locally integrable bounded paths such that $\delta M'' = X''$. Next, because $|X'|^2 \leq 2|X|^2 + 2|X''|^2$ the process $\beta(t) = \sum_{s\leq t} |X'(s)|^2$ takes finite values. Since ${}^pX = 0$ we have ${}^pY = -{}^p(X\mathbb{1}_{|X|\leq 1})$, $|{}^pY| \leq 1$ and $|X'| \leq 2$, which yields $\delta\beta(t) \leq 4$, proving that the increasing process β is locally integrable. Again, in view of the previous argument (1), there is a local martingale M' such that $\delta M' = X'$. The proof is ended by setting $M = M' + M''$.

Since any local-martingale M can (uniquely) expressed as the sum $M = M^c + M^d$, where M^c is a continuous local-martingale and M^d is a purely discontinuous local-martingale (with $M^d(0) = 0$), the purely discontinuous part M^d is uniquely determined by the jumps δM. So adding the property *purely discontinuous* to the above martingale, we have the uniqueness. Full details can be found in Jacod and Shirayaev [84, Theorem 4.56, Chapter 1, p. 56–57]. □

Let ν be a quasi-left continuous integer-valued random measure (in particu-

lar, a Poisson measure), i.e,

$$\nu(B\times]a,b],\omega) = \sum_{n=1}^{\infty} \mathbb{1}_{a_n(\omega)\in B}\mathbb{1}_{\tau_n(\omega)\in]a,b]},$$

$$\mathbb{E}\{\nu(\mathbb{R}^m_* \times \{t\})\} = \mathbb{E}\Big\{\sum_{n=1}^{\infty} \mathbb{1}_{\tau_n(\omega)=t}\Big\} = 0,$$

for every B in $\mathcal{B}(\mathbb{R}^m_*)$, $b > a \geq 0$ and $t \geq 0$, where $\{a_n : n \geq 1\}$ is a sequence of points in $\mathbb{R}^m_*$ such that a_n is $\mathcal{F}(n)$-measurable, and $\{\tau_n : n \geq 1\}$ is a (unordered) sequence of predictable stopping times. Then, the stochastic integral with respect to ν is (uniquely) defined for any predictable process $f(z,s,\omega)$ such that $F : t \mapsto \sqrt{\sum_n |f(a_n,\tau_n)|^2}$ is locally integrable, in particular if $\mathbb{E}\{|f(a_n,\tau_n)|^2\} < \infty$ for every $n \geq 1$. If ν is not quasi-left continuous (e.g., an extended Poisson measure) then the predictable projection of F may not vanish, i.e., ${}^pF(t) = \sum_n f(a_n,t)\,\mathbb{1}_{\tau_n=t}$, when every the (pathwise) series converges absolutely. Thus f is integrable with respect to ν if the (optional) process $F(t) - {}^pF(t)$ is locally integrable, see Jacod and Shirayaev [84, Definition 1.27, Chapter 2, p. 72].

For future reference, we conclude this subsection with the following summery of key properties and relations.

Let us go back to the case of a Poisson measure ν with Lévy measure (properly saying, intensity or characteristic measure) Π, i.e., $\Pi(B\times]s,t]) = \mathbb{E}\{\nu(B\times]s,t])\}$, $\Pi(B\times\{t\}) = 0$, for every $t > s > 0$ and Borel subset B of $\mathbb{R}^m_*$, and Π integrates the function $z \mapsto |z|^2 \wedge |z|$ on $\mathbb{R}^m_*\times]0,T]$, for every $T > 0$.

Next, we construct a local-martingale measure $\tilde{\nu} = \nu - \Pi$, and its associated purely jumps (which is quasi-continuous from the left, i.e., with no deterministic jumps) local-martingale process $L = (L_i)$ with values in $\mathbb{R}^m$,

$$L_i(t) = \int_{\mathbb{R}^m_*\times]0,t]} z_i\tilde{\nu}(\mathrm{d}z,\mathrm{d}s), \quad \forall\, t \geq 0,\ i = 1,\dots,m,$$

with predictable compensator

$$L^p_i(t) = \int_{\mathbb{R}^m_*\times]0,t]} z_i\Pi(\mathrm{d}z,\mathrm{d}s), \quad \forall\, t \geq 0,\ i = 1,\dots,m.$$

Usually, L is referred to as the *canonical* compensated Poisson (jump) process associated with the Poisson measure ν, and reciprocally, ν is referred to as the *canonical* Poisson measure associated with the compensated Poisson (jump) process L.

For a predictable process $f(x,s,\omega)$ satisfying the integrability condition (3.26) we can define the stochastic integral (a real-valued local-martingale)

$$I(t) = \int_{\mathbb{R}^m_*\times]0,t]} f(z,s)\tilde{\nu}(\mathrm{d}z,\mathrm{d}s) \quad \forall\, t \geq 0,$$

and $I(0) = 0$, as a cad-lag process (and quasi-continuous from the left). If the integrand takes the form $f(z, s, \omega) = \sum_{i=1}^{m} g_i(t, \omega) z_i$ then we can write

$$I(t) = \sum_{i=1}^{m} \int_{]0,t]} g_i(s) \mathrm{d}L_i(s) \quad t \geq 0.$$

Always, we have the following properties on their jumps:

$$I(t) - I(t-) = \delta I(t) = f(\delta L(t), t) \mathbb{1}_{\{|\delta L(t)|>0\}}, \quad \forall\, t > 0.$$

The stochastic integral process $I(t)$ is a locally integrable bounded variation process if and only if

$$P\big\{ \int_{\mathbb{R}^d_* \times (0,t]} |f(z,s)|\, \Pi(\mathrm{d}z, \mathrm{d}s) < \infty \big\} = 1, \quad \forall\, t > 0$$

or equivalently

$$P\big\{ \sum_{0<s\leq t} |\delta I(s)| < \infty \big\} = 1, \quad \forall\, t > 0,$$

and in this case we have

$$I(t) = \sum_{0<s\leq t} f(\delta L(s), s) \mathbb{1}_{\{|\delta L(s)|>0\}} - \int_0^t f(z,s) \Pi(\mathrm{d}z, \mathrm{d}s), \quad \forall\, t > 0,$$

where the series converges absolutely almost surely. It is clear that the separation of the stochastic integral into a series of jumps and Lebesgue-type integral is not possible in general. However, the definition allows a suitable limit $I(t) = \lim_{\varepsilon \to 0} I_\varepsilon(t)$, where $I_\varepsilon(t)$ is the stochastic integral (of finite jumps almost surely) associated with the Lévy measure $\Pi_\varepsilon(B \times]s,t]) = \Pi\big((B \cup \{|z| \geq \varepsilon\} \times]s,t]\big)$, which can be written as previously (actually the series of jumps becomes a stochastic finite sum). In any case, the series of the jumps squared is absolutely convergent almost surely, and the process

$$t \mapsto \sum_{0<s\leq t} [I(s) - I(s-)]^2 - \int_0^t |f(z,s)|^2 \Pi(\mathrm{d}z, \mathrm{d}s)$$

is a local-martingale.

Note that the integer measure ν_I on $\mathbb{R}_*$ induced by the jumps of $I(t)$, namely,

$$\nu_I(K \times]0,t]) = \sum_{0<s\leq t} \mathbb{1}_{\{f(\delta L(s),s) \in K\}}, \quad \forall t > 0,\ K \subset \mathbb{R}_*, \ \text{ compact},$$

with predictable compensator

$$\nu_I^p(K \times]0,t]) = \int_0^t \Pi\big(\{z \in \mathbb{R}^m_* : f(z,s) \in K\}, \mathrm{d}s\big),$$

yield the martingale measure $\tilde{\nu}_I = \nu - \nu^p$.

If we take an integrand $f(z,t,\omega)$ with values in $\mathbb{R}^n$ then the stochastic integral I will take values in $\mathbb{R}^n$ and its associated integer measure ν_I would be defined in $\mathbb{R}^n_*$.

Certainly, if we begin with a Lévy measure Π that integrates only $|z|^2 \wedge 1$ then we need to split the jumps into two classes (small and large) to express the above properties. Also, recall that if we begin with Lévy processes $\ell_i(t)$, $i = 1, \dots, m$ we may construct the integer measure ν (which is actually a standard Poisson measure) associated with the *jumps* of the $\mathbb{R}^m$-valued process $\ell = (\ell_1, \dots, \ell_m)$. The Lévy measure associated with (standard) Poisson measure ν or the Lévy m-dimensional process ℓ is the same (of the form $\pi(\mathrm{d}z)\mathrm{d}s$), and the canonical compensated Poisson process L has exactly the same jumps as ℓ, i.e., $\delta\ell(t) = \delta L(t)$, for every $t > 0$. Note that the Lévy measure $\pi(\mathrm{d}z)$ in $\mathbb{R}^m_*$ is not necessarily the product measure of the individual Lévy measures $\pi_i(\mathrm{d}z_i)$ in $\mathbb{R}_*$ of each ℓ_i, even if the ℓ_i are independent, one needs also to assume no simultaneous jumps. Actually, if ℓ_i are independent then $\pi(\mathrm{d}z) = \sum_i \pi_i(\mathrm{d}z_i)$, after identifying the measure $\pi_i(\mathrm{d}z_i)$ in $\mathbb{R}^1_*$ with the measure $\pi_i(\mathrm{d}z_i) \times 0_i$ in $\mathbb{R}^m_*$ where 0_i is the zero-measure in $(\mathrm{d}z_1, \dots, \mathrm{d}z_{i-1}, \mathrm{d}z_{i+1}, \dots, \mathrm{d}z_m)$.

3.2.3 Extension to Semi-martingales

Remark that the initial intension is to integrate a process $f(s)$ or $f(z,t)$ which is adapted (predictable) with respect to a Wiener process $w(s)$ or centered Poisson measure $\tilde{\nu}(\mathrm{d}z, \mathrm{d}s)$. This is to say that in most of the cases, the filtration $\{\mathcal{F}(t) : t \geq 0\}$ is generated by the Wiener process or the Poisson measure, which is completed for convenience. However, what is mainly used in the construction of the stochastic integral are the following conditions:

(a) the filtration $\mathbb{F} = \{\mathcal{F}(t) : t \geq 0\}$ is complete and right-continuous,

(b) the integrand f is predictable with respect to filtration $\mathbb{F}$,

(c) the integrator w (or $\tilde{\nu}$) is a (semi-)martingale with respect to filtration $\mathbb{F}$.

Thus we are interested in choosing the filtration $\mathbb{F}$ as large as possible, but preserving the (semi-)martingale character. e.g., the non-anticipative filtration $\mathbb{A}$, where $\mathcal{A}(t)$ is defined as the σ-algebra of all sets in $\mathcal{F}$ which are independent of either $w(t_1)-w(t_0), \dots, w(t_n)-w(t_{n-1})$ or $\tilde{\nu}(K_j \times]t_{i-1}, t_i])$, for any $j = 1, \dots, m$ and $t \leq t_0 < t_1 < \cdots < t_n$. Note that $\mathcal{A}(t)$ contains all null sets in $\mathcal{F}$ and the cad-lag property of w (or $\tilde{\nu}$) shows that $\mathcal{A}(t) = \bigcap_{s>t} \mathcal{A}(s)$. Because $w(t)$ (or $\tilde{\nu}(K \times]s,t])$) is independent of any future increment, the σ-algebra $\mathcal{F}(t)$ generated by $\{w(s) : s \leq t\}$ (or by $\{\tilde{\nu}(K \times]0,s]) : s \leq t\}$) is included in $\mathcal{A}(t)$. Moreover, since

$$\mathbb{E}\{w(t) \mid \mathcal{A}(s)\} = \mathbb{E}\{w(t) - w(s) \mid \mathcal{A}(s)\} + \mathbb{E}\{w(s) \mid \mathcal{A}(s)\} = \\ = \mathbb{E}\{w(t) - w(s)\} + w(s) = w(s),$$

the martingale character is preserved.

Actually, the cancelation is produced when the integrator is independent and has increment of zero-mean, even least, when the increments of the integrator are orthogonal to the integrand, e.g., $\mathbb{E}\{f(s)[w(t)-w(s)]\} = \mathbb{E}\{f(s)\}\mathbb{E}\{w(t)-w(s)\} = 0$ for $t > s$. Thus, define the class $\mathcal{E}^*$ of processes of the form $f(z,t,\omega) = f_{i-1,j}(\omega)$ if $t_{i-1} < t \le t_i$ and z belongs to K_j with some $i = 1,\dots,n$, and $j = 1,\dots,m$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers, K_j are disjoint sets with compact closure in $\mathbb{R}^m_*$ and $f_{i-1,j}$ is a bounded random variable which is *orthogonal* to $\tilde{\nu}(K_j\times]t_{i-1},t_i])$ (in particular $\mathcal{F}(t_{i-1})$-measurable) for any i, and $f(t,\omega) = 0$ otherwise, and an analogous definition for the Wiener process case. The stochastic integral is then initially defined on the class $\mathcal{E}^*$ and the extension procedure can be carried out successfully, we refer to Section 3.1 of the the previous chapter on *Random Orthogonal Measures*. In any case, remark that if f is a deterministic function then to define the stochastic integral we need the local L^2-integrability in time, e.g., an expression of the form $s \mapsto s^\alpha$ or $(z,s) \mapsto (z \wedge 1)s^\alpha$ is integrable as long as $\alpha > -1/2$.

Space of Semi-martingales

Let us now consider the space $\mathcal{S}^p(\Omega,\mathcal{F},P,\mathcal{F}_t,t\ge 0)$, $1 \le p \le \infty$ of p-integrable semi-martingale on $[0,\infty]$ is defined as the cad-lag processes X with a decomposition of the form $X = M + A^+ - A^-$ where M is a local martingale and A^+, A^- are adapted monotone increasing processes with $A^+(0) = A^-(0) = 0$, both relative to $(\mathcal{F}_t : t \ge 0)$ and such that the quantity

$$\|X\|_{S^p} = \inf_{X=M+A^+-A^-} \|M,A^+,A^-\|_{S^p},$$

where

$$\|M,A^+,A^-\|_{\mathbb{S}^p} = \mathbb{E}\Big\{\big[\sqrt{[M](\infty)} + |A^+(\infty)| + |A^-(\infty)|\big]^p\Big\}^{1/p},$$

is finite. This is a semi-norm and by means a of equivalence classes we define the *non-separable* Banach space $\mathbb{S}^p(\Omega,\mathcal{F},P,\mathcal{F}_t,t\ge 0)$ or $\mathbb{S}^p(\Omega,\mathbb{F},P)$ with $\mathbb{F} = \{\mathcal{F}_t, t \ge 0\}$.

Going back to the above definition of the semi-norm $\|X\|_{\mathbb{S}^p}$, if the square bracket process $\sqrt{[M](\infty,\omega)}$ is replaced with maximal process $M^*(\infty,\omega) = \sup_{t\ge 0}|M(t,\omega)|$ then we obtain an equivalent semi-norm.

This procedure can be localized, i.e., define $\mathcal{S}^p_{\rm loc}(\Omega,\mathcal{F},P,\mathcal{F}_t,t\ge 0)$ and the space of equivalence classes $\mathbb{S}^p_{\rm loc}(\Omega,\mathcal{F},P,\mathcal{F}_t,t\ge 0)$ as the spaces of semi-martingales X such that there is a sequence of stopping times $\tau_k \to \infty$ as $k \to \infty$ satisfying $X_k(\cdot) = X(\cdot\wedge\tau_k)$ belongs to $\mathcal{S}^p(\Omega,\mathcal{F},P,\mathcal{F}_t,t\ge 0)$, for any $k \ge 1$. Thus $\mathcal{S}^1_{\rm loc}(\Omega,\mathcal{F},P,\mathcal{F}_t,t\ge 0)$ is the space of special semi-martingales.

A further step is to consider $\mathcal{S}^0(\Omega,\mathcal{F},P,\mathcal{F}_t,t\ge 0)$ the space of all semi-martingales (including non-special) X on the closed real semi-line $[0,\infty]$, i.e., $X = M + A^+ - A^-$ where M is a local-martingale in $[0,\infty]$ and A^+, A^- are adapted monotone increasing processes with $A^+(0) = A^-(0) = 0$ and

$A^+(\infty)$, $A^-(\infty)$ are almost surely finite. With the topology induced by the semi-distance

$$\begin{aligned}
&\|X\|_{\mathbb{S}^0} = \inf_{X=M+A^+-A^-} \|M, A^+, A^-\|_{\mathbb{S}^0},\\
&\|M, A^+, A^-\|_{\mathbb{S}^0} = \mathbb{E}\{1 \wedge \big(\sqrt{[M](\infty)} + |A^+(\infty)| + \\
&\qquad\qquad + |A^-(\infty)|\big)\} + \sup_\tau \mathbb{E}\{|M(\tau) - M(\tau-)|\},
\end{aligned}$$

for any stopping time τ. Thus $\mathbb{S}^0(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$, after passing to equivalence classes, is a non-separable complete vector space. A closed non-separable subspace is the set $\mathbb{S}^p_{\rm c}(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$ of all continuous p-integrable semi-martingales, which admits a localized space denoted by $\mathbb{S}^p_{\rm c,loc}(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$. The reader may take a look at Protter [149, Section V.2, pp. 138–193] for others similar spaces of semi-martingales.

A companion (dual) space is the set $\mathcal{P}^p(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$ of p-integrable predictable processes X, i.e., besides being predictable we have

$$||X||_{\mathbb{P}^p} = \Big\{ \int_0^\infty \mathrm{d}t \int_\Omega |X(t,\omega)|^p \, P(d\omega) \Big\}^{1/p},$$

which yields the non-separable Banach space $\mathbb{P}^p(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$. Its localized spaces $\mathcal{P}^p_{\rm loc}(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$ and $\mathbb{P}^p_{\rm loc}(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$, $p \geq 1$, are defined by the conditions (1) X is a predictable process and (2) such that there is an increasing sequence of stopping times $\tau_k \to \infty$ as $k \to \infty$ such that the processes $X_k = \mathbb{1}_{]0,\tau_k]} X$ belong to $\mathcal{P}^p(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$, for any $k \geq 1$.

Note that the uncountable set of bounded and adapted left-continuous (having right-hand limit) processes is a dense subspace of $\mathbb{P}^p(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$. However, the set $\mathbb{P}^p_{\rm c}(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$ of bounded and continuous (adapted, p-integrable) processes is neither dense nor closed. We refer to Dellacherie and Meyer [32, Sections VII.3.96–105, pp. 308–324].

Back to the Stochastic Integral

It is clear by now that semi-martingales are *desirable integrators* while predictable processes are *desirable integrands*. Semi-martingales contain two type of (localized) processes, (1) a bounded variation process which is integrated following the classic measure theory and (2) a local-martingale which is the main study of stochastic integrals. To focus in the stochastic integral itself, the natural integrators (without localizing) are the so-called *quasi-martingales*, defined as an adapted cad-lag process X satisfying $\mathbb{V}\mathrm{ar}(X) = \sup\{\mathbb{V}\mathrm{ar}(X, \pi) : \pi\} < \infty$, where $\pi = \{t_0, t_1, \ldots, t_n\}$, $0 = t_0 < t_i < t_{i+1}$,

$$\mathbb{V}\mathrm{ar}(X, \pi) = \sum_{i=1}^n \big|\mathbb{E}\{X(t_i) - X(t_{i-1}) \,|\, \mathcal{F}(t_{i-1})\}\big| + |X(t_n)|. \tag{3.28}$$

It can be proved, e.g. see Rogers and Williams [153, Section VI.41, pp. 396–398]), that any quasi-martingale admits a representation $X = Y - Z$, where Y

and Z are two nonnegative super-martingales such that $\mathbb{V}\mathrm{ar}(X) = \mathbb{V}\mathrm{ar}(Y) + \mathbb{V}\mathrm{ar}(Z)$ and that if $X = \bar{Y} - \bar{Z}$ are two other nonnegative super-martingales then $\bar{Y} - Y = \bar{Z} - Z$ is also a nonnegative super-martingale.

Given a filtered probability space $(\Omega, P, \mathcal{F}, \mathcal{F}_t : t \geq 0)$, let $\mathcal{M}$, $\mathcal{O}$ and $\mathcal{P}$ be the *measurable*, *optional* and *predictable* σ-algebras on $[0,\infty)\times\Omega$. Now, a subset N of $[0,\infty)\times\Omega$ is called *evanescent* if $P\{\omega \in \Omega : (t,\omega) \in N\} = 0$ for every $t \geq 0$. We suppose that $\mathcal{M}$, $\mathcal{O}$ and $\mathcal{P}$ have been augmented with all evanescent sets.

For a given integrable monotone increasing (bounded variation) cad-lag process A, with its associated continuous and jump parts $A(t) = A^c(t) + [A(t+) - A(t-)]$, we may define a (signed) measure μ by the expression

$$\mu(X) = \mathbb{E}\Big\{ \int_{[0,\infty)} X(t)\mathrm{d}A(t)\Big\} = \\ = \mathbb{E}\Big\{ \int_0^\infty X(t)\mathrm{d}A^c(t) + \sum_{t\geq 0} X(t)\,[A(t+) - A(t-)]\Big\}$$

for any nonnegative $\mathcal{M}$ measurable process X. This measure vanishes on evanescent sets. Conversely, it can be proved (Doléans' Theorem, e.g., Rogers and Williams [153, Section VI.20, pp. 249–351]) that any bounded measure μ on $\mathcal{M}$, which vanishes on evanescent sets, can be represented (or disintegrated) as above for some process A as above. Furthermore, if μ satisfies

$$\mu(X) = \mu({}^oX) \quad \text{or} \quad \mu(X) = \mu({}^pX)$$

then A is optional or predictable.

Denote by $\mathcal{D}_0$ the vector space either (1) of all adapted cad-lag and bounded processes or (2) of all processes X of the form

$$X = \sum_{i=0}^n X_i\, \mathbb{1}_{[\tau_i,\tau_{i+1}[}, \quad 0 = \tau_0 \leq \tau_1 \leq \cdots \leq \tau_n \leq \tau_{n+1} = \infty,$$

for any n and stopping times τ_i. Now, if $A[\cdot]$ is a linear and positive functional on $\mathcal{D}_0$ satisfying the condition

$$P\{\lim_n \sup_{0\leq s\leq t} |X_n(s)|\} = 0,\ \forall t \geq 0 \ \text{ implies } \ \lim_n A(X_n) = 0, \tag{3.29}$$

then there should exist two integrable monotone increasing cad-lag processes A^o, A^p, with $A^o(-0) = 0$, A^o optional and purely jumps, and with A^p predictable, such that

$$A[X] = \mathbb{E}\Big\{ \int_{(0,\infty]} X(t-)\, dA^p(t) + \sum_{t\geq 0} X(t)\,[A^o(t) - A^o(t-)]\Big\},$$

for any X in $\mathcal{D}_0$, and the above representation is unique up to an evanescent set. Indeed, by means of condition (3.29) the functional $A[\cdot]$ can be extended

to a bounded positive measure which vanishes on evanescent sets and the result follows from the previous representation.

Similarly, an adapted process A, which is right-continuous in probability (not necessarily cad-lag), is a suitable integrator if and only if the set of random variables

$$\int X \mathrm{d}A = \sum_{i=0}^{n} X_i \left[A(\tau_{i+1}) - A(\tau_i)\right]$$

remains bounded (e.g., in probability or in L^2) for every elementary predictable process X satisfying

$$X = \sum_{i=0}^{n} X_i \, \mathbb{1}_{[\tau_i, \tau_{i+1}[}, \qquad |X| \leq 1.$$

For instance, the reader is referred to the book Bichteler [11, Chapter 2, pp. 43–86] for a carefully analysis on this direction.

Then, a desirable property for a linear positive function $M[\cdot]$ defined on $\mathcal{D}_0$ to be called stochastic integral is the following condition

$$\text{if } P\{\lim_n \sup_{0\leq s\leq t} |X_n(s)| \geq \delta\} = 0, \ \forall\, t \geq 0, \ \delta > 0$$
$$\text{then } P\{\lim_n \sup_{0\leq t\leq T} |M[X_n \, \mathbb{1}_{]0,t]}]| \geq \varepsilon\} = 0, \quad (3.30)$$

for every $T \geq 0$ and $\varepsilon > 0$, or even a weaker version of it.

For a given adapted cad-lag integrable real-valued process $\{Z(t) : t \geq 0\}$ we can define a functional $Z[\cdot]$ on $\mathcal{D}_0$ as follows:

$$Z[\sum_{i=0}^{n} X_i \, \mathbb{1}_{[\tau_i, \tau_{i+1}[}] = \sum_{i=0}^{n} X_i \left(Z(\tau_{i+1}) - Z(\tau_i)\right), \quad (3.31)$$

which can be initially defined on predictable rectangles $F\times]a, b]$, F in $\mathcal{F}(a)$ by means of

$$\lambda_Z(]a, b] \times F) = \mathbb{E}\{\mathbb{1}_F[Z(b) - Z(a)]\},$$
$$\lambda_Z(\{0\} \times F_0) = 0, \quad \forall\, F_0 \in \mathcal{F}(0), \quad (3.32)$$

and then extended by additivity. If the process Z is only locally integrable, we may suppress the last term with $\tau_{n+1} = +\infty$ or consider only (deterministic) times t_i instead of stopping times τ_i. If the functional $Z[\cdot]$ or equivalent the additive set function λ_Z is nonnegative, then λ_Z is called a *content*.

It is clear that $\lambda_Z \geq 0$ if Z is monotone increasing. However, $\lambda_Z = 0$ if Z is a martingale and $\lambda_Z \geq 0$ if Z is a sub-martingale. If $\{M(t) : t \geq 0\}$ is a square integrable then $\{M^2(t) : t \geq 0\}$ is a sub-martingale and hence $\lambda_{M^2} \geq 0$, moreover

$$\lambda_{M^2}(]a, b] \times F) = \mathbb{E}\{\mathbb{1}_F[M^2(b) - M^2(a)]\} =$$
$$= \mathbb{E}\{\mathbb{1}_F[M(b) - M(a)]^2\}, \quad \forall\, b > a \geq 0, \ F \in \mathcal{F}(a). \quad (3.33)$$

The extension of λ_{M^2} to a measure on $(\mathbb{R}^+ \times \Omega, \mathcal{P})$ is called Doléans measure. It can be proved (e.g. Chung and R.J. Williams [25, Theorem 2.16, Chapter 2, pp. 52–53]) that if the process Z is a positive sub-martingale then the content λ_Z can be uniquely extended to a σ-finite measure on $\mathcal{P}$. In particular this applies to λ_{M^2}.

Extension Argument

Denote by $\mathcal{E}$ the vector space of all processes of the form $X(t,\omega) = X_{i-1}(\omega)$ if $t_{i-1} < t \le t_i$ with $i = 1, \dots, n$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers and X_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $X(t,\omega) = 0$ otherwise. Elements in $\mathcal{E}$ are called elementary predictable processes. Given a square integrable $\{M(t) : t \ge 0\}$ we denote by μ_M its Doléans measure and define the stochastic integral as follows:

$$\begin{aligned}
&\int X(s)\mathrm{d}M(s) = \sum_{i=1}^{n} X_{i-1}[M(t_i) - M(t_{i-1})],\\
&\int_{(0,t]} X(s)\mathrm{d}M(s) = \sum_{i=1}^{n} X_{i-1}[M(t \wedge t_i) - M(t \wedge t_{i-1})],\\
&\int_{(a,b]} X(s)\mathrm{d}M(s) = \int_{(0,b]} f(s)\mathrm{d}M(s) - \int_{(0,a]} X(s)\mathrm{d}M(s),
\end{aligned} \tag{3.34}$$

for every $t \ge 0$ and $b > a \ge 0$. Note that

$$\int_{(a,b]} X(s)\mathrm{d}M(s) = \int X(s)\, \mathbb{1}_{(a,b]}(s)\mathrm{d}M(s),$$

for every $b > a \ge 0$. This definition (3.34) (defined up to an evanescent set) is independent of the particular representation used and the fact that X_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable implies

$$\mathbb{E}\Big\{\Big|\int X(s)\mathrm{d}M(s)\Big|^2\Big\} = \sum_{i=1}^{n} \mathbb{E}\{|X_{i-1}|^2[M^2(t_i) - M^2(t_{i-1})]\} = \\ = \int_{\mathbb{R}^+\times\Omega} |X|^2 \mathrm{d}\mu_M, \tag{3.35}$$

for every X in $\mathcal{E}$, and

$$\mathbb{E}\Big\{\Big[\int X(s)\mathrm{d}M(s)\Big]\Big[\int Y(s)\mathrm{d}M(s)\Big]\Big\} = \int_{\mathbb{R}^+\times\Omega} XY\mathrm{d}\mu_M, \tag{3.36}$$

for any two stochastic processes X and Y in $\mathcal{E}$.

Moreover the process $Z(t) = (X \diamond M)(t)$,

$$(X \diamond M)(t) = \int_{(0,t]} X(s)\mathrm{d}M(s), \quad \forall t \ge 0, \tag{3.37}$$

is (cad-lag) square integrable martingale, which is continuous if M is so. Since,

$$\mu_Z(]a,b]\times F) = \mathbb{1}_F[Z(b)-Z(a)]^2) = \\ = \mathbb{E}\{\mathbb{1}_F[\int_{(a,b]} X(s)\mathrm{d}M(s)]^2\} = \int_{(a,b]\times F}|X|^2\mathrm{d}\mu_M,$$

we deduce that

$$\mu_{X\diamond M}(B) = \int_B |X|^2\mathrm{d}\mu_M, \quad \forall\, B\in\mathcal{P}. \tag{3.38}$$

If X belongs to $\mathcal{E}$, F is a $\mathcal{F}(a)$-measurable set and τ a stopping time which takes only finitely many values then $\mathbb{1}_F X$ and $\mathbb{1}_{]0,\tau]}X$ belong to $\mathcal{E}$ and

$$\begin{aligned}&\int_{]a,b]}\mathbb{1}_F X(s)\mathrm{d}M(s) = \mathbb{1}_F\int_{]a,b]}X(s)\mathrm{d}M(s),\\ &[X\diamond M](\tau) = \int\mathbb{1}_{]0,\tau]}(s)X(s)\mathrm{d}M(s).\end{aligned} \tag{3.39}$$

It is also clear from the expression (3.34) that the jumps of $(X\diamond M)$ are produced only by jumps of the integrator M, i.e.,

$$(X\diamond M)(t) - (X\diamond M)(t-) = X(t)[M(t)-M(t-)], \quad \forall\, t>0, \tag{3.40}$$

except for a set of measure zero.

Denote by $\bar{\mathcal{E}}_M$ the L^2-closure of $\mathcal{E}$, i.e., the Hilbert space of all processes X for which there exists a sequence $(X_1, X_2, \ldots)$ of processes in $\mathcal{E}$ such that

$$\lim_{n\to\infty}\int_{\mathbb{R}^+\times\Omega}|X_n - X|^2\mathrm{d}\mu_M = 0.$$

Based on the isometry identity (3.35), and the maximal martingale inequality, for every $T\geq 0$,

$$\mathbb{E}\{\sup_{0\leq t\leq T}\big|\int_{(0,t]}X(s)\mathrm{d}M(s)\big|^2\} \leq 4\,\mathbb{E}\{\big|\int_0^T X(s)\mathrm{d}M(s)\big|^2\}, \tag{3.41}$$

this linear operation (called *stochastic integral*) can be extended to the closure $\bar{\mathcal{E}}_M$, preserving linearity and the properties (3.35), ..., (3.40). Moreover, (3.39) holds for any bounded $\mathcal{F}(a)$-measurable function f replacing $\mathbb{1}_F$ (even if a is a bounded stopping times) and any bounded stopping time τ.

In general, it is proved in Doob [33, Section IX.5, pp. 436–451] that any martingale M with orthogonal increments (i.e., a square-integrable martingale), the Hilbert space $\bar{\mathcal{E}}_M$ contains all adapted process X and square-integrable respect to the product measure $P(\mathrm{d}\omega)$ times the Lebesgue-Stieltjes measure $\mathrm{d}\mathbb{E}\{|M(t)-M(0)|^2\}$.

It is convenient to localize the above processes, i.e., we say that a measurable process X belongs to $\bar{\mathcal{E}}_{M,\mathrm{loc}}$ if and only if there exists a sequence of stopping

times $\{\tau_k : k \geq 1\}$ such that $\tau_k \to \infty$ almost sure and $\mathbb{1}_{]0,t\wedge\tau_k]}X$ belongs to $\bar{\mathcal{E}}_{M_k}$, for every $t > 0$, where $M_k = \{M(s \wedge \tau_k) : s \geq 0\}$. Therefore, the stochastic integral $X \diamond M$ is defined as the almost sure limit of the sequence $\{X_k \diamond M_k : k \geq 1\}$, with $X_k = \mathbb{1}_{]0,\tau_k]}X$. This should be validated by a suitable condition to make this definition independent of the choice of a localizing sequence, see Chung and Williams [25, Theorem 2.16, Chapter 2, pp. 23–48].

The use of the quadratic variation process is simple when dealing with a continuous square integrable martingale. The general case is rather technical. Anyway, a key point is the following: If $M = \{M(t) : t \geq 0\}$ is a locally square integrable martingale then there exists an increasing predictable process $\langle M\rangle$ such that $M^2 - \langle M\rangle$ is a local-martingale, which is continuous if and only if M is quasi-left continuous (e.g., Jacod and Shiryaev [84, Theorem 4.2, Chapter 1, pp. 38–39]). It is clear that we have, first for X in $\mathcal{E}$ and then for every X in $\mathcal{E}_M$, the relation

$$\langle X \diamond M\rangle(t) = \int_0^t |X(s)|^2 \mathrm{d}\langle M\rangle(s), \quad \forall t \geq 0, \tag{3.42}$$

so that the process

$$\Big[\int_{(0,t]} X(s)\mathrm{d}M(s)\Big]^2 - \int_0^t |X(s)|^2 \mathrm{d}\langle M\rangle(s), \quad \forall t \geq 0, \tag{3.43}$$

is a (cad-lag) local-martingale.

Lenglart's domination property (see inequality (2.10) in Chapter 3 or more details in Jacod and Shiryaev [84, Section 1.3c, pp. 35–36]) yields the useful estimate

$$P\Big\{\sup_{0\leq t\leq T}\Big|\int_{(0,t]} X(s)\mathrm{d}M(s)\Big| \geq \varepsilon\Big\} \leq \frac{\delta}{\varepsilon^2} + \\ + P\Big\{\int_0^T |X(s)|^2 \mathrm{d}\langle M\rangle(s) \geq \delta\Big\}, \tag{3.44}$$

for any positive numbers T, ε and δ. By means of this estimate, all properties (3.35), ..., (3.40), (3.43), (3.44) hold, except that the process (3.37) is now a (cad-lag, continuous whenever M is such) local square martingale. Moreover, the continuity property (3.30) is now verified.

Since any continuous local-martingale is a local square integral martingale, the stochastic integral is well defined. To go one step further and define the stochastic integral for any (cad-lag, not necessarily continuous and not necessarily local square integrable) local-martingale M, we need to define the (optional) quadratic variation, see (2.7) in Chapter 3 or for more detail see for instance Dellacherie and Meyer [32, Chapters V–VIII] or Liptser and Shiryayev [111],

$$[M](t) = \langle M^c\rangle(t) + A_M(t), \quad A_M(t) = \sum_{s\leq t}[M(s) - M(s-)]^2, \tag{3.45}$$

for any $t \geq 0$, where M^c is the continuous part of the (local) martingale M and the second term in the right-hand side A_M is an optional monotone increasing process null at time zero, not necessarily locally integrable, but such that $\sqrt{A_M}$ is locally integrable. It can be proved (see Rogers and Williams [153, Theorem 37.8, Section VI.7, pp. 389–391]) that the process $[M]$ given by (3.45) is the unique optional monotone increasing process null at time zero such that $M^2 - [M]$ is a local-martingale and $[M](t) - [M](t-) = [M(t) - M(t-)]^2$, for every $t > 0$.

On the other hand, a local-martingale admits a unique decomposition $M = M_0 + M^c + M^d$, where M_0 is a $\mathcal{F}(0)$-measurable random variable, M^c is a continuous local-martingale (null at $t = 0$) and M^d is a purely discontinuous local-martingale, i.e., $M^d(0) = 0$ and for every continuous local-martingale N the product $M^d N$ is a local-martingale. Let us show that for a given $\varepsilon > 0$, any local-martingale M admits a (non unique) decomposition $M = M_0 + M_\varepsilon^{'} + M_\varepsilon^{''}$, where M_0 is a $\mathcal{F}(0)$-measurable random variable, $M_\varepsilon^{'}$ is a (cad-lag, only the small jumps) local-martingale (null at $t = 0$) satisfying $|M_\varepsilon^{'}(t) - M_\varepsilon^{'}(t-)| \leq \varepsilon$ for every $t > 0$, and $M_\varepsilon^{''}$ is a (cad-lag, only the large jumps) local martingale (null at $t = 0$) which have local bounded variation. Indeed, set $\delta M(t) = M(t) - M(t-)$ and because M is a cad-lag process we can define $A(t) = \sum_{s \leq t} \delta M(s) \mathbb{1}_{|\delta M(s)| > \varepsilon/2}$, whose variation process $\mathrm{var}(A, t) = \sum_{s \leq t} |\delta M(s)| \mathbb{1}_{|\delta M(s)| > \varepsilon/2}$ is finite for almost every path. Setting $\tau_k = \inf\{t > 0 : \mathrm{var}(A, t) > k \text{ or } |M(t)| > k\}$ we obtain $\mathrm{var}(A, \tau_k) \leq k + |\delta M(\tau_k)|$, i.e, $\mathrm{var}(A, \tau_k) \leq 2k + |M(\tau_k)|$ so that the sequence of stopping times $\{\tau_k : k \geq 1\}$ is a reducing sequence for $\mathrm{var}(A, \cdot)$, proving that the process $\mathrm{var}(A, \cdot)$ is local integrable. Therefore A admits a dual predictable compensator A^p, see Definition 2.24 in Chapter 3. It is clear that $M_\varepsilon^{''} = A - A^p$ is a local-martingale with local bounded variation. A simple calculation show that $M_\varepsilon^{'} = M - A + A^p$ satisfies $|\delta M(t)| \leq \varepsilon$, for every $t > 0$. Moreover, since $M_\varepsilon^{''}$ is also a purely discontinuous martingale, i.e., $M_\varepsilon^{''}$ is orthogonal to any continuous local-martingale N, namely $M_\varepsilon^{''} N$ is a local martingale, see Jacod and Shiryaev [84, Section 1.4b, pp. 40–43]).

Thus, an essential fact needed to complete the definition of stochastic integral is that either a local-martingale or semi-martingale M admits a (non-unique) decomposition $M = M_0 + M_1 + M_2$, where M_0 is a $\mathcal{F}(0)$-measurable random variable, M_1 is a cad-lag process with locally bounded variation paths and M_2 is a local square integrable martingale, both null at time zero. Therefore, integration with respect to M_1 is pathwise (as in the classic measure theory) and integration with respect to M_2 is as above, via the martingale argument.

The only continuous local-martingale which has bounded variation paths is the constant process. However, there are (e.g., the Poisson process, after subtracting its compensator) (cad-lag, non-continuous) local martingale with bounded variation paths. Therefore there are two possible interpretations when taking those processes as integrators. This is resolved by using the predictable version representing the equivalence class of the integrand process. For instance, if X is a cad-lag process and M is a local-martingale with bounded variation paths, e.g., for a sequence $\{a_i, \tau_i : i \geq 1\}$ of stopping times $\tau_1 \leq \tau_2 \leq \cdots$, with

$\tau_i \to \infty$, and $\mathcal{F}(\tau_i)$-measurable random variables a_i we have

$$A(t) = \sum_{i=1}^{\infty} a_i \mathbb{1}_{\tau_i \le t}, \quad \forall\, t \ge 0, \qquad M = A - A^p,$$

where A^p is the dual compensator of A. The expression

$$\int_{]0,t]} X(t) \mathrm{d}A(t) = \sum_{i=1}^{\infty} X(\tau_i) a_i \mathbb{1}_{\tau_i \le t}, \quad \forall\, t \ge 0,$$

is pathwise interpreted (and well defined) in the Riemann-Stieltjes sense if and only if the process X is left-continuous at each jump time, i.e., $X(\tau_i) = X(\tau_i-)$, for every $i \ge 1$. On the other hand, the measure induced by A or by $A_- : t \mapsto A(t-)$ (its left-continuous version) is the same sum of Dirac measures so that the expression

$$\int_{]0,t]} X(t) \mathrm{d}A_-(t) = \sum_{i=1}^{\infty} X(\tau_i) a_i \mathbb{1}_{\tau_i \le t}, \quad \forall\, t \ge 0,$$

is pathwise interpreted (and well defined) in the Riemann-Stieltjes sense if and only if the process X is right-continuous at each jump time, i.e., $X(\tau_i) = X(\tau_i+)$, for every $i \ge 1$. In the Lebesgue-Stieltjes sense, it does not matter which version A or A_- is used to derived the measure, proving that a bounded process X is integrable if it is right (or left) continuous at τ_i for every $i \ge 0$.

The dual compensator A^p of a (cad-lag) process A with locally integrable bounded variation satisfied, see Definition 2.24 in Chapter 3,

$$\mathbb{E}\{\int_{[0,T_k)} X(t,\omega) \mathrm{d}A^p(t,\omega) = \mathbb{E}\{\int_{[0,T_k)} {}^pX(t,\omega) \mathrm{d}A(t,\omega),$$

for every $k \ge 1$ and for any bounded measurable process X, where the predictable projection ${}^pX$, is such that for any predictable stopping time τ we have $\mathbb{E}\{{}^pX \mathbb{1}_{\tau<\infty}\} = \mathbb{E}\{X \mathbb{1}_{\tau<\infty}\}$. The sequence of stopping times $\{T_k : k \ge 1\}$ localizes A, i.e., the process $t \mapsto A(t \wedge T_k)$ has integrable bounded variation (meaning in this case $\mathbb{E}\{A(T_k)\} < \infty$) and $T_k \to \infty$ almost surely. We deduce that the stochastic integral with respect to an integrator $A - A^p$ is always zero for any predictable process X. Recall that the stochastic integral is meaningful only for the predictable member representing a given equivalence class of processes used as integrand.

Therefore, we conclude that as long as the predictable (in particular any adapted left-hand continuous) version of the integrand (equivalence class) process is used, the pathwise and stochastic integral coincide.

Back to Integer Random Measures

Let ν be an integer-valued (random) measure, see Definition 2.25, and let ν^p be a good version of its compensator, see Theorem 2.26. For instance, if ν

is a extended Poisson measure then ν^p is a deterministic Radon measure on $\mathbb{R}^m_* \times [0,\infty)$ with $\nu^p(\mathbb{R}^m_* \times \{0\}) = 0$. Denote by ν_{qc} the quasi-continuous part of ν, i.e.,

$$\nu_{qc}(B\times]a,b]) = \nu(B\times]a,b]) - \nu^p_d(B\times]a,b]),$$
$$\nu^p_d(B\times]a,b]) = \sum_{a<s\le b} \nu^p(\{s\}\times B),$$

with $\nu^p_c = (\nu_{qc})^p$, where

$$\nu^p_c = \nu^p(B\times]a,b]) - \nu^p_d(B\times]a,b]),$$

is a good version of the compensator of ν_{qc}. The measure ν^p_d contains all non-predictable discontinuities, which are not handled with the stochastic integral, they must be treated pathwise, by means of the classic measure theory. For instance, if $\nu = \nu_X$ defined as the number of jumps associated to a (cad-lag) local-martingale (or semi-martingale) X, see (2.35) then ν^p_d is locally integrable. The integral with respect to the predictable discontinuous part $\nu_d = \nu - \nu_{qc}$ is part of the stochastic integral. Thus, using the (cad-lag and quasi-left continuous, purely discontinuous) local-martingale measure $\tilde\nu_{qc} = \nu_{qc} - \nu^p_c = \nu - \nu^p$, we proceed as in Section 3.2.2 to define the stochastic integral, essentially replacing the Lévy measure $m(\mathrm{d}s,\mathrm{d}z)$ by (continuous part of) the compensator ν^p_c. Thus, for a elementary predictable process f of the form $f(t,z,\omega) = f_{i-1,j}(\omega)$ if $t_{i-1} < t \le t_i$ and z belongs to K_j with $i = 1,\dots,n$, and $j = 1,\dots,m$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers, K_j are disjoint compact subsets of $\mathbb{R}^m_*$ and $f_{i-1,j}$ is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $f(t,\omega) = 0$ otherwise, we set

$$\int_{\mathbb{R}^m_*\times(0,\infty)} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s) = \sum_{i=1}^n\sum_{j=1}^m f_{i-1,j}\,\tilde\nu_{qc}(K_j\times]t_{i-1},t_i]),$$

and

$$\int_{\mathbb{R}^m_*\times(a,b]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s) = \int_{\mathbb{R}^m_*\times(0,\infty)} f(z,s)\,\mathbb{1}_{(a,b]}(s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s),$$

for every $b > a \ge 0$. The L^2-closure of all elementary predictable processes $\mathcal{E}$ is denoted by $\bar{\mathcal{E}}_\nu$, i.e., processes f such that there is a sequence $(f_1, f_2, \dots)$ of processes in $\mathcal{E}$ such that

$$\lim_{k\to\infty} \mathbb{E}\Big\{ \int_{\mathbb{R}^m_*\times(0,\infty)} |f_k(z,s) - f(z,s)|^2\,\nu^p_c(\mathrm{d}z,\mathrm{d}s)\Big\} = 0.$$

Note that we may use (indistinctly), ν_c^p or ν_{qc} in the above condition, both are random measure. Based on the isometry and estimate

$$\mathbb{E}\big\{\big|\int_{\mathbb{R}_*^m\times(0,T]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s)\big|^2\big\} = \mathbb{E}\big\{\int_{\mathbb{R}_*^m\times(0,T]} |f(z,s)|^2\,\nu_c^p(\mathrm{d}z,\mathrm{d}s)\big\},$$

$$\mathbb{E}\big\{\sup_{0\le t\le T}\big|\int_{\mathbb{R}_*^m\times(0,t]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s)\big|^2\big\} \le$$

$$\le 4\,\mathbb{E}\big\{\int_{\mathbb{R}_*^m\times(0,T]} |f(z,s)|^2\,\nu_c^p(\mathrm{d}z,\mathrm{d}s)\big\},$$

for every $T\ge 0$, the stochastic integral is defined in the Hilbert space $\bar{\mathcal{E}}_\nu$, which can be also extended to the localized space $\bar{\mathcal{E}}_{\nu,\mathrm{loc}}$. Therefore, the integral with respect to $\tilde\nu$ when it is not quasi-left continuous is defined by

$$\int_{\mathbb{R}_*^m\times]a.b]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) = \int_{\mathbb{R}_*^m\times]a,b]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s) + \\ + \int_{\mathbb{R}_*^m\times]a.b]} f(z,s)\,\nu_d^p(\mathrm{d}z,\mathrm{d}s), \quad (3.46)$$

where the second term is a pathwise Lebesgue-Stieltjes integral.

Taking the quasi-left continuous part $\tilde\nu_{qc}$, the process

$$f\diamond\tilde\nu_{qc} : t\mapsto \int_{\mathbb{R}_*^m\times(0,t]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s),$$

is a (local) martingale with predictable quadratic variation process

$$\langle f\diamond\tilde\nu_{qc}\rangle = \int_{\mathbb{R}_*^m\times(0,t]} |f(z,s)|^2\,\nu_c^p(\mathrm{d}z,\mathrm{d}s),$$

so that denoting by $\nu_{qc}(f)$ its associate integer-valued measure with (good predictable) compensator $\nu_{qc}^p(f)$ and local-martingale measure $\tilde\nu_{qc}(f)=\nu_{qc}(f)-\nu_{qc}^p(f)$ we have the substitution formula

$$\int_{\mathbb{R}_*^m\times(0,t]} g(z,s)\,\tilde\nu_{qc}^f(\mathrm{d}z,\mathrm{d}s) = \int_{\mathbb{R}_*^m\times(0,t]} g(z,s)\,f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s), \quad (3.47)$$

first for elementary predictable processes g, which is extended by continuity to any integrable processes f and g.

When the stochastic integral is defined for random measures associated to a semi-martingale, i.e., the integer-valued measure ν_M associated with a (cad-lag) local-martingale (or semi-martingale) M is the same as the one associated with its jumps part, $M^d = M - M^c$, i.e., $\nu_M = \nu_{M^c}$, a general form of the stochastic integral takes the form

$$\int_{]a,b]} X(s)\,\mathrm{d}M^c(s) + \int_{\mathbb{R}_*^m\times]a,b]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s) + \int_{\mathbb{R}_*^m\times]a,b]} f(z,s)\,\nu_d^p(\mathrm{d}z,\mathrm{d}s),$$

where the first (stochastic) integral is a continuous local-martingale, the second (stochastic) integral is a purely discontinuous local-martingale and the last term makes sense as a Lebesgue-Stieltjes pathwise integral. Note that integral with respect to ν_c^p or ν^p is part of the stochastic integral with respect to $\tilde{\nu}_{qc}$ or μ, respectively, i.e., if

$$P\{\int_{\mathbb{R}_*^m\times]a.b]} |f(z,s)|\,\nu(\mathrm{d}z,\mathrm{d}s) < \infty\} = 1$$

then we have

$$\begin{aligned}\int_{\mathbb{R}_*^m\times]a.b]} f(z,s)\,\nu(\mathrm{d}z,\mathrm{d}s) &= \\ &= \int_{\mathbb{R}_*^m\times]a,b]} f(z,s)\,\tilde{\nu}(\mathrm{d}z,\mathrm{d}s) + \int_{\mathbb{R}_*^m\times]a.b]} f(z,s)\,\nu^p(\mathrm{d}z,\mathrm{d}s) = \\ &= \int_{]a,b]\times\mathbb{R}_*^m} f(z,s)\,\tilde{\nu}_{qc}(\mathrm{d}z,\mathrm{d}s) + \int_{\mathbb{R}_*^m\times]a.b]} f(z,s)\,\nu_c^p(\mathrm{d}z,\mathrm{d}s),\end{aligned}$$

almost surely. Moreover, any integer-valued measure ν has the form

$$\nu(B\times]a,b]) = \sum_{i=1}^{\infty} \mathbb{1}_{a<\tau_i\le b}\,\mathbb{1}_{\zeta_i\in B}, \quad \forall\, b > a \ge 0,\ B\in\mathcal{B}(\mathbb{R}_*^m),$$

for some sequence $\{\tau_i,\zeta_i : i\ge 1\}$, where the stopping times τ_i cannot be ordered, i.e., it is not necessarily true that $\tau_i \le \tau_{i+1}$, and the $\mathbb{R}_*^m$-valued random variables ζ_i are $\mathcal{F}(\tau_i)$-measurable, but $\nu(\mathbb{R}_*^m\times\{0\}) = 0$ and $\nu(K\times]a,b]) < \infty$ for any $b > a \ge 0$ and any compact subset K of $\mathbb{R}_*^m$. Thus, we expect

$$\int_{\mathbb{R}_*^m\times]a.b]} f(z,s)\nu(\mathrm{d}z,\mathrm{d}s) = \sum_{i=1}^{\infty} \mathbb{1}_{a<\tau_i\le b}\,f(\zeta_i,\tau_i),$$

whenever the above series converges absolutely and f is a continuous process.

To integrate a general predictable process $f = f(s,z,\omega)$, we may proceed as follows: first we separate the integrable jumps (jumps of order 1) from the square integrable jumps (jumps of order 2), namely, first we define

$$f_1(s) = \sum_{i=1}^{\infty} \mathbb{1}_{\tau_i=s}\,f(\zeta_i,\tau_i),$$

whenever sum is absolutely convergent, i.e.,

$$\sum_{i=1}^{\infty} \mathbb{1}_{\tau_i=s}\,|f(\zeta_i,\tau_i)| < \infty,$$

and $f_1(s) = 0$ otherwise. The particular case where $f(z,t,\omega) = 0$ for any z such that $|z| < \varepsilon$, for some $\varepsilon = \varepsilon(\omega) > 0$ is the leading example, since the above series

becomes a finite sum. Recalling that the jump process $t \mapsto \sum_{i=1}^{\infty} \mathbb{1}_{\tau_i \le t} f_1(\tau_i)$ is a cad-lag process, so it has only a finite number of jumps greater than $\varepsilon > 0$ on any bounded time interval $[0, T]$, $T > 0$, we can set, for any $b > a \ge 0$

$$\int_{\mathbb{R}_*^m \times]a.b]} f(z,s)\, \nu^p(\mathrm{d}z, \mathrm{d}s) = \sum_{i=1}^{\infty} \mathbb{1}_{a<\tau_i \le b}\, f_1(\tau_i),$$

as a pathwise integral (defined as a finite sum or a convergent series, for each ω almost surely) with respect to measure ν^p (all locally integrable jumps), and we give a L^2-sense (it cannot be pathwise!) to

$$\int_{\mathbb{R}_*^m \times]a.b]} f(z,s)\, (\nu - \nu^p)(\mathrm{d}z, \mathrm{d}s) = \sum_{i=1}^{\infty} \mathbb{1}_{a<\tau_i \le b}\, [f(\zeta_i, \tau_i) - f_1(\tau_i)],$$

whenever the process

$$t \mapsto \sqrt{\sum_{i=1}^{\infty} \mathbb{1}_{\tau_i \le t}\, [f(\zeta_i, \tau_i) - f_1(\tau_i)]^2}$$

is (locally) integrable. The compensator argument is used to define a measure ν^p, which agrees with ν on predictable processes and such that $\tilde{\nu} = \nu - \nu^p$ is a local-martingale measure. Briefly, for each ω, we make use of a series with indices i such that either $\sum_i |\zeta_i| \wedge 1$ converges or such that the quadratic series $\sum_i |\zeta_i|^2 \wedge 1$ converges to define ν^p. All other indices are ignored.

Here, the martingale theory is used to define the stochastic integral with respect to $\tilde{\nu}$ for any predictable process (class of equivalence) $f(z,s)$ such that the monotone increasing process

$$t \mapsto \Big[\int_{\mathbb{R}_*^m \times]0.t]} |f(z,s)|^2\, \nu^p(\mathrm{d}z, \mathrm{d}s) \Big]^{1/2}$$

is (locally) integrable. Moreover, we can require only that the following process

$$t \mapsto \sqrt{\sum_{i=1}^{\infty} \mathbb{1}_{\tau_i \le t}\, [f(\zeta_i, \tau_i) - f_1(\tau_i)]^2}$$

be (locally) integrable.

For a neat and deep study, the reader may consult Chung and Williams [25], while a comprehensive treatment can be found in Dellacherie and Meyer [32, Chapters V–VIII], Jacod and Shiryaev [84, Chapters 1 and 2]), Rogers and Williams [153, Volume 2]). Also, a more direct approach to stochastic integrals can be found in the book Protter [149], covering even discontinuous martingales.

3.2.4 Vector Valued Integrals

Firstly, recall that any local-martingale M can be written in a unique form as the sum $M_0 + M^c + M^d$, where $M_0 = M(0)$ is a $\mathcal{F}$-measurable random variable, M^c

is a continuous local-martingale (and therefore locally square integrable) and M^d is a purely discontinuous local martingale, both $M^c(0) = M^d(0) = 0$. Also, any local-martingale M with $M(0) = 0$ (in particular a purely discontinuous local-martingale) can be written in a (non unique) form as the sum $M' + M''$, where both M' and M'' are local-martingale, the jumps of M'' are bounded by a constant a (i.e., $|\delta M''| \leq a$ so that M'' is locally square integrable) and M' has locally integrable bounded variation paths. The predictable projection of a local-martingale M is $(M(t-) : t > 0)$ so that a predictable local-martingale is actually continuous. Finally, a continuous or predictable local-martingale with locally bounded variation paths is necessarily a constant.

Secondly, recall the definitions of the predictable and the optional quadratic variation processes. Given real-valued local square integrable martingale M the predictable (increasing) quadratic variation process $t \mapsto \langle M \rangle(t)$ obtained via the Doob-Meyer decomposition Theorem 2.7 applied to $t \mapsto M^2(t)$ as a local sub-martingale of class (D). This is the only increasing predictable locally integrable process $\langle M \rangle$ such that $M^2 - \langle M \rangle$ is a martingale. However, the predictable quadratic variation process is generally used for continuous local martingales. For a real-valued (non necessarily continuous) local (non necessarily square integrable) martingale M, the optional (increasing) quadratic variation process $t \mapsto [M](t)$ is defined as $\langle M \rangle(t) + \sum_{s \leq t} |M(s) - M(s-)|^2$. This is the only increasing optional process (not necessarily locally integrable) $[M]$ such that $M^2 - [M]$ is a local-martingale and $\delta[M] = (\delta M)^2$. The increasing optional process $\sqrt{[M]}$ is locally integrable, and if $[M]$ is locally integrable then it is a local sub-martingale of class (D) and again via the Doob-Meyer decomposition we obtain a predictable increasing locally integrable $\langle M \rangle$ (called the compensator of $[M]$), which agrees with the predictable quadratic variation process previously defined for local square integrable martingales. Therefore, the predictable quadratic variation process $\langle M \rangle$ may not be defined for a discontinuous local-martingale, but the optional quadratic variation $[M]$ is always defined. The concept of integer-valued random measures is useful to interpret $[M]$ as the increasing process associated with the integer-valued measure ν_M derived from M. Thus $\langle M \rangle$ is the increasing predictable process (not necessarily integrable) associated with the predictable compensator ν_M^p of ν_M. If M is quasi-left continuous then $\langle M \rangle$ is continuous, and therefore locally integrable. Next, for any two real-valued local-martingale M and N the predictable and optional quadratic co-variation processes are defined by the formula $4\langle M, N \rangle = \langle M+N \rangle - \langle M-N \rangle$ and $4[M, N] = [M + N] - [M - N]$. Note that

$$E\Big\{ \int_{]a,b]} f(t) \mathrm{d}\langle M, N \rangle(t) \Big\} = E\Big\{ \int_{]a,b]} f(t) \mathrm{d}[M, N](t) \Big\},$$

for every predictable process such that the above integrals are defined.

An important role is played by the *Kunita-Watanabe inequality*, namely for any two real-valued local-martingales M and N and any two (extended) real-

valued measurable processes α and β we have the inequality

$$\int_0^t |\alpha(s)|\,|\beta(s)|\,|\mathrm{d}[M,N](s)| \le \sqrt{\int_0^t |\alpha(s)|\mathrm{d}[M](s)} \times \\ \times \sqrt{\int_0^t |\beta(s)|\mathrm{d}[N](s)}, \quad (3.48)$$

almost surely for every $t > 0$, where $|\mathrm{d}[M,N]|$ denotes the total variation of the signed measure $\mathrm{d}[M,N]$. Certainly, the same estimate is valid for the predictable quadratic co-variation process $\langle M,M\rangle$ instead of optional process $[M,N]$. The argument to prove estimate (3.48) is as follow. Since $[M+rN, M+rN] = [M] - 2r[M,N] + r^2[N]$ is an increasing process for every r, we deduce $(\mathrm{d}[M,N])^2 \le \mathrm{d}[M]\,\mathrm{d}[N]$. Next, Cauchy-Schwarz inequality yields (3.48) with $\mathrm{d}[M,N](s)$ instead of $|\mathrm{d}[M,N](s)|$. Finally, by means of the Radon-Nikodym derivative, i.e., replacing α by $\alpha = (\mathrm{d}[M,N]/|\mathrm{d}[M,N](s)|)\,\alpha$, we conclude. For instance, a full proof can be found in Durrett [40, Section 2.5, pp. 59–63] or Revuz and Yor [151, Proposition 1.15, Chapter, pp. 126–127].

Let $M = (M_1,\dots,M_d)$ a d-dimensional continuous local-martingale in a filtered space $(\Omega,\mathcal{F},P,\mathcal{F}(t) : t \ge 0)$, i.e., each component $(M_i(t) : t \ge 0)$, $i = 1,\dots,d$, is a local continuous martingale in $(\Omega,\mathcal{F},P,\mathcal{F}(t) : t \ge 0)$. Recall that the predictable quadratic co-variation $\langle M\rangle = (\langle M_i, M_j\rangle : i,j = 1,\dots,d)$ is a symmetric nonnegative matrix valued process. The stochastic integral with respect to M is defined for a d-dimensional progressively measurable process $f = (f_1,\dots,f_d)$ if for some increasing sequence of stopping times $\{\tau_n : n \ge 1\}$ with $\tau_n \to \infty$ we have

$$\mathbb{E}\Big\{\int_0^{\tau_n} \sum_{i,j=1}^d f_i(s)f_j(s)\mathrm{d}\langle M_i,M_j\rangle(s)\Big\} < \infty. \quad (3.49)$$

Based on (3.48), it is clear that, if each component f_i is locally square integrable with respect to $\langle M_i\rangle$, i.e.,

$$\mathbb{E}\Big\{\int_0^{\tau_n} |f_i(s)|^2\mathrm{d}\langle M_i\rangle(s)\Big\} < \infty,$$

then the above condition (3.49) is satisfied. However, the converse may be false, e.g., if $w = (w_1,w_2)$ is a two-dimensional standard Wiener process then set $M_1 = w_1$, $M_2 = kw_1 + (1-k)w_2$, where k is a $(0,1)$-valued predictable process. Choosing $f = (f_1,f_2) = (-\frac{k}{1-k}, \frac{1}{1-k})$, we have $\sum_{i,j} f_i f_j \mathrm{d}\langle M_i,M_j\rangle = \mathrm{d}\ell$, the Lebesgue measure, so we certainly have (3.49), but

$$\int_0^t |f_1(s)|^2\mathrm{d}\langle M_1\rangle(s) = \int_0^t \Big|\frac{k(s)}{1-k(s)}\Big|^2 \mathrm{d}s < \infty \quad \text{a.s. } \forall t > 0,$$

may not be satisfied.

For a n-dimensional *continuous* local-martingale $M = (M_1, \dots, M_n)$ and an adapted $\mathbb{R}^n$-valued (measurable) process $f = (f_1, \dots, f_n)$ we have the following estimate: for every $p > 0$ there exists a positive constant $C = C_p$ depending only on p, such that

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} \Big[\sum_{i=1}^n \Big| \int_0^t f_i(s) \mathrm{d}M_i(s) \Big|^2 \Big]^{p/2} \Big\} \le \\ \le C\, \mathbb{E}\Big\{ \Big[\sum_{i,j=1}^n \int_0^T f_i(s)\, f_j(s) \mathrm{d}\langle M_i(s), M_j(s) \rangle \Big]^{p/2} \Big\}. \quad (3.50)$$

for any stopping time T. In particular, for a standard n-dimensional Wiener process $(w(t) : t \ge 0)$, we can write

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} \Big| \int_0^t f(s) \mathrm{d}w(s) \Big|^p \Big\} \le C\, \mathbb{E}\Big\{ \Big[\int_0^T |f(s)|^2 \mathrm{d}s \Big]^{p/2} \Big\}. \quad (3.51)$$

This estimate follows from Burkhölder-Davis-Gundy inequality (as in (2.9) of Chapter 3), e.g., see Karatzas and Shreve [91, Section 3.3.D, pp. 163–167]. Note that we make take $C_1 = 3$ and $C_2 = 4$.

Regarding the stochastic integral with respect to a Poisson measure in the Polish space $\mathbb{R}^m_*$ (or more general in a Blackwell space), we should mention that the key elements are the compensated local martingale measure $\tilde\nu$ and the compensator ν^p, which is a predictable random measure. Both are constructed from an integer-valued random measure, which is naturally obtained from a optional locally integrable jump process or better a (purely discontinuous) local (sub-)martingale. A posteriori, we may use a predictable real-valued process $\gamma(z,t)$ on $\mathbb{R}^m_* \times (0,\infty)$ such that

$$P\Big\{ \int_{]0,t]} \mathrm{d}s \int_{\mathbb{R}^m_*} \gamma^2(z,s)\, \pi(\mathrm{d}z) < \infty \Big\} = 1, \quad \forall t > 0,$$

and use the stochastic integral to define a local-martingale measure

$$\tilde\nu_\gamma(B\times]a,b]) = \int_{\mathbb{R}^m_*\times]a,b]} \mathbb{1}_B\, \gamma(z,s)\, \tilde\nu(\mathrm{d}z, \mathrm{d}s),$$

for every B in $\mathcal{B}(\mathbb{R}^m_*)$ and $b > a \ge 0$, with a predictable quadratic variation (or *compensator*) given by

$$\nu^p_\gamma(B\times]a,b]) = \int_{\mathbb{R}^m_*\times]a,b]} \mathbb{1}_B\, \gamma^2(z,s)\, \nu^p(\mathrm{d}z, \mathrm{d}s),$$

for every B in $\mathcal{B}(\mathbb{R}^m_*)$, $b > a \ge 0$, and for the case of the Poisson measure $\nu^p(\mathrm{d}z, \mathrm{d}s) = \pi(\mathrm{d}z)\, \mathrm{d}s$. Thus ν^p_γ has a density $\delta = \gamma^2$ with respect to ν^p.

The estimate (3.50) is also valid for a Poisson integral, with a small correction, namely, for any p in $(0,2]$ there exists a positive constant $C = C_p$

(actually $C_p = (4-p)/(2-p)$ if $0 < p < 2$ and $C_2 = 4$) such that for any adapted (measurable) process $f(\zeta, s)$ (actually, the predictable version is used) we have

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} \Big| \int_{\mathbb{R}^m_*\times]0,t]} f(\zeta,s)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s)\Big|^p\Big\} \le \\ \le C\,\mathbb{E}\Big\{\Big[\int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |f(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big]^{p/2}\Big\}, \quad (3.52)$$

for every stopping time T. This follows immediately from estimate (2.8) of Chapter 3. The case $p > 2$ is a little more complicate and involves Itô formula as discussed in the next section.

For the sake of simplicity and to recall the fact that stochastic integral are defined in an L^2-sense, instead of using the natural notation $\bar{\mathcal{E}}_{M,\mathrm{loc}}$, $\bar{\mathcal{E}}_M$, $\bar{\mathcal{E}}_{\pi,\mathrm{loc}}$, $\bar{\mathcal{E}}_\pi$, $\bar{\mathcal{E}}_{\mathrm{loc}}$, $\bar{\mathcal{E}}$ of this Section 3.2 we adopt the following

Definition 3.7 (L^2-Integrand Space)**. (a)** Given a d-dimensional continuous square integrable martingale M with predictable quadratic variation process $\langle M\rangle$ in a filtered space $(\Omega,\mathcal{F},P,\mathcal{F}(t):t\ge 0)$, we denote by $L^2(M)$ or in long $L^2(\Omega,\mathcal{F},P,\mathcal{F}(t):t\ge 0,M,\langle M\rangle)$, the equivalence class with respect to the completion of product measure $P\times\langle M\rangle$ of $\mathbb{R}^d$-valued square integrable predictable processes X, i.e. (3.49) with $\tau_n=\infty$. This is regarded as a closed subspace of the Hilbert space $L^2([0,\infty)\times\Omega,\bar{\mathcal{P}},\langle M\rangle\times P)$, where $\bar{\mathcal{P}}$ is the $\langle M\rangle\times P$-completion of the predictable σ-algebra $\mathcal{P}$ as discussed at the beginning of this chapter.
(b) Given a $\mathbb{R}^m$-valued quasi-left continuous square integrable martingale M with integer-valued measure ν_M and compensated martingale random measure $\tilde\nu_M$ in the filtered space $(\Omega,\mathcal{F},P,\mathcal{F}(t):t\ge 0)$, we denote by $L^2(\tilde\nu_M)$ or $L^2(\Omega,\mathcal{F},P,\mathcal{F}(t):t\ge 0,M,\tilde\nu_M)$ the equivalence class with respect to the completion of product measure $\tilde\nu_M\times P$ of real-valued square integrable predictable processes X, i.e., as a closed subspace of the Hilbert space $L^2(\mathbb{R}^m_*\times[0,\infty)\times\Omega,\overline{\mathcal{B}(\mathbb{R}^m_*)\times\mathcal{P}},\tilde\nu_M\times P)$, where $\mathcal{B}(\mathbb{R}^m_*)$ is the Borel σ-algebra in $\mathbb{R}^m_*$ and the *bar* means completion with respect to the product measure $\tilde\nu_M\times P$. If an integer-valued random measure ν is initially given with compensated martingale random measure $\tilde\nu=\nu-\nu^p$, where ν^p is the predictable compensator satisfying $\nu^p(\mathbb{R}^m_*\times\{t\})=0$ for every $t\ge 0$, then we use the notation $L^2(\tilde\nu)$ or $L^2(\Omega,\mathcal{F},P,\mathcal{F}(t):t\ge 0,\tilde\nu_M)$. Moreover, the same applies if a predictable ν^p-locally integrable density δ is used, i.e., if $\tilde\nu$ and ν^p are replaced by $\tilde\nu_\delta=\sqrt{\delta}\,\tilde\nu$ and $\nu^p_\delta=\delta\,\tilde\nu$.
(c) Similarly, localized Hilbert spaces $L^2_{\mathrm{loc}}(\Omega,\mathcal{F},P,\mathcal{F}(t):t\ge 0,M,\langle M\rangle)$ or $L^2_{\mathrm{loc}}(M)$ and $L^2_{\mathrm{loc}}(\Omega,\mathcal{F},P,\mathcal{F}(t):t\ge 0,M,\tilde\nu_M)$ or $L^2_{\mathrm{loc}}(\tilde\nu_M)$ are defined. If M is only a local continuous martingale then X in $L^2_{\mathrm{loc}}(M)$ means that for some localizing sequence $\{\tau_n:n\ge 1\}$ the process $M_n:t\mapsto M(t\wedge\tau_n)$ is a square integrable martingale and $\mathbb{1}_{]0,\tau_n]}X$ belongs to $L^2(M_n)$, i.e, (3.49) holds for every $n\ge 1$. Similarly, if M is only a local quasi-left continuous square integrable martingale then X in $L^2_{\mathrm{loc}}(\tilde\nu_M)$ means that for some localizing sequence $\{\tau_n:n\ge 1\}$ the process $M_n:t\mapsto M(t\wedge\tau_n)$ is a square integrable martingale, with compensated

martingale random measure denoted by $\tilde{\nu}_{M_n}$, and $\mathbb{1}_{]0,\tau_n]}X$ belongs to $L^2(\tilde{\nu}_{M_n})$, i.e., the M and X share the same localizing sequence of stopping times. □

Note that we do not include the general case where M is a semi-martingale (in particular, local-martingales which are neither quasi-left continuous nor local square integrable), since the passage to include these situation is essentially a pathwise argument covered by the measure theory. If the predictable quadratic variation process $\langle M\rangle$ gives a measure equivalent to the Lebesgue measure $\mathrm{d}\ell$ then the spaces $L^2(M)$ and $L^2_{\mathrm{loc}}(M)$ are equals to $\mathbb{P}^p(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$ and $\mathbb{P}^p_{\mathrm{loc}}(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$, for $p=2$, as defined at the beginning of this Section 3.2 in the one-dimensional case. If M is a (local) quasi-left continuous square integrable martingale then we can write (uniquely) $M=M^c+M^d$, where M^c is the continuous part and M^d the purely discontinuous part with $M^d(0)=0$. Then, we may write $L^2_{\mathrm{loc}}(M^d)=L^2_{\mathrm{loc}}(\tilde{\nu}_{M^d})$, $L^2_{\mathrm{loc}}(M)=L^2_{\mathrm{loc}}(M^c)+L^2_{\mathrm{loc}}(M^d)$, and similarly without the localization. Furthermore, if predictable quadratic co-variation (matrix) process $\langle M\rangle$ or the predictable compensator ν^p is deterministic then the (local) space $L^2_{\mathrm{loc}}(M)$ or $L^2_{\mathrm{loc}}(\tilde{\nu})$ is characterized by the condition

$$P\Big\{\int_{]0,t]}\sum_{i,j=1}^{d} f_i(s)f_j(s)\mathrm{d}\langle M_i,M_j\rangle(s)<\infty\Big\}=1$$

or

$$P\Big\{\int_{\mathbb{R}^m_*\times]0,t]} f^2(z,s)\nu^p(\mathrm{d}z,\mathrm{d}s)<\infty\Big\}=1,$$

for every $t>0$. This applies even if the local-martingale M or the integer-valued random measure ν is not quasi-left continuous, in which case the predictable quadratic co-variation process $\langle M_i,M_j\rangle(s)$ may be discontinuous or the predictable compensator measure ν^p may not vanish on $\mathbb{R}^m_*\times\{t\}$ for some $t>0$, we must have $\nu^p(\mathbb{R}^m_*\times\{0\})=0$.

The Case of Semi-martingales

Another point to stress is the following fact. If M is a n-dimensional continuous local-martingale and f is a $d\times n$ matrix-valued process in $L^2_{\mathrm{loc}}(M)$, i.e., each columns vector $f_{i\cdot}=(f_{ik}:k=1,\ldots,n)$ belongs to $L^2_{\mathrm{loc}}(M)$, then we can define d-dimensional continuous local-martingale

$$(f\star M)_i(t)=\sum_{k=1}^{n}\int_0^t f_{ik}(s)\mathrm{d}M_k(s),\quad \forall t\geq 0,$$

and $i=1,\ldots,d$. The predictable quadratic co-variation process becomes

$$\langle (f\star M)_i,(f\star M)_j\rangle=\sum_{k,\ell=1}^{n} f_{ik}\langle M_k,M_\ell\rangle f_{j\ell}\,.$$

On the other hand, if $\tilde{\nu}$ is a local-martingale measure with a predictable compensator ν^p in $\mathbb{R}_*^m$ and g is a d vector-valued process in $L^2_{\rm loc}(\tilde{\nu})$, i.e., each component g_i belongs to $L^2_{\rm loc}(\tilde{\nu})$, then we can define d-dimensional purely discontinuous local martingale

$$(g \star \tilde{\nu})_i(t) = \int_{]0,t]} g_i(\zeta, s)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) \quad \forall t \geq 0,$$

and $i = 1, \ldots, d$. The local-martingale measure $\tilde{\nu}_{(g\star\tilde{\nu})}$ associated with $g \star \tilde{\nu}$ in $\mathcal{B}(\mathbb{R}_*^d)$ can be expressed as

$$\tilde{\nu}_{(g\star\tilde{\nu})}(B\times]a,b]) = \int_{\mathbb{R}_*^m\times]a,b]} \mathbb{1}_{\{g(\zeta,s)\in B\}}\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s)$$

with its predictable compensator $\nu^p_{(g\star\tilde{\nu})}$

$$\tilde{\nu}^p_{(g\star\tilde{\nu})}(B\times]a,b]) = \int_{\mathbb{R}_*^m\times]a,b]} \mathbb{1}_{\{g(\zeta,s)\in B\}}\nu^p(\mathrm{d}\zeta, \mathrm{d}s),$$

for every $b > a \geq 0$ and B in $\mathcal{B}(\mathbb{R}_*^d)$. In short we write $\tilde{\nu}_{(g\star\tilde{\nu})} = g\tilde{\nu}$ and $\tilde{\nu}^p_{(g\star\tilde{\nu})} = g\tilde{\nu}^p$. Note that the optional quadratic variation process is given by

$$[(g \star \tilde{\nu})_i, (g \star \tilde{\nu})_j](t) = \int_{\mathbb{R}_*^m\times]0,t]} g_i(\zeta, s)g_j(\zeta, s)\nu^p(\mathrm{d}\zeta, \mathrm{d}s),$$

for every $t \geq 0$.

Let $g(z, s)$ be a d-dimensional predictable process which is integrable in $\mathbb{R}_*^m$ with respect to the measure $\nu^p(\mathrm{d}z, \mathrm{d}s)$ almost surely, i.e.,

$$P\Big\{ \int_{\mathbb{R}_*^m\times]0,t]} |g(z, s)|\nu^p(\mathrm{d}z, \mathrm{d}s) < \infty\Big\} = 1, \quad \forall t > 0,$$

which is a classic pointwise integral in the Lebesgue-Stieltjes. Moreover, if $\{(\zeta_n, \tau_n) : n = 1, 2, \ldots\}$ are the atoms of ν (i.e., its associated point process) then

$$\mathbb{E}\Big\{ \int_{\mathbb{R}_*^m\times]0,t]} |g(z, s)|\nu^p(\mathrm{d}z, \mathrm{d}s)\Big\} = \\ = \mathbb{E}\Big\{ \int_{\mathbb{R}_*^m\times]0,t]} |g(z, s)|\nu(\mathrm{d}z, \mathrm{d}s)\Big\} = \mathbb{E}\Big\{ \sum_{0<\tau_n\leq t} |g(\zeta_n, \tau_n)|\Big\}.$$

Since

$$\sum_{0<\tau_n\leq t} |g(\zeta_n, \tau_n)|^2 \leq \max_{0<\tau_n\leq t} |g(\zeta_n, \tau_n)| \sum_{0<\tau_n\leq t} |g(\zeta_n, \tau_n)|,$$

the process $g(z,s)$ also belongs to $L^2_{\rm loc}(\tilde\nu)$ and we have

$$\int_{\mathbb{R}^m_*\times]0,t]} g(z,s)\nu(\mathrm{d}z,\mathrm{d}s) = \int_{\mathbb{R}^m_*\times]0,t]} g(z,s)\tilde\nu(\mathrm{d}z,\mathrm{d}s) + \\ + \int_{\mathbb{R}^m_*\times]0,t]} g(z,s)\nu^p(\mathrm{d}z,\mathrm{d}s),$$

for every $t>0$.

When comparing both stochastic integrals, with respect to (1) a continuous local-martingale (typically a Wiener process) and (2) a quasi-left continuous (cad-lag) purely jump local-martingale (typically a Poisson compensated-jump or martingale measure) we have two notations, which are different only in form. If $w=(w_k(t):t\geq 0,\,k\geq 1)$ is a (standard) Wiener process and $\sigma=(\sigma_k(s):s\geq 0,\,k\geq 1)$ is a adapted process then

$$(\sigma\star w)_t = \sum_k \int_0^t \sigma_k(s)\mathrm{d}w_k(s)$$

makes sense as long as

$$\sum_k \int_0^t |\sigma_k(s)|^2\mathrm{d}s<\infty,\quad \forall t\geq 0,$$

almost surely. On the other hand, if $\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s)$ is a (standard) Poisson martingale measure with Lévy measure and $\gamma=(\gamma(\zeta,s):s\geq 0,\,\zeta\in\mathbb{R}^m_*)$ is a adapted process then

$$(\gamma\star\tilde\nu)_t = \int_{\mathbb{R}^m_*\times]0,t]} \gamma(\zeta,s)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s)$$

makes sense as long as

$$\int_0^t \mathrm{d}s\int_{\mathbb{R}^m_*} |\gamma(\zeta,s)|^2\pi(\mathrm{d}\zeta)<\infty,\quad \forall t\geq 0,$$

almost surely. At this point it is clear the role of the parameters k and ζ in the integrands $\sigma_k(\cdot)$ and $\gamma(\zeta,\cdot)$, i.e., the sum in k and the integral in ζ with respect to the Lévy measure $m(\cdot)$. Moreover, the integrands σ and γ can be considered as ℓ^2-valued processes, i.e.,

$$\sum_k |\sigma_k|^2<\infty \quad\text{and}\quad \int_{\mathbb{R}^m_*} |\gamma(\zeta)|^2\pi(\mathrm{d}\zeta)<\infty,$$

so that the parameters k and ζ play similar roles. The summation in k can be converted to an integral and the separable locally compact and locally convex space $\mathbb{R}^m_*$ can be replaced by any Polish (or Backwell) space.

In general, if the (local) martingale measure $\tilde{\nu}$ is known then the Lévy measure $\pi(\cdot)$ is found as its predictable quadratic variation, and therefore ν is constructed as the integer measure associated with the compensated-jump process

$$\tilde{p}(t) = \int_{\mathbb{R}^m_* \times]0,t]} \zeta \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall t \geq 0.$$

Hence, the integer measure ν, the (local) martingale measure $\tilde{\nu}$ and the $\mathbb{R}^m$-valued compensated-jump process $\tilde{p}$ can be regarded as different viewpoints of the same concept. Each one of them completely identifies the others.

To conclude this section we mention that any quasi-left continuous (cadlag) semi-martingale X can be expressed in a unique way as $X(t) = X(0) + A(t) + M(t) + z \star \tilde{\nu}_X$, where $X(0)$ is a $\mathcal{F}(0)$-measurable random variable, $A(0) = M(0) = 0$, A is a continuous process with locally integrable bounded variation paths, M is a continuous local-martingale, and $z \star \tilde{\nu}_X$ is the stochastic integral of the process $(z, t, \omega) \mapsto z$ with respect to the local-martingale measure $\tilde{\nu}_X$ associated with X.

Integral Relative to Lévy Processes

As mentioned early, a Wiener process in $\mathbb{R}^d$ has a canonical construction from a its characteristics (i.e., its variance, a symmetric positive definite matrices, and its drift or mean, a vector). Similarly, a Poisson measure on $\mathbb{R}^m_*$ has a canonical construction from its characteristics (i.e., its Lévy –rather intensity– measure $\mathbb{R}^m_*$), and a Poisson Martingale measure is obtained. The stochastic integrals with respect to a Wiener or a Martingale (Poisson) measure are not pathwise, since pathwise integrals are with respect to processes having a local bounded variation.

Also, a Lévy process in $\mathbb{R}^d$ has a canonical construction from its characteristics (i.e., its Lévy measure in $\mathbb{R}^d_*$ and the part corresponding to its 'Wiener' part in $\mathbb{R}^d$). If $m(\mathrm{d}x)$ denotes the given Lévy measure in $\mathbb{R}^d_*$, i.e., $m(\mathrm{d}x)$ integrate the function $x \mapsto 1 \wedge |x|^2$, then the jumps of the Lévy process $\ell(t)$ define its point process $p(t) = \ell(t) - \ell(t-)$ in $\mathbb{R}^d_*$, its Poisson (random) measure $p(\mathrm{d}x, \mathrm{d}t) = \sum_{s \in \mathrm{d}t} \mathbb{1}_{\ell(t)-\ell(t-) \in \mathrm{d}x}$ on $\mathbb{R}^d_*$, and its Poisson Martingale measure (or compensated Poisson random measure) $\tilde{p}(\mathrm{d}x, \mathrm{d}t) = p(\mathrm{d}x, \mathrm{d}t) - m(\mathrm{d}x)\,\mathrm{d}t$ on $\mathbb{R}^d_*$. In the previous sections, the stochastic integral with respect to a Martingale measure was established, i.e.,

$$\int_{\{x \in \mathbb{R}^d_* : |x| < 1\} \times]0,t]} f(x,s)|x|^2 \tilde{p}(\mathrm{d}x, \mathrm{d}s)$$

is well defined for any predictable and bounded function f, and also, the (regular) theory of measure/integration can be applied to

$$\int_{\{x \in \mathbb{R}^d_* : |x| \geq 1\} \times]0,t]} g(x,s) p(\mathrm{d}x, \mathrm{d}s) = \sum_{0 < s \leq t} g(p(s), s) \mathbb{1}_{|p(s)| \geq 1},$$

which is a finite sum (depending on ω) and well defined for any predictable and bounded function g, actually an integral with respect to a compounded Poisson process (very similar to an integral with respect to a Poisson process). This is probably why the stochastic integrals are defined with respect to a Wiener process and a Poisson Martingale measure, instead of directly with respect to a Lévy process (regarded as a semi-martingale). For example, the reader should check Applebaum [1], Bichteler [11], Ishikawa [76], Protter [149], among others, to have a wider point of view.

3.3 Stochastic Differential

One of the most important tools used with stochastic integrals is the *change-of-variable rule* or better known as *Itô's formula*. This provides an integral-differential calculus for the sample paths.

To motivate our discussion, let us recall that at the end of Subsection 3.2.1 we established the identity

$$\int_{(0,t]} w(s)\mathrm{d}w(s) = \frac{w^2(t)}{2} - \frac{t}{2}, \quad \forall t \geq 0,$$

for a real-valued standard Wiener process $(w(t) : t \geq 0)$, where the presence of new term, $t/2$, is noted, with respect to the classic calculus.

In general, Fubini's theorem proves that given two processes X and Y of locally bounded variation (cad-lag) we have the integration-by-part formula

$$X(b)Y(b) - X(a)Y(a) = \int_{(a,b]} X(t-)\mathrm{d}Y(t) + \\ + \int_{(a,b]} Y(t-)\mathrm{d}X(t) + \sum_{a<t\leq b} \delta X(t)\,\delta Y(t), \quad (3.53)$$

where $X(t-)$ and $Y(t-)$ are the left-limits at t and δ is the jump-operator, e.g., $\delta X(t) = X(t) - X(t-)$. Since the integrand $Y(t-)$ is left-continuous and the integrator $X(t)$ is right-continuous as above, the pathwise integral can be interpreted in the Riemann-Stieltjes sense or the Lebesgue-Stieltjes sense, indistinctly. Consider, for example, a Poisson process with parameter $c > 0$, i.e., $X = Y = (p(t) : t \geq 0)$, we have

$$\int_{(0,t]} p(s-)\mathrm{d}p(s) = \frac{p^2(t)}{2} - \frac{p(t)}{2}, \quad \forall t \geq 0,$$

because all jumps are equals to 1. However, strictly in the Lebesgue-Stieltjes sense we write

$$\int_{(0,t]} p(s)\mathrm{d}p(s) = \frac{p^2(t)}{2} + \frac{p(t)}{2}, \quad \forall t \geq 0.$$

Recall that the stochastic integral is initially defined as the L^2-limit of Riemann-Stieltjes sums, where the integrand is a predictable (essentially, left-continuous having right-limits) process and the integrator is a (local) square integrable martingale. The (local) bounded variation integral can be defined by either way with a unique value, as long as the integrand is the predictable member of its equivalence class of processes. Thus, as mentioned at the end of Subsection 3.2.3, the stochastic integral with respect to the compensated Poisson process (or martingale) $\bar{p}(t) = p(t) - ct$ satisfies

$$\int_{(0,t]} \bar{p}(s)\mathrm{d}\bar{p}(s) = \int_{(0,t]} \bar{p}(s-)\mathrm{d}\bar{p}(s), \quad \forall t \geq 0,$$

the expression in left-hand side is strictly understood only as a stochastic integral, because it makes non sense as a pathwise Riemann-Stieltjes integral and does not agree with one in the pathwise Lebesgue-Stieltjes sense. However, the expression in right-hand side can be interpreted either as a pathwise Riemann-Stieltjes integral or as a stochastic integral. Note that the processes $(p(t) : t \geq 0)$ and $(p(t-) : t \geq 0)$ belong to the same equivalence class for the $\mathrm{d}t \times P(\mathrm{d}\omega)$ measure, under which the stochastic integral is defined.

We may calculate the stochastic integral as follows. For a given partition $\pi = (0 = t_0 < t_1 < \cdots < t_n = t)$ of $[0,t]$, with $\|\pi\| = \max_i(t_i - t_{i-1})$, consider the Riemann-Stieltjes sums

$$S_\pi = \sum_{i=1}^{n} \bar{p}(t_{i-1})[\bar{p}(t_i) - \bar{p}(t_{i-1})] = \int_{]0,t]} \bar{p}_\pi(s)\mathrm{d}\bar{p}(s) =$$

$$= \int_{]0,t]} \bar{p}_\pi(s)\mathrm{d}p(s) - c\int_0^t \bar{p}_\pi(s)\mathrm{d}s,$$

for the predictable process $\bar{p}_\pi(s) = \bar{p}(t_{i-1})$ for any s in $]t_{i-1}, t_i]$. Since $\bar{p}_\pi(s) \to \bar{p}(s-)$ as $\|\pi\| \to 0$, we obtain

$$\int_{(0,t]} \bar{p}(s-)\mathrm{d}\bar{p}(s) = \int_{(0,t]} \bar{p}(s-)\mathrm{d}p(s) - c\int_0^t \bar{p}(s-)\mathrm{d}s,$$

which is a martingale null at time zero. For instance, because $\mathbb{E}\{p(t)\} = ct$ and $\mathbb{E}\{[p(t) - ct]^2\} = ct$ we have $\mathbb{E}\{p^2(t)\} = c^2t^2 + ct$, and therefore

$$\mathbb{E}\Big\{\int_{(0,t]} p(s-)\mathrm{d}p(s) - c\int_0^t p(s-)\mathrm{d}s\Big\} = 0,$$

as expected.

Given a smooth real-valued function $\varphi = \varphi(t,x)$ defined on $[0,T] \times \mathbb{R}^d$ and a $\mathbb{R}^d$-valued semi-martingale $\{M(t) : t \geq 0\}$ we want to discuss the stochastic chain-rule for the real-valued process $\{\varphi(t, M(t)) : t \geq 0\}$. If φ is complex-valued then we can tread independently the real and the imaginary parts.

For a real-valued Wiener process $(w(t) : t \geq 0)$, we have deduced that

$$w^2(t) = 2\int_{(0,t]} w(s)\mathrm{d}w(s) + t, \quad \forall t \geq 0,$$

so that the standard chain-rule does not apply. This is also seen when Taylor formula is used, say taking mathematical expectation in

$$\varphi(w(t)) = \varphi(0) + \varphi'(0)w(t) + \varphi''(0)\frac{w^2(t)}{2} + \int_0^1 \varphi'''(sw(t))\frac{w^3(t)}{6}\mathrm{d}s,$$

we obtain

$$\mathbb{E}\varphi(w(t)) = \varphi(0) + \varphi''(0)\frac{t}{2} + \int_0^1 \mathbb{E}\{\varphi'''(sw(t))\frac{w^3(t)}{6}\}\mathrm{d}s,$$

where the error-term integral can be bounded by $2t^{3/2}\sup|\varphi|$. The second order derivative produces a term of order 1 in t.

Given a (cad-lag) locally integrable bounded variation process $A = (A(t) : t \geq 0)$ and a locally integrable process $X = (X(t) : t \geq 0)$ with respect to A, we can define the pathwise Lebesgue-Stieltjes integral

$$(X \star A)(t) = \int_{]0,t]} X(s)\mathrm{d}A(s), \quad \forall\, t \geq 0,$$

which produces a new (cad-lag) locally integrable bounded variation process $X \star A = ((X \star A)(t) : t \geq 0)$. The substitution formula establishes that for any locally integrable process $Y = (Y(t) : t \geq 0)$ with respect to $X \star A$, the process $YX = (Y(t)X(t) : t \geq 0)$ is locally integrable process with respect to A and

$$\int_{]0,t]} Y(s)\mathrm{d}(X \star A)(s) = \int_{]0,t]} Y(s)X(s)\mathrm{d}A(s), \tag{3.54}$$

for every $t \geq 0$. Certainly, if the processes X and Y are left-continuous then the above integral can be interpreted in the (pathwise) Riemann-Stieltjes sense. Moreover, if both processes X and Y are predictable and A is adapted then the $\star$ symbol, representing the pathwise Lebesgue-Stieltjes, can be replaced by the $\diamond$ symbol, representing the stochastic integral relative to an adapted (cad-lag) process with locally integrable bounded variation.

Similarly, given a (cad-lag) local-martingale $M = (M(t) : t \geq 0)$ and a locally integrable predictable process $X = (X(t) : t \geq 0)$ relative to M (i.e., there is a reducing sequence of stopping times $(\tau_n : n \geq 0)$ for both processes X and M, simultaneously), we can define the stochastic integral which produces a new (cad-lag) local-martingale $X \diamond M = ((X \diamond M)(t) : t \geq 0)$. Let $Y = (Y(t) : t \geq 0)$ be a locally integrable predictable process relative to $X \diamond M$ (i.e., there is another reducing sequence of stopping times $(\bar{\tau}_n : n \geq 0)$ for both processes Y and $X \diamond M$). The stochastic substitution formula says that the predictable process

$YX = (Y(t)X(t) : t \geq 0)$ is locally integrable with respect to M admitting the (minimum) reducing sequence $(\tau_n \wedge \bar{\tau}_n : n \geq 0)$ and

$$\int_{]0,t]} Y(s)\mathrm{d}(X \diamond M)(s) = \int_{]0,t]} Y(s)X(s)\mathrm{d}M(s), \tag{3.55}$$

for every $t \geq 0$.

The first step in the proof of the above stochastic substitution formula is to observe that by taking the minimum localizing sequence $(\tau_n \wedge \bar{\tau}_n : n \geq 0)$ it suffices to show the result for an L^2-martingales M. Secondly, it is clear that equality (3.55) holds for any elementary predictable processes Y and that because of the isometry

$$\int_{]0,t]} Y^2(s)\mathrm{d}[X \diamond M](s) = \int_{]0,t]} Y^2(s)X^2(s)\mathrm{d}[M](s), \quad \forall\, t \geq 0,$$

for every $t \geq 0$, where $[\cdot]$ denotes the (optimal) quadratic variation of a martingale (as in Section 3.2.3), the process YX is integrable with respect to M. Finally, by passing to the limit we deduce that (3.55) remains valid almost surely for every $t \geq 0$. Since both sides of the equal sign are cad-lag processes, we conclude. A detailed proof can be found in Chung and Williams [25, Theorem 2.12, Section 2.7, pp. 48–49].

Let M be a (real-valued) square integrable martingale with its associated optional and predictable integrable monotone increasing processes $[M]$ and $\langle M\rangle$. Recall that $M^2 - [M]$ and $M^2 - \langle M\rangle$ are uniformly integrable martingale, $[M](t) = \langle M_c\rangle(t) + \sum_{s\leq t}[M(s) - M(s-)]^2$, where M_c is the continuous part of M. Moreover, if $\langle M\rangle$ is continuous (i.e., the martingale is quasi-left continuous) and ${}^{\mathrm{p}}\mathrm{var}_2(M,\pi)$ denotes the predictable quadratic variation operator defined by

$${}^{\mathrm{p}}\mathrm{var}_2(M,\pi_t) = \sum_{i=1}^{m} \mathbb{E}\{|M(t_i) - M(t_{i-1})|^2 \,|\, \mathcal{F}(t_{i-1})\}, \tag{3.56}$$

for $\pi_t = (0 = t_0 < t_1 < \cdots < t_m = t)$, then ${}^{\mathrm{p}}\mathrm{var}_2(M,\pi)$ converges in L^1 to $\langle M\rangle$ as the mesh (or norm) of the partition $\|\pi_t\| = \max_k(t_i - t_{i-1})$ goes to zero, see Theorem 2.8 in Chapter 3. Another key point is the study of the variation of M, as defined by the operator

$$\mathrm{var}_\ell(M,\pi_t) = \sum_{i=1}^{m} |M(t_i) - M(t_{i-1})|^\ell, \tag{3.57}$$

as the mesh $\|\pi\|$ vanishes, the cases $\ell = 2$ (quadratic variation) and $\ell = 4$ are of particular interest. As we have seen, the quadratic variation plays an important role in the stochastic integral.

Proposition 3.8 (Quadratic Variation Convergence)**.** *If M is a (real-valued) continuous square integrable martingale then for every $\varepsilon, t > 0$ there exists $\delta > 0$ such that for any partition π_t of the interval $[0,t]$ with $\|\pi_t\| < \delta$ we have $P\{|\mathrm{var}_2(M,\pi_t) - \langle M\rangle(t)| > \varepsilon\} < \varepsilon$.*

Proof. We only give some details for the case when M is continuous and bounded in L^4. Indeed, the martingale property yields

$$\sum_{i=k+1}^{m} \mathbb{E}\{[M(t_i) - M(t_{i-1})]^2 \,|\, \mathcal{F}(t_{i-1})\} = \\ = \sum_{i=k+1}^{m} \mathbb{E}\{M^2(t_i) - M^2(t_{i-1}) \,|\, \mathcal{F}(t_{i-1})\} \le \mathbb{E}\{M^2(t_m)| \,\mathcal{F}(t_k)\},$$

so that

$$\sum_{k=1}^{m-1} \sum_{i=k+1}^{m} \mathbb{E}\{[M(t_i) - M(t_{i-1})]^2 \, [M(t_k) - M(t_{k-1})]^2\} = \\ = \sum_{k=1}^{m-1} \mathbb{E}\Big\{[M(t_k) - M(t_{k-1})]^2 \sum_{i=k+1}^{m} \mathbb{E}\{[M(t_i) - M(t_{i-1})]^2 \,|\, \mathcal{F}(t_{i-1})\}\Big\} \le \\ \le \sum_{k=1}^{m-1} \mathbb{E}\{[M(t_k) - M(t_{k-1})]^2 \sum_{i=k+1}^{m} \mathbb{E}\{M^2(t_m)| \,\mathcal{F}(t_k)\}\} = \\ = \mathbb{E}\{M^2(t_m)\} \sum_{k=1}^{m-1} \mathbb{E}\{[M(t_k) - M(t_{k-1})]^2\}.$$

Since

$$\sum_{k=1}^{m} \mathbb{E}\{[M(t_k) - M(t_{k-1})]^4\} \le \\ \le \mathbb{E}\Big\{\Big(\max_i [M(t_i) - M(t_{i-1})]^2\Big) \sum_{k=1}^{m} [M(t_k) - M(t_{k-1})]^2\Big\} \le \\ \le \big(\mathbb{E}\{\max_i [M(t_i) - M(t_{i-1})]^4\}\big)^{\frac{1}{2}} \big(\mathbb{E}\big\{\big[\sum_{k=1}^{m} [M(t_k) - M(t_{k-1})]^2\big]^2\big\}\big)^{\frac{1}{2}},$$

we deduce

$$\mathbb{E}\{[\mathrm{var}_2(M, \pi_t)]^2\} = \sum_{k=1}^{m} \mathbb{E}\{[M(t_k) - M(t_{k-1})]^4\} + \\ + 2 \sum_{k=1}^{m-1} \sum_{i=k+1}^{m} \mathbb{E}\{[M(t_i) - M(t_{i-1})]^2 \, [M(t_k) - M(t_{k-1})]^2\} \le \\ \le 2\mathbb{E}\{M^2(t)\} \, \mathbb{E}\{[\mathrm{var}_2(M, \pi_t)]\} + \mathbb{E}\{\max_i [M(t_i) - M(t_{i-1})]^4\},$$

after using Hölder inequality. This shows that

$$\sup_{0<s\le t} \mathbb{E}\{|M(s)|^4\} < \infty \quad \Rightarrow \quad \mathbb{E}\{[\mathrm{var}_2(M, \pi_t)]^2\} < \infty, \tag{3.58}$$

and if M is continuous then $\mathbb{E}\{\mathrm{var}_4(M,\pi_t)\} \to 0$ as $\|\pi_t\| \to 0$.

Therefore, because $M^2 - \langle M\rangle$ is a martingale we also have

$$\begin{aligned}
&\mathbb{E}\{[\mathrm{var}_2(M,\pi_t) - \langle M\rangle(t)]^2\} = \\
&\quad = \sum_{k=1}^{m} \mathbb{E}\{[(M(t_k) - M(t_{k-1}))^2 - (\langle M\rangle(t_k) - \langle M\rangle(t_{k-1}))]^2\} \leq \\
&\quad \leq 2\sum_{k=1}^{m} \mathbb{E}\{[M(t_k) - M(t_{k-1})]^4 - [\langle M\rangle(t_k) - \langle M\rangle(t_{k-1})]^2\} \leq \\
&\quad \leq 2\mathbb{E}\{\mathrm{var}_4(M,\pi_t)\} + 2\mathbb{E}\{\langle M\rangle(t) \max_i [\langle M\rangle(t_i) - \langle M\rangle(t_{i-1})]\},
\end{aligned}$$

which proves that $\mathrm{var}_2(M,\pi)$ converges in L^2 to $\langle M\rangle$, whenever M is continuous and belongs to L^4. □

For instance, a complete proof of this result can be found in Chung and Williams [25, Theorem 4.1, Section 4.3, pp. 76–79] or Karatzas and Shreve [91, Theorem 5.8, Chapter 1, pp. 32–34].

3.3.1 Itô's processes

Let $(w(t) : t \geq 0)$ be a n-dimensional standard Wiener process in a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, i.e., with $w(t) = (w_1(t), \dots, w_n(t))$ we have $w_k(t)$ and $w_k(t)w_\ell(t) - \mathbb{1}_{k=\ell}t$ are continuous martingales null at time zero (i.e., $w_i(0) = 0$) relative to the filtration $(\mathcal{F}_t : t \geq 0)$, for any $k, \ell = 1, \dots, n$. Thus $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t) : t \geq 0)$ is called a n-dimensional (standard) Wiener space.

A $\mathbb{R}^d$-valued stochastic process $(X(t) : t \geq 0)$ is called a d-dimensional Itô's process if there exist real-valued adapted processes $(a_i(t) : t \geq 0,\ i = 1, \dots, d)$ and $(b_{ik}(t) : t \geq 0,\ i = 1, \dots, d,\ k = 1, \dots, n)$ such that for every $i = 1, \dots, d$ we have

$$\begin{aligned}
&\mathbb{E}\Big\{ \int_0^{\tau_r} \Big[|a_i(t)| + \sum_{k=1}^{n} |b_{ik}(t)|^2 \Big] \mathrm{d}t \Big\} < \infty, \quad \forall\, r = 1, 2, \dots, \\
&X_i(t) = X_i(0) + \int_0^t a_i(s)\mathrm{d}s + \sum_{k=1}^{n} \int_0^t b_{ik}(s)\mathrm{d}w_k(s), \quad \forall\, t \geq 0,
\end{aligned} \tag{3.59}$$

in some n-dimensional Wiener space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t) : t \geq 0)$, where $\{\tau_r : r \geq 1\}$ is a non-decreasing sequence of stopping times satisfying $\tau_r \to \infty$ almost surely. In short we write $\mathrm{d}X(t) = a(t)\mathrm{d}t + b(t)\mathrm{d}w(t)$, for every $t \geq 0$, with a in L^1_{loc} and b in L^2_{loc}. Note that for a Wiener process or in general for a continuous local martingale M, we write the stochastic integral

$$\int_{]0,t]} f(s)\mathrm{d}M(s) = \int_{(0,t]} f(s)\mathrm{d}M(s) = \int_0^t f(s)\mathrm{d}M(s),$$

indistinctly. Remark that any Itô process is a (special) semi-martingale, and a quasi-martingale whenever a and b belong to $L^1([0,T]\times\Omega)$ and $L^2([0,T]\times\Omega)$, for any $T > 0$, respectively.

Theorem 3.9 (Itô formula)**.** *Let $(X(t) : t \geq 0)$ be a d-dimensional Itô's process in a given Wiener space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t) : t \geq 0)$, i.e, (3.59), and $\varphi = \varphi(t, x)$ be a real-valued smooth function on $[0, \infty) \times \mathbb{R}^d$, i.e., C^1 in the first variable t on $[0, \infty)$ and C^2 in the second variable x on $\mathbb{R}^d$. Then $(\varphi(t, X(t)) : t \geq 0)$ is a (real-valued) Itô's process and*

$$\varphi(t, X(t)) = \varphi(0, X(0)) + \int_0^t A(s, X)\varphi(s, X(s))\mathrm{d}s + \\ + \sum_{k=1}^n \int_0^t B_k(s, X)\varphi(s, X(s))\mathrm{d}w_k(s), \quad \forall\, t \geq 0, \quad (3.60)$$

where the linear differential operators $A(s, X)$ and $B(s, X) = (B_k(s, X) : k = 1, \ldots, n)$ are given by

$$A(s, X)\varphi(t, x) = \partial_t\varphi(t, x) + \sum_{i=1}^d a_i(s)\, \partial_i\varphi(t, x) + \\ + \frac{1}{2} \sum_{i,j=1}^d \Big(\sum_{k=1}^n b_{ik}(s) b_{jk}(s) \Big)\, \partial^2_{ij}\varphi(t, x),$$

and

$$B_k(s, X)\varphi(t, x) = \sum_{i=1}^d b_{ik}(s)\partial_i\varphi(t, x),$$

for any $s, t \geq 0$ and x in $\mathbb{R}^d$, with ∂_t, ∂_i and $\partial^2_{i,j}$ denoting the partial derivatives with respect to the variable t, x_i and x_j.

Proof. The first step is to localize, i.e., setting

$$T_r = \tau_r \wedge \inf\big\{t \geq 0 : |X(t)| \geq r\big\}$$

we have a non-decreasing sequence of stopping times satisfying $T_r \to \infty$ almost surely. Moreover, if $X_n(t) = X(t \wedge T_n)$ then X_n is a processes with values in the compact ball of radius r and therefore the processes $A(s) = A(s, X)\varphi(s, X_n(s))$ and $B_k(s) = B_k(s, X)\varphi(s, X_n(s))$ are in L^1 and L^2, respectively, i.e.,

$$\mathbb{E}\Big\{ \int_0^{T_r} \Big[|A(t)| + \sum_{k=1}^n |B_k(t)|^2\Big]\mathrm{d}t\Big\} < \infty, \quad \forall\, r = 1, 2, \ldots,$$

so that the right-hand side of the so-called Itô formula or rule (3.60) is an real-valued Itô's process. This shows that without loss of generality, we may assume that the function φ has a compact support. Furthermore, details on the proof are only provided for the one-dimensional case, i.e., $d = 1$ and $n = 1$, with $X(t) = X(0) + A(t) + B(t)$ and

$$A(t) = \int_0^t a(s)\mathrm{d}s, \qquad B(t) = \int_0^t b(s)\mathrm{d}w(s), \quad (3.61)$$

$a(s)$ and $b(s)$ are predictable (actually, adapted is sufficient) processes such that

$$|B(t)| + \int_0^t [|a(s)| + |b(s)|^2]\mathrm{d}s \le C,$$

for any $t \ge 0$ and some deterministic constant $C > 0$.

The second step is to apply Taylor formula for a smooth real-valued function $\varphi = \varphi(x)$ on $\mathbb{R}$, with a partition $\pi = (0 = t_0 < t_1 < \cdots < t_m = t)$ of $[0, t]$,

$$\varphi(X(t)) - \varphi(X(0)) = \sum_{k=1}^m [\varphi(X(t_k)) - \varphi(X(t_{k-1})] = \\ = \sum_{k=1}^m [X(t_k) - X(t_{k-1})]\varphi'_k + \frac{1}{2}\sum_{k=1}^m [X(t_k) - X(t_{k-1})]^2 \varphi''_k, \quad (3.62)$$

where $X(t) = X(0) + A(t) + B(t)$ satisfying (3.61),

$$\varphi'_k = \varphi'(X(t_{k-1})), \qquad \varphi''_k = \int_0^1 \varphi''((1-s)X(t_{k-1}) + sX(t_k))\mathrm{d}s,$$

and the mesh (or norm) $\|\pi\| = \max_i(t_i - t_{i-1})$ is destined to vanish.

Considering the predictable process $\varphi'_\pi(s) = \varphi'(X(t_{k-1}))$ for s belonging to $]t_{k-1}, t_k]$, we check that

$$\sum_{k=1}^m [X(t_k) - X(t_{k-1})]\varphi'_k = \int_{]0,t]} \varphi'_\pi(s)\mathrm{d}A(s) + \int_{]0,t]} \varphi'_\pi(s)\mathrm{d}B(s),$$

which converges in $L^1 + L^2$ (or pathwise for the first term and L^2 for the second term) to

$$\int_{]0,t]} \varphi'(X(s))\mathrm{d}A(s) + \int_{]0,t]} \varphi'(X(s))\mathrm{d}B(s)$$

where the first integral is in the Riemann-Stieltjes (or Lebesgue-Stieltjes) sense and the second term is a stochastic integral. By means of the substitution formula (3.54) and (3.55), the above limit can be rewritten as

$$\int_{]0,t]} \varphi'(X(s))a(s)\mathrm{d}s + \int_{]0,t]} \varphi'(X(s))b(s)\mathrm{d}w(s),$$

where the first integral is now in the Lebesgue sense, which agrees with the stochastic sense if a predictable version of the integrand is used.

To handle the quadratic variation in (3.62), we note that

$$[X(t_k) - X(t_{k-1})]^2 = -2[A(t_k) - A(t_{k-1})]\,[B(t_k) - B(t_{k-1})] + \\ + [A(t_k) - A(t_{k-1})]^2 + [B(t_k) - B(t_{k-1})]^2,$$

and for any $k \geq 1$,

$$|\varphi''(X(t_{k-1})) - \varphi''_k| \leq \max_k \rho(\varphi'', |X(t_k) - X(t_{k-1})|),$$

where $\rho(\varphi'', r)$ is the modulus of continuity of φ'', i.e.,

$$\rho(\varphi'', r) = \sup_{|x-y|\leq r} |\varphi''(x) - \varphi''(y)|.$$

Therefore

$$\sum_{k=1}^{m} [X(t_k) - X(t_{k-1})]^2 \varphi''_k = \\ = \sum_{k=1}^{m} \varphi''(X(t_{k-1}))[B(t_k) - B(t_{k-1})]^2 + \mathbf{o}(\|\pi\|)$$

where

$$|\mathbf{o}(\|\pi\|)| \leq \max_k \big\{\rho(\varphi'', |X(t_k) - X(t_{k-1})|)\big\} \big\{ \sum_{k=1}^{m} [B(t_k) - B(t_{k-1})]^2 \big\} + \\ + \max_k \big\{ [2|B(t_k) - B(t_{k-1})| + |A(t_k) - A(t_{k-1})|] \, |\varphi''_k| \big\} \times \\ \times \big\{ \sum_{k=1}^{m} |A(t_k) - A(t_{k-1})| \big\},$$

i.e., $\mathbf{o}(\|\pi\|)$ is bounded by a deterministic constant and $\mathbf{o}(\|\pi\|) \to 0$ as $\|\pi\| \to 0$, almost surely.

Since $\bar{\varphi}''_k = \varphi''(X(t_{k-1}))$ is $\mathcal{F}(t_k)$-measurable and

$$B^2(t) - \int_0^t |b(s)|^2 \mathrm{d}s,$$

is a martingale, we have

$$\mathbb{E}\big\{ \{ \sum_{k=1}^{m} [(B(t_k) - B(t_{k-1}))^2 - \int_{t_{k-1}}^{t_k} |b(s)|^2 \mathrm{d}s] \, \bar{\varphi}''_k \}^2 \big\} = \\ = \mathbb{E}\big\{ \sum_{k=1}^{m} \big[(B(t_k) - B(t_{k-1}))^2 - \int_{t_{k-1}}^{t_k} |b(s)|^2 \mathrm{d}s \big]^2 \, |\bar{\varphi}''_k|^2 \big\},$$

which is bounded by the expression

$$\big(\max_i \mathbb{E}\{|\bar{\varphi}''_k|^2\} \big) \, \mathbb{E}\big\{ \sum_{k=1}^{m} \big[(B(t_k) - B(t_{k-1}))^2 - \int_{t_{k-1}}^{t_k} |b(s)|^2 \mathrm{d}s \big]^2 \big\}.$$

In view of Proposition 3.8, we deduce that

$$\mathbb{E}\Big\{\Big|\sum_{k=1}^{m}[B(t_k)-B(t_{k-1})]^2\varphi''_k-\int_{]0,t]}|b(s)|^2\bar\varphi''_\pi(s)\mathrm{d}s\Big|^2\Big\}\to 0,$$

as $\|\pi\|\to 0$, where $\varphi''_\pi(s)=\varphi''(X(t_{k-1}))=\bar\varphi''_k$ for any s in $]t_{k-1},t_k]$.

Thus, we have establish the one-dimensional Itô formula for a (real-valued) smooth function with compact support $\varphi(x)$, which conclude the proof. □

Note the short vector notation for Itô formula when $\varphi=\varphi(x)$, namely,

$$\mathrm{d}\varphi(X(t))=\nabla\varphi(X(t))\mathrm{d}X(t)+\frac{1}{2}\,\mathrm{Tr}[b(t)b^*(t)\nabla^2\varphi(x)]\mathrm{d}t \tag{3.63}$$

for every $t\geq 0$, where ∇ is the gradient operator and $\nabla^2\varphi$ is the matrix of second derivatives.

From the above proof, it is clear also that several generalizations of Itô formula are possible. Note that it is not necessary to separate the t variable, since we may add one more dimension with $a(s)=1$ and $b(s)=0$ to pass from $\varphi(x)$ to $\varphi(t,x)$. By reviewing the previous steps and remarking the use of the continuity and the quadratic variation of the martingale M, we can show the following rule.

Theorem 3.10. *Let $(X_i(t):t\geq 0)$ be a continuous semi-martingale in a given filtered space $(\Omega,\mathcal{F},P,\mathcal{F}_t:t\geq 0)$, for each $i=1,\dots,d$, and $\varphi=\varphi(x)$ be a real-valued C^2 function on $\mathbb{R}^d$. Then $(\varphi(X(t)):t\geq 0)$, $X(t)=(X_1(t),\dots,X_d(t))$ is a continuous semi-martingale and*

$$\begin{aligned}\varphi(X(t))=\varphi(X(0))&+\sum_{i=1}^{d}\int_{]0,t]}\partial_i\varphi(X(s))\mathrm{d}X_i(t)+\\&+\sum_{i,j=1}^{d}\int_{]0,t]}\partial^2_{ij}\varphi(X(s))\mathrm{d}\langle X_i,X_j\rangle(s),\quad\forall\,t\geq 0,\end{aligned} \tag{3.64}$$

where ∂_i and ∂^2_{ij} denote partial derivatives, and $\langle X_i,X_j\rangle(s)$ is the only predictable process with locally integrable bounded variation such that the expression $X_iX_j-\langle X_i,X_j\rangle$ is a martingale. □

We can also extend the integration-by-part formula (3.53) for two (cad-lag) real-valued semi-martingales $X=V_X+M_X$ and $Y=V_Y+M_Y$ where V_X, V_Y have locally bounded variation and M_X, M_Y are continuous local-martingales as follows

$$\begin{aligned}X(t)Y(t)-X(0)Y(0)=\langle M_X,M_Y\rangle(t)&+\int_{(0,t]}X(s-)\mathrm{d}Y(s)+\\&+\int_{(0,t]}Y(s-)\mathrm{d}X(s)+\sum_{0<s\leq t}\delta V_X(s)\,\delta V_Y(s),\end{aligned} \tag{3.65}$$

for every $t \geq 0$, where $X(t-)$ and $Y(t-)$ are the left limits at t, and δ is the jump-operator, e.g., $\delta X(t) = X(t) - X(t-)$. Note that the correction term satisfies

$$\langle M_X, M_Y \rangle(t) + \sum_{0<s\leq t} \delta V_X(s)\, \delta V_Y(s) = [X,Y](t),$$

i.e., it is equal to the optional quadratic covariation process $[X,Y]$ associated with the semi-martingale X and Y.

As seen in (3.51) of the previous section, for a standard n-dimensional Wiener process $(w(t) : t \geq 0)$, for any adapted (measurable) process $f(s)$ and for any stopping time T, we can write

$$\mathbb{E}\Big\{ \sup_{0\leq t\leq T} \Big| \int_0^t f(s)\mathrm{d}w(s)\Big|^p \Big\} \leq C_p\, \mathbb{E}\Big\{ \Big[\int_0^T |f(s)|^2 \mathrm{d}s \Big]^{p/2} \Big\}. \tag{3.66}$$

for some constant positive C_p. Actually, for p in $(0,2]$ the proof is very simple (see (2.8) of Chapter 3) and $C_p = (4-p)/(2-p)$ if $0 < p < 2$ and $C_2 = 4$. However, the proof for $p > 2$ involves Burkhölder-Davis-Gundy inequality. An alternative is to use Itô formula for the function $x \mapsto |x|^p$ and the process

$$X(t) = \int_0^t f(s)\mathrm{d}w(s), \quad \forall t \geq 0$$

to get

$$\mathbb{E}\{|X(t)|^p\} = \frac{p(p-1)}{2}\, \mathbb{E}\Big\{ \int_0^t |X(s)|^{p-2} |f(s)|^2 \mathrm{d}s \Big\}.$$

By means of the Doob's maximal inequality, for some constant $\tilde{C}_p$ depending only on p we have

$$\mathbb{E}\{ \sup_{0\leq t\leq T} |X(t)|^p\} \leq \tilde{C}_p\, \mathbb{E}\Big\{ \Big(\sup_{0\leq t\leq T} |X(t)|^{p-2} \Big) \Big(\int_0^T |f(s)|^2 \mathrm{d}s \Big) \Big\}$$

and in view of Hölder inequality with exponents $p/2$ and $p/(p-2)$, we deduce the desired estimate (3.66). Similarly, we can treat the multidimensional case.

3.3.2 Discontinuous Local Martingales

Let $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t) : t \geq 0)$ be a n-dimensional (standard) Wiener space and $(p(B,]0,t]) : B \in \mathbb{R}_0^m,\ t \geq 0)$ be an independent (standard) Poisson measure with (intensity) Lévy measure $\pi(B) = \mathbb{E}\{p(B,]0,t])\}/t$, which satisfies [1]

$$\int_{\mathbb{R}_*^m} \frac{|\zeta|^2}{1+|\zeta|} \pi(\mathrm{d}\zeta) < \infty,$$

[1] the Polish space $\mathbb{R}_0^m = \mathbb{R}^m \setminus \{0\}$ may be replaced by a general Backwell space.

and martingale measure $\tilde{p}(B,]0,t]) = p(B,]0,t]) - t\pi(B)$, as discussed in Sections 2.7 and 3.2.2. This is referred to as a (standard) Wiener-Poisson space. Clearly, a non-standard Wiener-Poisson space corresponds to a Poisson measure with (deterministic) intensity $\Pi(\mathrm{d}\zeta, \mathrm{d}s)$, which is not necessarily absolutely continuous (in the second variable $\mathrm{d}s$) with respect to the Lebesgue measure $\mathrm{d}s$, but $\Pi(\mathbb{R}^m_*, \{t\}) = 0$ for every $t \geq 0$. Also, an extended Wiener-Poisson space corresponds to an extended Poisson measure with (deterministic) intensity $\Pi(\mathrm{d}\zeta, \mathrm{d}s)$, which may have atoms of the form $\mathbb{R}^m_* \times \{t\}$. In any case, the deterministic intensity $\Pi(\mathrm{d}\zeta, \mathrm{d}s) = \mathbb{E}\{p(\mathrm{d}\zeta, \mathrm{d}s)\}$ is the (predictable) compensator of the optional random measure p.

So, a (standard) Wiener-Poisson space with Lévy measure $\pi(\cdot)$ is denoted by $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{p}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0)$, and the (local) martingale measure $\tilde{p}$ is identified with the $\mathbb{R}^m$-valued compensated-jump (Poisson) process

$$\tilde{p}(t) = \int_{\mathbb{R}^m_* \times]0,t]} \zeta \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad t \geq 0,$$

which induces, on the canonical space $D = D([0,\infty[, \mathbb{R}^m)$ of cad-lag functions, a probability measure $P_{\tilde{\nu}}$, namely,

$$P_{\tilde{\nu}}(B) = P\Big\{\tilde{p}(\cdot) \in B\Big\}, \quad \forall B \in \mathcal{B}(D).$$

with the characteristic function (or Fourier transform) given by

$$\mathbb{E}\Big\{ \exp\Big[\mathtt{i} \int_{\mathbb{R}^m_* \times]0,t]} (z \cdot \zeta) \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s)\Big]\Big\} = \\ = \exp\Big[- t \int_{\mathbb{R}^m_*} \big(1 - \mathrm{e}^{\mathtt{i}\, z \cdot \zeta} + \mathtt{i}\, z \cdot \zeta\big) \pi(\mathrm{d}\zeta)\Big],$$

for every $t \geq 0$ and z in $\mathbb{R}^m$. Also note that the Wiener process w induces a probability measure P_w on the canonical space $C = C([0,\infty[, \mathbb{R}^n)$ of continuous functions, namely,

$$P_w(B) = P\big\{w(\cdot) \in B\big\}, \quad \forall B \in \mathcal{B}(C).$$

and its the characteristic function (or Fourier transform) is given by

$$\mathbb{E}\big\{ \exp\big[\mathtt{i}\, \xi \cdot w(t)\big]\big\} = \exp\big(- t \frac{|\xi|^2}{2}\big),$$

for every $t \geq 0$ and ξ in $\mathbb{R}^n$. Therefore, a *canonical* (standard) Wiener-Poisson space with Lévy measure $\pi(\cdot)$ is a probability measure $P = P_w \times P_{\tilde{p}}$ on the Polish space $C([0,\infty[, \mathbb{R}^n) \times D([0,\infty[, \mathbb{R}^m)$. In this case, the projection map $(\omega_1, \omega_2) \mapsto \big(\omega_1(t), \omega_2(t)\big)$ on $\mathbb{R}^n \times \mathbb{R}^m$, for every $t \geq 0$, is denoted by $\big(X_w(t,\omega), X_{\tilde{p}}(t,\omega)\big)$, and under the probability P the canonical process $(X_w(t) : t \geq 0)$ is a n-dimensional (standard) Wiener process and the canonical process $X_{\tilde{p}}(t)$ is a $\mathbb{R}^m$-valued compensated-jump Poisson process with Lévy measure $\pi(\cdot)$ on $\mathbb{R}^m_*$.

The filtration $(\mathcal{F}_t : t \geq 0)$ is generated by the canonical process X_w and $X_{\tilde{p}}$ and completed with null sets with respect to the probability measure P. Note that since the Wiener process is continuous and the compensated-jump Poisson process is purely discontinuous, they are orthogonal (with zero-mean) so that they are independent, i.e., the product form of $P = P_w \times P_{\tilde{p}}$ is a consequences of the statistics imposed on the processes w and $\tilde{p}$.

Definition 3.11 (Itô process with jumps)**.** A $\mathbb{R}^d$-valued stochastic process $(X(t) : t \geq 0)$ is called a d-dimensional *Itô's process with jumps* if there exist real-valued adapted processes $(a_i(t) : t \geq 0,\ i = 1, \dots, d)$, $(b_{ik}(t) : t \geq 0,\ i = 1, \dots, d,\ k = 1, \dots, n)$ and $(\gamma_i(\zeta, t) : t \geq 0,\ \zeta \in \mathbb{R}^m_*)$, such that for every $i = 1, \dots, d$ and any $r = 1, 2, \dots$, we have

$$\mathbb{E}\Big\{ \int_0^{\tau_r} \Big[|a_i(t)| + \sum_{k=1}^n |b_{ik}(t)|^2 + \int_{\mathbb{R}^m_*} |\gamma_i(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \Big] \mathrm{d}t \Big\} < \infty, \tag{3.67}$$

and

$$X_i(t) = X_i(0) + \int_0^t a_i(s)\mathrm{d}s + \sum_{k=1}^n \int_0^t b_{ik}(s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}^m_* \times]0,t]} \gamma_i(\zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq 0,$$

in some (standard) Wiener-Poisson space

$$(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{p}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0),$$

with Lévy measure π, where $\{\tau_r : r \geq 1\}$ is a non-decreasing sequence of stopping times satisfying $\tau_r \to \infty$ almost surely. In short we write

$$\mathrm{d}X(t) = a(t)\mathrm{d}t + b(t)\mathrm{d}w(t) + \int_{\mathbb{R}^m_*} \gamma(\zeta, t)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}t),$$

for every $t \geq 0$, with a in L^1_{loc}, b in L^2_{loc} and γ in $L^2_{\mathrm{loc},\pi}$. The local-martingale measure $\tilde{p}(\mathrm{d}\zeta, \mathrm{d}t) = p(\mathrm{d}\zeta, \mathrm{d}t) - E\{p(\mathrm{d}\zeta, \mathrm{d}t)\}$ is also referred to as the compensated jumps (martingale) measure. If the compensator has the form $\Pi(\mathrm{d}\zeta, \mathrm{d}s) = E\{p(\mathrm{d}\zeta, \mathrm{d}t)\}$ then the local integrability assumption on the coefficients γ_i should be changed accordingly, and γ_i should be progressively measurable. Moreover, if $\Pi(\mathbb{R}^m_* \times \{t\}) \neq 0$ for some t, then γ_i must be predictable. □

Note that any Itô process with jumps is a quasi-left continuous (cad-lag) semi-martingale, and a quasi-martingale whenever a, b and γ belong to the spaces $L^1(]0, T[\times\Omega)$, $L^2(]0, T[\times\Omega)$ and $L^2_\pi(\mathbb{R}^m_* \times]0, T[\times\Omega)$, for any $T > 0$, respectively. Condition (3.67) is equivalent to

$$P\Big\{ \int_0^t \Big[|a(s)| + \mathrm{Tr}[b(s)b^*(s)] + \int_{\mathbb{R}^m_*} |\gamma(\zeta, s)|^2 \pi(\mathrm{d}\zeta) \Big] \mathrm{d}s < \infty \Big\} = 1, \tag{3.68}$$

for every $t \geq 0$, where $\mathrm{Tr}[\cdot]$ denotes the trace of a matrix and $|\cdot|$ is the Euclidean norm of a vector in $\mathbb{R}^m$. Again, for non-standard case, we modify all conditions accordingly to the the use of $\Pi(\mathrm{d}\zeta, \mathrm{d}s)$ in lieu of $\pi(\mathrm{d}\zeta)\mathrm{d}s$.

Theorem 3.12 (Itô formula with jumps)**.** *Let $(X(t) : t \geq 0)$ be a d-dimensional Itô's process with jumps in a Wiener-Poisson space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{p}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0)$ with Lévy measure $\pi(\mathrm{d}\zeta)$, i.e., (3.67), and let $\varphi = \varphi(x)$ be a real-valued twice continuously differentiable function on $\mathbb{R}^d$, satisfying*

$$\mathbb{E}\Big\{ \int_0^{T_r} \mathrm{d}t \int_{\mathbb{R}^m_*} \big[|\varphi\big(X(t) + \gamma(\zeta, t)\big) - \varphi\big(X(t)\big)|^2 + \varphi\big(X(t) + \gamma(\zeta, t)\big) - \\ - \varphi\big(X(t)\big) - \gamma(\zeta, t) \cdot \nabla\varphi\big(X(t)\big)\big]\pi(\mathrm{d}\zeta)\Big\} < \infty, \quad (3.69)$$

for some increasing sequence $\{T_r : r \geq 1\}$ of stopping times such that $T_r \to \infty$ almost surely. Then $(\varphi(X(t)) : t \geq 0)$ is a (real-valued) Itô's process with jumps and

$$\varphi(X(t)) = \varphi(X(0)) + \int_0^t A(s, X)\varphi(X(s))\mathrm{d}s + \\ + \sum_{k=1}^n \int_0^t B_k(s, X)\varphi(X(s))\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}^m_* \times]0,t]} C(\zeta, s, X)\varphi(X(s))\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq 0, \quad (3.70)$$

where the linear integro-differential operators $A(s, X)$, $B(s, X) = (B_k(s, X) : k = 1, \ldots, n)$ and $C(\zeta, s, X)$ are given by

$$A(s, X)\varphi(x) = \sum_{i=1}^d a_i(s)\, \partial_i\varphi(x) + \frac{1}{2} \sum_{i,j=1}^d \big(\sum_{k=1}^n b_{ik}(s)b_{jk}(s)\big)\, \partial^2_{ij}\varphi(x) + \\ + \int_{\mathbb{R}^m_*} [\varphi(x + \gamma(\zeta, s)) - \varphi(x) - \sum_{i=1}^d \gamma_i(\zeta, s)\, \partial_i\varphi(x)]\pi(\mathrm{d}\zeta),$$

and

$$B_k(s, X)\varphi(x) = \sum_{i=1}^d b_{ik}(s)\, \partial_i\varphi(x),$$
$$C(\zeta, s, X)\varphi(x) = \varphi(x + \gamma(\zeta, s)) - \varphi(x),$$

for any $s \geq 0$ and x in $\mathbb{R}^d$, with ∂_i, ∂_{ij}, denoting the first and second partial derivatives with respect to the i and j, and ∇ being the gradient operator.

Proof. First, we replace the coefficients $a(s)$, $b(s)$ and $\gamma(\zeta, s)$ by

$$a(s)\mathbb{1}_{s \leq \tau}, \qquad b(s)\mathbb{1}_{s \leq \tau}, \qquad \gamma(\zeta, s)\mathbb{1}_{s \leq \tau}\mathbb{1}_{\varepsilon < |\zeta| \leq 1/\varepsilon},$$

where $\tau = \tau_r$ and $\varepsilon > 0$. We could use elementary predictable processes a, b and γ to force concrete a calculation. Thus we can pass to the limit in $r \to \infty$ and $\varepsilon \to 0$ to revalidate (3.70), as long as the smooth function φ satisfies (3.69).

The continuity of the semi-martingale was needed in the proof of Theorem 3.9. Nevertheless, when $\gamma(\zeta, s) = 0$ for any $|\zeta| \leq \varepsilon$, the integer-valued measure of the Poisson measure used to integrate has bounded variation and the stochastic integral relative to the local-martingale measure becomes a pathwise Lebesgue-Stieltjes integral. Then we can use the integration by parts formula (3.65) to get

$$X(t)Y(t) - X(0)Y(0) = \int_{(0,t]} X(s-)\mathrm{d}Y(s) + \\ + \int_{(0,t]} Y(s-)\mathrm{d}X(s) + [X,Y](t), \quad \forall\, t \geq 0, \quad (3.71)$$

where $[X, Y]$ is the optional quadratic co-variation process. Actually, we may apply (3.65) for jumps with bounded variation and as ε vanishes, we deduce the validity of (3.71) for any two (real-valued) Itô's processes with jumps X and Y.

Note that

$$[X,Y](t) = \langle X^c, Y^c\rangle(t) + \sum_{0<s\leq t} \big(X(s) - X(s-)\big)\big(Y(s) - Y(s-)\big) = \\ = \langle X^c, Y^c\rangle(t) + \int_{\mathbb{R}^m_* \times]0,t]} \gamma^X(\zeta, s)\, \gamma^Y(\zeta, s)\, p(\mathrm{d}\zeta, \mathrm{d}s),$$

where $\langle \cdot, \cdot \rangle$ is the optional quadratic co-variation process, X^c and Y^c are the continuous parts of X and Y, e.g.,

$$X^c(t) = \int_0^t a^X(s)\mathrm{d}s + \int_0^t b^X(s)\mathrm{d}w(s),$$

and ν is the integer-valued measure, i.e., $\tilde{\nu}(\cdot,]0, t]) = \nu(\cdot,]0, t]) - t\,\pi(\cdot)$. We can rewrite (3.65) explicitly as

$$X(t)Y(t) - X(0)Y(0) = \int_{(0,t]} X(s-)\mathrm{d}Y^c(s) + \\ + \int_{(0,t]} Y(s-)\mathrm{d}X^c(s) + \langle X^c, Y^c\rangle(t) + \\ + \int_{\mathbb{R}^m_* \times]0,t]} [X(t)\gamma^Y(\zeta, s) + Y(t)\gamma^X(\zeta, s)]\, \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s) + \\ + \int_{\mathbb{R}^m_* \times]0,t]} \gamma^X(\zeta, s)\, \gamma^Y(\zeta, s)\, p(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq 0.$$

In particular, if $X = Y$ we get

$$X^2(t) - X^2(0) = 2\int_{(0,t]} X(s-)\mathrm{d}Y^c(s) + \langle X^c\rangle(t) + \\ + 2\int_{\mathbb{R}_*^m\times]0,t]} X(t)\gamma(\zeta,s)\,\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s) + \int_{\mathbb{R}_*^m\times]0,t]} \gamma^2(\zeta,s)\,p(\mathrm{d}\zeta,\mathrm{d}s),$$

for every $t \geq 0$, which exactly reproduces Itô formula (3.70) for $\varphi(x) = x^2$.

Iterating this argument, we check the validity of (3.70) for any multi-dimensional polynomial function $\varphi(x_1,\ldots,x_d)$, and by density, for any smooth function $\varphi(x)$.

Finally, for any smooth function satisfying (3.69) we may let $r \to \infty$ and $\varepsilon \to 0$ to conclude. □

Note that we also have

$$X(t)Y(t) - X(0)Y(0) = \int_{(0,t]} X(s-)\mathrm{d}Y(s) + \\ + \int_{(0,t]} Y(s-)\mathrm{d}X(s) + \langle X,Y\rangle(t), \quad \forall\, t \geq 0, \quad (3.72)$$

i.e., in the integration by parts the optional quadratic variation $[X,Y]$ may be replaced by the predictable quadratic variation $\langle X,Y\rangle$ associated with the whole *quasi-left continuous* square integrable semi-martingales X and Y. Also for a function $\varphi = \varphi(t,x)$, we do not need to require C^2 in the variable t. Also, when $\varphi = \varphi(x)$, we could use a short vector notation

$$\mathrm{d}\varphi(X(t)) = \nabla\varphi(X(t))\mathrm{d}X^c(t) + [\varphi \diamond_\gamma \tilde{p}](\cdot,\mathrm{d}t)(t,X(t)) + \\ + \big[\frac{1}{2}\,\mathrm{Tr}[b(t)b^*(t)\nabla^2\varphi(x)] + [\varphi \bullet_\gamma \pi](t,X(t))\big]\mathrm{d}t, \quad (3.73)$$

for every $t \geq 0$, where

$$[\varphi \diamond_\gamma \tilde{p}(\cdot,\mathrm{d}t)](t,x) = \int_{\mathbb{R}_*^m} [\varphi(x+\gamma(\zeta,t)) - \varphi(x)]\,\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t),$$

$$[\varphi \bullet_\gamma \pi](t,x) = \int_{\mathbb{R}_*^m} [\varphi(x+\gamma(\zeta,t)) - \varphi(x) - \gamma(\zeta,t)\cdot\nabla\varphi(x)]\pi(\mathrm{d}\zeta),$$

and ∇ and $\mathrm{Tr}[\cdot]$ are the gradient and trace operator, respectively. The above calculation remains valid for a Poisson measure not necessarily standard, i.e., the intensity or Lévy measure has the form $\Pi(\mathrm{d}\zeta,\mathrm{d}t) = \mathbb{E}\{p(\mathrm{d}\zeta,\mathrm{d}t)\}$ and $\Pi(\mathbb{R}_*^m \times \{t\}) = 0$ for every $t \geq 0$. For an extended Poisson measure, the process is no longer quasi-left continuous and the rule (3.70) needs a jump correction term, i.e., the expression $X(s)$ is replaced by $X(s-)$ inside the stochastic integrals. For instance, the reader may consult Bensoussan and Lions [6, Section 3.5, pp. 224–244] or Gikhman and Skorokhod [62, Chapter II.2, pp. 215–272] for more details on this approach.

Semi-martingale Viewpoint

In general, the integration by parts formula (3.71) is valid for any two semi-martingales X and Y, and we have the following general Itô formula for semi-martingales, e.g., Chung and Williams [25, Theorems 38.3 and 39.1, Chapter VI, pp. 392–394], Dellacherie and Meyer [32, Sections VIII.15–27, pp. 343–352], Jacod and Shiryaev [84, Theorem 4.57, Chapter 1, pp. 57–58].

Theorem 3.13. *Let $X = (X_1, \dots, X_d)$ be a d-dimensional semi-martingale and φ be a complex-valued twice-continuously differentiable function on $\mathbb{R}^d$. Then $\varphi(X)$ is a semi-martingale and we have*

$$\varphi(X(t)) = \varphi(X(0)) + \sum_{i=1}^{d} \int_{]0,t]} \partial_i \varphi(X(s-)) \mathrm{d}X_i(s) + \\ + \frac{1}{2} \sum_{i,j=1}^{d} \int_{]0,t]} \partial^2_{ij} \varphi(X(s-)) \mathrm{d}\langle X_i^c, X_j^c \rangle(s) + \\ + \sum_{0<s\le t} \Big\{ \varphi(X(s)) - \varphi(X(s-)) - \sum_{i=1}^{d} \partial_i \varphi(X(s-)) \delta X(s) \Big\},$$

where ∂_i and ∂^2_{ij} denotes partial derivatives, $\delta X(s) = [X_i(s) - X_i(s-)]$ and $X(s-)$ is the left limit at s and X_i^c is the continuous part. □

First remark that

$$\int_{]0,t]} \partial^2_{ij} \varphi(X(s-)) \mathrm{d}\langle X_i, X_j \rangle(s) = \sum_{0<s\le t} \partial^2_{ij} \varphi(X(s-)) \big[\delta X(s)\big]^2 + \\ + \int_{]0,t]} \partial^2_{ij} \varphi(X(s-)) \mathrm{d}\langle X_i^c, X_j^c \rangle(s),$$

where the integrals and series are absolutely convergent. Hence, the above formula can be rewritten using the predictable quadratic variation $\langle X_i, X_j \rangle$, i.e., the predictable processes obtained via the Doob-Meyer decomposition when X is locally square integrable or in general the predictable projection of the optional quadratic variation $[X_i, X_j]$.

Let X be a (special) quasi-left continuous semi-martingale written in the canonical form

$$X(t) = X(0) + X^c(t) + A(t) + \int_{\mathbb{R}^d_* \times]0,t]} z \tilde{\nu}(\mathrm{d}z, \mathrm{d}s), \forall t \ge 0,$$

where X^c is the continuous (local-martingale) part, A is the predictable locally bounded variation (and continuous) part, and $\tilde{\nu}$ is the compensated (local-martingale) random measure associated with the integer-valued measure $\nu = \nu_X$ of the process X with compensator ν^p. Then

$$\mathrm{d}X_i(s) = \mathrm{d}X_i^c(s) + \int_{\mathbb{R}^d_*} z_i \tilde{\nu}(\mathrm{d}z, \mathrm{d}s),$$

so that

$$\sum_{i=1}^{d}\int_{]0,t]}\partial_i\varphi(X(s-))\mathrm{d}X_i(s)=\sum_{i=1}^{d}\int_{]0,t]}\partial_i\varphi(X(s-))\mathrm{d}X_i^c(s)+$$
$$+\sum_{i=1}^{d}\int_{\mathbb{R}_*^d\times]0,t]}z_i\partial_i\varphi(X(s-))\tilde{\nu}(\mathrm{d}z,\mathrm{d}s),$$

and the jump part can be written as

$$\sum_{0<s\le t}\big[\varphi(X(s))-\varphi(X(s-))-\sum_{i=1}^{d}\partial_i\varphi(X(s-))\delta X(s)\big]=$$
$$=\int_{\mathbb{R}_*^d\times]0,t]}\big[\varphi(X(s-)+z)-\varphi(X(s-))-\sum_{i=1}^{d}z_i\partial_i\varphi(X(s-))\big]\nu(\mathrm{d}z,\mathrm{d}s),$$

for every $t\ge 0$. Moreover, because $\nu^p(\mathbb{R}_*^m\times\{t\})=0$ for any $t\ge 0$, we can substitute $X(s-)$ for $X(s)$ in the above stochastic integral. Thus, combining the above jump parts we see that the expression (3.70) of Theorem 3.12 remains valid for any quasi-left continuous integer measure $\nu(\mathrm{d}z,\mathrm{d}s)$ with a local-martingale measure $\tilde{\nu}(\mathrm{d}z,\mathrm{d}s)$ and compensator $\nu^p(\mathrm{d}z,\mathrm{d}s)$, which replaces the deterministic product measure $\pi(\mathrm{d}z)\times\mathrm{d}s$. The case of interest for us is when the predictable compensator measure $\nu^p(\mathrm{d}z,\mathrm{d}s)$ has a density with respect to the Lebesgue measure, i.e.,

$$\nu^p(B,]0,t])=\int_0^t\mathtt{M}(B,s)\mathrm{d}s,\quad\forall B\in\mathcal{B}(\mathbb{R}_*^d),\ t\ge 0,$$

where the intensity kernel $\mathtt{M}$ is such that for every fixed B, the function $s\mapsto\mathtt{M}(B,s)$ defines a predictable process, while $B\mapsto\mathtt{M}(B,s)$ is a (random) measure for every fixed s. It is clear that Itô formula is suitable modified.

• *Remark* 3.14. In particular, Theorem 3.13 can be formulated as follows. Let $X=(X_1,\dots,X_d)$ be a semi-martingale, M be local-martingale and g, a and $\mathtt{M}$ be local integrable predictable processes such that

$$X(t)-X(0)-M(t)=\int_0^t g(s)\mathrm{d}s,\quad\forall t\ge 0,$$
$$\langle M_i^c,M_j^c\rangle(t)=\int_0^t a_{ij}(s)\mathrm{d}s,\quad\forall t\ge 0,$$
$$\nu_M^p(B,]0,t])=\int_0^t\mathtt{M}(B,s)\mathrm{d}s,\quad\forall B\in\mathcal{B}(\mathbb{R}_*^d),\ t\ge 0,$$

where M^c is the continuous part of M and ν_M^p is the compensator of the integer

measure ν_M associated with M. Then

$$\varphi(X(t),t) = \varphi(X(0),0) + \int_0^t \big[(\partial_s + A_X)\varphi(X(s-),s)\big]\mathrm{d}s + \\ + \sum_{i=1}^d \int_0^t \partial_i\varphi(X(s-),s)\mathrm{d}M^c(s) + \\ + \int_{\mathbb{R}^d_*\times]0,t]} \big[\varphi(X(s-)+z,s) - \varphi(X(s-),s)\big]\tilde{\nu}_M(\mathrm{d}z,\mathrm{d}s),$$

where

$$(\partial_s + A_X)\varphi(\cdot,s) = \partial_s\varphi(\cdot,s) + \sum_{i=1}^d g_i(s)\partial_i\varphi(\cdot,s) + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(s)\partial^2_{ij}\varphi(\cdot,s) + \\ + \int_{\mathbb{R}^d_*} \big[\varphi(\cdot+z,s) - \varphi(\cdot,s) - \sum_{i=1}^d z_i\partial_i\varphi(\cdot,s)\big]\mathtt{M}(\mathrm{d}z,s),$$

for every bounded function $\varphi(x,t)$ in $\mathbb{R}^d \times [0,\infty)$, which is twice continuously differentiable in x, once continuously differentiable in t with all derivatives bounded. In general, if the semi-martingale $X = V + M$, where V is a continuous process with local bounded variation and M a locally square-integrable martingale then $\phi(X(t)) = \phi(X(0)) + V_\phi(t) + M_\phi(t)$ is a semi-martingale with

$$V_\phi(t) = \int_0^t \nabla\phi(X(s-))\cdot\mathrm{d}V(s) + \frac{1}{2}\int_0^t \mathrm{Tr}\left(D^2\phi(X(s-))\mathrm{d}\langle M^c\rangle(s)\right] + \\ + \int_{\mathbb{R}^d_*\times]0,t]} \big[\phi(X(s-)+z) - \phi(X(s-)) - z\cdot\nabla\phi(X(s-))\big]\nu^p_M(\mathrm{d}z,\mathrm{d}s)$$

and

$$M_\phi(t) = \int_0^t \nabla\phi(X(s-))\cdot\mathrm{d}M^c(s) + \\ + \int_{\mathbb{R}^d_*\times]0,t]} \big[\phi(X(s-)+z) - \phi(X(s-))\big]\tilde{\nu}_M(\mathrm{d}z,\mathrm{d}s),$$

for any bounded twice continuously differentiable ϕ with all derivative bounded. This is usually referred to as the Itô formula for semi-martingales, which can be written as above, by means of the associated integer measure, or as in Theorem 3.13. □

• *Remark* 3.15. In general, if $\{x(t) : t \geq 0\}$ is a real-valued predictable process with local bounded variation (so $x(t+)$ and $x(t-)$ exist for every t) and $\{y(t) : t \geq 0\}$ is a (cad-lag) semi-martingale then we have

$$\mathrm{d}\big(x(t)y(t)\big) = x(t)\mathrm{d}y(t) + y(t-)\mathrm{d}x(t),$$
$$\mathrm{d}[x,y](t) = \big(x(t+) - x(t-)\big)\mathrm{d}y(t),$$
$$\mathrm{d}|y(t)|^2 = 2y(t-)\mathrm{d}y(t) + \mathrm{d}[y,y](t),$$

with the above notation. By the way, note that $\mathrm{d}x(t) = \mathrm{d}x(t+)$ and $x(t)\mathrm{d}y(t) = x(t-)\mathrm{d}y(t)$. □

Approximations and Comments

A double sequence $\{\tau_m(n) : n, m \geq 0\}$ of stopping times is called a *Riemann sequence* if $\tau_m(0,\omega) = 0$, $\tau_m(n,\omega) < \tau_m(n+1,\omega) < \infty$, for every $n = 0, 1, \ldots, N_m(\omega)$ and as $m \to 0$ we have

$$\sup_n\{\tau_m(n+1,\omega) \wedge t - \tau_m(n,\omega) \wedge t\} \to 0, \quad \forall\, t > 0,$$

for every ω, i.e., the mesh or norm of the partitions or subdivisions restricted to each interval $[0,t]$ goes to zero. A typical example is the dyadic partition $\tau_m(n) = n2^{-m}$, $m = 1, 2, \ldots$, and $n = 0, 1, \ldots, 2^m$, which is deterministic. We have the following general results:

Theorem 3.16 (Riemann sequence)**.** *Let X be a semi-martingale, Y be a cadlag adapted process and $\{\tau_m(n) : n, m \geq 0\}$ be a Riemann sequence. Then the sequence of Riemann-Stieltjes sums, $m \geq 0$,*

$$\sum_n Y(\tau_m(n)-)\big(X(\tau_m(n+1) \wedge t) - X(\tau_m(n) \wedge t)\big)$$

converges in probability, uniformly on each compact interval, to the stochastic integral

$$\int_{]0,t]} Y(s-)\mathrm{d}X(s).$$

Moreover, if Y is also a semi-martingale then the optional process

$$t \mapsto \sum_n \big(X(\tau_m(n+1) \wedge t) - X(\tau_m(n) \wedge t)\big) \times \\ \times \big(Y(\tau_m(n+1) \wedge t) - Y(\tau_m(n) \wedge t)\big)$$

converges in probability, uniformly on each compact interval, to the optional quadratic covariation process $[X, Y]$.

Proof. For instance to prove the first convergence, it suffices to see that the above Riemann-Stieltjes sums are equal to the stochastic integral

$$\int_{]0,t]} Y_m(s)\mathrm{d}X(s),$$

where $Y_m(s) = Y(\tau_m(n))$ for any s in the stochastic interval $\rrbracket \tau_m(n), \tau_m(n+1) \rrbracket$, is clearly a predictable left continuous process for each $m \geq 0$.

The proof of the second convergence is essentially based on the integration by part formula (3.71), which actually can be used to define the optional quadratic covariation process.

For instance, a full proof can be found in Jacod and Shiryaev [84, Proposition 4.44 and Theorem 4.47, Chapter 1, pp. 51–52]. □

The estimate (3.52) of the previous section for for Poisson integral, namely, for any p in $(0,2]$ there exists a positive constant $C = C_p$ (actually $C_p = (4-p)/(2-p)$ if $0 < p < 2$ and $C_2 = 4$) such that for any adapted (measurable) process $f(\zeta, s)$ (actually, the predictable version is used) we have

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} \Big| \int_{\mathbb{R}^m_* \times]0,t]} f(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)\Big|^p \Big\} \le \\ \le C\,\mathbb{E}\Big\{ \Big[\int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |f(\zeta,s)|^2 \pi(\mathrm{d}\zeta)\Big]^{p/2}\Big\}, \quad (3.74)$$

for every stopping time T. The case $p > 2$ is a little more complicate and involves Itô formula. Indeed, for the sake of simplicity let us consider the one-dimensional case, use Itô formula with the function $x \mapsto |x|^p$ and the process

$$X(t) = \int_{\mathbb{R}^m_* \times]0,t]} f(\zeta,s)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s), \quad \forall t \ge 0$$

to get

$$\mathbb{E}\{|X(t)|^p\} = \mathbb{E}\Big\{ \int_0^t \mathrm{d}s \int_{\mathbb{R}^m_*} \Big[|X(s) + f(\zeta,s)|^p - |X(s)|^p - \\ - p\,|X(s)|^{p-2}X(s)f(\zeta,s)\Big]\pi(\mathrm{d}\zeta)\Big\} = p(p-1)\times \\ \times \mathbb{E}\Big\{ \int_0^t \mathrm{d}s \int_0^1 (1-\theta)\mathrm{d}\theta \int_{\mathbb{R}^m_*} |X(s) + \theta f(\zeta,s)|^{p-2}|f(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big\}.$$

The integrand is bounded as follows

$$|X(s) + \theta f(\zeta,s)|^{p-2}|f(\zeta,s)|^2 \le 2^{p-2}\big[|X(s)|^{p-2}|f(\zeta,s)|^2 + |f(\zeta,s)|^p\big],$$

and by means of the Doob's maximal inequality, we deduce

$$\mathbb{E}\{ \sup_{0\le t\le T} |X(t)|^p\} \le \tilde{C}_p \Big[\mathbb{E}\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |f(\zeta,s)|^p \pi(\mathrm{d}\zeta)\Big\} + \\ + \mathbb{E}\Big\{ \Big(\sup_{0\le t\le T} |X(t)|^{p-2}\Big)\Big(\int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |f(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big)\Big\}\Big],$$

for some constant $\tilde{C}_p$ depending only on p. Hence, the simple inequality for any $a, b, \varepsilon \ge 0$,

$$ab \le \frac{p-2}{p}(\varepsilon a)^{p/(p-2)} + \frac{2}{p}\big(\frac{a}{\varepsilon}\big)^{p/2}$$

and the Hölder inequality yield the following variation of (3.74): for any $p > 2$

there exists a constant $C = C_p$ depending only on p such that

$$\mathbb{E}\Big\{ \sup_{t\le T} \Big| \int_{\mathbb{R}^m_* \times]0,t]} f(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)\Big|^p \Big\} \le C\mathbb{E}\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |f(\zeta,s)|^p \pi(\mathrm{d}\zeta)\Big\} + \\ + \mathbb{E}\Big\{ \Big[\int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |f(\zeta,s)|^2 \pi(\mathrm{d}\zeta)\Big]^{p/2}\Big\}, \quad (3.75)$$

for any adapted (measurable) process $f(\zeta,s)$ and any stopping time T.

• *Remark* 3.17. These estimates for the moments of a stochastic integral can be partially generalized to some other type of integral, e.g., let M be a d-dimensional continuous square integrable martingale with predictable quadratic covariation process $\langle M_i, M_j\rangle = \mathrm{d}\ell$ if $i = j$ and $\langle M_i, M_j\rangle = 0$ if $i \ne j$, where ℓ is a continuous nondecreasing adapted process satisfying

$$\mathbb{E}\{\ell(t) - \ell(s) \,|\, \mathcal{F}(s)\} \le h(t-s),$$

for every $t \ge s \ge 0$ and for some monotone function h form $[0,\infty)$ into itself. Using the integration by part formula

$$[\ell(t) - \ell(s)]^k = k \int_s^t [\ell(t) - \ell(r)]^{k-1}\, \mathrm{d}\ell(r)$$

and by induction on k, we can show that

$$\mathbb{E}\big\{ [\ell(t) - \ell(s)]^k \,|\, \mathcal{F}(s)\big\} \le k!\, [h(t-s)]^k,$$

for every $t \ge s \ge 0$ and any $k \ge 1$. Similarly, by means of Itô formula, the sub-martingale inequality and by induction, we can prove that for every positive integer k there exists a constant $C = C(k,d)$, depending only on k and the dimension d, such that

$$\mathbb{E}\big\{ \sup_{s\le r\le t} |M(r) - M(s)|^k \,|\, \mathcal{F}(s)\big\} \le C(k,d)\, [h(t-s)]^{k/2},$$

for every $t \ge s \ge 0$ and any $k \ge 1$. □

It is clear that the above Itô calculus can be extended to non deterministic smooth functions, i.e., predictable processes $\varphi(t,x,\omega)$ which are continuously differentiable in t and twice-continuously differentiable in x. The rule given in this section is unchanged. As we may expect, if for each x the process $t \mapsto \varphi(t,x,\omega)$ is a local-martingale (which has not bounded variation paths) then Itô calculus applies and another bracket $[\cdot,\cdot]$ with respect to this process should appear.

• *Remark* 3.18. In a given complete filtered space, an adapted increasing (locally integrable) cad-lag process A is called *natural* if for every bounded (not necessarily continuous) martingale M we have

$$\int_{[0,t]} M(s)\mathrm{d}A(s) = \int_{[0,t]} M(s-)\mathrm{d}A(s), \quad \forall\, t \ge 0.$$

This is equivalent to the concept of *predictable* process. On the other hand, a quasi left continuous (increasing or martingale) cad-lag process is also called *regular*. It turns out that an adapted increasing cad-lag process is continuous if and only if it is natural and regular. The reader is referred to the books Kallenberg [88] and Yeh [181] for a comprehensive treatment. □

• *Remark* 3.19. The operational Itô formula is better understood in its simplest product form, i.e., let X and Y be two d-dimensional Itô processes with jumps (see Definition 3.11), namely

$$\mathrm{d}X(t) = a^X(t)\mathrm{d}t + b^X(t)\mathrm{d}w(t) + \int_{\mathbb{R}^m_*} \gamma^X(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \geq 0,$$
$$\mathrm{d}Y(t) = a^Y(t)\mathrm{d}t + b^Y(t)\mathrm{d}w(t) + \int_{\mathbb{R}^m_*} \gamma^Y(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \geq 0,$$

then

$$\mathrm{d}\big(X_i(t)Y_j(t)\big) = X_i(t-)\mathrm{d}Y_j(t) + \big(\mathrm{d}X_i(t)\big)Y_j(t-) + \\ + \sum_k b^X_{ik}(t)b^Y_{jk}(t)\mathrm{d}t + \int_{\mathbb{R}^m_*} \gamma^X_i(\zeta,t)\gamma^Y_j(\zeta,t)p(\mathrm{d}\zeta,\mathrm{d}t),$$

for any $t \geq 0$. Note the independent role of the diffusion and jumps coefficients. Moreover, the last (jump) integral is not a *pure* stochastic integral, it is with respect to $p(\mathrm{d}\zeta,\mathrm{d}t)$ which can be written as $\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t) + \pi(\mathrm{d}\zeta)\mathrm{d}t$. We can go further and make explicit each term, i.e.,

$$X_i(t)\mathrm{d}Y_j(t) = X_i(t-)\mathrm{d}Y_j(t) = X_i(t)a^Y_j(t)\mathrm{d}t + X_i(t)b^Y_j(t)\mathrm{d}w(t) + \\ + \int_{\mathbb{R}^m_*} X_i(t)\gamma^Y_j(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t),$$

where $X_i(t)$ goes inside the stochastic integral indistinctly as either $X_i(t)$ or it predictable projection $X_i(t-)$. □

Similarly to above Remark 3.19, a operational (generalized) Itô formula can be written for processes driven by local-martingales. Let $M = M^c + M^d$ be a quasi-left continuous local square-integrable martingale in $\mathbb{R}^n$ written as the sum of a continuous local-martingale $\{M^c_i : i = 1,\dots,n\}$ with predictable variation process $\{\langle M^c_i\rangle : i = 1,\dots,n\}$, satisfying $\langle M^c_i, M^c_j\rangle = 0$ if $i \neq j$, and a purely discontinuous local-martingale $\{M^d_i : i = 1,\dots,n\}$ which yields an integer measure $\nu_{_M}$ with compensator $\nu^p_{_M}$ and (local) martingale measure $\tilde{\nu}_{_M} = \nu_{_M} - \nu^p_{_M}$. Note that

$$\int_{]0,t]} \alpha(s)\mathrm{d}M^d_i(s) = \int_{\mathbb{R}^d\times]0,t]} \alpha(s)\zeta_i\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}s).$$

Also let $\{V^c_i : i = 1,\dots,d\}$ be a local bounded variation continuous process, non-anticipating with respect to M. Now, if X and Y are two d-dimensional

processes of the form

$$\mathrm{d}X(t) = a^X(t)\mathrm{d}V^c(t) + b^X(t)\mathrm{d}M^c(t) + \int_{\mathbb{R}^m_*} \gamma^X(\zeta,t)\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \geq 0,$$
$$\mathrm{d}Y(t) = a^Y(t)\mathrm{d}V^c(t) + b^Y(t)\mathrm{d}M^c(t) + \int_{\mathbb{R}^m_*} \gamma^Y(\zeta,t)\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \geq 0,$$

then

$$\mathrm{d}\big(X_i(t)Y_j(t)\big) = X_i(t-)\mathrm{d}Y_j(t) + \big(\mathrm{d}X_i(t)\big)Y_j(t-) + \\ + \sum_k b^X_{ik}(t)b^Y_{jk}(t)\mathrm{d}\langle M^c_k\rangle(t) + \int_{\mathbb{R}^m_*} \gamma^X_i(\zeta,t)\gamma^Y_j(\zeta,t)\nu_{_M}(\mathrm{d}\zeta,\mathrm{d}t),$$

for any $t \geq 0$. In particular, in term of the purely jumps (local) martingale M^d_k, i.e., $\gamma_i(\zeta,t) = \sum_k c_{ik}(t)\zeta_k$ for both processes, we have

$$\int_{\mathbb{R}^m_*} \gamma^X_i(\zeta,t)\gamma^Y_j(\zeta,t)\nu_{_M}(\mathrm{d}\zeta,\mathrm{d}t) = \\ = \frac{1}{2}\sum_{k,\ell}\int_{]0,t]} \big(c^X_{ik}(s)c^Y_{j\ell}(s) + c^X_{i\ell}(s)c^Y_{jk}(s)\big)\mathrm{d}[M^d_k, M^d_\ell](s),$$

where $[M^d_k, M^d_\ell]$ is the optional quadratic (matrix) variation, i.e.,

$$[M^d_k, M^d_\ell](t) = \sum_{s\leq t}\big(M^d_k(s) - M^d_k(s-)\big)\big(M^d_\ell(s) - M^d_\ell(s-)\big),$$

Hence, if c^X_{ik} and $c^Y_{j\ell}$ are cad-lag then

$$\int_{\mathbb{R}^m_*} \gamma^X_i(\zeta,t)\gamma^Y_j(\zeta,t)\nu_{_M}(\mathrm{d}\zeta,\mathrm{d}t) = \frac{1}{2}\sum_{k,\ell}\sum_{0<s\leq t}\big(c^X_{ik}(s-)c^Y_{j\ell}(s-) + \\ + c^X_{i\ell}(s-)c^Y_{jk}(s-)\big)\big(M^d_k(s) - M^d_k(s-)\big)\big(M^d_\ell(s) - M^d_\ell(s-)\big).$$

Moreover, if each coordinate is orthogonal to each other (i.e., $[M^d_i, M^d_j] = 0$, for $i \neq j$), equivalent to the condition that there are no simultaneous jumps of M^d_i and M^d_j, then only the terms $k = \ell$ and the $1/2$ is simplified. Clearly, there is only a countable number of jumps and

$$\mathbb{E}\Big\{\sum_{0<s\leq t\wedge\tau_n}\big[\big(c^X_{ik}(s-)\big)^2 + \big(c^Y_{jk}(s-)\big)^2\big]\big(M^d_k(s) - M^d_k(s-)\big)^2\Big\} < \infty,$$

for every $t > 0$, where $\{\tau_n\}$ is some sequence the stopping times increases to ∞ almost surely, i.e., the above series is absolutely convergence (localized) in the L^2-sense. If c^X_{ik} or c^Y_{jk} is not cad-lag, then a predictable version should be used in the series. Furthermore, if the initial continuous martingale M^c do not have

orthogonal components then we may modify the drift and reduce to the above case, after using Gram-Schmidt orthogonalization procedure, or alternatively, we have a double (symmetric) sum,

$$\frac{1}{2}\sum_{k,\ell}[b^X_{ik}(t)b^Y_{j\ell}(t)+b^X_{i\ell}(t)b^Y_{jk}(t)]\mathrm{d}\langle M^c_k,M^c_\ell\rangle(t)$$

instead of the single sum in k. On the other hand, to include discontinuous process V or a non-necessarily quasi-left continuous local-martingale, we need to carefully consider possible deterministic jumps. Indeed, denoting by δ the jump operator, i.e., $\delta X(t)=\big(X(t)-X(t-)\big)$ for a cad-lag process X, the relation

$$\delta\big(X_i(t)Y_j(t)\big)=\big(\delta X_i(t)\big)Y_j(t-)+X_i(t-)\big(\delta Y_j(t)\big)\big)+\big(\delta X_i(t)\big)\big(\delta Y_j(t)\big)$$

shows the general expression

$$\begin{aligned}\mathrm{d}\big(X_i(t)Y_j(t)\big)&=X_i(t-)\mathrm{d}Y_j(t)+\big(\mathrm{d}X_i(t)\big)Y_j(t-)+\\&\quad+\sum_k b^X_{ik}(t)b^Y_{jk}(t)\mathrm{d}\langle M^c_k\rangle(t)+\mathrm{d}\Big(\sum_{s\in]0,t]}\big(\delta X_i(s)\big)\big(\delta Y_j(s)\big)\Big),\end{aligned}$$

which makes sense as a stochastic integral after compensating the jumps. Since the jumps of $X_i(t)$ or $X_i(t)$ are due only to $V^d(t)=\sum_{0<s\le t}\delta V(s)$ and $M^d(t)$, we have

$$\begin{aligned}&\big(\delta X_i(t)\big)\big(\delta Y_j(t)\big)=\\&\quad=\sum_{k,\ell}\big(a^X_{ik}(s)\delta V_k(s)+c^X_{ik}(s)\delta M^d_k(s)\big)\big(a^Y_{j\ell}(s)\delta V_\ell(s)+c^X_{j\ell}(s)\delta M^d_\ell(s)\big).\end{aligned}$$

Hence, without loss of generality, it seems better to take $V=V^c$ continuous and put all jumps into the integer measure ν, which may not be quasi-left continuous. This is the case of a special semi-martingale $S(t)$, $S(0)=0$, written in its canonical form as $V+M^c+M^d$, where $V=V^c$ if S is quasi-left continuous. Essentially, this discontinuity (of V) imposes (implicitly) the condition that the drift must be continuous at each predictable jump (jumps non switchable to M^d, e.g., deterministic jumps), otherwise, the integrability of the drift with respect to a discontinuous V may be an issue, i.e., in the Stieltjes-Riemann sense may be not integrable and in the Stieltjes-Lebesgue sense may yield distinct values, depending on whether $a(s)$, $a(s+)$ or $a(s-)$ is used. This never arrive in the stochastic integral.

• *Remark* 3.20. Let X be a 1-dimensional Itô processes with jumps (see Definition 3.11), namely

$$\mathrm{d}X(t)=a(t)\mathrm{d}t+b(t)\mathrm{d}w(t)+\int_{\mathbb{R}^m_*}\gamma(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t),\quad\forall t\ge 0,$$

with $X(0)=0$, and such that almost surely we have $\gamma(\zeta,t)>-1$ or equivalently $\inf\big\{\delta X(t)\,:\,t>0\big\}>-1$, where $\delta X(t)=X(t)-X(t-)$ is the jump at time t.

Based on the inequalities $r - \ln(1+r) \geq 0$ if $r > -1$ and $r - \ln(1+r) \leq r^2/2$ if $r \geq 0$, we deduce that the infinite product $\prod_{0\leq s\leq t} \big[1+\delta X(s)\big]\mathrm{e}^{-\delta X(s)}$ is almost surely finite and positive. Moreover, for every $t \geq 0$, either the exponential expression

$$\mathcal{E}_X(t) = \exp\Big\{X(t) - \frac{1}{2}\int_0^t \sum_{k=1}^n |b_k(s)|^2 \mathrm{d}s\Big\} \prod_{0\leq s\leq t} \big[1+\delta X(s)\big]\mathrm{e}^{-\delta X(s)},$$

or the log-differential expression

$$\mathrm{d}\ln\big(\mathcal{E}_X(t)\big) = \big[a(t) - \frac{1}{2}|b(t)|^2\big]\mathrm{d}t + \int_{\mathbb{R}_*^m} \ln\big(1+\gamma(\zeta,t)\big)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t) + \\ + \int_{\mathbb{R}_*^m} \big[\ln\big(1+\gamma(\zeta,t)\big) - \gamma(\zeta,t)\big]\pi(\mathrm{d}\zeta)$$

defines a 1-dimensional Itô processes with jumps satisfying

$$\mathrm{d}\mathcal{E}_X(t) = \mathcal{E}_X(t-)\,\mathrm{d}X(t),$$

which is called exponential martingale. Recall that $\tilde{p} = -\pi$ so that if γ has a finite π-integral (i.e., the jumps are of bounded variation) then

$$\int_{\mathbb{R}_*^m} \ln\big(1+\gamma(\zeta,t)\big)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t) + \int_{\mathbb{R}_*^m} \big[\ln\big(1+\gamma(\zeta,t)\big) - \gamma(\zeta,t)\big]\pi(\mathrm{d}\zeta) = \\ = \int_{\mathbb{R}_*^m} \ln\big(1+\gamma(\zeta,t)\big)p(\mathrm{d}\zeta,\mathrm{d}t) - \int_{\mathbb{R}_*^m} \gamma(\zeta,t)\pi(\mathrm{d}\zeta),$$

as formally expected. For instance, see Applebaum [1, Chapter 5, pp 246-291] or Jacod and Shiryaev [84, Section III.3, pp. 152–166]. □

3.3.3 Non-Anticipative Processes

The concept of non-anticipative or non-anticipating is rather delicate, and usually it means adapted or strictly speaking, if a process is adapted then it should be non-anticipative. For instance, a random process x is called non-anticipative with respect to a Markov process y if the past of x is independent of the future of y given the present of y, this means that given a realization y of a Markov process in some probability space $(\Omega, \mathcal{F}, P)$ with values in a topological space Y then any process x with values in some topological space X is called non-anticipative with respect to y if for any bounded Borel measurable functions f, g and h and times $s_1 < \cdots < s_n \leq t \leq t_1 < t_2 < \cdots < t_n$, we have

$$\mathbb{E}\big\{f(x_{s_1},\ldots,x_{s_n})\,g(y_t)\,h(y_{t_1},\ldots,y_{t_n})\big\} = \\ = \mathbb{E}\big\{\mathbb{E}\{f(x_{s_1},\ldots,x_{s_n})\,|\,y_t\}\,g(y_t)\,\mathbb{E}\{h(y_{t_1},\ldots,y_{t_n})\,|\,y_t\}\big\},$$

where n is arbitrary. Note that the three functions f, g and h may be taken only bounded continuous, as long as the Baire σ-algebra (the one generated

by continuous functions on X and Y) and the Borel σ-algebra coincide, e.g., if (X,d) is a metric space then $F=\{x\in X : d(x,X)=\inf_{u\in F} d(x,u)=0\}$ for any closed subset F of X, so $x\mapsto d(x,F)$ is continuous, and so both σ-algebras coincide. Since Jakubowski topology is weaker that a metrizable topology, the Baire and the Borel σ-algebras coincide in this case too. Usually, X and Y are some $\mathbb{R}^n$ and the processes x and y are at least stochastically right continuous. It is convenient to take a cad-lag version of x and y if possible.

On the other hand, if y is a random process with independent increments and $y_0=0$, then a non-anticipative process x is a process such that the past of x is independent of the increments of y given the present of y, i.e, for any bounded Borel measurable functions f, g and h and times $s_1<\cdots<s_n\le t\le t_1<t_2<\cdots<t_n$, we have

$$\begin{aligned}\mathbb{E}\big\{f(x_{s_1},\dots,x_{s_n})\,g(y_t)\,h(y_{t_2}-y_{t_1},\dots,y_{t_n}-y_{t_{n-1}})\big\}=\\ =\mathbb{E}\big\{\mathbb{E}\{f(x_{s_1},\dots,x_{s_n})\,|\,y_t\}\,g(y_t)\,\mathbb{E}\{h(y_{t_2}-y_{t_1},\dots,y_{t_n}-y_{t_{n-1}})\,|\,y_t\}\big\},\end{aligned}$$

where n is arbitrary. In any case, note that (contrary to the adapted case) if x_1 and x_2 are non-anticipative then the cartesian product (x_1,x_2) is not necessarily non-anticipative. Recall that y is a process of independent increments (i.e., $y(t_1),\dots,y(t_n)$ are independent of $y(s_2)-y(s_1)$, for any $t_1<\cdots<t_n<s_1<s_2$), if and only if $y=m+a$, where m is a semi-martingale (and a process of independent increments) and a is a deterministic cad-lag process (e.g., see Jacod and Shiryaev [84, Theorem II.5.1, p. 114]).

Perhaps a better concept is the following:

Definition 3.21. For a given a process y in a probability space $(\Omega,\mathcal{F},P)$ we define the *non-anticipative* filtration $\mathbb{A}=\{\mathcal{A}(t) : t\ge 0\}$, where $\mathcal{A}(t)$ is the σ-algebra composed by all sets in $\mathcal{F}$ which are independent of $y_{t_1}-y_{t_0},\dots y_{t_n}-y_{t_{n-1}}$, for any $t\le t_0<t_1<\cdots<t_n$, and $n\ge 1$. So a measurable process x is non-anticipative with respect to y if it is adapted to $\mathbb{A}$, i.e., if for any bounded Borel measurable functions f and g we have

$$\begin{aligned}\mathbb{E}\{f(x_{s_1},\dots,x_{s_n})\,g(y_{t_1}-y_{t_0},\dots,y_{t_n}-y_{t_{n-1}})\}=\\ =\mathbb{E}\{f(x_{s_1},\dots,x_{s_n})\}\,\mathbb{E}\{g(y_{t_1}-y_{t_0},\dots,y_{t_n}-y_{t_{n-1}})\},\end{aligned}$$

for any times $s_1<\cdots<s_n\le t_0<t_1<\cdots<t_n$. □

Clearly, once the non-anticipative filtration $\mathbb{A}$ has been defined, the concept of a non-anticipative process reduces to being adapted to the non-anticipative filtration $\mathbb{A}$. However, a good part for this concept is the fact of being a finite-dimensional property, i.e., if x' and y' two processes in another probability space $(\Omega',\mathcal{F}',P')$ with the same (joint) finite-dimensional distributions as x and y then x' is also non-anticipative with respect to y'.

Alternatively, if y is a random process with orthogonal (or uncorrelated) increments and $y_0=0$, then any random process x which is orthogonal (or uncorrelated) to the increments of y could be called weakly non-anticipative,

i.e., if

$$\mathbb{E}\big\{x_s \cdot (y_{t_2} - y_{t_1})\big\} = \mathbb{E}\{x_s\} \cdot \mathbb{E}\{(y_{t_2} - y_{t_1})\},$$

for any $0 \le s \le t_1 < t_2$, where the $\cdot$ denotes the scalar product. Certainly, an orthogonal process x is weakly non-anticipative if x_t belongs to the closed linear span of the variables $y_{s_1}, \dots, y_{s_n}$, with $0 \le s_1 < \dots < s_n \le t$. All this means that any information on x does not help to gain some extra information on the characteristics of y. However, the following concept seems better for martingales.

Recall that for a $\mathbb{R}^d$-valued stochastic process y, the martingale property reads as follows:

$$\mathbb{E}\big\{\big(y(t) - y(s)\big)\, f\big(y(s_1), \dots, y(s_n)\big)\big\} = 0,$$

for any bounded continuous functions f and any times $s_1 < \dots < s_n \le s \le t$. This is a property finite-dimensional (i.e., any other stochastic process y' satisfies the above martingale properties provided $\mathbb{E}\big\{f\big(y(s_1), \dots, y(s_n)\big)\big\} = \mathbb{E}'\big\{f\big(y'(s_1), \dots, y'(s_n)\big)\big\}$, for any bounded continuous functions f and any times $s_1 < \dots < s_n$), which makes sense for processes satisfying $\mathbb{E}\{|y(t)|\} < \infty$ for every $t \ge 0$ (or for a time-localization, as in the case of local martingales). However, most of the useful results for martingale processes requires a separable martingale, and separability is not finite-dimensional property.

Thus, of particular interest for us is the case when y is a (local) martingale.

Definition 3.22. Let y be a $\mathbb{R}^d$-valued (separable) martingale (with zero mean) in some probability space $(\Omega, \mathcal{F}, P)$. A process x is called *weakly non-anticipative* with respect to y if for any bounded continuous functions f and g and any times $s_1 < \dots < s_n \le s \le t$ and $s'_1 < \dots < s'_n \le s$, we have

$$\mathbb{E}\big\{\big(y(t) - y(s)\big)\, f\big(x(s_1), \dots, x(s_n)\big)\, g\big(y(s'_1), \dots, y(s'_n)\big)\big\} = 0.$$

If y is a martingale relative to a filtration $\mathbb{F} = (\mathcal{F}_t : t \ge 0)$ then we say that x is weakly non-anticipative with respect to y (and $\mathbb{F}$) if for any bounded continuous functions f and any times $s_1 < \dots < s_n \le s \le t$, we have

$$\mathbb{E}\big\{\big(y(t) - y(s)\big)\, f\big(x(s_1), \dots, x(s_n)\big)\, z_s\big\} = 0.$$

where z_s is any bounded $\mathcal{F}_s$-measurable function. Clearly, this notion extends to local-martingales or semi-martingales. This means that the stochastic process x does not change the martingale property of y. □

It is clear that weakly non-anticipative is a finite-dimensional distribution property when the filtration is not mentioned, i.e., if x' and y' two processes in another probability space $(\Omega', \mathcal{F}', P')$ with the same finite-dimensional distributions and y' being integrable, then y is a martingale and x is non-anticipative with respect to y if and only if then x' is non-anticipative with respect to y'. Also, if $\mathcal{F}(x,t)$ denotes the σ-algebra generated by the random variables $x(s)$,

$0 \le s \le t$, then x is non-anticipative with respect to y if $\mathcal{F}(x,t) \vee \mathcal{F}(y,t)$ is orthogonal to the increments $y(b) - y(a)$, for any $b > a \ge t$, where $\mathcal{F}(x,t) \vee \mathcal{F}(y,t)$ is the minimal σ-algebra containing both $\mathcal{F}(x,t)$ and $\mathcal{F}(y,t)$.

Recall that a general (local) martingale is a (local) integrable process y satisfying the martingale property, namely,

$$\mathbb{E}\{y(t) \,|\, \mathcal{F}(y,s)\} = y(s), \quad \forall t \ge s \ge 0,$$

or equivalently

$$\mathbb{E}\big\{\big(y(t) - y(s)\big)\, f\big(y(s_1), \dots, y(s_n)\big)\big\} = 0, \quad \forall 0 \le s_1 < \cdots < s_n \le s < t,$$

and any arbitrary bounded continuous function f. Note that when the prefix *general* (or separable) is used, we mean that no particular version (or that a separable version) has been chosen.

Thus, if x is an adapted process to a martingale y relative to the filtration $\mathbb{F}$ then $\mathcal{F}_t$ contains $\mathcal{F}(x,t) \vee \mathcal{F}(y,t)$ and x results non-anticipative with respect to y and $\mathbb{F}$. Note that if x_1 and x_2 are two weakly non-anticipative processes then the cartesian product (x_1, x_2) is not necessarily weakly non-anticipative, clearly, this is not the case for adapted processes. Conversely, if x is weakly non-anticipative with respect to a general (local) martingale y we deduce that x is certainly adapted to $\mathcal{F}(t) = \mathcal{F}(x,t) \vee \mathcal{F}(y,t)$ and also that y satisfies the martingale property relative to $\mathcal{F}(t)$, instead of just $\mathcal{F}(y,t)$. Moreover, if y is cad-lag then the martingale property holds for $\mathcal{F}^+(t) = \cap_{\varepsilon>0}\mathcal{F}(t+\varepsilon)$.

Now, if we assume that y is a general martingale (non necessarily cad-lag) with $t \mapsto \mathbb{E}\{y(t)\}$ cad-lag (which is a finite-dimensional distribution property) then there is a cad-lag version of y, still denoted by y, where the above argument applies. Therefore, starting with a process x weakly non-anticipative with respect to y (satisfying the above conditions) we obtain a filtration $\{\mathcal{F}^+(t) : t \ge 0\}$ such that x is adapted and y is a (local) martingale. If the function $t \mapsto \mathbb{E}\{y(t)\}$ is continuous then the process y has also a cag-lad version (left continuous having right-hand limit) which is denoted by y_-, with $y_-(0) = y(0)$ and $y_-(t) = \lim_{\varepsilon \to 0} y(t-\varepsilon)$, $t > 0$. In this case, x is also weakly non-anticipative with respect to y_-, since any version of y can be used.

Recall that with the above notation, a process x is progressively measurable if $(t,\omega) \mapsto x(t,\omega)$, considered as defined on $[0,T] \times \Omega$ is measurable with respect to the product σ-algebra $\mathcal{B}([0,T]) \times \mathcal{F}(x,T)$ or $\mathcal{B}([0,T]) \times \mathcal{F}(T)$, if the family of increasing σ-algebra $\{\mathcal{F}(t) : t \ge 0\}$ is a priori given. Progressively measurability and predictability are not a finite-dimensional distribution property, but for a given filtration and assuming that x is adapted and stochastically left continuous, we can obtain a predictable version of x. Similarly, if x is adapted and stochastically right continuous then there exists a progressively measurable version.

Suppose that x and y are two weakly non-anticipative processes with respect to M, which is a cad-lag square-integrable martingale. Let M_c and $\nu_{_M}$ be their associated continuous part and integer measure, with predictable covariance $\ell_{_M} = \langle M_c \rangle$, martingale measure $\tilde{\nu}_{_M}$ and predictable jump compensator $\nu_{_{M,p}} =$

$\pi_{_M}\mathrm{d}\varrho_{_M}$, where $\pi_{_M}$ is a Levy measure and $\varrho_{_M}$ is a predictable continuous increasing process. If

$$P\Big\{\int_0^t |x(s)|^2\mathrm{d}\ell_{_M}(s) < \infty\Big\} = 1$$

and

$$P\Big\{\int_0^t \mathrm{d}\varrho_{_M}(s)\int_{\mathbb{R}_*^m} |y(\zeta,s)|^2\pi_{_M}(\mathrm{d}\zeta) < \infty\Big\} = 1$$

then the stochastic integrals

$$\int_0^t x(s)\mathrm{d}M_c(s) \qquad \text{and} \qquad \int_{\mathbb{R}_*^m\times(0,t]} y(\zeta,s)\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}s)$$

can be defined. Now, assume that in some other probability space there are processes $(x', y', M', \ell'_{_M}, \varrho'_{_M})$ having the same finite-dimensional distribution, where M' is cad-lag, $\ell'_{_M}$ and $\varrho'_{_M}$ continuous (and increasing), and x and y are almost surely integrable with respect to $\mathrm{d}\ell'_{_M}$ and $\mathrm{d}\pi_{_M}\mathrm{d}\varrho'_{_M}$, respectively. Thus, M' is a cad-lag martingale and $(x, y, \ell'_{_M}, \varrho'_{_M})$ is weakly non-anticipative with respect to M', hence, for a suitable filtration $\mathbb{F}$ the process M' remains a martingale and x and y adapted processes, $\ell'_{_M}$ and $\varrho'_{_M}$ are predictable processes. Then the associate continuous martingale M'_c and integer measure $\nu'_{_M}$ have predictable covariance $\langle M_c\rangle = \ell'_{_M}$ and predictable jump compensator $\nu'_{_{M'},p} = \pi_{_M}\mathrm{d}\varrho'_{_M}$, where $\ell'_{_M}$ and $\mathrm{d}\varrho'_{_M}$ are continuous. Hence, the stochastic integrals

$$\int_0^t x'(s)\mathrm{d}M'_c(s) \qquad \text{and} \qquad \int_{\mathbb{R}_*^m\times(0,t]} y'(\zeta,s)\tilde{\nu}_{_{M'}}(\mathrm{d}\zeta,\mathrm{d}s)$$

are defined and have the same finite-dimensional distributions. In this sense, the stochastic integral are preserved if the characteristics of the integrand and integrator are preserved.

3.3.4 Functional Representation

First we recall a basic result (due to Doob) about functional representation, e.g., see Kallenberg [88, Lemma 1.13, pp. 7-8]. Given a probability space, let b and m be two random variables with values in B and M, respectively, where $(B, \mathcal{B})$ is a Borel space (i.e., a measurable space isomorphic to a Borel subset of $[0, 1]$, e.g., a Polish space) and $(M, \mathcal{M})$ is a measurable space. Then b is m-measurable (i.e., measurable with respect to the σ-algebra generated by m) if and only if there exists a measurable function h from M into B such that $b = h(m)$.

In general, a Wiener-Poisson space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}_*^m,\ t \geq 0)$, with Lévy measure $\pi(\cdot)$ is composed by a complete filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, the stochastic process $(w(t) : t \geq 0)$ is a n-dimensional (standard) Wiener space and $(\nu(B,]0, t]) : B \in \mathbb{R}_*^m, t \geq 0)$ is an

independent (standard) Poisson measure with (intensity) Lévy measure $\pi(B) = \mathbb{E}\{\nu(B,]0,t])\}/t$, which satisfies

$$\int_{\mathbb{R}^m_*} \frac{|\zeta|^2}{1+|\zeta|}\pi(\mathrm{d}\zeta) < \infty,$$

with martingale measure $\tilde{\nu}(B,]0,t]) = \nu(B,]0,t]) - t\pi(B)$. This martingale measure $\tilde{\nu}$ is identified with the $\mathbb{R}^m$-valued (Poisson) compensated-jump process

$$\tilde{p}(t) = \int_{\mathbb{R}^m_* \times]0,t]} \zeta\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad t \geq 0,$$

in the sense that given the Poisson integer measure ν we obtain the Poisson martingale measure $\tilde{\nu}$, which yields the Poisson compensated-jump process $\tilde{p}$, and conversely, starting from a Poisson compensated-jump process $\tilde{p}$ we may define a Poisson integer measure

$$\nu(B,]0,t]) = \sum_{0<s\leq t} \mathbb{1}_{\{\tilde{p}(s)-\tilde{p}(s-)\in B\}},$$

which yields the Poisson martingale measure $\tilde{\nu}$. Thus, only the p and $\tilde{p}$ is used instead of ν and $\tilde{\nu}$, i.e., the Poisson jump-compensated process $\tilde{p}$ and the Poisson martingale measure $\tilde{p}$ are used indistinctive, and differentiated from the context.

• *Remark* 3.23. Using $\tilde{p}$ instead of $\tilde{\nu}$ in the setting of the stochastic integral results in an integrand of the form

$$\gamma_i(\zeta, t) = \sum_j \tilde{\gamma}_i(t)\zeta_j,$$

i.e., particular cases, but sufficiently general for all considerations. □

It should be clear that a Wiener-Poisson space could be called a Gauss-Poisson space or a Lévy space since $\ell = w + \tilde{p}$ is a (centered) Lévy process, where w is its continuous or Gaussian part and $\tilde{p}$ is its purely jumps or Poisson part. We prefer to emphasize the fact that a Wiener process and a Poisson measure are the driven objects. Recalling that any continuous martingale is orthogonal to any purely discontinuous martingale (with respect to a common filtration), we deduce that the processes $\phi(w) - \phi(0)$ and $\psi(\tilde{p}) - \mathbb{E}\{\psi(\tilde{p})\}$ are orthogonal martingales for any smooth functions ϕ and ψ, i.e., w and $\tilde{p}$ (or ν) are independent. Then, as long as the filtration $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$ is given and w, $\tilde{p}$ (or ν) are martingales, the independence of the Wiener process and the Poisson measure is granted.

As mentioned early, the canonical Wiener-Poisson measure P is defined on canonical sample space

$$\mathbb{C}_n \times \mathbb{D}_m = C([0,\infty), \mathbb{R}^n) \times D([0,\infty), \mathbb{R}^m)$$

as having characteristic measure

$$\mathbb{E}\big\{\exp\big[\mathtt{i}\,\xi \cdot x(t)\big]\big\} = \exp\Big\{-t\Big[\frac{|\xi_1|^2}{2} + \int_{\mathbb{R}^m_*}\big(1 - \mathrm{e}^{\mathtt{i}\,\xi_2\cdot\zeta} + \mathtt{i}\,\xi_2 \cdot \zeta\big)\pi(\mathrm{d}\zeta)\Big]\Big\},$$

for every $t \geq 0$ and $\xi = (\xi_1, \xi_2)$ in $\mathbb{R}^n \times \mathbb{R}^m$, where $x(t)$ is the projection (or evaluation) map from $\mathbb{C}_n \times \mathbb{D}_m$ into $\mathbb{R}^n \times \mathbb{R}^m$.

The canonical sample space $\mathbb{C}_n \times \mathbb{D}_m$ is a Polish space (with the locally uniform convergence in the first variable and the Skorokhod topology in the second variable) and a probability measure is then defined on the Borel σ-algebra $\mathcal{B}(\mathbb{C}_n \times \mathbb{D}_m)$ which coincides with the σ-algebra generated by the projections, i.e., $\mathcal{F}^0_\infty = \sigma(x_t : t \geq 0)$. Also, we have the (uncompleted) filtration $\mathbb{F}^0 = \{F^0_t : t \geq 0\}$ generated by the projection maps x, i.e., $\mathcal{F}^0_t = \sigma(x_s : 0 \leq s \leq t)$. This filtration induces a predictable σ-algebra $\mathcal{P}^0$ on $\mathbb{R}^+ \times \mathbb{C}_n \times \mathbb{D}_m$, i.e., $\mathcal{P}^0$ is the σ algebra generated by the sets of the form $\{0\} \times F_0$ or $(s,t] \times F_s$, for any F_s in $\mathcal{F}^0_s$, $t > s \geq 0$. Because we are working on the sample space of cad-lag processes, the predictable σ-algebra $\mathcal{P}^0$ is not the same as the *optional* σ-algebra $\mathcal{O}^0$ (also called well-measurable), generated by sets of the form $\{0\} \times F_0$ and $[s,t) \times F_s$ for any F_s in $\mathcal{F}^0_s$, any $t > s \geq 0$. Similarly, the σ-algebra $\mathcal{M}^0$ of *progressively measurable* sets is composed by all subsets A of $\Omega \times [0,\infty)$ such that $A \cap (\Omega \times [0,t])$ belongs to $\mathcal{F}^0(t) \times \mathcal{B}([0,t])$ for every $t \geq 0$. Clearly, on the sample space on $\mathbb{C}_k$ we have $\mathcal{P}^0 = \mathcal{O}^0 = \mathcal{M}^0$, while on $\mathbb{D}_k$ we have only $\mathcal{O}^0 = \mathcal{M}^0$ as expected. Sometimes, this predictable σ-algebra $\mathcal{P}^0$ is universally completed, i.e., one universally complete $\mathcal{F}^0_t$ to $\mathcal{F}^u_t$ and then $\mathcal{P}^u$ is constructed. We proceed similarly with $\mathcal{O}^0$ and $\mathcal{M}^0$ to get $\mathcal{O}^u$ and $\mathcal{M}^u$. The interested reader is referred to the book Bichteler [11], where various measurability questions are treated in great details.

• *Remark* 3.24. Let $(\Omega, \mathcal{F}, P)$ be a probability space with $\mathcal{F}$ not necessarily completed with respect to P. If y is a cad-lag process (i.e., a random variable with values in some $\mathbb{D}_k$) and $\mathcal{F}^0_t(y)$ denotes the σ-algebra generated by the random variables $\{y(s) : 0 \leq s \leq t\}$ then the filtration $\mathbb{F}^0(y) = \{\mathcal{F}^0_t(y) : t \geq 0\}$ is not necessarily neither right-continuous nor complete. However, if y is a Lévy process and we add all null sets then we obtain a complete (relative to $\mathcal{F}$) right-continuous filtration, i.e, if $\mathcal{N}$ denotes the σ-algebra of all the P-null sets in $\mathcal{F}$ then $\mathcal{F}_t(y) = \mathcal{F}^0_t(y) \vee \mathcal{N}$ satisfies $\mathcal{F}_t(y) = \cap_{s>t}\mathcal{F}_s(y)$, for any $t \geq 0$, see Proposition 2.3 in Chapter 3. In particular, if y is a Lévy process and z is a $\mathbb{R}^k$-valued stochastic process which is predictable, optional or progressively measurable relative to $\mathbb{F}(y) = \{\mathcal{F}_t(y) : t \geq 0\}$ then there exists a version of z which is predictable, optional or progressively measurable relative to $\mathbb{F}^0(y)$, and so $P\{z(t) = h(t, y|_{[0,t]})\} = 1$, for every $t \geq 0$, for some measurable function h from $\mathbb{R}^+ \times \mathbb{D}_k$ endowed with the σ-algebra $\mathcal{P}^0$, $\mathcal{O}^0$ or $\mathcal{M}^0$ into $\mathbb{R}^k$, where $y|_{[0,t]}$ means the random variable $\omega \mapsto y(\cdot \wedge t, \omega)$. □

Now we are ready to discuss the following

Definition 3.25. A *non-anticipating functional* is any Borel measurable function f from $\mathbb{C}_n \times \mathbb{D}_m$ into $\mathbb{C}_k \times \mathbb{D}_\ell$ such that the mapping $x \mapsto f(x)(t)$ with values in $\mathbb{R}^{k+\ell}$ is $\mathcal{F}^0_t$-measurable, for every $t \geq 0$. Similarly, a measurable function from $(\mathbb{R}^+ \times \mathbb{C}_n \times \mathbb{D}_m, \mathcal{P}^0)$ into $\mathbb{R}^{k+\ell}$ is called a *predictable functional*. Moreover, if the universally completed σ-algebra $\mathcal{F}^u_t$ or $\mathcal{P}^u$ is used instead of $\mathcal{F}^0_t$ or $\mathcal{P}^0$, then the prefix *universally* is added, e.g., an universally predictable functional. □

Because non-anticipating functionals take values in some $\mathbb{C}_k \times \mathbb{D}_\ell$, the notions of optional, progressively measurable and adapted functional coincide. Actually, another name for non-anticipating functionals could be progressively measurable or optional functionals. Furthermore, we may consider predictable functionals defined on $E \times \mathbb{R} \times \mathbb{C}_n \times \mathbb{D}_m$ or $\mathbb{R} \times \mathbb{C}_n \times \mathbb{D}_m \times E$, for any Polish space E, in particular $E = \mathbb{R}_*^m$ or $E = \mathbb{R}^d$. Clearly the identity map is a non-anticipating functional and the following function

$$(t, x) \mapsto x_-(t), \quad \text{where} \quad x_-(0) = 0, \quad x_-(t) = \lim_{s \to t^-} x(t), \quad t > 0,$$

is a predictable functional. Perhaps another typical example is the (stochastic) integral of a simple integrand, i.e., if $0 = t_0 < t_1 < t_2 < \cdots < t_n$ are given real numbers and g_i is a (real-valued) measurable function in $(\mathbb{C}_n \times \mathbb{D}_m, \mathcal{F}^0_{t_{i-1}})$, for every $i = 1, \ldots, n$, then

$$x \mapsto z, \qquad z(t) = \sum_{i=1}^{n} g_i(x)[x(t \wedge t_i) - x(t \wedge t_{i-1})], \ t \geq 0, \tag{3.76}$$

defines a non-anticipating functional, and $z(t) = z(t_n)$ if $t \geq t_n$. Moreover, if t_i are stopping times relative to the uncompleted filtration $\mathbb{F}^0$ then g_i should be (real-valued) $\mathcal{F}^0(t_{i-1})$-measurable functions. Furthermore, if f is a non-anticipating functional then the mapping $(t, x) \mapsto f_-(t, x)$ defined as $f_-(t, x) = f(x_-(t))$ is a predictable functional.

• *Remark* 3.26. Once a probability P is given in $\mathbb{C}_n \times \mathbb{D}_m$ we complete the predictable σ-algebra, i.e., we may complete first the filtration and then we generate the predictable σ-algebra. Thus, an integrand of stochastic integrals is a predictable process y, which is identified with its equivalence class, relative to the measure $\mathrm{d}t \times P(\mathrm{d}\omega)$, for the Wiener process, and to the measure $\pi(\mathrm{d}\zeta) \times \mathrm{d}t \times P(\mathrm{d}\omega)$, for the Poisson measure. In this case, any adapted (and measurable) process has a predictable process belonging to the same equivalence class, moreover, once a predictable (respect to the completed filtration) representative of the equivalence class has been chosen, there is a version which is predictable with respect to uncompleted filtration, i.e., a predictable functional. Hence, in the case of the canonical Wiener-Poisson integrals, any integrands may be assumed to be a predictable functionals. □

On the canonical Wiener-Poisson space, the filtration $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$ is the minimal completed filtration (and right-continuous) such that canonical process x is adapted. However, given a Wiener-Poisson space, the filtration is also assumed given and it may not be the one generated by the Wiener process w and the Poisson measure ν. Therefore, if in a given Wiener-Poisson space the filtration results to be the one generated by the Wiener process w and the Poisson measure ν, then we can consider the image measure and reduce to the canonical Wiener-Poisson space.

Suppose that on the canonical Wiener-Poisson space with Lévy measure π, we are given some real-valued adapted processes $(a_i(t) : t \geq 0, \ i = 1, \ldots, d)$,

$(b_{ik}(t) : t \geq 0,\ i = 1, \dots, d,\ k = 1, \dots, n)$ and $(\gamma_i(\zeta, t) : t \geq 0,\ \zeta \in \mathbb{R}_0^m)$, such that for every $i = 1, \dots, d$ and any $r = 1, 2, \dots$, we have

$$\int_0^T \Big[|a_i(t)| + \sum_{k=1}^n |b_{ik}(t)|^2 + \int_{\mathbb{R}_*^m} |\gamma_i(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \Big] \mathrm{d}t < \infty, \tag{3.77}$$

P-almost surely for any $T > 0$. This means that a_i, b_{ik} and γ_j are real-valued predictable functionals $a_i(t, w, \tilde{p})$, $b_{ik}(t, w, \tilde{p})$ and $\gamma_i(\zeta, t, w, \tilde{p})$. Hence, an Itô process with jumps takes the form

$$X_i(t) = \int_0^t a_i(s, w, \tilde{p}) \mathrm{d}s + \sum_{k=1}^n \int_0^t b_{ik}(s, w, \tilde{p}) \mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}_*^m \times]0,t]} \gamma_i(\zeta, s, w, \tilde{p}) \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq 0, \tag{3.78}$$

for any $i = 1, \dots, d$. We may use the notation $X(t) = X(t, \omega, w, \tilde{p})$, with ω in $\Omega = \mathbb{C}_n \times \mathbb{D}_m$, or just $X = X(w, \tilde{p})$ to emphasize the dependency on the Wiener process and the Poisson measure $\tilde{p}$.

Proposition 3.27. *Any Itô process with jumps of the form (3.78) is a non-anticipating functional on the canonical Wiener-Poisson space, namely, $X = F(w, \tilde{p})$, for some non-anticipating functional. Moreover, if $(\Omega', P', \mathbb{F}', w', p')$ is another Wiener-Poisson space then*

$$P'\big\{ X'(w', \tilde{p}') = F(w', \tilde{p}') \big\} = 1,$$

i.e., the stochastic integral is a non-anticipating functional on the Wiener-Poisson space.

Proof. This means that we should prove that any process of the form (3.78) is indistinguishable from a non-anticipating functional. As usual, by a localization argument, we may assume that the predictable functional coefficients satisfy

$$\int_0^T \mathbb{E}\Big\{ |a_i(t)| + \sum_{k=1}^n |b_{ik}(t)|^2 + \int_{\mathbb{R}_*^m} |\gamma_i(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \Big\} \mathrm{d}t < \infty.$$

Now, if the coefficients are piecewise constant (i.e., simple or elementary functions) then (as noted early) the stochastic integral is a non-anticipating functional.

In general, by a monotone class argument (or merely, by the proper definition of the stochastic integral) we may find a sequence of elementary predictable functionals a^k, b^k and γ^k such that

$$\mathbb{E}\Big\{ \int_0^T \Big[|a^k(t) - a(t)| + |b^k(t) - b(t)|^2 + \\ + \int_{\mathbb{R}_*^m} |\gamma^k(\zeta, t) - \gamma(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \Big] \mathrm{d}t \Big\} \to 0,$$

for any $T > 0$. Then, by passing to a subsequence if necessary, we have

$$\sup_{0\le t\le T} |X^k(t, w, \tilde{p}) - X(t, w, \tilde{p})| \to 0,$$

outside of a set N with $P(N) = 0$, for any $T > 0$, where $X^k(t, w, \tilde{p})$ denotes the stochastic integral with elementary integrands a^k, b^k and γ^k.

Hence, if F_k is a non-anticipating functional satisfying $X^k(w, \tilde{p}) = F_k(w, \tilde{p})$ then define

$$F(w, \tilde{p}) = \begin{cases} \lim_k F_k(w, \tilde{p}) & \text{in} \quad \Omega \setminus N, \\ 0 & \text{in} \quad N, \end{cases}$$

where the limit is uniformly on $[0, T]$, any $T > 0$. Actually, we can use the convergence in L^2-sup-norm to define the non-anticipating functional F. Thus $X = F(w, \tilde{p})$.

This procedure gives an approximation independent of the particular Wiener process and Poisson measure used, so that the same approximation yields the equality $X'(w', \tilde{p}') = F(w', \tilde{p}')$, P'-almost surely. □

Now, let η and ξ be two cad-lag non-anticipative processes relative to $(w, \tilde{p})$, see Definition 3.22, and assume that each component η_i of η is non-decreasing. The non-anticipative property imply that if $\mathbb{F}_{\eta,\xi} = \mathbb{F}(w, \tilde{p}, \eta, \xi)$ is the minimum completed filtration such that $(w, \tilde{p}, \eta, \xi)$ is adapted to, then $(w, \tilde{p})$ is a martingale, i.e., $(\Omega, P, \mathbb{F}_{\eta,\xi}, w, \tilde{p})$ is a Wiener-Poisson space. Moreover, any $\mathbb{F}_{\eta,\xi}$-adapted process y can be represented by a predictable functional, i.e., $y(t) = y(t, w, \tilde{p}, \eta, \xi)$, P-almost surely, for almost every t, where $(t, w, \tilde{p}, \eta, \xi) \mapsto y$ is a measurable function from $\mathbb{R} \times \mathbb{C}_n \times \mathbb{D}_{m+r+d}$ into $\mathbb{R}^{k+\ell}$.

Proposition 3.28. *Let us assume that a_{ik}, b_{ik} and γ_i are real-valued predictable functional on $\mathbb{C}_n \times \mathbb{D}_{m+r+d}$ as above. Then the stochastic integral*

$$X_i(t) = \xi_i(t) + \sum_{j=1}^{r} \int_0^t a_{ij}(s)\mathrm{d}\eta_k(s) + \sum_{k=1}^{n} \int_0^t b_{ik}(s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}_*^m \times]0,t]} \gamma_i(\zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \ge 0, \quad (3.79)$$

defines a non-anticipating functional, i.e., $X = F(w, \tilde{p}, \eta, \xi)$. Moreover, if the process η is also a non-anticipating functional $\eta(w, \tilde{p}, \xi)$ then $X = G(w, \tilde{p}, \xi)$. Furthermore, if $H(w, \tilde{p}, \xi)$ denotes a non-anticipating functional corresponding to a deterministic process ξ, then for any Wiener-Poisson space $(\Omega', P', \mathbb{F}', w', p')$ with a cad-lag process ξ' independent of $(w', \tilde{p}')$ the stochastic integral process like (3.79) is clearly defined and denoted by $X'(t)$, and we have $X'(t) = H(w', \tilde{p}', \xi')$, P'-almost surely.

Proof. The arguments are essentially the same as in previous Proposition 3.27. Note that the functional $G(w', \tilde{p}', \xi')$ depends on the distribution P_ξ on $\mathbb{D}_d$.

Perhaps we should make some comments on the functional H. Indeed, if the coefficients are simple (or elementary) functions then the stochastic integral takes the form

$$X(t) = \xi(t) + \sum_{i=1}^{n} a^{i-1}[\eta(t \wedge t_i-) - \eta(t \wedge t_{i-1})] + \\ + \sum_{i=1}^{n} b^{i-1}[w(t \wedge t_i) - w(t \wedge t_{i-1})] + \sum_{i=1}^{n}\sum_{j=1}^{m} \gamma^{i-1,j} \tilde{p}(K_j \times]t_{i-1}, t \wedge t_i]),$$

where a^i, b^i and γ^i are themselves predictable functionals depending on some parameter integer k. This defines a approximating functional $H_k(w, \tilde{p}, \xi)$, having the desired properties, which are preserved (P- or P'-) almost surely as k goes to infinite. □

Certainly, an important particular case is when the process $\xi(\cdot)$ is actually equal to a $\mathbb{R}^d$-valued random variable ξ, which is independent of the Wiener process and the Poisson measure p.

Chapter 4

Stochastic Calculus II

This is the second chapter dedicated to the stochastic integral. In the first section, stochastic integration is reconsidered as an extension of Stieltjes-Riemann integral, using continuity in probability and with an emphasis on the integrand/integrator processes, like in Protter [149]. This requires a quick refresh on quasi-martingales and Stieltjes integral, to cover martingale integrals and then non-martingale integrals, like Stratonovich) stochastic integrals. Section 2 is a systematic discussion on the quadratic variation (or variance) process associated with a local-martingale, mainly on estimates for the stochastic integral processes. Finally, Section 3 is an introduction to random fields and stochastic flows, focusing on an extension of Itô formula and the homeomorphic property of stochastic ordinary differential equation, which are only mentioned.

4.1 Other Stochastic Integrals

First we recall some key facts about possibly discontinuous martingales and then we discuss Stratonovich (and other) stochastic integrals.

4.1.1 Refresh on Quasi-Martingales

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a complete (relative to $\mathcal{F}$), right-continuous (not necessarily quasi-left continuous) filtration $\mathbb{F} = \{\mathcal{F}(t) : t \geq 0\}$. Recall that an adapted cad-lag process X is called a quasi-martingale if $\mathbb{E}\{|X(t)|\} < \infty$ and ${}^{\mathrm{p}}\mathrm{Var}(X)(t) < \infty$, for every $t \geq 0$, where the *conditional variation* is defined by

$$
\begin{aligned}
&{}^{\mathrm{p}}\mathrm{var}_{\varpi}(X)(t,\cdot) = \sum_i \left|\mathbb{E}\{X(t_{i+1}\wedge t) - X(t_i \wedge t)\,|\,\mathcal{F}(t_i\wedge t)\}\right|,\\
&{}^{\mathrm{p}}\mathrm{Var}(X) = \sup\left\{{}^{\mathrm{p}}\mathrm{Var}_{\varpi}(X) \,:\, \varpi\right\}, \qquad {}^{\mathrm{p}}\mathrm{Var}_{\varpi}(X) = \mathbb{E}\{{}^{\mathrm{p}}\mathrm{var}_{\varpi}(X)\},
\end{aligned}
$$

where the supremum is taken over all (deterministic) partitions $\varpi = \{0 = t_0 < t_1 < \cdots < t_{n-1} < t_n < \cdots\}$ of $[0,\infty)$ with norm $|\varpi| = \sup\{t_i - t_{i-1} \,:\, i \geq 1\}$.

An adapted processes X is a quasi-martingale if and only it can be decomposed as the difference $X = Y - Z$ of two positive (cad-lag) super-martingales Y and Z, or equivalently, it is a special semi-martingale, which yields the decomposition $X = M + A$ with M a local-martingale and A a predictable local integrable finite variation process, i.e., $A = A_+ - A_-$, both predictable, local integrable and monotone increasing. In particular if X is an adapted local integrable monotone increasing (or finite variation) process then $X = M + A$, where M is a local-martingale and A is a predictable local integrable monotone increasing (or finite variation) process. The process A is called the predictable (jumps) compensator of X. Note that the essential different between quasi-martingales and semi-martingales is the integrability of the large jumps.

If X is a semimartingale then the optional quadratic variation of X is defined as

$$[X](t) = X^2(t) - X^2(0) - 2\int_{]0,t]} X(s-)\mathrm{d}X(s), \quad t \geq 0,$$

or equivalently, as

$$[X]_\varpi(t) = \sum_{i=1}^{n} \big|X(t_{i+1} \wedge t) - X(t_i \wedge t)\big|^2, \qquad [X](t) = \lim_{|\varpi|\to 0} [X]_\varpi(t).$$

However, the predictable quadratic variation $\langle X\rangle$ is the predictable (jumps) compensator of $[X]$, i.e., $\langle X\rangle$ is the unique predictable process with local integrable finite variation (increasing) vanishing at 0 such that $[X] - \langle X\rangle$ is a local-martingale or equivalently $X^2 - \langle X\rangle$ is a local-martingale. Because $[X]$ is an adapted increasing process we may define its continuous part

$$[X]^c(t) = [X](t) - \sum_{s\leq t} \delta[X](s),$$

where δ is the jump operator, $\delta Y(0) = 0$,

$$\delta Y(t) = Y(t+) - Y(t-), \quad t > 0,$$

defined for any process Y having no discontinuities of second kind.

For any quasi-martingale X we have:

(1) if X is continuous then $[X]$ and $\langle X\rangle$ are (the same) continuous processes,

(2) if X has local integrable finite variation then $[X]^c = 0$,

(3) if $[X] = 0$ then $X = X(0)$,

(4) if X is a local-martingale satisfying $\langle X\rangle = 0$ then $X = X(0)$,

(5) X is quasi-left continuous if and only if $\langle X\rangle$ is continuous.

Moreover, any quasi-martingale X has a unique decomposition $X(0) + V_p(t) + M_c(t) + M_d(t)$, where $V_p(0) = M_c(0) = M_d(0) = 0$, V_p is a predictable process with local integrable finite variation, M_c is a continuous local-martingale and

M_d is a local-martingale satisfying $[M_d]^c = 0$, also (1) $[X]^c = \langle M_c \rangle$, (2) V_p is continuous if $M_d = 0$, and (3) if X has also local integrable finite variation then $M_c = 0$.

Note that $[X](t) = \sum_{s \le t}(X(t) - X(t-))^2$ for any process X of local bounded variation, and we have $\langle X \rangle = 0$ if X^2 is a local martingale. In particular if $X = N$ is a Poisson process then $X = V_p + M_d$, where $V_p(t) = \mathbb{E}\{X(t)\}$ is continuous, and $\sum_{s \le t} M_d(s) = X(t)$, $[X] = [M_d] = \sum_{s \le t}(X(t) - X(t-))^2$ and $\langle X \rangle = \langle M_d \rangle = \mathbb{E}\{X(t)\}$. In general, the sum of jumps $\sum_{s \le t} X(s)$ of a local-martingale satisfying $[X]^c = 0$ may not be defined (i.e., the series of jumps may not be pathwise convergent) or it may converge not necessarily to X. The local-martingale M_d contains predictable and unpredictable jumps, and $\langle M_d \rangle$ contains only the predictable jumps, but if $\langle M_d \rangle = 0$ then $M_d = 0$. Note that the square-bracket $[\cdot]$ is defined for any semi-martingale (and so for any quasi-martingale), while the angle-bracket $\langle \cdot \rangle$ is only define for local-martingale.

4.1.2 Refresh on Stieltjes integrals

Let us consider the pathwise Riemann-Stieltjes integral for bounded variation integrator and integrand, which is defined as a limit on partitions of a compact interval $[a, b]$. Typically, the integral exists for a continuous integrand f and a bounded variation integrator g (or conversely), but if fails to exists if both f and g are discontinuous on the same side (either right or left). The integration by part formula is granted if one of the integral exists, namely,

$$f(b)g(b) - f(a)g(a) = \int_a^b f(t)\mathrm{d}g(t) + \int_a^b g(t)\mathrm{d}f(t).$$

However, we have

$$f(b)g(b) - f(a)g(a) = \int_{]a,b]} f(t)\mathrm{d}g(t) + \int_{]a,b]} g(t-)\mathrm{d}f(t).$$

in the Lebesgue-Stieltjes sense, if both f and g are only right-continuous with finite variation. Indeed, if V is a cad-lag process with locally bounded variation and X is a cad-lag process then, for any $b > a \ge 0$ we have

$$\int_a^b X_-(t)\mathrm{d}V(t) = \int_a^b X_-(t)\mathrm{d}V^c(t) + \sum_{a \le t < b} X_-(t)\delta V(t),$$
$$\int_a^b X(t)\mathrm{d}V_-(t) = \int_a^b X(t)\mathrm{d}V^c(t) + \sum_{a < t \le b} X(t)\delta V(t),$$

where $X_-(t) = X(t-)$, $V_-(t) = V(t-)$, for every $t > 0$, and V^c is the continuous part of V, i.e.,

$$V(t) = V^c(t) + \sum_{0 < s \le t} \delta V(s).$$

Note that $X = X_- + \delta X$ and that we may replace X with X_- as the integrand of $\mathrm{d}V^c$. Nevertheless, if $X = U$ is a cad-lag process with locally bounded variation then we can rewrite the integration by part formula as

$$U(b)V(b) - U(a)V(a) = \int_{]a,b]} U_-(t)\mathrm{d}V(t) + \\ + \int_{]a,b]} V_-(t)\mathrm{d}U(t) + \sum_{a<t\leq b} \delta U(t)\delta V(t), \tag{4.1}$$

where all series are absolutely convergent and all integrals are considered pathwise, in either Riemann-Stieltjes (without including any possible jump at a, but including any possible jump at b) or Lebesgue-Stieltjes sense, i.e., if μ_V denotes the Lebesgue-Stieltjes measure generated by the cad-lag path function $t \mapsto V(t)$ then

$$\int_{]a,b]} X_-(t)\mathrm{d}V(t) = \lim_{\alpha\to a^+,\, \beta\to b^+} \int_\alpha^\beta X_-(t)\mathrm{d}V(t) = \int_{]a,b]} X_-(t)\mu_V(\mathrm{d}t),$$

actually, this is the definition of the integral in Riemann-Stieltjes sense over the semi-open interval $]a, b]$ for cag-lad (left continuous with right limits) integrands and cad-lag integrators.

A cad-lag process is integrable for the (signed) Lebesgue-Stieltjes measure μ_V (which can be expressed as the difference of two measures, the positive and negative variations) and for any $b > a \geq 0$, we have

$$\int_{]a,b]} X(t)\mu_V(\mathrm{d}t) = \int_{]a,b]} X_-(t)\mu_V(t) + \int_{]a,b]} \delta X(t)\mu_V(\mathrm{d}t) = \\ = \int_{]a,b]} X_-(t)\mathrm{d}V(t) + \sum_{a<t\leq b} \delta X(t)\delta V(t),$$

Note that $\mu_V = \mu_V^c + \mu_V^d$, where μ_V^c is the continuous part of μ_V, i.e., when all atoms have been removed (or equivalently, the measure associated with the continuous part V^c of V). Moreover, $\mu_V^c = \mu_V^a + \mu_V^s$, where μ_V^a is absolutely continuous with respect to the Lebesgue measure and μ_V^s is singular (i.e., there exists a Borel measurable set S of Lebesgue measure zero such that for any measurable set N with Lebesgue measure zero we have $\mu_V^c(N \smallsetminus S) = 0$, and then we define $\mu_V^s(A) = \mu_V^c(A \cap S)$, for any measurable set A). Thus, any set of one point $\{t\}$ is μ_V^c-negligible and so is any countable set, i.e., $\delta X = 0$ μ_V^c-almost surely and the integral of X and X_- relative to μ_V^c coincide. It is cleat that, both X_- and X (and any bounded Borel measurable process) are integrable with respect to μ_V, but to recall that the integration by part formula (4.1) should be written with X_-, we use the Riemann-Stieltjes sense over the semi-open interval $]a, b]$. Certainly, the notation $\mathrm{d}V$ actually means $\mathrm{d}\mu_V$, when the integrands are not cag-lad processes.

On the other hand, if $F\colon \mathbb{R} \to \mathbb{R}$ is a locally Lipschitz function and V is a cad-lag process with locally bounded variation then $t \mapsto F\big(V(t)\big)$ is also a cad-lag process with locally bounded variation. Moreover, if f is also continuously

differentiable then we have the change of variable formula

$$F\big(V(b)\big) - F\big(V(a)\big) = \int_{]a,b]} F'\big(V_-(t)\big)\mathrm{d}V(t) + \\ + \sum_{a<t\le b} \big\{F\big(V(t)\big) - F\big(V(t-)\big) - F'\big(V(t-)\big)\delta V(t)\big\}, \tag{4.2}$$

where F' denotes the derivative of F. Since F' is locally bounded and V has locally bounded variation, the above series can be written as

$$\sum_{a<t\le b} \delta F\big(V(t)\big) + \sum_{a<t\le b} F'\big(V_-(t)\big)\delta V(t)$$

and both series are absolutely convergent. Clearly, the change of variable (4.2) is usually written as

$$F\big(V(b)\big) - F\big(V(a)\big) = \int_{]a,b]} F'\big(V_-(t)\big)\mathrm{d}V^c(t) + \sum_{a<t\le b} \delta F\big(V(t)\big),$$

and we note that $\delta F\big(V(t)\big) > 0$ if and only if $\delta V(t) > 0$.

At this point, it is important to recognize that to capture the jumps of a $\mathbb{R}^m$-valued cad-lag process X we need to study its associate integer measure ν_X, which is defined as the extension of

$$\nu_X(K\times]a,b]) = \sum_{a<s\le b} \mathbb{1}_{\delta X(s)\in K}, \qquad \text{(a finite sum)},$$

for any compact set K in $\mathbb{R}^m_* = \mathbb{R}^m \setminus \{0\}$ and any $b \ge a \ge 0$. If X is cad-lag with bounded variation then X and ν_X are equivalent in the sense that from

$$X(t) = X(0) + \int_{\mathbb{R}^m_*} z\nu_X(]0,t], \mathrm{d}x)$$

we can reconstruct X from ν_X. However, if the process X is not necessarily of bounded variation then we need to make sense to the limits

$$X(t) = X(0) + \lim_{\varepsilon\to 0} \int_{\mathbb{R}^m_*} z\mathbb{1}_{\{|z|\ge\varepsilon\}}\nu_X(]0,t], \mathrm{d}z)$$

to be able to reconstruct X. Clearly, this limit makes sense as a stochastic integral if X is a local-martingale.

4.1.3 Square-Brackets and Angle-Brackets

If X and Y are two semi-martingales then we define the square-bracket by

$$[X,Y] = XY - \int_{]0,\cdot]} X(s-)\mathrm{d}Y(s) - \int_{]0,\cdot]} Y(s-)\mathrm{d}X(s),$$

when $X(0) = Y(0) = 0$, or by polarization as

$$[X,Y] = \big([X+Y] - [X] - [Y]\big)/2 = \big([X+Y] - [X-Y]\big)/4.$$

Similarly with the angle-bracket $\langle X,Y\rangle$, which is defined only for local-martingales. If X and Y are local square-integrable martingales then $[X,Y]$ is the unique cad-lag adapted process with integrable finite variation and vanishing at 0 such that (1) $XY - [X,Y]$ is a local-martingale and (2) $\delta[X,Y] = \delta X\,\delta Y$, while $\langle X,Y\rangle$ is the unique cad-lag predictable process with integrable finite variation and vanishing at 0 such that $XY - \langle X,Y\rangle$ is a local-martingale. For any quadratic pure jump semi-martingale X, i.e., satisfying $[X]^c = 0$, and any semi-martingale Y we have

$$[X,Y](t) = \sum_{s\le t} \delta X(s)\,\delta Y(s), \quad \forall t > 0.$$

A local-martingale X is called purely discontinuous if $X(0) = 0$ and $\langle X,Y\rangle = 0$ for any continuous local-martingale Y. Then (1) a local-martingale X vanishing at 0 is purely discontinuous if and only if $[X]^c = 0$, (2) a local-martingale with local finite variation and $X(0) = 0$ is purely discontinuous, (3) a continuous local-martingale which is purely discontinuous is indeed null, and (4) a predictable local-martingale is a continuous martingale.

Let X be a quasi-martingale and V be an adapted process with local integrable finite variation and $V(0) = 0$: we have (a)

$$[X,V] = \int_{]0,\cdot]} \delta X(s)\mathrm{d}V(s), \quad XV = \int_{]0,\cdot]} V_-(s)\mathrm{d}X(s) + \int_{]0,\cdot]} X(s)\mathrm{d}V(s),$$

and (b) if V is predictable then

$$[X,V] = \int_{]0,\cdot]} \delta V(s)\mathrm{d}X(s), \quad XV = \int_{]0,\cdot]} V(s)\mathrm{d}X(s) + \int_{]0,\cdot]} X_-(s)\mathrm{d}V(s).$$

Hence, we also have (c) if X is a local-martingale and V is predictable then the optional covariance or square-bracket $[X,V]$ is a local-martingale, and (d) we have $[X,V](t) = \sum_{s\le t}\delta X(t)\delta V(t)$ and so $[X,V] = 0$, if at least one (either X or V) is continuous. There are several useful estimates involving local-martingales, e.g., Davis-Burkhölder-Gundy inequality for local-martingales vanishing at the initial time, namely, for any $p \ge 1$ there exist constants $C_p > c_p > 0$ (recall that $C_1 = 3$ and $C_2 = 4$) such that for any stopping time T and any local martingale M with $M(0) = 0$, we have

$$c_p\,\mathbb{E}\{\big([M](T)\big)^{p/2}\} \le \mathbb{E}\{\sup_{t\le T}|M(t)|^p\} \le C_p\,\mathbb{E}\{\big([M](T)\big)^{p/2}\},$$

and Lenglart domination estimate, namely, for any cad-lag adapted process X dominated by an increasing cad-lag process A with $A(0) = 0$ (i.e., $\mathbb{E}\{|X(\tau)|\} \le \mathbb{E}\{A(\tau)\}$ for any bounded stopping time τ) we have

$$P\big\{\sup_{t\le T}|X(t)| \ge \varepsilon\big\} \le \frac{1}{\varepsilon}\Big[\eta + \mathbb{E}\big\{\sup_{t\le T}|A(t) - A(t-)|\big\}\Big] + P\big\{A(T) \ge \eta\big\},$$

for any positive constants ε, η and any stopping time T, and if A is predictable, we may drop the term with the jumps. However, for any p in $(0, 2]$ there exist a constant $C_p > 0$ (with $C_1 = 3$ and $C_2 = 4$) such that

$$\mathbb{E}\{\sup_{t\leq T}|M(t)|^p\} \leq C_p\,\mathbb{E}\{\big(\langle M\rangle(T)\big)^{p/2}\},$$

for any stopping time T and any local-martingale M with $M(0) = 0$.

Let X and Y be two semi-martingales, and a and b be two adapted cag-lad (left continuous with right limits) then

$$\Big[\int_{]0,\cdot]} a(s)\mathrm{d}X(s), \int_{]0,\cdot]} b(s)\mathrm{d}Y(s)\Big] = \int_{]0,\cdot]} a(s)b(s)\mathrm{d}[X,Y](s),$$

and similarly with the angle-bracket $\langle\cdot,\cdot\rangle$, where the last integral is in either the Riemann-Stieltjes or Lebesgue-Stieltjes sense. Suppose that X_i, $i = 1,\ldots,n$ and Y_j, $j = 1,\ldots,m$ are semi-martingales, and that $\varphi(x)$ and $\psi(y)$ are smooth real-valued functions then Itô formula shows that $\varphi(X)$ and $\psi(Y)$ are also semi-martingales and

$$\langle\varphi(X),\psi(Y)\rangle(t) = \sum_{ij}\int_0^t \partial_i\varphi(X(s))\partial_j\psi(Y)(s)\mathrm{d}\langle X_i,Y_j\rangle(s),$$

$$[\varphi(X),\psi(Y)]^c(t) = \sum_{ij}\int_0^t \partial_i\varphi(X(s))\partial_j\psi(Y)(s)\mathrm{d}[X_i,Y_j]^c(s),$$

$$[\varphi(X),\psi(Y)](t) - [\varphi(X),\psi(Y)]^c(t) = \sum_{s\leq t}\delta\varphi(X)(s)\delta\psi(Y)(s),$$

for any $t > 0$. Hence, let ν_{XY} denote the integer measure associated with the (joint) jumps of the $\mathbb{R}^{n+m}$-valued process (X, Y), namely,

$$\nu_{XY}(B,]a,b]) \text{ is the number of jumps } \big(\delta X(s),\delta Y(s)\big) \text{ in } B \text{ within the interval }]a,b], \text{ for any } B \text{ in } \mathcal{B}(\mathbb{R}^{n+m}_*) \text{ with } B\cap\{0\} = \emptyset \text{ and } 0 \leq a < b,$$

with a predictable jumps compensator $\nu^p_{XY}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}t)$. Thus, the jumps part of the optional quadratic covariation, i.e., $[X,Y] - [X,Y]^c$, can be expressed as

$$\sum_{s\leq t}\delta\varphi(X)(s)\delta\psi(Y)(s) = \int_{\mathbb{R}^{n+m}\times]0,t]}\big[\varphi(X(s-)+x) - \varphi(X(s-))\big]\times \\ \times\big[\psi(Y(s-)+y) - \psi(Y(s-)\big]\nu_{XY}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}t).$$

The continuous part of a semi-martingale X is defined as the unique continuous semi-martingale X^c satisfying $[X - X^c, Z] = 0$, for any continuous semi-martingale Z. Then we have $[X^c, Y] = [X,Y]^c$. On the other hand, the processes X and Y are quasi-left continuous if only if there are no predictable jumps, i.e., $\nu^p_{XY}(\mathbb{R}^{n+m}_*\times\{t\}) = 0$, for any $t \geq 0$, or equivalently the predictable covariation $\langle X,Y\rangle$ is continuous. Note that if the jumps of X and Y have the form

$$X(t) = X^c(t) + \int_{Z\times]0,t]}\gamma^X(z,s)\tilde{\nu}(\mathrm{d}z,\mathrm{d}s), \quad \forall t \geq 0,$$

and similarly for Y, with the same martingale measure $\tilde{\nu}$ and continuous processes X^c and Y^c, then

$$\int_{\mathbb{R}_*^{n+m}\times]0,t]} h(x,y,s)\nu_{XY}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}s) = \\ = \int_{Z\times]0,t]} h(\gamma^X(z,s),\gamma^Y(z,s),s)\nu(\mathrm{d}z,\mathrm{d}s),$$

for any positive Borel measurable function h, and similarly for the predictable jump compensator measure.

In particular, let $M = M_c + M_d$ be a quasi-left continuous local square-integrable martingale in $\mathbb{R}^n$ written as the sum of a continuous local-martingale $\{M_{c,i} : i = 1,\dots,n\}$ with predictable variation process $\{\langle M_{c,i}\rangle : i = 1,\dots,n\}$, and a purely discontinuous local-martingale $\{M_{d,i} : i = 1,\dots,n\}$ which yields an integer measure $\nu_{_M}$ with compensator $\nu_{_M}^p$ and martingale measure $\tilde{\nu}_{_M} = \nu_{_M} - \nu_{_M}^p$. Note that

$$\int_{]0,t]} \alpha(s)\mathrm{d}M_{d,i}(s) = \int_{\mathbb{R}^d\times]0,t]} \alpha(s)\zeta_i\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}s), \quad i = 1,\dots,d,$$

and

$$\sum_{s\le t} h\big(s,\delta M_d(s)\big) = \int_{\mathbb{R}^d\times]0,t]} h(s,\zeta)\nu_{_M}(\mathrm{d}\zeta,\mathrm{d}s),$$

for any predictable integrable processes α and h. Thus, if X is a d-dimensional processes of the form

$$\mathrm{d}X(t) = a^X(t)\mathrm{d}V^c(t) + b^X(t)\mathrm{d}M_c(t) + \int_{\mathbb{R}_*^m} \gamma^X(\zeta,t)\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \ge 0,$$

where V^c is an adapted continuous process with local integrable finite variation, and φ is real-valued smooth functions then Itó formula shows that the semi-martingales $\varphi(t) = \varphi(t,X(t))$ can be expressed in term of continuous part M_c, the compensated integer (or martingale) measure $\tilde{\nu}_{_M}$ and some continuous locally bounded variation processes V_φ, i.e.,

$$\mathrm{d}\varphi(t) = \mathrm{d}V_\varphi(t) + b^\varphi(t)\mathrm{d}M_c(t) + \int_{\mathbb{R}_*^m} \gamma^\varphi(\zeta,t)\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \ge 0,$$

where

$$\mathrm{d}V_\varphi(t) = \partial_t\varphi(t,\cdot)\mathrm{d}t + \partial_x\varphi(t,\cdot)\mathrm{d}V^c(t) + \\ + \sum_{i,j,k} b_{ik}^X(t)b_{jk}^X(t)\partial_{ij}\varphi(t,\cdot)\mathrm{d}\langle M_{c,k}\rangle(t) + \\ + \int_{\mathbb{R}_*^m\times]0,t]} \big[\varphi(t,\cdot+\gamma^X(\zeta,t)) - \varphi(t,\cdot) - \gamma^X(\zeta,t)\partial_x\varphi(t,\cdot)\big]\nu_{_M}^p(\mathrm{d}\zeta,\mathrm{d}t),$$

and

$$b_k^{\varphi}(t) = \sum_i \partial_i \varphi(t,\cdot) b_{ik}^X(t), \qquad \gamma^{\varphi}(\zeta,t) = \varphi(t,\cdot + \gamma^X(\zeta,t)) - \varphi(t,\cdot),$$

the dot $\cdot$ is replaced by $M(t-)$. Thus

$$\langle \varphi_c, M_{c,k} \rangle = \sum_{i,j} \int_0^{\cdot} \partial_i \varphi(t, M(t-)) b_{ij}^X(t) \mathrm{d}\langle M_{c,j}, M_{c,k} \rangle(t),$$

$$\delta\varphi(t) = \int_{\mathbb{R}_*^m} \big(\varphi(t, M(t-) + \gamma^X(\zeta,t)) - \varphi(t, M(t-))\big) \nu_{_M}(\mathrm{d}\zeta, \{t\}),$$

$$[\varphi, M_{d,k}] = \int_{\mathbb{R}_*^m \times]0,\cdot]} \big(\varphi(t, M(t-) + \gamma^X(\zeta,t)) - \varphi(t, M(t-))\big) \zeta_k \nu_{_M}(\mathrm{d}\zeta, \mathrm{d}t),$$

which give full information on the covariance of φ and M. These calculations are neat for the particular case where M^c is a standard Wiener process and $\tilde{\nu}_{_M}$ is a compensated Poisson integer measure with predictable compensator $\nu_{_M}^p(\mathrm{d}\zeta, \mathrm{d}t) = \pi(\mathrm{d}\zeta)\mathrm{d}t$.

For instance, the reader is referred to Dellacherie and Meyer [32, Sections VI.37–42, pp. 105–112], He et al. [68, Chapter VIII, pp. 209–223], Jacod and Shiryaev [84, Chapter 1, pp. 1–63], Kallenberg [88, Chapter 26, pp. 515–536], Protter [149, Chapter III, pp. 43–86] for more detail on the above statements.

4.1.4 Martingales Integrals

After refreshing the above martingales concepts, we may reframe the stochastic integral defined previously for a quasi-martingale X. Recall that a sequence of general processes $\{u_k\}$ is said to converge to u *uniformly on compacts in probability* abbreviated "ucp", if for any $\varepsilon > 0$ there exists $K > 0$ such that $P\{\sup_{0\le t\le 1/\varepsilon} |u_k(t) - u(t)| \ge \varepsilon\} \le \varepsilon$ for any $k \ge K$. Given a filtration $\mathbb{F} = \{\mathcal{F}(t) : t \ge 0\}$, denote by $\mathtt{S} = \mathtt{S}(\mathbb{F})$, $\mathtt{L} = \mathtt{L}(\mathbb{F})$ and $\mathtt{D} = \mathtt{D}(\mathbb{F})$ the vector space of simple predictable processes (i.e., $H(t) = h_{i-1}$ for t on the stochastic interval $\rrbracket \tau_{i-1}, \tau_i \rrbracket$, for $i = 1, 2, \ldots, n$, where $\tau_0 = 0$, $\tau_{i-1} \le \tau_i$ are finite stopping times and h_i is a $\mathbb{R}^d$-valued $\mathcal{F}(\tau_i)$-measurable random variable), adapted cag-lad (left continuous with right limits) processes, and cad-lag (right continuous with left limits). If we add the ucp-convergence and version of processes are considered equals, then we use $\mathtt{S}_{\mathrm{ucp}}$, $\mathtt{L}_{\mathrm{ucp}}$ and $\mathtt{D}_{\mathrm{ucp}}$. Actually, any process X in $\mathtt{L}$ (or in $\mathtt{S}$) are technically defined on $(0,\infty)$, but we assume $X(0) = X(0+)$ and so X is defined on $[0,\infty)$; otherwise, we may decide to work on the whole space $(-\infty,+\infty)$ or to explicitly introduce a value at time $t = 0$. Moreover a better notation would be $\mathtt{S}_{\mathrm{ucp}}(\mathbb{F})$, $\mathtt{L}_{\mathrm{ucp}}(\mathbb{F})$ and $\mathtt{D}_{\mathrm{ucp}}(\mathbb{F})$ to recall the dependency on the filtration $\mathbb{F}$, however this is implicitly assumed.

Remark that a modification (also called a version) of an element in $\mathtt{S}$, $\mathtt{L}$ or $\mathtt{D}$ does not necessarily belongs to $\mathtt{S}$, $\mathtt{L}$ or $\mathtt{D}$, it belongs to $\mathtt{S}_{\mathrm{ucp}}$, $\mathtt{L}_{\mathrm{ucp}}$ or $\mathtt{D}_{\mathrm{ucp}}$. We may have an element u in $\mathtt{L}_{\mathrm{ucp}}$ and an element v in $\mathtt{D}_{\mathrm{ucp}}$ such that u is a version of v, i.e., not any version of a given element in $\mathtt{L}_{\mathrm{ucp}}$ (or $\mathtt{D}_{\mathrm{ucp}}$) can be considered an

element in L (or D). On the other hand, we are allowed to modified an element (in any of the three topological vector spaces S, L or D) on a evanescent set and still remain in the same space.

Note that $\mathtt{L}_{\rm ucp}$ and $\mathtt{D}_{\rm ucp}$ are complete metric spaces, and let us prove that $\mathtt{S}_{\rm ucp}$ is dense in $\mathtt{L}_{\rm ucp}$. Indeed, given a positive number η and a u in $\mathtt{L}_{\rm ucp}$ we define an increasing sequence of stopping times $0 = T_0^\eta < T_1^\eta < T_2^\eta < \cdots$ by recurrence

$$T_{k+1}^\eta = \inf\{s > T_k^\eta : |u(s+) - u(T_k^\eta+)| > \eta\},$$

where $T_{k+1}^\eta = \infty$ if $|u(s+) - u(T_k^\eta+)| \le \eta$, for every $s \ge T_k^\eta$. Because $t \mapsto u(t+)$ is cad-lag, the sequence T_k^η is almost surely increasing to infinite, i.e., $P\{T_k^\eta \le r\} \to 0$ as $k \to \infty$, for every positive constants r. Clearly, $|u(s+) - u(T_k^\eta+)| \le \eta$, for any s such that $T_k^\eta \le s < T_{k+1}^\eta$, and by continuity, we have $|u(s) - u(T_k^\eta+)| \le \eta$, if $T_k^\eta < s \le T_{k+1}^\eta$. Hence, define $u_k^\eta(t) = u(n \wedge T_i^\eta+)$ if $k \wedge T_i^\eta < t \le k \wedge T_{i+1}^\eta$ with $i = 0, 1, \ldots, k$ to have

$$P\Big\{ \sup_{0\le t\le r} |u_k^\eta(t) - u(t)| > \eta \Big\} \le P\big\{ k \wedge T_k^\eta \le r \big\}.$$

Therefore, u_k^η belongs to $\mathtt{S}_{\rm ucp}$, and for suitable η and k we construct a sequence convergent to u.

For any H in $\mathtt{S}_{\rm ucp}$ and X in $\mathtt{D}_{\rm ucp}$ we define the simple integral

$$\Sigma(H, \mathrm{d}X)(t) = \int_{]0,t]} H(s)\mathrm{d}X(s) = \sum_{i=1}^n h_i X(\tau_i \wedge t) - X(\tau_{i-1} \wedge t),$$

if $H = \sum_{i=1}^n h_{i-1} \mathbb{1}_{]\!]\tau_{i-1}, \tau_i]\!]}$. Now if X is a quasi-martingale then this linear operator $\Sigma(\cdot, \mathrm{d}X)$ is continuous from $\mathtt{S}_{\rm ucp}$ into $\mathtt{D}_{\rm ucp}$ and therefore it can be uniquely extended to $\mathtt{L}_{\rm ucp}$, i.e., for any H in $\mathtt{L}_{\rm ucp}$ there exists a sequence H_k in $\mathtt{S}_{\rm ucp}$ such that $H_k \to H$ in $\mathtt{L}_{\rm ucp}$ and $\Sigma(H_k, \mathrm{d}X) \to \Sigma(H, \mathrm{d}X)$ in $\mathtt{D}_{\rm ucp}$. Actually, the continuity property can be proved directly or by means of Lenglart dominate estimate, namely, for any positive constants ε, η, any stopping time T, and any H in $\mathtt{L}_{\rm ucp}$, we have (recall δ is the jump operator)

$$P\Big\{ \sup_{t\le T} \big|\Sigma(H, \mathrm{d}A)(t)\big| \ge \varepsilon \Big\} \le P\big\{ A(T) \ge \eta \big\} + \\ + \frac{1}{\varepsilon}\Big[\eta + \mathbb{E}\Big\{ \sup_{t\le T} |H(t)|\, |\delta A(t)| \Big\} \Big],$$

if A is an adapted increasing integrable process, and

$$P\Big\{ \sup_{t\le T} \big|\Sigma(H, \mathrm{d}M)(t)\big|^2 \ge \varepsilon \Big\} \le \frac{\eta}{\varepsilon} + P\big\{ \Sigma(|H|^2, \mathrm{d}\langle M\rangle)(T) \ge \eta \big\},$$

if M is a local-martingale with predictable variance $\langle M\rangle$.

Moreover, if $\varpi = \{t_i\}$ is a partition of $[0, \infty)$, $t_0 = 0$, $t_{i-1} < t_i$, $t_i \to \infty$, and $|\varpi| = \sup_i(t_i - t_{i-1})$ (possible of stopping times) and H is an element of $\mathtt{L}_{\rm ucp}$, then we may define $H_{\varpi,n}(t) = H(t_{i-1})$ for t in $(t_{i-1}, t_i]$, $i = 1, \ldots, n$. It is

clear that $H_{\varpi,n}$ belongs to $\mathtt{S}_{\rm ucp}$, and $\sup_{0\le t\le T}|H_{\varpi,n}(t)|\le C(T)$, almost surely for a constant $C(T)$ independent of ϖ, n, and $H_{\varpi,n}(t)\to H(t)$, almost surely, for every $t>0$. Hence, after using Lenglart dominate estimate, we deduce that $\Sigma(H_{\varpi,n},\mathrm{d}M)\to\Sigma(H,\mathrm{d}M)$ in $\mathtt{L}_{\rm ucp}$, as $n\to\infty$ and $|\varpi|\to 0$.

Note that if X is a process with locally bounded variation belonging to $\mathtt{D}_{\rm ucp}$ and H is any process in $\mathtt{L}_{\rm ucp}$ then $\Sigma(H,\mathrm{d}M)$ coincides with the (pathwise) Riemann-Stieltjes (or Lebesgue-Stieltjes) integral.

Clearly, this technique can be extended simple integral relative to martingale measures $\tilde\nu$, e.g.,

$$\Sigma(H,\mathrm{d}\tilde\nu)(t)=\int_{\mathbb{R}^m_*\times]0,t]}H(\zeta,s)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s)=$$
$$=\sum_{i=1}^n\sum_{j=1}^k H_{ij}\tilde\nu(B_j\times]\tau_{i-1}\wedge t,\tau_i\wedge t]),$$

where $H=\sum_{i=1}^n\sum_{j=1}^k H_{ij}\mathbb{1}_{\rrbracket\tau_{i-1},\tau_i\rrbracket}\mathbb{1}_{B_j}$, and B_j is a Borel set separated from the origin, i.e., the closure $\bar B_j\cap\{0\}=\emptyset$. As discussed early in this chapter, the cad-lag processes $\Sigma(H,\mathrm{d}X)$ and $\Sigma(H,\mathrm{d}\tilde\nu)$ are local-martingales with

$$[\Sigma(H,\mathrm{d}X)](t)=\int_{]0,t]}|H(s)|^2\mathrm{d}[X](s),$$
$$[\Sigma(H,\mathrm{d}\tilde\nu)](t)=\int_{\mathbb{R}^m_*\times]0,t]}|H(\zeta,s)|^2\nu(\mathrm{d}\zeta,\mathrm{d}s),$$

and

$$\langle\Sigma(H,\mathrm{d}X)\rangle(t)=\int_{]0,t]}|H(s)|^2\mathrm{d}\langle X\rangle(s),$$
$$\langle\Sigma(H,\mathrm{d}\tilde\nu)\rangle(t)=\int_{\mathbb{R}^m_*\times]0,t]}|H(\zeta,s)|^2\nu^p(\mathrm{d}\zeta,\mathrm{d}s),$$

where ν^p is the predictable compensator of martingale measure $\tilde\nu$, i.e., for any fixed Borel set B separated from the origin, the process $t\mapsto\nu^p(B,]0,t])$ is the compensator of the local-martingale $t\mapsto\tilde\nu(B,]0,t])$, or $\tilde\nu$ is the martingale measure corresponding to an integer measure ν with predictable jumps compensator ν^p. Note that

$$\int_{\mathbb{R}^m_*}|H(\zeta,t)|\nu(\mathrm{d}\zeta,\{t\})\quad\text{replaces}\quad|H(t)|\,|\delta A(t)|\quad\text{and}$$
$$\int_{\mathbb{R}^m_*\times]0,T]}|H(\zeta,t)|^2\nu^p(\mathrm{d}\zeta,\mathrm{d}t)\quad\text{replaces}\quad\Sigma(|H|^2,\mathrm{d}\langle M\rangle)(T)$$

in Lenglart dominate estimate.

It should be clear that besides the probability measure P, the initial filtration $\mathbb{F}=\{\mathcal{F}(t):t\ge 0\}$ plays a fundamental role in the above construction. Perhaps,

a full notation for spaces $\mathtt{S}_{\rm ucp}$, $\mathtt{L}_{\rm ucp}$ and $\mathtt{D}_{\rm ucp}$ should includes the filtration and the probability, e.g., $\mathtt{D}_{\rm ucp}(P,\mathbb{F})$. However, if another filtration $\mathbb{G}=\{\mathcal{G}(t):t\geq 0\}$ is given and

$$H\in\mathtt{L}_{\rm ucp}(P,\mathbb{F})\cap\mathtt{L}_{\rm ucp}(P,\mathbb{G}),\qquad X\in\mathtt{D}_{\rm ucp}(P,\mathbb{F})\cap\mathtt{D}_{\rm ucp}(P,\mathbb{G})$$

then H can be approximate in $\mathtt{S}_{\rm ucp}$ with respect to both filtrations, which implies that the limit $\Sigma(H,\mathrm{d}X)$ is independent of the particular filtration used. Certainly, if the limit exists for a probability P then also it exits for any other probability Q which is absolutely continuous with respect to P.

4.1.5 Non-Martingales Integrals

Consider a partition ϖ of $[0,\infty)$ and for any two cad-lag processes X and Y define the symmetric square-bracket along ϖ,

$$[X,Y]_\varpi(t)=\sum_i\big(X(t_i\wedge t)-X(t_{i-1}\wedge t)\big)\big(Y(t_i\wedge t)-Y(t_{i-1}\wedge t)\big),$$

as well as the bilinear expressions (integrals along ϖ)

$$\begin{aligned}
\Sigma^-_\varpi(X,\mathrm{d}Y)(t)&=\sum_i X(t_{i-1}\wedge t)\big(Y(t_i\wedge t)-Y(t_{i-1}\wedge t)\big),\\
\Sigma^+_\varpi(X,\mathrm{d}Y)(t)&=\sum_i X(t_i\wedge t)\big(Y(t_i\wedge t)-Y(t_{i-1}\wedge t)\big),\\
\Sigma^\circ_\varpi(X,\mathrm{d}Y)(t)&=\sum_i \big(X(t_i\wedge t)+X(t_{i-1}\wedge t)\big)\big(Y(t_i\wedge t)-Y(t_{i-1}\wedge t)\big)/2,
\end{aligned}$$

which are finite sums of non-zero terms. Note the relations

$$\begin{aligned}
\Sigma^\circ_\varpi(X,\mathrm{d}Y)(t)&=\tfrac{1}{2}\big(\Sigma^-_\varpi(X,\mathrm{d}Y)(t)+\Sigma^+_\varpi(X,\mathrm{d}Y)(t)\big),\\
\Sigma^+_\varpi(X,\mathrm{d}Y)(t)-\Sigma^-_\varpi(X,\mathrm{d}Y)(t)&=[X,Y]_\varpi(t)=[Y,X]_\varpi(t),
\end{aligned}$$

and

$$\begin{aligned}
\Sigma^+_\varpi(X,\mathrm{d}Y)(t)+\Sigma^-_\varpi(Y,\mathrm{d}X)(t)&=Y(t)X(t)-X(0)Y(0),\\
\Sigma^\mp_\varpi(X,\mathrm{d}Y)(t)+\Sigma^\mp_\varpi(Y,\mathrm{d}X)(t)\pm[X,Y]_\varpi(t)&=Y(t)X(t)-X(0)Y(0),
\end{aligned}$$

where we use the telescopy sum

$$\sum_{i=1}^n a_i(b_i-b_{i-1})+\sum_{i=1}^n b_{i-1}(a_i-a_{i-1})=a_nb_n-a_0b_0,$$

valid for any numbers a_i and b_i.

For any cad-lag process X, we can consider the cag-lad process X_- defined as the left-hand limits, i.e., $X_-(t)=X(t-)$. If δ is the jump operator then we have $\delta X=\delta X_-$, $X_-=X-\delta X$, and $X=X_-+\delta X$. Hence we have

$\Sigma^-_\varpi(X_-, \mathrm{d}Y) \to \Sigma(X_-, \mathrm{d}Y)$ in $\mathtt{D}_{\mathrm{ucp}}$ as $|\varpi| \to 0$, for any X in $\mathtt{D}_{\mathrm{ucp}}$ and for any quasi-martingale Y.

If X and Y are quasi-left continuous then for any t there exist a null set N_t such that $\delta X(t,\omega) = 0$ and $\delta Y(t,\omega) = 0$, for any ω in $\Omega \setminus N_t$. Thus,

$$[X,Y]_\varpi(t) + \Sigma^-_\varpi(X_-, \mathrm{d}Y)(t) + \Sigma^-_\varpi(Y_-, \mathrm{d}X)(t) = \\ = Y(t)X(t) - X(0)Y(0), \quad (4.3)$$

almost surely, for each t. In particular, this proves that $[X,Y]_\varpi \to [X,Y]$ in $\mathtt{D}_{\mathrm{ucp}}$ and that $\Sigma^\pm_\varpi(X_-, \mathrm{d}Y)$ and $\Sigma^\pm_\varpi(X, \mathrm{d}Y)$ have a common limit $\mathtt{D}_{\mathrm{ucp}}$, as $|\varpi| \to 0$, for any quasi-left continuous quasi-martingales X and Y.

Our interest is on processes where the jumps are only due to a local martingale, i.e., the finite variation part of X can be chosen continuous. Now, let π be a Lévy measure in $\mathbb{R}^d_*$, $\gamma(z)$ be a (deterministic) function in $L^2(\mathbb{R}^m_*, \pi)$ and X be a real-valued Itô process with jumps,

$$X(t) = \int_0^t a^X(s)\mathrm{d}v(s) + \int_0^t b^X(s)\mathrm{d}w(s) + \int_{\mathbb{R}^m_* \times]0,t]} c^X(\zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad (4.4)$$

for any $t \geq 0$, where v is a d-dimensional adapted *continuous* process with local integral finite variation, w is a d-dimensional standard Wiener process independent of the compensated Poisson point process $\tilde{p}$ with Levy measure π in $\mathbb{R}^m_*$, and the coefficients suitable predictable processes, i.e., a^X is locally integrable with respect to the variation process $|\mathrm{d}v|$, b^X is locally square integrable, and c^X is jointly locally square integrable relative to $\pi(\mathrm{d}\zeta) \times \mathrm{d}t$. Choose $Y = w_k$ or $Y = \tilde{p}_\gamma$, where

$$\tilde{p}_\gamma(b) - \tilde{p}_\gamma(a) = \tilde{p}(\gamma,]a,b]) = \int_{\mathbb{R}^m_* \times]a,b]} \gamma(z)\tilde{p}(\mathrm{d}z, \mathrm{d}t), \quad \forall b > a \geq 0. \quad (4.5)$$

Thus, the expressions $[X, w_k]_\varpi$, $[X, \tilde{p}_\gamma]_\varpi$, $\Sigma^\pm_\varpi(X, \mathrm{d}w_k)$, $\Sigma^\circ_\varpi(X, \mathrm{d}w_k)$, $\Sigma^\pm_\varpi(X, \mathrm{d}\tilde{p}_\gamma)$ and $\Sigma^\circ_\varpi(X, \mathrm{d}\tilde{p}_\gamma)$ are adapted quasi-left continuous and cad-lag processes, and we may replace X by X_- without any modifications. By means of Itô formula we can calculate the predictable and optional covariances

$$[X, v_k](t) = \langle X, v_k\rangle(t) = 0, \qquad [X, w_k](t) = \langle X, w_k\rangle(t) = \int_0^t b^X_k(t)\mathrm{d}t,$$
$$\langle X, \tilde{p}_\gamma\rangle(t) = \int_0^t \mathrm{d}s \int_{\mathbb{R}^d_*} c^X(z,s)\gamma(z)\pi(\mathrm{d}z),$$
$$[X, \tilde{p}_\gamma](t) = \int_{\mathbb{R}^d_* \times]0,t]} c^X(z,s)\gamma(z)p(\mathrm{d}z, \mathrm{d}s),$$

and, for instance, Theorem 3.16 shows that

$$\lim_{|\varpi|\to 0}[X,w_k]_\varpi = [X,w_k], \qquad \lim_{|\varpi|\to 0}[X,\tilde p_\gamma]_\varpi = [X,\tilde p_\gamma],$$
$$\lim_{|\varpi|\to 0}\Sigma^-_\varpi(X,\mathrm{d}w_k) = \int_0^\cdot X(t)\mathrm{d}w_k(t),$$
$$\lim_{|\varpi|\to 0}\Sigma^-_\varpi(X_-,\mathrm{d}\tilde p_\gamma) = \int_{\mathbb{R}^d_*\times]0,\cdot]} X(t-)\gamma(z)\tilde p(\mathrm{d}z,\mathrm{d}t),$$

where the limits are uniformly on compacts in probability (i.e., in the ucp sense). Moreover, because the limits of the two last term of the left-hand side of the equality (4.3) converges to the stochastic integrals, we re-establish the convergence of the square-bracket to the optional covariation. Clearly, for the adapted continuous process having local bounded variation v we have

$$\lim_{|\varpi|\to 0}[X,v_k]_\varpi = 0,$$
$$\lim_{|\varpi|\to 0}\Sigma^-_\varpi(X,\mathrm{d}v_k) = \int_0^\cdot X(t-)\mathrm{d}v_k(t),$$

where the integral is pathwise, in either Riemann-Stieltjes or Lebesgue-Stieltjes sense.

If $\{\gamma_j : j\ge 1\}$ is an orthonormal basis in $L^2(\mathbb{R}^m_*,\pi)$ then the jumps of X given by (4.4) can be expressed as

$$\int_{\mathbb{R}^m_*\times]0,t]} c^X(\zeta,s)\tilde p(\mathrm{d}\zeta,\mathrm{d}s) = \sum_j \int_{]0,t]} c^X_j(s)\mathrm{d}\tilde p_j(s),$$

where

$$c^X_j(s) = \int_{\mathbb{R}^m_*} c^X(\zeta,s)\gamma_j(\zeta)\pi(\mathrm{d}\zeta), \qquad \tilde p_j(t) = \int_{\mathbb{R}^m_*\times]0,t]}\gamma_j(\zeta)\tilde p(\mathrm{d}\zeta,\mathrm{d}s),$$

$c^X_j(s)$ are predictable processes and $\tilde p_i(s)$ are purely discontinuous martingales, and

$$\mathrm{d}\langle\tilde p_i,\tilde p_j\rangle(t) = \Big(\int_{\mathbb{R}^m_*}\gamma_i(\zeta)\gamma_j(\zeta)\pi(\mathrm{d}\zeta)\Big)\mathrm{d}t,$$

i.e., $\langle\tilde p_i,\tilde p_j\rangle(t) = t$ if $i=j$ and $\langle\tilde p_i,\tilde p_j\rangle = 0$ otherwise. Thus, we may rewrite X as

$$X(t) = \int_0^t a^X(s)\mathrm{d}v(s) + \int_0^t b^X(s)\mathrm{d}w(s) + \sum_j\int_{]0,t]} c^X_j(s)\mathrm{d}\tilde p_j(s),$$

for any $t\ge 0$. Formally, we have $\tilde p = \sum_j \gamma_j\tilde p_j$, but

$$\mathbb{E}\big\{|\sum_j\gamma_j\tilde p_j(t)|^2_\pi\big\} = \mathbb{E}\big\{\sum_j|\tilde p_j(t)|^2\big\} = \mathbb{E}\big\{t\sum_j|\gamma_j|^2_\pi\big\} = \infty, \quad \forall t>0,$$

i.e., the series cannot be considered as $L^2(\mathbb{R}^m_*,\pi)$-valued martingale. However, for as given convergent sequence of strictly positive numbers $\{\kappa_i\}$ we may consider the Hilbert subspace

$$H = H_{\kappa,\gamma,\pi} = \big\{h \in L^2(\mathbb{R}^m_*,\pi) : \sum_i \kappa_i \Big| \int_{\mathbb{R}^m_*} h(\zeta)\gamma_i(\zeta)\pi(\mathrm{d}\zeta)\Big|^2 < \infty\big\}.$$

Hence, we may regard the series $\tilde p(s) = \sum_j \gamma_j \tilde p_j(s)$ and $c^x(s) = \sum_j \gamma_j c^x_j(s)$ as processes with values in H,

$$\|\tilde p(s)\|^2_H = \sum_j \kappa_j |\tilde p_j(s)|^2 \le \Big(\sum_j \kappa_j\Big) s,$$
$$\|c^x(s)\|^2_H = \sum_j \kappa_j |c^x_j(s)|^2 \le \Big(\sum_j \kappa_j\Big) \int_{\mathbb{R}^m_*} |c^x(\zeta,s)|^2 \pi(\mathrm{d}\zeta),$$

and $\tilde p$ is a local-martingale, while $c^x(s)$ is predictable with values in the dual space H', via the functional Riesz representation, and the duality inclusion $H \subset L^2(\mathbb{R}^m_*,\pi) \subset H'$. Therefore, the stochastic integral with respect to the (local) martingale measure $\tilde p$ can be regarded as an stochastic integral with respect to a (local) martingale with values in the Hilbert space H and a predictable process with values in its dual space H'. Nevertheless, we may define

$$\langle\!\langle X,\tilde p\rangle\!\rangle = \sum_j \langle X,\tilde p_j\rangle = \sum_j \int_0^\cdot c^x_j(s)\mathrm{d}s =$$
$$= \sum_j \int_0^\cdot \mathrm{d}s \int_{\mathbb{R}^m_*} c^x(\zeta,s)\gamma_j(\zeta)\pi(\mathrm{d}\zeta),$$
$$[\![X,\tilde p]\!] = \sum_j [X,\tilde p_j] = \sum_j \sum_{s\le\cdot} \delta X(s)\delta\tilde p_j(s) =$$
$$= \sum_j \int_{\mathbb{R}^m_*\times]0,\cdot]} c^x(\zeta,t)\gamma_j(\zeta)p(\mathrm{d}z,\mathrm{d}t),$$

if the coefficients are sufficiently smooth (in time) to make the above series convergent. Since the integrand is predictable,

$$\mathbb{E}\Big\{\int_{\mathbb{R}^m_*\times]0,T]} c^x(\zeta,t)\gamma_j(\zeta)p(\mathrm{d}\zeta,\mathrm{d}t)\Big\} =$$
$$= \mathbb{E}\Big\{\int_0^T \mathrm{d}t \int_{\mathbb{R}^m_*} c^x(\zeta,t)\gamma_j(\zeta)\pi(\mathrm{d}z)\Big\} = \mathbb{E}\Big\{\int_0^T c^x_j(t)\mathrm{d}t\Big\},$$

for any $T > 0$.

Now, recall that a Poisson measure p is a sum of (random) Dirac measures, i.e., $p(K,]a,b]) = \sum_{a<s\le t} \mathbb{1}_{\delta p(s)\in K}$ where $\delta p(s)$ denotes the jumps at time s (i.e., the Poisson point process originating the Poisson measure p), and assume

that γ_j is π-integrable so that

$$\int_{\mathbb{R}_*^m\times]0,T]}\gamma_j(\zeta)p(\mathrm{d}\zeta,\mathrm{d}t)= \\ =\int_{\mathbb{R}_*^m\times]0,T]}\gamma_j(\zeta)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t)+T\mathbb{E}\Big\{\int_{\mathbb{R}_*^d}\gamma_j(\zeta)\pi(\mathrm{d}\zeta)\Big\}$$

can be defined. Therefore, the integer measure ν_j induced by the pathwise integral of γ_j over p, i.e.,

$$p(\gamma_j,t)=\int_{\mathbb{R}_*^d\times]0,t]}\gamma_j(\zeta)p(\mathrm{d}\zeta,\mathrm{d}s),\qquad \nu_j(K,]a,b])=\sum_{a<s\le b}\mathbb{1}_{\{\delta p(\gamma_j,t)\in K\}}$$

are defined, and the jump satisfy $\delta\nu_j(t)=\delta p(\gamma_j,t)=\gamma_j(\delta p(t))$. Hence the integer measure ν_j is indeed a Poisson measure with Lévy measure $\pi_j(\mathrm{d}\zeta)=\gamma_j(\zeta)\pi(\mathrm{d}\zeta)$. Moreover, $t\mapsto\nu_j(\mathbb{R}_*^d,]0,t])$ is a composed (real valued) Poisson process with the finite measure π_j on $\mathbb{R}_*^d$ as parameter.

Thus, the stochastic integral with respect to either the initial Poisson measure $\{p(K,]a,b])\}$ or its associate Poisson point process $\{\delta p(t)\}$ can be written as an orthogonal series either $\{\nu_j\}$ or $\{\delta p(\gamma_j,t)\}$, i.e., with $\tilde{\nu}_j=\nu_j-\pi_j(\mathrm{d}\zeta)\mathrm{d}t$,

$$\int_{\mathbb{R}_*^m\times]0,t]}c^X(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)=\sum_j\int_{\mathbb{R}_*^d\times]0,t]}c^X(\zeta,s)\tilde{\nu}_j(\mathrm{d}\zeta,\mathrm{d}s).$$

and

$$\int_{\mathbb{R}_*^d\times]0,t]}c^X(\zeta,s)\nu_j(\mathrm{d}\zeta,\mathrm{d}s)=\int_{\mathbb{R}_*^d\times]0,t]}c^X(\zeta,s)\gamma_j(\zeta)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s).$$

Note that we may write $\tilde{p}_j(t)=\tilde{p}(\gamma_j,]0,t])$ and in the proper meaning for the jumps operator δ we have $\delta\tilde{p}_j(t)=\delta p(\gamma_j,t)$.

Sometimes, it is convenient to use the following

Definition 4.1. Let X be Itô processes with jumps as above, satisfying (4.4). We define the *backward*, *forward* and *symmetric* (or Stratonovich) stochastic integrals in term of the Itô stochastic integral as follows

$$\int_{]0,T]}X(t-)\mathrm{d}^-w_k(t)=\int_{]0,T]}X(t-)\mathrm{d}w_k(t),$$
$$\int_{]0,T]}X(t-)\mathrm{d}^+w_k(t)=\int_{]0,T]}X(t-)\mathrm{d}w_k(t)+\int_0^T b_k(t)\mathrm{d}t,$$
$$\int_{]0,T]}X(t-)\mathrm{d}^\circ w_k(t)=\int_{]0,T]}X(t-)\mathrm{d}w_k(t)+\frac{1}{2}\int_0^T b_k(t)\mathrm{d}t$$

and in general, for any two given semimartingales M and N we define the *backward*, *forward* and *symmetric* (or Stratonovich) stochastic integrals in term

of the Itô stochastic integral as follows

$$\int_{]0,T]} M(t-)\mathrm{d}^{-}N(t) = \int_{]0,T]} M(t-)\mathrm{d}N(t),$$
$$\int_{]0,T]} M(t-)\mathrm{d}^{+}N(t) = \int_{]0,T]} M(t-)\mathrm{d}N(t) + \big[M,N\big](T),$$
$$\int_{]0,T]} M(t-)\mathrm{d}^{\circ}N(t) = \int_{]0,T]} M(t-)\mathrm{d}N(t) + \frac{1}{2}\big[M,N\big](T).$$

Clearly, this take place in a probability space $(\Omega,\mathcal{F},P)$, with a completed (relative to $\mathcal{F}$), right-continuous and quasi-left continuous filtration $\mathbb{F}$. □

Remark that because the martingales are quasi-continuous and the local finite variation part is continuous, we are allow to use the square bracket $[\cdot,\cdot]$ instead of the angular bracket $\langle\cdot,\cdot\rangle$ as usually, without to much complication in the calculations, since jumps are deduced from the martingale measure $\tilde{\nu}$.

Thus, if M and N are two local-martingales with values in $\mathbb{R}^d$ and $\mathbb{R}$ (not necessarily continuous, but quasi-continuous and relative to the same filtered space) and associated martingale measures $\tilde{\nu}_{M}(\mathrm{d}z,\mathrm{d}t) = c^{M}(z,t)\tilde{\nu}(\mathrm{d}z,\mathrm{d}t)$ and $\tilde{\nu}_{N}(\mathrm{d}z,\mathrm{d}t) = c^{N}(z,t)\tilde{\nu}(\mathrm{d}z,\mathrm{d}t)$, for some integer measure ν in $\mathbb{R}^m_*$, then

$$\int_{]0,T]} \varphi\big(M(t-)\big)\mathrm{d}^{\circ}N(t) = \int_{]0,T]} \varphi\big(M(t-)\big)\mathrm{d}N(t) + \frac{1}{2}\big[\varphi(M),N\big](T),$$

for any smooth function $\varphi(x)$, and

$$\big[\varphi(M),N\big](T) = \sum_{i=1}^{d}\int_0^T \partial_i\varphi\big(M(t-)\big)\mathrm{d}\big[M_i,N\big]^{c}(t) +$$
$$+ \int_{\mathbb{R}^m_*\times]0,T]} \big[\varphi\big(M(t-)+c^{M}(z,t)\big) - \varphi\big(M(t-)\big)\big]c^{N}(z,t)\nu(\mathrm{d}z,\mathrm{d}t),$$

where $\partial_i\varphi$ denotes the derivative in x, and ν is the common integer measure. Clearly, the predictable covariance $\langle\varphi(M),N\rangle$ has an expression similar to the above with ν^p replacing the ν. In general, we may use the integer measure ν_{XY} in $\mathbb{R}^{d+1}_*$ associated with the (purely discontinuous part of the) $\mathbb{R}^{d+1}$-valued local-martingale (M,N), where we replace the integer measure $\nu(\mathrm{d}z,\mathrm{d}t)$ with $\nu_{XY}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}t)$ and the coefficients $c^{M}(z,t)$ and $c^{N}(z,t)$ with the variables x and y. In this case, the variable (x,y) belongs to $\mathbb{R}^{d+1}_*$ and the integral should be in $\mathbb{R}^{d+1}_*$. However, because ν_{XY} is an integer measure and the integrand function $[\varphi(\cdot+x)-\varphi(x)]y$ vanishes if $x=0$ or $y=0$, the integral is only on the region $\{(x,y)\in\mathbb{R}^{m+1}_* : x\neq 0,\ y\neq 0\}\times]0,T]$ as expected, i.e., when both martingales have jumps simultaneously.

It is clear that the *vector-form* is deduced from the above definition and the *operational* Itô rule becomes

$$X(b)Y(b) - X(a)Y(a) = \int_{]a,b]} X(t-)\mathrm{d}^{\circ}Y(t) + \int_{]a,b]} Y(t-)\mathrm{d}^{\circ}X(t),$$

i.e., as the deterministic case, with all the jumps incorporated into the integral. Note that the processes X and Y are cad-lag and quasi-left continuous, and that the bounded variation part v is a continuous process. In general,

$$\varphi(X(T)) - \varphi(X(0)) = \int_{]0,T]} \partial_x \varphi(X(t-)) \mathrm{d}X(t) + \int_{]0,T]} \partial_x \varphi(X(t-)) \mathrm{d}[X]^c(t) + \\ + \sum_{0<t\leq T} \Big(\varphi(X(t)) - \varphi(X(t-)) - \nabla \varphi(X(t-)) \cdot \delta X(t) \Big),$$

for any a smooth function φ. Clearly, we have

$$\sum_{0<t\leq T} \Big(\varphi(X(t)) - \varphi(X(t-)) - \partial_x \varphi(X(t-)) \delta X(t) \Big) = \\ = \int_{\mathbb{R}^d_* \times]0,T]} \big[\varphi(X(t-) + z) - \varphi(X(t-)) - z \cdot \nabla \varphi(X(t-)) \big] \nu_X(\mathrm{d}z, \mathrm{d}t),$$

where ν_X is the integer measure associated with X in $\mathbb{R}^d_*$. Clearly, $\nu_X = \tilde{\nu}_X + \nu^p_X$, where $\tilde{\nu}_X$ is the martingale measure (yielding the martingale stochastic integral) and ν^p_X is its predictable jump compensator.

The square-bracket $[\cdot, \cdot]$ and the integer measures can be defined for any cad-lag processes, non necessarily semimartingales. The previous relations between the backward, forward and symmetric integrals with the quadratic variation are essential for this analysis. The interested reader may consult for instance, Chao and Chou [21], Errami et al. [44], Fisk [49], Föllmer [53], Meyer [130], among others.

4.2 Quadratic Variation Arguments

One way of establishing Itô formula requires a more detailed discussion on the predictable quadratic variation defined via Doob-Meyer Decomposition (Theorem 2.7) for any local square integrable local-martingale (this includes any local-martingale with continuous paths) and the optional quadratic variation defined via the orthogonal decomposition into a continuous and a purely discontinuous martingale, applicable to any local-martingale.

4.2.1 Recall on Martingales Estimates

For future reference, recall that if M is a real valued martingale the Jensen's inequality implies that the absolute value $X = |M|$ is a sub-martingale. Thus, let us summarize the key estimates for a nonnegative cad-lag sub-martingale $X = \{X(t) : t \geq 0\}$:

(a) Doob's sup-estimate (or maximal inequality)

$$\mathbb{E}\big\{ \big(\sup_{t\leq T} X(t) \big)^p \big\} \leq \Big(\frac{p}{p-1} \Big)^p \mathbb{E}\big\{ \big(X(T) \big)^p \big\}, \tag{4.6}$$

for any bounded stopping time, and

$$\mathbb{E}\big\{\big(\sup_{t\geq 0} X(t)\big)^p\big\} \leq \Big(\frac{p}{p-1}\Big)^p \sup_{t\geq 0}\mathbb{E}\big\{\big(X(t)\big)^p\big\}, \tag{4.7}$$

which requires M bounded in L^p with $p>1$. However, the estimates

$$\varepsilon P\big\{\sup_{t\leq T} X(t)\geq \varepsilon\big\} \leq \mathbb{E}\big\{X(T)\,\mathbb{1}_{\sup_{t\leq T} X(t)\geq\varepsilon}\big\} \leq \mathbb{E}\big\{X(T)\big\}, \tag{4.8}$$

for any ε and T positive numbers, and

$$\mathbb{E}\big\{X(\tau)\big\} \leq 3\sup_{t\geq 0}\mathbb{E}\big\{X(t)\big\} \tag{4.9}$$

for any finite stopping time τ, hold true.

(b) Davis-Burkhölder-Gundy inequality for (cad-lag) local martingales vanishing at the initial time, namely

$$c_p\,\mathbb{E}\big\{\big([M](T)\big)^{p/2}\big\} \leq \mathbb{E}\{\sup_{t\leq T}|M(t)|^p\} \leq C_p\,\mathbb{E}\big\{\big([M](T)\big)^{p/2}\big\}, \tag{4.10}$$

valid for any $T\geq 0$ and $p\geq 1$ and some universal constants $C_p > c_p > 0$ independent of the filtered space, T and the local martingale M. In particular, we can take $C_1 = C_2 = 4$ and $c_1 = 1/6$. Moreover, a stopping time τ can be used in lieu of the time T and the above inequality holds true. It is clear that Davis-Burkhölder-Gundy inequality requires to calculate a priori the optional quadratic variation. However, if M is a square integrable local-martingale and $\langle M\rangle$ is its predictable quadratic variation (given via Doob-Meyer Decomposition) then

$$\mathbb{E}\{\sup_{t\leq T}|M(t)|^p\} \leq \Big(\frac{4-p}{2-p}\Big)\mathbb{E}\big\{\big(\langle M\rangle(T)\big)^{p/2}\big\}, \tag{4.11}$$

valid for any $T\geq 0$ and $0<p<2$. Furthermore, this bound holds for any $p\geq 1$ and some suitable constant $C_p>0$, provided the (square integrable) local-martingale M is continuous. Recall that a continuous local-martingale M is always a square integrable local-martingale and in this case $[M]=\langle M\rangle$, and Davis-Burkhölder-Gundy inequality (4.10) holds for any $p>0$. Also note that if $1<p\leq 2$ then Doob's maximal inequality (4.8) with p yields

$$c_p\mathbb{E}\{\sup_{t\leq T}|M(t)|^p\} \leq \mathbb{E}\{|M(T)|^p\} \leq \big(\mathbb{E}\{|M(T)|^2\}\big)^{p/2} = \mathbb{E}\big\{\big(\langle M\rangle(T)\big)\big\},$$

because $M^2-\langle M\rangle$ is a martingale, i.e., the constant (4.11) becomes 4 for $p=2$,

(c) Lenglart's inequality for dominate processes, i.e., if X and A are two cad-lag adapted processes such that A is monotone increasing and $\mathbb{E}\{|X_\tau|\}\leq\mathbb{E}\{A_\tau\}$, for every bounded stopping time τ, then for every stopping time τ and constants $\varepsilon,\eta>0$ we have

$$P\big\{\sup_{t\leq\tau}|X_t|\geq\varepsilon\big\} \leq \frac{1}{\varepsilon}\Big[\eta+\mathbb{E}\big\{\sup_{t\leq\tau}(A_t-A_{t-})\big\}\Big] + P\big\{A_\tau\geq\eta\big\}, \tag{4.12}$$

and if A is also predictable then the term with the jump $(A_t - A_{t-})$ is removed from the above estimate. This becomes

$$P\big\{\sup_{t\le\tau}|M_t|\ge\varepsilon\big\}\le\frac{\eta}{\varepsilon^2}+P\big\{\langle M\rangle_\tau\ge\eta\big\},\quad\forall\varepsilon,\eta>0, \tag{4.13}$$

for any stopping time τ, where $X=M^2$ and M is a (cad-lag) local integrable martingale having a continuous predictable quadratic variation $\langle M\rangle$.

(d) If M is a square integrable local-martingale with $\mathbb{E}\{M(0)\}=0$, then for any $t>s\ge 0$

$$\begin{aligned}\mathbb{E}\big\{\big(M(t)-M(s)\big)^2\big|\mathcal{F}(s)\big\}&=\\&=\mathbb{E}\big\{M^2(t)\big|\mathcal{F}(s)\big\}-2M(s)\mathbb{E}\big\{M(t)\big|\mathcal{F}(s)\big\}+M^2(s)=\\&=\mathbb{E}\big\{M^2(t)\big|\mathcal{F}(s)\big\}-M^2(s)=\mathbb{E}\big\{M^2(t)-M^2(s)\big|\mathcal{F}(s)\big\},\end{aligned}$$

which is usually called the orthogonal increment property. Moreover, if M has continuous paths with bounded variation then for a partition $\varpi=\{t_0<t_1<\cdots<t_n=t\}$ write

$$\begin{aligned}\mathbb{E}\{M^2(t)\}=\mathbb{E}\big\{\sum_{i=1}^n[M^2(t_i)-M^2(t_{i-1})]\big\}&=\\&=\mathbb{E}\big\{\sum_{i=1}^n\big(M(t_i)-M(t_{i-1})\big)^2\big\}\le\\&=\mathbb{E}\big\{\big(\sup_i|M(t_i)-M(t_{i-1})|\big)\sum_{i=1}^n|M(t_i)-M(t_{i-1})|\big\},\end{aligned}$$

to deduce $M(t)=0$ for every $t>0$, as the mesh of the partition $|\varpi|\to 0$. Also the following limit

$$\sup_{0<t\le T}\Big|\sum_{i=1}^n[M(t\wedge t_i)-M(t\wedge t_{i-1})]^p\big\}-[M](t)\Big|\to 0$$

holds in probability, for every $T>0$, as the mesh $|\varpi|\to 0$, with a partition $\varpi=\{t_0<t_1<\cdots<t_n=T\}$. At least this convergence is clear for local-martingales with continuous paths.

• *Remark* 4.2. It is perhaps interesting to note that the Doob's maximal inequality is a direct consequence of (4.8) by using the following claim: If a and p be strict positive constants and β be a function from $\mathbb{R}_+$ into itself such that

$$\frac{b^{-p}}{a^{-p}-\beta(b)}=C>0,\quad\text{for some}\quad b>0.$$

Then for any pair of nonnegative random variables X, Y satisfying

$$P\{X>at,\,Y<bt\}\le\beta(b)P\{X\ge r\},\quad\forall t,b>0,$$

the inequality $\mathbb{E}\{X^p\} \le C\mathbb{E}\{Y^p\}$ also holds true. To validate this claim, first use the elementary inequality $P\{X > at\} - P\{Y \ge bt\} \le P\{X > at,\, Y < bt\}$ to obtain $P\{X > at\} - P\{Y \ge bt\} \le \beta(b)P\{X \ge t\}$, for every $t, b > 0$. Since

$$\int_0^\infty t^{p-1}\mathbb{1}_{\{Z\ge t\}}\mathrm{d}t = p^{-1}\mathbb{E}\{Z^p\}, \quad p > 0$$

for any nonnegative random variable Z, integrating the previous inequality (after multiplying by t^{p-1}) yields $\mathbb{E}\{(X/a)^p\} - \mathbb{E}\{(Y/b)^p\} \le \mathbb{E}\{X^p\}$ or equivalently $(a^{-p} - \beta(b))\mathbb{E}\{X^p\} \le b^{-p}\mathbb{E}\{Y^p\}$, which rearranged is the desired inequality. Moreover, estimate (4.8) follows form the key upcrossing estimate, which yields also the following fact: For any super-martingale real-valued martingale X and any $t < T$ we have

$$\varepsilon P\{\sup_{t\le s\le T} X_s \ge \varepsilon\} \le \mathbb{E}\{X(t)\} + \mathbb{E}\{X^-(T)\},$$
$$\varepsilon P\{\inf_{t\le s\le T} X_s \le -\varepsilon\} \le \mathbb{E}\{X^-(T)\},$$

for every $\varepsilon > 0$, where $X^- = -\min\{X, 0\}$ is the negative part. Certainly, this may be applied to a sub-martingale Y by saying that $X = -Y$ is then a super-martingale. Combining all, we have

$$\varepsilon P\{\sup_{t\le s\le T} X_s \ge 3\varepsilon\} \le 4\mathbb{E}\{X(0)\} + 3\mathbb{E}\{X(T)\},$$

when X is a sub- or super-martingale. □

4.2.2 Estimates for Stochastic Integrals

Recall the stochastic integral relative to a Wiener process (or a continuous local-martingale) or relative to a Poisson measure (or a square local-martingale quasi-continuous)

$$I(t, f, \mathrm{d}w) = \int_0^t f(s)\mathrm{d}w(s) \quad \text{or} \quad I(t, g, \mathrm{d}\tilde\nu) = \int_{\mathbb{R}_*^m\times]0,t]} g(s,\zeta)\tilde\nu(\mathrm{d}\zeta, \mathrm{d}s),$$

where the predictable/optimal quadratic variations are known, namely,

$$\langle I(\cdot, f, \mathrm{d}w)\rangle = \int_0^\cdot |f(s)|^2\mathrm{d}s \quad \text{or} \quad \langle I(\cdot, g, \mathrm{d}\tilde\nu)\rangle = \int_0^\cdot \mathrm{d}s\int_{\mathbb{R}_*^m} |g(s,\zeta)|^2\pi(\mathrm{d}\zeta),$$

$[I(\cdot, f, \mathrm{d}w)] = \langle I(\cdot, f, \mathrm{d}w)\rangle$, and

$$[I(\cdot, g, \mathrm{d}\tilde\nu)] = \langle I(\cdot, g, \mathrm{d}\tilde\nu)\rangle + \int_{\mathbb{R}_*^m\times]0,\cdot]} |g(s,\zeta)|^2\nu(\mathrm{d}\zeta, \mathrm{d}s).$$

The Doob's maximal inequality (isometric equality without the sup for $p = 2$) with $p > 1$ and $C_p = [p(p-1)]^p$

$$\mathbb{E}\big\{\sup_{0\le t\le T} |I(t, f, \mathrm{d}w)|^p\big\} \le C_p\mathbb{E}\big\{|I(T, f, \mathrm{d}w)|^p\big\}, \quad p > 1$$
$$\mathbb{E}\big\{\sup_{0\le t\le T} |I(t, g, \mathrm{d}\tilde\nu)|^p\big\} \le C_p\mathbb{E}\big\{|I(T, g, \mathrm{d}\tilde\nu)|^p\big\}, \quad p > 1$$

which yield, only for $1 < p \le 2$

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} |I(t,f,\mathrm{d}w)|^p \Big\} \le C_p \Big(\mathbb{E}\Big\{ \int_0^T |f(s)|^2 \mathrm{d}s \Big\} \Big)^{p/2},$$
$$\mathbb{E}\Big\{ \sup_{0\le t\le T} |I(t,g,\mathrm{d}\tilde\nu)|^p \Big\} \le C_p \Big(\mathbb{E}\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}_*^m} |g(s,\zeta)|^2 \pi(\mathrm{d}\zeta) \Big\} \Big)^{p/2},$$

are sufficient (actually, $p = 2$ suffices) to develop the stochastic integral theory. A posteriori, the upper bound of Davis-Burkhölder-Gundy inequality (4.10) and (4.11) yield

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} |I(t,f,\mathrm{d}w)|^p \Big\} \le C_p \mathbb{E}\Big\{ \Big(\int_0^T |f(s)|^2 \mathrm{d}s \Big)^{p/2} \Big\}, \tag{4.14}$$

for any $p > 0$, and

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} |I(t,g,\mathrm{d}\tilde\nu)|^p \Big\} \le C_p \mathbb{E}\Big\{ \Big(\int_0^T \mathrm{d}s \int_{\mathbb{R}_*^m} |g(s,\zeta)|^2 \pi(\mathrm{d}\zeta) \Big)^{p/2} \Big\}, \tag{4.15}$$

only for $0 < p \le 2$. However, $p > 2$ this last estimate becomes

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} |I(t,g,\mathrm{d}\tilde\nu)|^p \Big\} \le C_p \Big(\mathbb{E}\Big\{ \Big[\int_0^T \mathrm{d}s \int_{\mathbb{R}_*^m} |g(s,\zeta)|^2 \pi(\mathrm{d}\zeta) \Big]^{p/2} \Big\} + \\ + \mathbb{E}\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}_*^m} |g(s,\zeta)|^p \pi(\mathrm{d}\zeta) \Big\} \Big), \quad p > 2. \tag{4.16}$$

Indeed, to check this bound, use Itö formula with the function $x \mapsto |x|^p$ and the process $X(t) = I(t,g,\mathrm{d}\tilde\nu)$ to obtain

$$c_p \mathbb{E}\{|X(t)|^p\} \le \int_0^t \mathrm{d}s \int_{\mathbb{R}_*^m} \big[|X(s-)|^{p-2} |g(s,\zeta)|^2 + |g(s,\zeta)|^p \big] \pi(\mathrm{d}\zeta),$$

for some constant $c_p > 0$ depending only on p, and satisfying

$$2\big[|a+b|^p - |a|^p - p|a|^{p-2}ab \big] = p(p-1)|a+\theta b|^{p-2}|b|^2 \le \\ \le \frac{2}{c_p} \big[|a|^{p-2}|b|^2 + |b|^p \big], \quad \forall a, b \ge 0,$$

for some θ in $[0,1]$. Next, Hölder inequality with q such that $q(1-2/p) = 1$, $q' = p/2$, yields

$$\int_0^t \mathrm{d}s \int_{\mathbb{R}_*^m} |X(t-)|^{p-2} |g(s,\zeta)|^2 \pi(\mathrm{d}\zeta) \le \Big(\mathbb{E}\Big\{ \sup_{0<s\le t} |X(s-)|^p \Big\} \Big)^{1-p/2} \times \\ \times \Big(\mathbb{E}\Big\{ \Big[\int_0^t \mathrm{d}s \int_{\mathbb{R}_*^m} |g(s,\zeta)|^2 \pi(\mathrm{d}\zeta) \Big]^{p/2} \Big\} \Big)^{2/p},$$

and the elementary inequality $ab \le a^{q'}/q' + b^q/q$, $a, b \ge 0$, implies

$$\int_0^t \mathrm{d}s \int_{\mathbb{R}_*^m} |X(t-)|^{p-2} |g(s,\zeta)|^2 \pi(\mathrm{d}\zeta) \le \frac{c_p}{2} \mathbb{E}\Big\{ \sup_{0<s\le t} |X(s-)|^p \Big\} + \\ + C_p' \mathbb{E}\Big\{ \Big[\int_0^t \mathrm{d}s \int_{\mathbb{R}_*^m} |g(s,\zeta)|^2 \pi(\mathrm{d}\zeta) \Big]^{p/2} \Big\},$$

with c_p the same constant as above and some suitable constant C_p'. Collection all pieces and combining with Doob's maximal inequality, estimate (4.16) follows with $C_p = \max\{2C_p'/c_p, 2/c_p\}$.

Since both stochastic integrals $I(\cdot, f, \mathrm{d}w)$ and $I(\cdot, g, \mathrm{d}\tilde{\nu})$ are locally square integrable, the predictable quadratic variations are defined and Lenglart's inequality (4.13) holds true, i.e.,

$$P\Big\{ \sup_{0\le t\le T} |I(t,f,\mathrm{d}w)| \ge \varepsilon \Big\} \le \frac{\eta}{\varepsilon^2} + P\Big\{ \int_0^T |f(s)|^2 \mathrm{d}s \ge \eta \Big\}, \tag{4.17}$$

$$P\Big\{ \sup_{0\le t\le T} |I(t,g,\mathrm{d}\tilde{\nu})| \ge \varepsilon \Big\} \le \frac{\eta}{\varepsilon^2} + P\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}_*^m} |g(s,\zeta)|^2 \pi(\mathrm{d}\zeta) \ge \eta \Big\},$$

for every $\varepsilon, \eta > 0$.

It is also clear that these estimates for stochastic integrals with respect to a Wiener process and a Poisson measure extent to more general situations, e.g., with respect to a local-martingale and with respect to a point process (or an integer random measure).

4.2.3 Quadratic Variations for Continuous SIs

Denote by $\varpi \wedge t$ the partition determined by the points $t \wedge t_i$, $i = 0, 1, \ldots$, where $\varpi = \{0 = t_0 < t_1 < \cdots < t_n < t_{n+1} < \cdots\}$ is a partition of the whole semi-line $[0, \infty)$, with $t_n \to \infty$ and mesh $|\varpi| = \sup_i (t_i - t_{i-1})$. Actually, a sequence $\{\varpi_n, n = 1, 2, \ldots\}$ of partitions with mesh $|\varpi_n| \to 0$ is used, which is referred to as $|\varpi| \to 0$. In a filtered probability space $(\Omega, \mathbb{F}, P)$, suppose given a predictable cag-lad (continuous from the left and having limits from the right) piecewise constant real-valued process h_ϖ, $h_\varpi(s) = h_\varpi(t_i)$ for $t_{i-1} < s \le t_i$, $h_\varpi(t_i)$ is a $\mathcal{F}(t_{i-1})$-measurable random variable, $i = 1, 2, \ldots$, and consider the 'Riemann' sum

$$K_\varpi(t, h_\varpi, \mathrm{d}^2 w) = \sum_{i=1}^{\infty} h_\varpi(t_{i-1}) [w(t \wedge t_i) - w(t \wedge t_{i-1})]^2, \tag{4.18}$$

where $\{w(t) : t \ge 0\}$ is a standard real-valued Wiener process. The above series is a finite sum for each t, and that only the restriction of h_ϖ to the partition $\varpi \wedge t$ actually intervenes in the expression of $K(t, h_\varpi, \mathrm{d}^2 w)$. The typical case is when h is a predictable process continuous from the left and $h_\varpi(s) = h_\varpi(t_i)$ for $t_{i-1} < s \le t_i$, so that if h also belongs to L^1 then $h_\varpi \to h$ in L^1 as $|\varpi| \to 0$.

Note that as in standard Riemann sums, if ϖ' is a refinement of ϖ, i.e., $\varpi' \subset \varpi$, then (a) the process h_φ can be considered as a particular process $h_{\varpi'} = h_\varpi$, and (b) $K(t, \varpi, h_\varpi, \mathrm{d}^2 w) = K(t, \varpi', h_{\varpi'}, \mathrm{d}^2 w)$, and as a consequence, if ϖ' and ϖ'' are two partitions, then the partition $\varpi = \varpi' \cup \varpi''$ is a refinement of both of them and

$$K_{\varpi'}(t, h_{\varpi'}, \mathrm{d}^2 w) + K_{\varpi''}(t, h_{\varpi''}, \mathrm{d}^2 w) = K_\varpi(t, h_\varpi, \mathrm{d}^2 w),$$

where $h_\varpi = h_{\varpi'} + h_{\varpi''}$. Also note that

$$\mathbb{E}\Big\{ K_\varpi(t, h_\varpi, \mathrm{d}^2 w) \Big\} = \mathbb{E}\Big\{ \int_0^t h_\varpi(s), \mathrm{d}s \Big\}, \quad \forall t \geq 0,$$

because h_ϖ is predictable and the increments of w are independent.

Comparing with the stochastic integral

$$I_\varpi(t, h_\varpi, \mathrm{d}w) = \sum_{i=1}^{\infty} h_\varpi(t_{i-1})[w(t \wedge t_i) - w(t \wedge t_{i-1})],$$

note that

$$|I_\varpi(t, h_\varpi, \mathrm{d}w)|^2 \neq K_\varpi(t, h_\varpi^2, \mathrm{d}^2 w), \quad \text{but}$$
$$\mathbb{E}\{|I_\varpi(t, h_\varpi, \mathrm{d}w)|^2\} = \mathbb{E}\{K_\varpi(t, h_\varpi^2, \mathrm{d}^2 w)\},$$

and that the mapping $(h_\varpi, w) \mapsto K_\varpi(t, h_\varpi, \mathrm{d}^2 w)$ is linear in h_ϖ and quadratic in w, and that $h_\varpi \geq 0$ implies $K_\varpi(t, h_\varpi, \mathrm{d}^2 w) \geq 0$. Contrary to the stochastic integral, there is not a visible cancellation of the mixed terms when calculating $\mathbb{E}\{|K_\varpi(t, h_\varpi, \mathrm{d}^2 w)|^2\}$.

Nevertheless, our interest is the limit of the 'Riemman like sum' process $t \mapsto K_\varpi(t, h_\varpi, \mathrm{d}^2 w)$, as the mesh of the partition $|\varpi|$ vanishes. Because $0 \leq t \leq T$ for a fixed real number T, the partition $\varpi = \varpi \wedge T$, and without loss of generality, we refer to a partition on the bounded interval $[0, T]$.

QV Definition of Wiener Integrals

The following result could be used to show the existence of the optional/predictable quadratic variation of a stochastic integral with respect to a Wiener process without invoking Doob-Meyer Decomposition (Theorem 2.7).

Theorem 4.3. *On a given filtered probability space $(\Omega, \mathbb{F}, P)$, let $\{w(t) : t \geq 0\}$ be a standard real-valued Wiener process and $\{h(t) : t \geq 0\}$ be a predictable locally integrable process on $[0, T]$, for a fixed real number $T > 0$. With the previous notation, if $\varpi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ is a deterministic partition with mesh $|\varpi| \to 0$ and $\{h_\varpi(t) : t \geq 0\}$ is a sequence of predictable processes, with $h_\varpi(t)$ equal to a $\mathcal{F}(t_{i-1})$ random variable on the interval $]t_{i-1}, t_i]$, such that*

$$P\Big\{ \int_0^T |h_\varpi(t) - h(t)| \mathrm{d}t \geq \varepsilon \Big\} \to 0 \quad \textit{as} \quad |\varpi| \to 0,$$

for every $\varepsilon > 0$, then

$$\sup_{0\le t\le T}\Big|\sum_{i=1}^{n} h_\varpi(t_{i-1})[w(t\wedge t_i)-w(t\wedge t_{i-1})]^2-\int_0^t h(s)\mathrm{d}s\Big|\to 0,$$

and for any nonnegative numbers α and β,

$$\sum_{i=1}^{n}|h_\varpi(t_{i-1})|\,|w(t_i)-w(t_{i-1})|^{2\alpha}|t_i-t_{i-1}|^{\beta}\to 0,\quad \alpha+\beta>1,$$

both in probability.

Proof. First by using the positive and negative parts of h_ϖ and h, we may assume h_ϖ and h nonnegative, without any lost of generality.

Next, by localization, define $h^r_\varpi(t)=h_\varpi(t\wedge\tau_r)$ and $h^r(t)=h(t\wedge\tau_r)$, where the stopping times

$$\tau_r=\inf\Big\{t\in[0,T]:\int_0^t|h(s)|\mathrm{d}s\ge r\Big\},$$
$$\tau_\varpi=\inf\Big\{t\in[0,T]:\int_0^t|h(s)-h_\varpi(s)|\mathrm{d}s\ge 1\Big\},$$

satisfy $P\{\tau_r<T\}\to 0$ as $r\to\infty$ and $P\{\tau_\varpi<T\}\to 0$ as $|\varpi|\to\infty$, because h is a locally integrable process and $h_\varpi\to h$ in probability. Moreover, the sequence h_ϖ converges to h and so, there exists a $\delta>0$ such that if $|\varpi|<\delta$ then

$$\int_0^t\big[|h_\varpi(s)|+|h(s)|\big]\mathrm{d}s\le\int_0^t\big[|h_\varpi(s)-h(s)|+2|h(s)|\big]\mathrm{d}s\le$$
$$\le 1+2\int_0^t|h(s)|\mathrm{d}s,$$

and

$$\sum_{i=1}^{n}|h^r_\varpi(t_{i-1})|(t_i-t_{i-1})\le(1+2r),\quad\text{as long as}\quad|\varpi|<\delta,$$

for every $0\le t<\tau_r\wedge\tau_\varpi$.

Therefore, assuming the results valid for 'bounded' data, the inequalities

$$P\Big\{\sup_{0\le t\le T}\Big|K_\varpi(t,h_\varpi,\mathrm{d}^2w)-\int_0^t h(s)\mathrm{d}s\Big|\ge\varepsilon\Big\}\le P\{\tau_r\wedge\tau_\varpi<T\}$$
$$+P\Big\{\sup_{0\le t\le T}\Big|K_\varpi(t,h^r_\varpi,\mathrm{d}^2w)-\int_0^t h^r(s)\mathrm{d}s\Big|\ge\varepsilon\Big\},$$

and

$$P\Big\{\Big|\sum_{i=1}^{n} h_\varpi(t_{i-1})|w(t_i)-w(t_{i-1})|^{2\alpha}|t_i-t_{i-1}|^{\beta}\Big|\geq\varepsilon\Big\}\leq P\{\tau_r\wedge\tau_\varpi<T\}+$$
$$+P\Big\{\Big|\sum_{i=1}^{n} h^r_\varpi(t_{i-1})|w(t_i)-w(t_{i-1})|^{2\alpha}|t_i-t_{i-1}|^{\beta}\Big|\geq\varepsilon\Big\}$$

imply the desired converges, after taking first limit as the mesh of the partition $|\varpi|\to 0$ and then as $r\to\infty$.

After all this simplification, we are reduced to the case where the processes h and h_ϖ are also nonnegative and satisfy the bound

$$\int_0^T \big[h(s)+h_\varpi(s)\big]\mathrm{d}s\leq C<\infty,\quad \text{a.s.},$$

for some constant $C>0$.

Now, because $w(t)-w(s)$ is a normal distributed random variable with $\mathbb{E}\{w(t)-w(s)\}=0$ and $\mathbb{E}\{|w(t)-w(s)|^2\}=(t-s)$, for any positive real number $p>0$ there exists a constant depending C_p such that $\mathbb{E}\{|w(t)-w(s)|^{2p}\}\leq C_p|t-s|^p$, for Wiener process and any times $t\geq s\geq 0$. Thus, based on this estimate and the independence of $h_\varpi(t_{i-1})$ and $|w(t_i)-w(t_{i-1})|$ we have

$$\mathbb{E}\Big\{\sum_{i=1}^{n} h_\varpi(t_{i-1})|w(t_i)-w(t_{i-1})|^{2\alpha}|t_i-t_{i-1}|^{\beta}\Big\}\leq$$
$$\leq C_\alpha\sum_{i=1}^{n}\mathbb{E}\big\{h_\varpi(t_{i-1})\big\}(t_i-t_{i-1})^{\alpha+\beta}\leq$$
$$\leq C_\alpha\Big(\int_0^T\mathbb{E}\big\{h_\varpi(t)\big\}\mathrm{d}t\Big)\max_i\{(t_i-t_{i-1})^{\alpha+\beta-1}\},$$

which shows the convergence to zero, since the mesh $|\varpi|=\max_i\{(t_i-t_{i-1})\}$ vanishes.

Consider the process

$$t\mapsto M_\varpi(t)=\Big\{\sum_{i=1}^{n} h_\varpi(t_{i-1})[w(t\wedge t_i)-w(t\wedge t_{i-1})]^2-\int_0^t h_\varpi(s)\mathrm{d}s\Big\},$$

and let us check that $M_\varpi(t)$ is a martingale. Suppose $0\leq s<t\leq T$ and note that

$$M_\varpi(t)=M_\varpi(s)+$$
$$+\Big\{\sum_{i=1}^{n} h_\varpi(t_{i-1})[w(t\wedge t_i\vee s)-w(t\wedge t_{i-1}\vee s)]^2-\int_s^t h_\varpi(s)\mathrm{d}s\Big\},$$

$M_\varpi(s)$ is $\mathcal{F}(s)$-measurable, and

$$\mathbb{E}\Big\{\varphi\sum_{i=1}^{n} h_\varpi(t_{i-1})[w(t\wedge t_i\vee s)-w(t\wedge t_{i-1}\vee s)]^2\Big\}=\mathbb{E}\Big\{\varphi\int_s^t h_\varpi(s)\mathrm{d}s\Big\},$$

for any $\mathcal{F}(s)$-measurable random variable φ. This shows that $\mathbb{E}\{M_\varpi(t)|\mathcal{F}(s)\}=M_\varpi(s)$, i.e., M_ϖ is a martingale. Therefore, the equality

$$E\Big\{\sum_{i=1}^{n} h_\varpi(t_{i-1})[w(t\wedge t_i)-w(t\wedge t_{i-1})]^2\Big\}=\mathbb{E}\Big\{\int_0^t h_\varpi(s)\mathrm{d}s\Big\},\quad \forall t$$

yields the estimate

$$P\Big\{\sup_{0\le t\le T}\Big|\sum_{i=1}^{n} h_\varpi(t_{i-1})[w(t\wedge t_i)-w(t\wedge t_{i-1})]^2-\int_0^t h_\varpi(s)\mathrm{d}s\Big|\ge\varepsilon\Big\}\le$$
$$\le\frac{2}{\varepsilon}\mathbb{E}\Big\{\int_0^T h_\varpi(s)\mathrm{d}s\Big\},\quad(4.19)$$

after using Doob's sup-estimate (4.8). Note that Lenglart's inequality (4.12) could also be used.

Now, if ϖ' and ϖ'' are two partitions with $h_{\varpi'}$ and $h_{\varpi''}$ being their corresponding predictable piecewise constant processes, then construct a predictable piecewise constant process $h_\varpi(t)=h_{\varpi'}(t)-h_{\varpi''}(t)$ corresponding to the union of those partitions $\varpi=\varpi'\cup\varpi''$ to check that

$$\sum_{i=1}^{n'} h_{\varpi'}(t'_{i-1})[w(t'_i)-w(t'_{i-1})]^2-\sum_{i=1}^{n''} h_{\varpi''}(t''_{i-1})[w(t''_i)-w(t''_{i-1})]^2=$$
$$=\sum_{i=1}^{n} h_\varpi(t_{i-1})[w(t_i)-w(t_{i-1})]^2,$$

which yields

$$\mathbb{E}\Big\{\Big|\sum_{i=1}^{n'} h_{\varpi'}(t'_{i-1})[w(t'_i)-w(t'_{i-1})]^2-\sum_{i=1}^{n''} h_{\varpi''}(t''_{i-1})[w(t''_i)-w(t''_{i-1})]^2\Big|\Big\}\le$$
$$\le\mathbb{E}\Big\{\int_0^T|h_{\varpi'}(t)-h_{\varpi''}(t)|\mathrm{d}t\Big\}.$$

Moreover, as in the case of the Riemann sums and integral, the study of the limit as the mesh of the partition goes to zero is reduced to increasing sequences $\{\varpi_k\}$ of partitions with the mesh $|\varpi_k|\to 0$, where increasing means $\varpi_k\subset\varpi_{k+1}$, for any k.

Recall the notation

$$K_\varpi(t,h_\varpi,\mathrm{d}^2w)=\sum_{i=1}^{n} h_\varpi(t_{i-1})[w(t\wedge t_i)-w(t\wedge t_{i-1})]^2,$$

for any generic partition $\varpi = \{t_0 < t_1 < \cdots < t_n = T\}$ and any time t in $[0,T]$. It is then clear that the union $\bigcup_k \varpi_k$ is dense in $[0,T]$ and $K_{\varpi_k}(t, h_{\varpi_k}, \mathrm{d}^2 w) \geq K_{\varpi_k}(s, h_{\varpi_k}, \mathrm{d}^2 w)$, for any $s < t$ in ϖ_k. These properties combined with sup-martingale estimate (4.19) and the fact that $h_{\varpi'} \to h$ and $h_{\varpi''} \to h$, prove that the process $t \mapsto K_\varpi(t, h_\varpi, \mathrm{d}^2 w)$ converges, as the mesh $\varpi \to 0$, to some process denoted by $K(t, h, \mathrm{d}^2 w)$, i.e.,

$$\varepsilon P\Big\{ \sup_{0\leq t\leq T} |K_\varpi(t, h_\varpi, \mathrm{d}^2 w) - K(t, h, \mathrm{d}^2 w)| \geq \varepsilon \Big\} \leq$$
$$\leq \mathbb{E}\big\{ |K_\varpi(T, h_\varpi, \mathrm{d}^2 w) - K(T, h, \mathrm{d}^2 w)| \big\} \to 0, \quad \forall \varepsilon > 0,$$

as the mesh $|\varpi| \to 0$. Moreover, the limiting process $t \mapsto K(t, h, \mathrm{d}^2 w)$ is continuous, integrable and monotone increasing.

On the other hand, since the process

$$t \mapsto M_\varpi(t) = K_\varpi(t, h_\varpi, \mathrm{d}^2 w) - \int_0^t h_\varpi(s)\mathrm{d}s, \quad 0 \leq t \leq T$$

is a continuous martingale, the limiting process

$$t \mapsto M(t) = K(t, h, \mathrm{d}^2 w) - \int_0^t h(s)\mathrm{d}s, \quad 0 \leq t \leq T,$$

is a continuous martingale, with $M(0) = 0$ and paths with bounded variation. Hence $M(t) = 0$, for every t, which means that $K(t, h, \mathrm{d}^2 w)$ is the desired limit. □

• *Remark* 4.4. From the above proof, it should be clear that because the Wiener process has continuous paths, the limiting process $K(t, h, \mathrm{d}^2 w)$ can be identified with the integral of the process h, after using the fact that a continuous martingale having paths with bounded variation is necessarily constant. However, if a Lévy process is used instead of the Wiener process w, then the paths of $K(t, h, \mathrm{d}^2 w)$ could only be cad-lag. In this case, if it is known that $K(t, h, \mathrm{d}^2 w)$ is predictable then the process

$$t \mapsto M(t) = K(t, h, \mathrm{d}^2 w) - \int_0^t h(s)\mathrm{d}s$$

is a (cad-lag) predictable martingale having paths with bounded variation, and thus, it is necessarily constant, e.g., see Rogers and Williams [153, Theorem VI.19.4, pp. 347–348]. Also, note there is not problem in allowing an stochastic partition, as long as the times t_i are stopping times, i.e., the partition ϖ could be taken not deterministic, but assuming that the t_i are stopping times. In this case, the convergence remains true, and the construction of a $\varpi = \varpi' \cup \varpi''$ is complicate. □

• *Remark* 4.5. Typically, given and adapted (locally) integrable process h and a partition $\varpi = \{t_0 < t_1 < \cdots < t_n = T\}$ the predictable process h_ϖ defined

by $h_\varpi(t) = 0$ for any t in $[t_0, t_1]$ and

$$h_\varpi(t) = \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} h(s)\mathrm{d}s, \quad \forall t \in]t_i, t_{i+1}],$$

for $i = 1, 2, \ldots, n-1$. Note that $t \mapsto h_\varpi(t+)$ is a cad-lag version of h_ϖ, while h_ϖ is a cag-lad (continuous from the left with limit from the right) process, which converges to h in the $L^1(\Omega\times]0,T[)$ norm. If a bounded approximation is required then the truncation, i.e., $h_\varpi \wedge (1/|\varpi|)$ is useful. Clearly, the expression $K_\varpi(t, h_\varpi, \mathrm{d}^2 w)$ is quadratic in $\mathrm{d}^2 w$ and linear in h_ϖ, and continuous in the $L^1(\Omega\times]0,T[)$ norm for h_ϖ, which means that by truncating h_ϖ, it can also be assumed bounded by a deterministic constant, if necessary. □

Connection with Two independent Wiener Processes

This generalize the previous arguments.

Corollary 4.6. *As in Theorem 4.3, but suppose given three optional processes f, g and h such that f and g are locally square integrable, h is locally integrable and the products hf^2 and hg^2 are locally integrable, and the sequence $\{h_\varpi(t) : t \geq 0\}$ converges only in probability in the sense that*

$$P\Big\{ \int_0^T |h_\varpi(t) - h(t)| \big[1 + |f(t)|^2 + |g(s)|^2\big] \mathrm{d}t \geq \varepsilon \Big\} \to 0, \quad \textit{as} \quad |\varpi| \to 0,$$

for every $\varepsilon > 0$, then

$$\sup_{0\leq t\leq T} \Big| \sum_{i=1}^n h_\varpi(t_{i-1}) \Big(\int_{t\wedge t_{i-1}}^{t\wedge t_i} f(s)\mathrm{d}w(s) \Big) \Big(\int_{t\wedge t_{i-1}}^{t\wedge t_i} g(s)\mathrm{d}w(s) \Big) - \\ - \int_0^t f(s)g(s)h(s)\mathrm{d}s \Big| \to 0,$$

and for any nonnegative numbers α and β with $\alpha + \beta > 1$,

$$\sum_{i=1}^n |h_\varpi(t_{i-1})| \Big| \int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s) \Big|^{2\alpha} \Big(\int_{t_{i-1}}^{t_i} |g(s)|^2 \mathrm{d}s \Big)^\beta \to 0,$$

both in probability. Moreover, if w_1 is a real-valued standard Wiener process independent of w then

$$\sup_{0\leq t\leq T} \Big| \sum_{i=1}^n h_\varpi(t_{i-1}) \Big(\int_{t\wedge t_{i-1}}^{t\wedge t_i} f(s)\mathrm{d}w(s) \Big) \Big(\int_{t\wedge t_{i-1}}^{t\wedge t_i} g(s)\mathrm{d}w_1(s) \Big) \Big| \to 0,$$

in probability, as $|\varpi| \to 0$.

Proof. First, use the stopping time

$$\tau_r = \inf \Big\{ t \geq 0 : \int_0^t \big[1 + |h(s)|\big] \big[1 + |f(s)|^2 + |g(s)|^2\big] \mathrm{d}s \geq r \Big\},$$

and proceed by localization as in Theorem 4.3 to be able to assume without any loss of generality that the processes h and h_ϖ are also nonnegative and together with f and g satisfy the bound

$$\int_0^T \big[1+h(s)+h_\varpi(s)\big]\big[1+|f(s)|^2+|g(s)|^2\big]\mathrm{d}s \le C<\infty, \quad \text{a.s.},$$

for some deterministic constant $C=C_T>0$, as $|\varpi|\to 0$.

Second, consider the expression

$$B_\varpi(t,h_\varpi,f\mathrm{d}w,g\mathrm{d}w) = \\ =\sum_{i=1}^n h_\varpi(t_{i-1})\Big(\int_{t\wedge t_{i-1}}^{t\wedge t_i} f(s)\mathrm{d}w(s)\Big)\Big(\int_{t\wedge t_{i-1}}^{t\wedge t_i} g(s)\mathrm{d}w(s)\Big),$$

to check that it is linear in each of the last three arguments h_ϖ, $f\mathrm{d}w$ and $g\mathrm{d}w$. Hence, the parallelogram equality yields

$$4B_\varpi(t,h_\varpi,f\mathrm{d}w,g\mathrm{d}w) = B_\varpi(t,h_\varpi,(f+g)\mathrm{d}w,(f+g)\mathrm{d}w) - \\ - B_\varpi(t,h_\varpi,(f-g)\mathrm{d}w,(f-g)\mathrm{d}w),$$

this means that only when $f=g$ can be discussed, without any loss of generality. Moreover, with the notation as in Theorem 4.3, this is $B_\varpi(t,h_\varpi,f\mathrm{d}w,f\mathrm{d}w)=K_\varpi(t,h_\varpi,f\mathrm{d}^2w)$.

At this point, note that $h_\varpi(t_{i-1})$ is not longer independent of the stochastic integral on $[t_{i-1},t_i]$, but

$$\mathbb{E}\Big\{h_\varpi(t_{i-1})\Big(\int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s)\Big)^2\Big\} = \\ =\mathbb{E}\Big\{h_\varpi(s)\mathbb{E}\Big[\Big(\int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s)\Big)^2\,\Big|\,\mathcal{F}(t_{i-1})\Big]\Big\} = \\ =\mathbb{E}\Big\{h_\varpi(s)\mathbb{E}\Big[\Big(\int_{t_{i-1}}^{t_i} |f(s)|^2\mathrm{d}s\Big)\,\Big|\,\mathcal{F}(t_{i-1})\Big]\Big\},$$

which shows the equality

$$\mathbb{E}\Big\{\sum_{i=1}^n h_\varpi(t_{i-1})\Big(\int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s)\Big)^2\Big\} = \mathbb{E}\Big\{\int_0^t h_\varpi(s)|f(s)|^2\mathrm{d}s\Big\},$$

and the bound

$$\mathbb{E}\big\{\sup_{0\le t\le T}\big|K_\varpi(t,h_\varpi,f\mathrm{d}^2w)\big|\big\} \le 4\mathbb{E}\Big\{\int_0^T h_\varpi(s)|f(s)|^2\mathrm{d}s\Big\},$$

if need it. Therefore, as $|\varpi|\to 0$, choose a sequence of predictable processes f_ϖ constant on each time-subinterval $]t_{i-1},t_i]$, such that $f_\varpi\to f$ in $L^2(P\times\mathrm{d}s)$, and replace f with f_ϖ to check that

$$P\big\{\sup_{0\le t\le T}\big|K_\varpi(t,h_\varpi,f_\varpi\mathrm{d}^2w)-K_\varpi(t,h_\varpi,f\mathrm{d}^2w)\big|\ge\varepsilon\big\}\to 0, \quad \forall\varepsilon>0,$$

and $K_\varpi(t, h_\varpi, f_\varpi \mathrm{d}^2 w) = K_\varpi(t, h_\varpi f_\varpi^2, \mathrm{d}^2 w)$. Hence, invoke Theorem 4.3 to conclude the first part.

Regarding the last convergence to zero, note that by means of the elementary inequality

$$\begin{aligned}\Big|\int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s)\Big|^{2\alpha}\Big(\int_{t_{i-1}}^{t_i}|g(s)|^2\mathrm{d}s\Big)^{\beta} \leq \\ \leq \Big(\Big|\int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s)\Big|^2 + \Big|\int_{t_{i-1}}^{t_i}|g(s)|^2\mathrm{d}s\Big|\Big)^{\alpha+\beta} \leq \\ \leq C_{\alpha,\beta}\Big[\Big(\Big|\int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s)\Big|^2\Big)^{\alpha+\beta} + \Big|\int_{t_{i-1}}^{t_i}|g(s)|^2\mathrm{d}s\Big|^{\alpha+\beta}\Big],\end{aligned}$$

for some constant $C_{\alpha,\beta} > 0$, and the martingale estimate

$$\begin{aligned}\mathbb{E}\Big\{\Big(\int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s)\Big)^{2(\alpha+\beta)}\,\Big|\,\mathcal{F}(t_{i-1})\Big\} \leq \\ \leq C_\alpha \mathbb{E}\Big\{\Big(\int_{t_{i-1}}^{t_i}|f(s)|^2\mathrm{d}s\Big)^{\alpha+\beta}\,\Big|\,\mathcal{F}(t_{i-1})\Big\},\end{aligned}$$

for some constant $C_\alpha > 0$, we are reduced to prove that

$$\mathbb{E}\Big\{\sum_{i=1}^n |h_\varpi(t_{i-1})|\Big(\int_{t_{i-1}}^{t_i}\big[|f(s)|^2 + |g(s)|^2\big]\mathrm{d}s\Big)^{\alpha+\beta}\Big\} \to 0,$$

as the mesh $|\varpi| \to 0$.

To this purpose, fix a path and in view of the absolute continuity (in t), for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\sup_i \Big\{\int_{t_{i-1}}^{t_i}\big[|f(s)|^2 + |g(s)|^2\big]\mathrm{d}s\Big\} < \varepsilon,$$

for any partition with mesh $|\varpi| < \delta$. Therefore,

$$\begin{aligned}\sum_{i=1}^n |h_\varpi(t_{i-1})|\Big(\int_{t_{i-1}}^{t_i}\big[|f(s)|^2 + |g(s)|^2\big]\mathrm{d}s\Big)^{\alpha+\beta} \leq \\ \leq \varepsilon^{\alpha+\beta-1}\Big\{\sum_{i=1}^n |h_\varpi(t_{i-1})|\int_{t_{i-1}}^{t_i}\big[|f(s)|^2 + |g(s)|^2\big]\mathrm{d}s\Big\},\end{aligned}$$

which shows the convergence to zero, for each fixed ω, almost surely. Since the integrand

$$\sum_{i=1}^n |h_\varpi(t_{i-1})|\Big(\int_{t_{i-1}}^{t_i}\big[|f(s)|^2 + |g(s)|^2\big]\mathrm{d}s\Big)^{\alpha+\beta} \leq$$

can be bounded by

$$\Big(\int_0^T \big[|f(s)|^2+|g(s)|^2\big]\mathrm{d}s\Big)^{\alpha+\beta-1}\Big(\int_0^T |h_\varpi(s)|\big[|f(s)|^2+|g(s)|^2\big]\mathrm{d}s\Big),$$

which in dominated by the deterministic constant $C_T^{\alpha+\beta}$, we can take the limit as $|\varpi\to 0$ inside the expectation $\mathbb{E}\{\cdot\}$ and the second part is proved.

Now, go back two independent Wiener processes w and w_1, where we have to verify that

$$P\big\{\sup_{0\le t\le T}|B_\varpi(t,h_\varpi,f\mathrm{d}w,g\mathrm{d}w_1)|\ge\varepsilon\big\}\to 0,\quad \forall\varepsilon>0,$$

with the previous notation. Again, note that the expression $B(t,\cdot,\cdot,\cdot)$ is linear in each variable and as seen early, the bound

$$\mathbb{E}\Big\{\sum_{i=1}^n |h_\varpi(t_{i-1})|\Big|\int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s)\Big|\Big|\int_{t_{i-1}}^{t_i} g(s)\mathrm{d}w_1(s)\Big|\Big\}\le$$
$$\le\mathbb{E}\Big\{\int_0^T |h_\varpi(s)|\big[|f(s)|^2+|g(s)|^2\big]\mathrm{d}s\Big\}.$$

allow the approximation of f and g with predictable processes f_ϖ and g_ϖ constant on $]t_{i-1},t_i]$, and the problem is reduced to prove that

$$\sum_{i=1}^n h_\varpi(t_{i-1})f_\varpi(t_{i-1})g_\varpi(t_{i-1})\big(w_k(t_i)-w_k(t_{i-1})\big)\times$$
$$\times\big(w_l(t_i)-w_l(t_{i-1})\big)\to 0.$$

Moreover, the processes satisfy

$$\int_0^T\big[1+h_\varpi(s)\big]\big[1+|f_\varpi(s)|^2+|g_\varpi(s)|^2\big]\mathrm{d}s\le C<\infty,\quad\text{a.s.},$$

for some deterministic constant $C=C_T>0$, as $|\varpi|\to 0$.

To show this convergence, remark that because the two Wiener processes are independent, the factor

$$\sigma_\varpi(t_{i-1})=h_\varpi(t_{i-1})f_\varpi(t_{i-1})\big(w(t_i)-w(t_{i-1})\big)g_\varpi(t_{i-1}),$$

for any s in $]t_{t_{i-1},t_i}]$, is independent of the increment $w_1(t_i)-w_1(t_{i-1})$, and as in the stochastic integral, this yields the equality

$$\mathbb{E}\Big\{\Big|\sum_{i=1}^n\sigma_\varpi(t_{i-1})\big(w_1(t_i)-w_1(t_{i-1})\big)\Big|^2\Big\}=\mathbb{E}\Big\{\sum_{i=1}^n|\sigma_\varpi(t_{i-1})|^2(t_i-t_{i-1})\Big\},$$
$$|\sigma_\varpi(t_{i-1})|^2=|h_\varpi(t_{i-1})|^2\,|f_\varpi(t_{i-1})|^2\big(w(t_i)-w(t_{i-1})\big)^2|g_\varpi(t_{i-1})|^2.$$

and rearranging the factors,

$$\mathbb{E}\Big\{\Big|\sum_{i=1}^{n}\sigma_\varpi(t_{i-1})\big(w_1(t_i)-w_1(t_{i-1})\big)\Big|^2\Big\}=$$
$$=\mathbb{E}\Big\{\sum_{i=1}^{n}|h_\varpi(t_{i-1})|^2|f_\varpi(t_{i-1})|^2|g_\varpi(t_{i-1})|^2(t_i-t_{i-1})^2\Big\}.$$

The sum inside the expectation is bounded by

$$\Big(\sup_i\int_{t_{i-1}}^{t_i}|h_\varpi(s)||f_\varpi(s)|^2\mathrm{d}s\Big)\Big(\int_0^T|h_\varpi(s)||g_\varpi(s)|^2\mathrm{d}s\Big)\le C_T^2,$$

and vanishes as the mesh $|\varpi|\to 0$, which complete the proof. □

• *Remark* 4.7. Instead of the interval $[0,T]$ we could consider a stochastic interval $[\tau_1\wedge T,\tau_2\wedge T[$ with two stopping times $\tau_1\le\tau_2$ and both convergences remain valid within this interval. Also, for the convergence to zero, instead of the stochastic integral in $\mathrm{d}w$ on the interval $[t_{i-1},t_i]$, we could use an expression of the form

$$\Big(\int_{t_{i-1}}^{t_i}f_1(s)\mathrm{d}w_1(s)\Big)^{2\alpha_1}\Big(\int_{t_{i-1}}^{t_i}f_2(s)\mathrm{d}w_2(s)\Big)^{2\alpha_2}\cdots$$
$$\cdots\Big(\int_{t_{i-1}}^{t_i}f_m(s)\mathrm{d}w_m(s)\Big)^{2\alpha_m}$$

with d independent real-valued Wiener processes $w_1,\dots,w_d$, and d locally square integrable processes $f_1,\dots,f_d$, provided that $2\alpha_1+\cdots+2\alpha_d+\beta>1$. Equivalently, suppose w is a d-dimensional Wiener process, f is a matrix of suitable size, and replace the absolute value $|\cdot|$ with a Euclidean norm. Note that in Corollary 4.6, the process h_ϖ can be considered part of f or g, and $B_\varpi(t,h_\varpi,f\mathrm{d}w,g\mathrm{d}w)$ is bilinear in $(f\mathrm{d}w,g\mathrm{d}w)$ and linear in h_ϖ. □

• *Remark* 4.8. Therefore, Theorem 4.3 and Corollary 4.6 yield the following: For real-valued Itô differentials of the form $\mathrm{d}M=g^M\mathrm{d}t+\sum_k\sigma_k^M\mathrm{d}w_k$ and $\mathrm{d}N=g^N\mathrm{d}t+\sum_k\sigma_k^M\mathrm{d}w_k$, where (w_k) is a finite number of independent real-valued standard Wiener processes, we have for any real-valued predictable bounded process h

$$\sum_{i=1}^{n}h(t_{i-1})\big(M(t_i)-M(t_{i-1})\big)\big(N(t_i)-N(t_{i-1})\big)\to$$
$$\to\sum_k\int_0^T h(t)\sigma_k^M(t)\sigma_k^N(t)\mathrm{d}t,$$

while for any nonnegative numbers α and β such that $\alpha+\beta>1$,

$$\sum_{i=1}^{n}|h(t_{i-1})|\big|M(t_i)-M(t_{i-1})\big|^{2\alpha}\big|N(t_i)-N(t_{i-1})\big|^{2\beta}\to 0,$$

as the mesh of the partition $0 = t_0 < t_1 < \cdots < t_n = T$ vanishes. Both convergences are in probability, and actually, if T is replaced by any time t in $[0, T]$, then the convergence is uniform in t. All processes g^M, σ_k^M, g^N, σ_k^N, and h are assumed to be predictable, g^M and g^N are locally integrable, σ_k^M and σ_k^N are locally square-integrable, h, $(\sigma_k^M)^2 h$ and $(\sigma_k^N)^2 h$ are locally integrable and if $h(0) = 0$ and $h_\varpi(s) = h(t_{i-1})$ for $t_{i-1} < s \le t_i$ then

$$\int_0^T |h(s) - h_\varpi(s)| \big[1 + |\sigma_k^M(s)|^2 + |\sigma_k^N(s)|^2\big] \mathrm{d}s \to 0 \quad \text{as} \quad |\varpi| \to 0,$$

in probability, for every k. □

Deterministic QV

As a transition to the quadratic variation for discontinuous processes, and before going further, let us recall some properties of cad-lag / cag-lad functions:

(1) If the oscillation of a function f on an interval I is given by $\mathrm{osc}(f, I) = \sup\{|f(s) - f(t)| : s, t \in I\}$ or equivalently as $\mathrm{osc}(f, I) = \sup_I f - \inf_I f$, then essentially by the definition of right- and left-hand limits follows that for any real-valued cad-lag (or cag-lad) function f on $[0, T]$ and for every $\varepsilon > 0$ there exists a partition ϖ of the form $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$ with mesh $|\varpi| < \varepsilon$ such that $\mathrm{osc}(f, [t_{i-1}, t_i[) < \varepsilon$ (or $\mathrm{osc}(f,]t_{i-1}, t_i]) < \varepsilon$), for every $i = 1, \dots, n$. In other words, the cad-lag piecewise constant function

$$f_\varpi(0) = f(0), \quad \text{and} \quad f_\varpi(t) = f(t_{i-1}), \quad \forall t \in [t_{i-1}, t_i[$$

satisfies $|f(t) - f_\varpi(t)| < \varepsilon$ (or $|f(t) - f_\varpi(t-)| < \varepsilon$), for every t in $[0, T]$. Hence, if a finite number of jumps are removed, namely,

$$g_\varpi(0) = f(0), \quad \text{and} \quad g_\varpi(t) = \sum_{i=1}^{n} \big(f(t_i) - f(t_i-)\big) \mathbb{1}_{t_i \le t}, \quad \forall t > 0,$$

then the function $t \mapsto f(t) - g_\varpi(t)$ (or $t \mapsto f(t) - g_\varpi(t-)$) is certainly cad-lag (or cag-lad) in $[0, T]$ and not necessarily continuous on (t_{i-1}, t_i), but it is continuous at each t_i, so that the oscillation within any closed interval $[t_{i-1}, t_i]$ is smaller than the given ε, i.e., $\mathrm{osc}(f - g_\varpi, [t_{i-1}, t_i]) < \varepsilon$. Moreover, if $\varpi' \supset \varpi$ is a subordinate partition then the above hold true with ϖ replaced by ϖ'.

(2) Any real-valued function with a bounded variation has a countable number of discontinuities and it can be written as the different of a cad-lag increasing and a cag-lad increasing functions, both cad-lag if f is itself cad-lag. A real-valued cad-lag function f is called a *purely jump function* if it is the sum of its jumps, i.e., if $f(t) = f(0) + \sum_{0<s\le t} \delta f(s)$, where $\delta f(s) = f(s) - f(s-)$. Since f is a cad-lag function, the piecewise constant function defined as the finite sum

$$f_n(t) = f(0) + \sum_{0<s\le t} \delta f(s) \mathbb{1}_{\{n|\delta f(s)| \ge 1\}}, \quad 0 \le t \le T,$$

converges to f, provided the series of jumps is absolutely convergent, i.e., if f has a bounded variation on $[0,t]$, for every $t>0$. The jumps $\delta f \neq 0$ of a cad-lag function f are countable, and the sequence of jumps $\{\delta f(t) \neq 0 : t>0\}$ (called a point function) defines a counting measure

$$f(B\times]0,t]) = \sum_{0<s\leq t} \mathbb{1}_{\{\delta f(s)\in B\}}, \quad 0\leq t\leq T$$

on $\mathbb{R}_* \times [0,T]$, with $\mathbb{R}_* = \mathbb{R}\setminus\{0\}$. Thus, the concept of a purely jumps function makes a neat sense only when the function has bounded variation, but for a generic cad-lag function, the function f_n may converge to f after being compensated on $[0,T]$, i.e., for every $\varepsilon>0$ there exists η sufficiently large such that the conditions $n\geq\eta$ and $0\leq a<b\leq a+1/\eta\leq T$ imply $\operatorname{osc}(f_n-f,[a,b])<\varepsilon$. Nevertheless, instead of adding all jumps, the point function or the counting measure could be analyzed, even if not every point function (or counting measure) corresponds to some cad-lag function.

(3) A cad-lag function f has finite quadratic variation if for any sequence of partitions ϖ, $0=t_0<t_1<\cdots<t_n=T$, with mesh $|\varpi|\to 0$ the ϖ-quadratic variation

$$[f]^{\varpi}(t) = \sum_{i=1}^{n} |f(t\wedge t_i) - f(t\wedge t_{i-1})|^2, \quad 0\leq t\leq T, \tag{4.20}$$

(or equivalently, just for $t=T$) is bounded. Note that if f is a continuous function with bounded variation on $[0,T]$ then the estimate

$$\sum_{i=1}^{n} |f(t\wedge t_i) - f(t\wedge t_{i-1})|^2 \leq$$
$$\leq \max_i \big\{\operatorname{osc}(f,[t_{i-1},t_i])\big\} \sum_{i=1}^{n} |f(t\wedge t_i) - f(t\wedge t_{i-1})|$$

shows that $\langle f\rangle^{\varpi}(t) = [f]^{\varpi}\to 0$ as $|\varpi|\to 0$, for every $0\leq t\leq T$. More general, if f is a continuous function and g is a function with bounded variation, the same type of estimate, namely,

$$\sum_{i=1}^{n} |f(t\wedge t_i) - f(t\wedge t_{i-1})|\,|g(t\wedge t_i) - g(t\wedge t_{i-1})| \leq$$
$$\leq \max_i \big\{\operatorname{osc}(f,[t_{i-1},t_i])\big\} \sum_{i=1}^{n} |g(t\wedge t_i) - g(t\wedge t_{i-1})| \tag{4.21}$$

show that the quadratic co-variation vanishes, i.e.,

$$\sum_{i=1}^{n} \big[f(t\wedge t_i) - f(t\wedge t_{i-1})\big]\big[g(t\wedge t_i) - g(t\wedge t_{i-1})\big] = \langle f,g\rangle^{\varpi}(t)\to 0,$$

as $|\varpi| \to 0$, for any t in $[0,T]$. However, if f is only a cad-lag function with bounded variation on $[0,T]$ then separate the jumps $f = f_c + f_d$ with $f_d(t) = \sum_{0<s\le t} \delta f(s)$ so that f_c is continuous, $\sum_{0<s\le T} |\delta f(s)| < \infty$ and use the equality

$$\big|f(t_i) - f(t_{i-1})\big|^2 = \big|f_c(t_i) - f_c(t_{i-1})\big|^2 + \big|f_d(t_i) - f_d(t_{i-1})\big|^2 + \\ + 2\big[f_c(t_i) - f_c(t_{i-1})\big]\big[f_d(t_i) - f_d(t_{i-1})\big]$$

to deduce that

$$[f]^{\varpi}(t) \to \sum_{0<s\le t} |\delta f(s)|^2 \quad \text{as} \quad |\varpi| \to 0,$$

uniformly in $0 \le t \le T$.

(4) If f is only cag-lad in (4.21) and g is also cad-lag then quadratic co-variation satisfies

$$\lim_{|\varpi'|\to 0,\, \varpi' \supset \varpi} \sum_{i=1}^{n'} \big|f(t\wedge t'_i) - f(t\wedge t'_{i-1})\big|\big|g(t\wedge t'_i) - g(t\wedge t'_{i-1})\big| \le \\ \le |\varpi| \sup_{\varpi' \supset \varpi} \sum_{i=1}^{n'} \big|g(t'_i) - g(t'_{i-1})\big|, \quad (4.22)$$

i.e., it vanishes as $|\varpi| \to 0$. Indeed, given a partition ϖ there exists another partition $\varpi'' \supset \varpi$ such that $\mathrm{osc}(f - f_{\varpi''}, [t''_{i-1}, t''_i]) < |\varpi|$, where $f_{\varpi''}$ is the cag-lad function

$$f_{\varpi''}(t) = \sum_{i=1}^{n''} \delta f(t''_i) \mathbb{1}_{t''_i < t}, \quad \delta f(t) = (f(t) - f(t-), \quad \forall t > 0,$$

with $t_0 = t''_0 = 0$ and $t_n = t''_{n''} = T$. Remark that $[f - f_{\varpi''}, g] = [f,g] - [f_{\varpi''}, g]$, $f_{\varpi''}(t) - f_{\varpi''}(t''_{i-1}) = \delta f(t''_{i-1})$ for any $t''_{i-1} < t \le t''_i$, so that for any other partition $\varpi' \supset \varpi''$, t'_{i-1} is not in ϖ'' implies $f_{\varpi''}(t'_i) - f_{\varpi''}(t'_{i-1}) = 0$, to deduce

$$\sum_{i=1}^{n'} \big|f_{\varpi''}(t\wedge t'_i) - f_{\varpi''}(t\wedge t'_{i-1})\big|\big|g(t\wedge t'_i) - g(t\wedge t'_{i-1})\big| \le \\ \le \sup_{1\le i\le n''-1,\ t''_i \le s < t''_i + |\varpi'|} \big\{|g(s) - g(t''_i)|\big\} \sum_{i=1}^{n''} \big|\delta f_{\varpi''}(t''_i)\big|$$

for any partition $\varpi' \supset \varpi''$, and any t in $[0,T]$. Hence, the right-hand side vanishes as $|\varpi'| \to 0$, and (4.22) follows after invoking (4.21).

(5) As mentioned early, a point function (or a counting measure) is a sequence of times and jumps given in the form $\{\delta f(s) : s \ge 0\}$ such that the jumps $\delta f(s) \ne 0$ for only a countable number of times s. A particular (and key

example) is the point function associated with jumps of a cad-lag function f, i.e., $\delta f(s) = f(s) - f(s-)$. It is clear that the set $\ell^p(]0,T])$, $1 \le p \le \infty$ of all point functions satisfying

$$\|\delta f\|_p^p = \sum_{0<s\le T} |\delta f(s)|^p < \infty$$

is a Banach space. In term of a counting measure νf and a Borel set $B \subset \mathbb{R}_*$, $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$, the notation $\nu f(B\times]0,t])$ means the value (a non-negative integer or the symbol ∞) of series $\sum_{0<s\le t} \mathbb{1}_{\{\delta f(s)\in B\}}$. Thus

$$\|\nu f\|_p^p = \int_{\mathbb{R}_*\times]0,T]} |\zeta|^p \nu f(\mathrm{d}\zeta, \mathrm{d}s).$$

and νf is an integer-valued measure on $\mathbb{R}_*\times]0,T]$. Note that in this context, δf (or νf) denotes a point function (or a counting measure), but if a cad-lag function f is given then δf (or νf) denotes its associated point function (or counting measure) as defined by the previous expressions, and clearly, a multidimensional case with $\mathbb{R}_*^m$ instead of $\mathbb{R}_*$ can be studied with almost no changes. For $p = 2$,

$$[\delta f, \delta g] = \sum_{0<s\le T} \big(\delta f(s)\big)\big(\delta g(s)\big)$$

is a scalar product in $\ell^2(]0,T])$, which becomes a Hilbert space. Also, adding the convention $f(0-) = 0$, the jump at $s = 0$ becomes $f(0)$ and the space $\ell^p([0,T])$ makes sense with obvious changes. For an element δf (or νf) in either $\ell^p(]0,T])$ or $\ell^p([0,T])$, the typical approximation by a finite number of bounded jumps, namely, $\{\delta f_n : s \ge 0\}$ including only jumps with size $|\delta f(s)|$ in the interval $[1/n, n]$, yields a convergence in norm, i.e., if $\|\delta f\|_p < \infty$ then $\|\delta f - \delta f_n\|_p \to 0$ as $n \to \infty$. If δf point function then the countable set of jump-times is $S_f = \{s : \delta f(s) \ne 0\}$, and thus, the scalar product $[\delta f, \delta g] = 0$ whenever S_f and S_g are disjoint. This implies that there exists an uncountable set of orthogonal point functions in $\ell^2(]0,T])$ or $\ell^2([0,T])$, so that these spaces are non separable, and similarly for ℓ^p. It should be clear that the finite sum

$$f_\eta(t) = \sum_{0<s\le t} \delta f(s) \mathbb{1}_{\{1/\eta\le|\delta f(s)|\le\eta\}}, \quad \forall t > 0, \quad \forall \eta \ge 1, \tag{4.23}$$

is defined for any given point function (or counting measure) δf in $\ell^p([0,T])$, $1 \le p < \infty$, but, as $\eta \to 0$ the limit may exist (or even can be compensated) even if the corresponding series does not converges absolutely.

(6) If f is a continuous function then its jumps δf vanish, i.e., the previous analysis on the space $\ell^p(]0,T])$ is meaningful only for the jumps. However, as mentioned early, a cad-lag function f is said to have finite quadratic variation if for any sequence of partitions $\varpi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ with mesh $|\varpi| \to 0$ the quadratic variation relative to ϖ on $[0,t]$ is given by (4.20)

is bounded, and if the limit exists $[f](t)$ then it is called the quadratic variation of f. Certainly, the p-variation is defined by exchanging the power 2 with p, $1 \leq p < \infty$. Alternatively, if a cad-lag function is approximate by cad-lag piecewise constant functions f_ϖ for a sequence of partitions, (e.g., $f_\varpi(s) = f_{t_{i-1}}$ for any $t_{i-1} \leq s < t_i$), then the numerical sequence $\|\delta f_\varpi\|_p$ is not necessarily convergent (or bounded) as the mesh $|\varpi| \to 0$. Moreover, if f is continuous and this series converges as $|\varpi| \to 0$ then f has finite p-variation. A typical candidate for a continuous function with finite p-variation is a Hölder continuous function of exponent $\alpha = 1/p$. Also, the quadratic co-variance of two cad-lag functions f and g is the limit (it exits and it is finite) of

$$[f,g]^\varpi = \sum_{i=1}^{n} \big(f(t_i) - f(t_{i-1})\big)\big(g(t_i) - g(t_{i-1})\big), \quad \forall t \in]0,T].$$

as the mesh $|\varpi| \to 0$. As mentioned early, if f is a continuous function with bounded variation on $[0,T]$ then the quadratic variation $\langle f \rangle = 0$, and if f is continuous and g has bounded variation then $[f,g] = 0$. Even if a purely jump cad-lag function g make proper sense only when g has bounded variation. However, if a point function δg is initially given then the approximation g_η as in (4.23) capture the meaning of a purely jump cad-lag function, and if ϖ is a sequence of partitions with $|\varpi| \to 0$ and η is another sequence of number with $\eta \to \infty$ (i.e, jumps of vanishing size) then there exits a subsequence ϖ' and η' such that

$$\lim_{|\varpi'|\to 0,\, \eta' \to \infty} \sum_{i=1}^{n'} \big|f(t'_i) - f(t'_{i-1})\big|\big|g_{\eta'}(t'_i) - g_{\eta'}(t'_{i-1})\big| = 0, \tag{4.24}$$

if the point function δg and the continuous function f have finite quadratic variations. Indeed, fix k large and use Hölder inequality to deduce

$$\begin{aligned}
\sum_{i=1}^{n} \big|f(t_i) - f(t_{i-1})\big|\big|g_\eta(t_i) - g_\eta(t_{i-1})\big| \leq \\
\leq \sum_{i=1}^{n} \big|f(t_i) - f(t_{i-1})\big|\big|g_k(t_i) - g_k(t_{i-1})\big| + \\
+ \Big(\sum_{i=1}^{n} \big|f(t_i) - f(t_{i-1})\big|^2\Big)^{1/2} \Big(\sum_{i=1}^{n} \big|\bar g_{\eta,k}(t_i) - \bar g_{\eta,k}(t_{i-1})\big|^2\Big)^{1/2},
\end{aligned}$$

where $\bar g_{\eta,k} = g_\eta - g_k$, with $\eta > k$, i.e., $\bar g_{\eta,k}$ contains only jump of size δg with $1/\eta \leq |\delta g(s)| < 1/k$ or $\eta \geq |\delta g(s)| > k$. Since the function $\bar g_{\eta,k}$ has only a finite number of jumps, they are separated by some positive number $\vartheta = \vartheta(\eta)$, so that for a subsequence of partitions ϖ' with mesh $|\varpi'| < \vartheta$ the sub-interval $[t'_{i-1}.t'_i]$ can capture only one jump $\delta g(s)$, i.e.,

$$\sum_{i=1}^{n} \big|g_{\eta,k}(t'_i) - g_{\eta,k}(t'_{i-1})\big|^2\Big) \leq \sum_{0<s\leq T} |\delta g(s)|^2 \mathbb{1}_{\{|\delta g(s)|<1/k \text{ or } |\delta g(s)|>k\}}.$$

Hence, given $\varepsilon > 0$ there exists $k = k(\varepsilon, g, f)$ sufficiently large so that

$$\Big(\sum_{i=1}^{n} \big|f(t_i) - f(t_{i-1})\big|^2\Big)^{1/2} \Big(\sum_{i=1}^{n} \big|\bar g_{\eta,k}(t_i) - \bar g_{\eta,k}(t_{i-1})\big|^2\Big)^{1/2} \le \varepsilon,$$

for every partition ϖ with mesh $|\varpi| < \vartheta(\eta)$ and for every $\eta > k$. Note that if $\tau(i,\eta)$ is the i jump of size either $|\delta g(s)| \ge \eta$ or $|\delta g(s)| \le 1/\eta$ then $\vartheta(\eta) = \min_i\{(\tau_i - \tau_{i-1})\}$. This proves that for any subsequence ϖ' and η' satisfying $|\varpi'| < \vartheta(\eta')$ we have

$$\sum_{i=1}^{n'} \big|f(t'_i) - f(t'_{i-1})\big|\big|g_\eta(t'_i) - g_\eta(t'_{i-1})\big| \le$$

$$\le \varepsilon + \sum_{i=1}^{n'} \big|f(t'_i) - f(t'_{i-1})\big|\big|g_k(t'_i) - g_k(t'_{i-1})\big|,$$

Hence, in view of (4.21), the sum on the right-hand-term vanishes as $|\varpi| \to 0$ and (4.24) follows. Remark that if g_η has not a bounded variation as $\eta \to \infty$ then for any partition ϖ

$$\lim_{\eta\to\infty} \sum_{i=1}^{n} \big|f(t_i) - f(t_{i-1})\big|\big|g_\eta(t_i) - g_\eta(t_{i-1})\big| = \infty,$$

which implies that if ϖ is a sequence of partitions with $|\varpi| \to 0$ and η is another sequence of number with $\eta \to \infty$ then there exits a subsequence ϖ' and η' such that $[f, g_{\eta'}]^{\varpi'} \to \infty$.

4.2.4 Quadratic Variations for Discontinuous SIs

Let p be a Poisson measure with Lévy measure π on $\mathbb{R}^m_*$ defined on the same filtered Probability space $(\Omega, \mathbb{F}, P)$ where the Wiener process w was given, for instance, if Lévy measure is finite (i.e., $\pi(\mathbb{R}^m_*) < \infty$) then p is a compound Poisson process (with values in $\mathbb{R}^m$), i.e., for two sequences $\{\zeta_k\}$ and $\{\tau_k\}$ of independent random variables, all ζ_k with distribution $\pi/\pi(\mathbb{R}^m_*)$ and all τ_k with $\pi(\mathbb{R}^m_*)$-exponential distribution, we have the expression

$$p(t) = \sum_{k=1}^{\infty} \zeta_k \mathbb{1}_{t\ge\theta_k} = \sum_{0<s\le t} \delta p(s), \qquad \forall t \ge 0,$$

where $\theta_k = \tau_1 + \cdots + \tau_k$, and $\delta p(s) = p(s) - p(s-)$ is the jump at time s. Its counting random measure is denoted by p and given by

$$p(K,]a, b]) = \sum_{k=1}^{\infty} \mathbb{1}_{\{\zeta_k\in K\}} \mathbb{1}_{a<\theta_k\le b} = \sum_{a<s\le b} \mathbb{1}_{\{\delta p(s)\in K\}},$$

for any compact subset K of $\mathbb{R}^m_*$, and any numbers $b > a \ge 0$.

The martingale measure $\tilde{p}(K,]a,b]) = p(K,]a,b]) - (b-a)\pi(K)$ is use to define the compensated Poisson process (with values in $\mathbb{R}^m$)

$$\tilde{p}(t) = \int_{\mathbb{R}^m_* \times]0,t]} \zeta \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall t \geq 0,$$

via the stochastic integral. In general, if γ is a function in $L^2(\mathbb{R}^m_*, \pi)$ then the real-valued compensated Poisson process $\tilde{p}(\gamma)$ is defined as

$$\tilde{p}(\gamma, t) = \int_{\mathbb{R}^m_* \times]0,t]} \gamma(\zeta) \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s),$$

which corresponds to the Lévy measure π_γ in $\mathbb{R}_*$, which is the image measure of π under γ. In any way, the jump δp at time t can always be defined in term of the Poisson measure p and form the so-called Poisson point process, i.e., the Poisson point process δp is equivalent to the Poisson measure p, and moreover, as seen below, they are equivalent to the compensated Poisson process $\tilde{p}$.

Alternatively, begin with a compensated Poisson process with Lévy measure π, i.e., an optional cad-lag process $\{\tilde{p}(t), t \geq 0\}$ with characteristic function

$$\mathbb{E}\{\mathrm{e}^{\mathrm{i}\, z \cdot \tilde{p}(t)}\} = \exp\Big(t \int_{\mathbb{R}^m_*} \big[\mathrm{e}^{-z\cdot\zeta} - 1 - z \cdot \zeta\big] \mathrm{d}\pi(\mathrm{d}\zeta)\Big), \quad \forall z \in \mathbb{R}^m, t \geq 0,$$

to define its associated counting measure (or Poisson measure) using the fact that the compensated Poisson process $\tilde{p}$ and the compound Poisson process p (when it is defined) have the jumps, i.e., $\delta\tilde{p}(s) = \delta p(s)$ for every $s > 0$, and

$$p(B,]a,b]) = \sum_{a<s\leq b} \mathbb{1}_{\{\delta\tilde{p}(s)\in B\}} \quad \text{or} \quad p(\gamma, B,]a,b]) = \sum_{a<s\leq b} \mathbb{1}_{\{\gamma(\delta\tilde{p}(s))\in B\}},$$

which is a (random) finite sum for each Borel set B separated from the origin (i.e., with a positive distance to the origin) in $\mathbb{R}^m_*$ or $\mathbb{R}_*$, and any numbers $b > a \geq 0$. However, if the function γ belongs to $L^1(\mathbb{R}^m_*)$ then the pathwise integral

$$p(\gamma, t) = \sum_{0<s\leq t} \gamma\big(\delta\tilde{p}(s)\big) = \int_{\mathbb{R}^m_* \times]0,t]} \gamma(\zeta) p(\mathrm{d}\zeta, \mathrm{d}s),$$

is defined as an absolutely convergence (random) series (with the convention that $\gamma(0) = 0$, and only non-zero term count), which is a compound Poisson process with the Lévy finite measure

$$\pi_\gamma(B) = \int_{\mathbb{R}^m_*} \mathbb{1}_{\{\gamma(\zeta)\in B\}} \pi(\mathrm{d}\zeta),$$

for any Borel set B in $\mathbb{R}_*$. Note that the function γ for the counting measure $p(\gamma, \mathrm{d}\zeta, \mathrm{d}s)$, is initially regarded as mapping $\mathbb{R}^m_*$ into $\mathbb{R}^m$ or $\mathbb{R}$, but only $\{\zeta \in$

$\mathbb{R}^m_* : \gamma(\zeta) \neq 0\}$ is the effective domain, i.e, the process $t \mapsto p(\gamma, t)$ has an actual jumps only if $\delta p(t) \neq 0$ and $\gamma(\delta p(t)) \neq 0$.

In general, recall that for any predictable square integrable process h with values in $L^2(\mathbb{R}^m_*, \pi)$ the stochastic integral

$$I(t, h, \mathrm{d}\tilde{p}) = \int_{\mathbb{R}^m_* \times]0,t]} h(\zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall t \geq 0,$$

is defined, and if h is integrable with values in $L^1(\mathbb{R}^m_*, \pi)$ then the pathwise integral, which is an absolutely convergence (random) series,

$$I(t, h, \mathrm{d}p) = \int_{\mathbb{R}^m_* \times]0,t]} h(\zeta, s)p(\mathrm{d}\zeta, \mathrm{d}s) = \sum_{0<s\leq t} h\big(\delta p(s), s\big), \tag{4.25}$$

for every $t \geq 0$, is also defined. It should be clear that the pathwise integral (4.25) with respect to the Poisson measure p (or Poisson point process δp) makes sense only for a predictable (or optional) processes h with values locally in $L^1(\pi, \mathrm{d}s)$ and the resulting integral $t \mapsto I(t, h, \mathrm{d}p)$ defines an optional process having paths with locally integrable bounded variation. Moreover, in this case, the stochastic integral (4.25) with respect to the Poisson martingale measure $\tilde{p}$ (or the compensated Poisson process $\tilde{p}$) is defined only when h is predictable locally square-integral process, and because h also takes values locally in $L^1(\pi, \mathrm{d}s)$, the stochastic integral $t \mapsto I(t, h, \mathrm{d}\tilde{p})$ can be separated into two pathwise integrals,

$$I(t, h, \mathrm{d}\tilde{p}) = I(t, h, \mathrm{d}p) - \int_0^t \mathrm{d}s \int_{\mathbb{R}^m_*} h(\zeta, s)\pi(\mathrm{d}\zeta), \quad \forall t \geq 0,$$

and in this case, the stochastic integral is an optional locally square martingale having paths with locally integrable bounded variation. Also recall the convention $\gamma(0) = 0$ in the series notation $\sum_{0<s\leq t}$ and that only non-zero term count in the series.

Another key point is the fact that any purely discontinuous square-integrable martingale it L^2-orthogonal to any continuous square-integrable martingale, which means that

$$\mathbb{E}\big\{[I(t, h, \mathrm{d}\tilde{p}) - I(s, h, \mathrm{d}\tilde{p})][I(t, f, \mathrm{d}w) - I(s, f, \mathrm{d}w)]\big\} = 0, \quad \forall t > s \geq 0,$$

provided that h and g are predictable square-integrable processes. Moreover, the stochastic integrals are defined for non-anticipative processes, even if they are not predictable with respect to the filtration where the integrator is considered, i.e., the integrand should be taken adapted to the filtration generated by the integrator, and this include integrands that are independent of the integrator.

QV Definition for Poisson Integrals (part 1)

Even if the optional quadratic variation of a purely jump local-square integrable martingale M is easily defined as $[M](t) = \sum_{0<s\leq t}(M(s) - M(s-))^2$, the following calculation help to understand better the situation. As mentioned early,

genuine stochastic integrals correspond to (local-) martingale which are quasi-left continuous, and in general, this is imposed directly into the filtration $\mathbb{F}$. Recall that quasi-left continuous (for a right-continuous martingale or process) means that there is not predictable jumps. In other words, either $\mathcal{F}(\tau-) = \mathcal{F}(\tau)$ for every predictable times (i.e., τ is a limit of a strictly increasing sequence of stopping times) or the predictable compensator ν_X^p of the jumps of the process X satisfies $\nu_X^p(\mathbb{R}_*^m \times \{t\}) = 0$ almost surely, for every deterministic time t. The expression

$$\langle I(\cdot, h, \mathrm{d}\tilde{p})\rangle = \int_0^{\cdot} \mathrm{d}s \int_{\mathbb{R}_*^m} |h(\zeta, s)|^2 \pi(\mathrm{d}\zeta)$$

is the continuous process referred to as the predictable quadratic variation, while $[I(\cdot, h, \mathrm{d}\tilde{p})] = I(\cdot, h^2, \mathrm{d}p)$ is referred to as the optional quadratic variation, for a stochastic integral with respect to a Poisson martingale measure. Certainly, $\langle I(\cdot, h, \mathrm{d}\tilde{p})\rangle$ is the predictable compensator of the local-martingale $I(\cdot, h^2, \mathrm{d}p)$.

Theorem 4.9. *In a given filtered probability space $(\Omega, \mathbb{F}, P)$, let w be a real-valued standard Wiener process and p be a Poisson measure p with Lévy measure π in $\mathbb{R}_*^m$, with compensated Poisson process $\tilde{p}$, counting Poisson measure or Poisson point process δp. Suppose that for a partition ϖ, a predictable piecewise constant process h_ϖ is given as in Theorem 4.3, with $h_\varpi \to h$. Suppose h, σ and γ are predictable processes such that h is locally integrable with values in $\mathbb{R}$, σ and γ are locally square-integrable with values in $\mathbb{R}$ and in $L^2(\mathbb{R}_*^m, \pi)$, and $h\sigma^2$ and $h\gamma^2$ are also locally integrable. If h_ϖ is a predictable piecewise constant process as in Theorem 4.3 associated with a partition ϖ such that*

$$P\Big\{ \int_0^T |h_\varpi(t) - h(t)|\big[1 + |\sigma(t)|^2\big]\mathrm{d}t + \\ + \int_0^T |h_\varpi(t) - h(t)|\mathrm{d}t \int_{\mathbb{R}_*^m} |\gamma(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \geq \varepsilon \Big\} \to 0,$$

as the mesh of the partition $|\varpi| \to 0$, for every $\varepsilon > 0$, then

$$\sup_{0\leq t\leq T} \Big| \sum_{i=1}^{\infty} h_\varpi(t_{i-1}) \Big[\int_{t\wedge t_{i-1}}^{t\wedge t_i} \sigma(s)\mathrm{d}w(s) \Big] \times \\ \times \Big[\int_{\mathbb{R}_*^m \times]t\wedge t_{i-1}, t\wedge t_i]} \gamma(\zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s) \Big] \Big| \to 0, \quad (4.26)$$

in probability, as $|\varpi| \to 0$.

Proof. First, use the stopping time

$$\tau_r = \inf \Big\{ t \geq 0 : \int_0^t \big[|h(s)| + 1\big]\big[1 + |\sigma(s)|^2\big]\mathrm{d}s + \\ + \int_0^t \big[|h(s)| + 1\big]\mathrm{d}s \int_{\mathbb{R}_*^m} |\gamma(\zeta, s)|^2 \pi(\mathrm{d}\zeta) \geq r \Big\},$$

to reduce to the case where the processes h, σ and γ are bounded with values in $\mathbb{R}$ and $L^2(\mathbb{R}^m_*, \pi)$, as in proof of Theorem 4.3.

Second, let us denote by $B_\varpi(t, h_\varpi, \sigma \mathrm{d}w, \gamma \mathrm{d}\tilde{p})$ the bilinear expression in (4.26), and note that if γ belongs to $L^1(\mathbb{R}^m_*, \pi) \cap L^2(\mathbb{R}^m_*, \pi)$ then the stochastic integral in $\mathrm{d}\tilde{p}$ becomes a pathwise integral and yields an optional process having path with bounded variation on any time-interval $[0, T]$, thus with $t_n = T$,

$$\sup_{0\le t\le T} |B_\varpi(t, h_\varpi, \sigma \mathrm{d}w, \gamma \mathrm{d}\tilde{p})| \le \sup_{i\le n} \sup_{t_{i-1}<t\le t_i} \Big| \int_{t_{i-1}}^t \sigma(s) \mathrm{d}w(s) \Big| \times \\ \times \sum_{i=1}^n |h_\varpi(t_{i-1})| \Big| \int_{\mathbb{R}^m_* \times]t_{i-1}, t_i]} \gamma(\zeta, s) \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s) \Big|,$$

where the $\sum_i$ is bounded almost surely and the continuity of the stochastic integral in $\mathrm{d}w$ for almost every paths ensures that the $\sup_i \sup_{t_{i-1}<t\le T\wedge t_i}$ vanishes as the mesh of the partition $|\varpi| \to 0$. This prove the almost surely convergence of (4.26), which implies the convergence in probability, for the particular case when γ belongs to $L^1(\mathbb{R}^m_*, \pi) \cap L^2(\mathbb{R}^m_*, \pi)$.

Next, note that working with the positive and negative parts of h there is not loss of generality when assuming that $h \ge 0$. Hence, if $h \ge 0$ and $\gamma_k = \gamma \mathbb{1}_{\{k\gamma \ge 1\}}$ with $k \to \infty$, then the process $t \mapsto \gamma_k(s, \cdot)$ belongs to has values in $L^1(\mathbb{R}^m_*, \pi) \cap L^2(\mathbb{R}^m_*, \pi)$,

$$B_\varpi(t, h_\varpi, \sigma \mathrm{d}w, \gamma \mathrm{d}\tilde{p}) = B_\varpi(t, h_\varpi, \sigma \mathrm{d}w, \gamma_k \mathrm{d}\tilde{p}) + \\ + B_\varpi\big(t, h_\varpi, \sigma \mathrm{d}w, (\gamma - \gamma_k) \mathrm{d}\tilde{p}\big).$$

and each term in the expression of $B_\varpi(t, h, \sigma \mathrm{d}w, (\gamma - \gamma_k) \mathrm{d}\tilde{p})$ can be bounded by the product

$$(A) = \sup_{t_{i-1}<t\le t_i} \Big| \int_{t_{i-1}}^t \sqrt{h_\varpi(s)}\, \sigma(s) \mathrm{d}w(s) \Big| \times \\ \times \sup_{t_{i-1}<t\le t_i} \Big| \int_{\mathbb{R}^m_* \times]t_{i-1}, t]} \sqrt{h_\varpi(s)} [\gamma(\zeta, s) - \gamma_k(\zeta, s)] \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s) \Big|.$$

Therefore, apply Hölder inequality and the $L^2 - \sup$ inequality for stochastic integral to bound the expectation of above product by

$$\mathbb{E}\{(A)\} = \Big(\mathbb{E}\Big\{ \int_{t_{i-1}}^{t_i} |h_\varpi(s)|\, |\sigma(s)|^2 \mathrm{d}s \Big)^{1/2} \times \\ \times \Big(\mathbb{E}\Big\{ \int_{t_{i-1}}^{t_i} |h_\varpi(s)| \mathrm{d}s \int_{\{k|\gamma|<1\}} |\gamma(\zeta, s)|^2 \pi(\mathrm{d}\zeta) \Big\} \Big)^{1/2},$$

and use the elementary inequality $2ab \le \varepsilon a^2 + b^2/\varepsilon$ to check that

$$\mathbb{E}\{(A)\} = \varepsilon\mathbb{E}\Big\{ \int_{t_{i-1}}^{t_i} |h_\varpi(s)|\,|\sigma(s)|^2 \mathrm{d}s + \\ + \frac{1}{\varepsilon}\mathbb{E}\Big\{ \int_{t_{i-1}}^{t_i} |h_\varpi(s)|\mathrm{d}s \int_{\{k|\gamma|<1\}} |\gamma(\zeta,s)|^2 \pi(\mathrm{d}\zeta)\Big\},$$

for every $\varepsilon > 0$.

Take this back to the expression of $B(t, h, \sigma \mathrm{d}w, (\gamma - \gamma_k)\mathrm{d}\tilde{p})$ to deduce

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} \big|B_\varpi\big(t, h_\varpi, \sigma \mathrm{d}w, (\gamma-\gamma_k)\mathrm{d}\tilde{p}\big)\big|\Big\} \le \varepsilon\mathbb{E}\Big\{ \int_0^T |h_\varpi(s)|\,|\sigma(s)|^2 \mathrm{d}s + \\ + \frac{1}{\varepsilon}\mathbb{E}\Big\{ \int_0^T |h_\varpi(s)|\mathrm{d}s \int_{\{k|\gamma|<1\}} |\gamma(\zeta,s)|^2 \pi(\mathrm{d}\zeta)\Big\}.$$

Because $h_\varpi\sigma^2$ and $h_\varpi\gamma^2$ are uniformly (in ϖ) integrable, we can take limit first as $k \to \infty$ and second as $\varepsilon \to 0$ to establish that

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} \big|B_\varpi\big(t, h_\varpi, \sigma \mathrm{d}w, (\gamma-\gamma_k)\mathrm{d}\tilde{p}\big)\big|\Big\} \to 0,$$

as $k \to \infty$, uniformly with respect to the partition ϖ. Hence, invoke the second step, i.e., the case with bounded variation, to deduce

$$\sup_{0\le t\le T} \big|B_\varpi(t, h_\varpi, \sigma \mathrm{d}w, \gamma \mathrm{d}\tilde{p})\big| \to 0, \ \text{ as } \ |\varpi| \to 0,$$

in probability.

Finally, make use of localization as in Theorem 4.3 to limit the analysis only up to the stopping time τ_r and then to conclude the argument, after remarking that $P\{\tau_r < T\} \to 0$ as $r \to \infty$. □

• *Remark* 4.10. Note that with the notation of Theorem 4.9, the process $t \mapsto B_\varpi(t, h_\varpi, \sigma \mathrm{d}w, \gamma \mathrm{d}\tilde{p})$ is a martingale, this expression is linear in h_ϖ and

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} \big|B_\varpi(t, h_\varpi, \sigma \mathrm{d}w, \gamma \mathrm{d}\tilde{p})\big|\Big\} \le \mathbb{E}\Big\{ \int_0^T |h_\varpi(s)|\,|\sigma(s)|^2 \mathrm{d}s\Big\} + \\ + \mathbb{E}\Big\{ \int_0^T |h_\varpi(s)|\mathrm{d}s \int_{\mathbb{R}^m_*} |\gamma(\zeta,s)|^2 \pi(\mathrm{d}\zeta)\Big\}.$$

Among other things, this allows to truncate h_ϖ and so to suppose that $|h_\varpi| \le C_\varpi$, for deterministic constant. Another possibility is to consider h_ϖ part of σ and γ, i.e., to reduce to the case $h_\varpi = 1$ and to use σ_ϖ and γ_ϖ such that $h_\varpi\sigma\gamma = \sigma_\varpi\gamma_\varpi$, and in view of the continuity, to truncate σ_ϖ and γ_ϖ so that

$|\sigma_\varpi|+|\gamma_\varpi| \le C_\varpi$, for some deterministic constant, without any loss of generality. Similarly, if h has left-continuous paths and $h_\varpi(t_i) = h(t_i)$ then the term

$$\sup_\varpi \sum_{i=1}^n |h(t_{i-1})| \Big| \int_{\mathbb{R}^m_* \times]t_{i-1},t_i]} \gamma(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)\Big| \le$$

can be bounded by

$$\le \sup_\varpi \sum_{i=1}^n |h(t_{i-1})| \Big(\int_{\mathbb{R}^m_* \times]t_{i-1},t_i]} |\gamma(\zeta,s)| p(\mathrm{d}\zeta,\mathrm{d}s) + \\ + \int_{t_{i-1}}^{t_i} \mathrm{d}s \int_{\mathbb{R}^m_*} |\gamma(\zeta,s)|\pi(\mathrm{d}\zeta)\Big),$$

i.e., bounded by

$$\le \Big(\int_{\mathbb{R}^m_* \times]0,T]} |h(s)||\gamma(\zeta,s)|\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s) + \\ + 2\int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |h(s)|\,|\gamma(\zeta,s)|\pi(\mathrm{d}\zeta)\Big),$$

which yields

$$\mathbb{E}\Big\{ \sup_\varpi \sum_{i=1}^n |h(t_{i-1})| \Big| \int_{\mathbb{R}^m_* \times]t_{i-1},t_i]} \gamma(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)\Big|\Big\} \le \\ \le 2\mathbb{E}\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |h(s)|\,|\gamma(\zeta,s)|\pi(\mathrm{d}\zeta)\Big\},$$

and

$$\mathbb{E}\Big\{ \Big[\sup_\varpi \sum_{i=1}^n |h(t_{i-1})| \Big| \int_{\mathbb{R}^m_* \times]t_{i-1},t_i]} \gamma(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)\Big|\Big]^2\Big\} \le \\ \le 2\mathbb{E}\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |h(s)|^2|\gamma(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big\} + \\ + 4\mathbb{E}\Big\{ \Big(\int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |h(s)|\,|\gamma(\zeta,s)|\pi(\mathrm{d}\zeta)\Big)^2\Big\}.$$

Hence, estimating the L^2-modulus of continuity for the stochastic integral with respect to the Wiener process, we could deduce that the convergence (4.26) take place in L^1, under the above assumptions. □

Corollary 4.11. *As in the Theorem 4.9 if two independent Poisson measure p_1 and p_2 with Lévy measures π_1 and π_2 on $\mathbb{R}^m_*$ are given, as well as predictable*

processes h, γ_1, γ_2 and h_ϖ satisfying

$$P\Big\{\int_0^T |h_\varpi(t)-h(t)|\Big[1+ \\ +\int_{\mathbb{R}_*^m}|\gamma_1(\zeta,t)|^2\pi_1(\mathrm{d}\zeta)+\int_{\mathbb{R}_*^m}|\gamma_2(\zeta,t)|^2\pi_2(\mathrm{d}\zeta)\Big]\geq\varepsilon\Big\}\to 0,$$

as the mesh of the partition $|\varpi|\to 0$, for every $\varepsilon>0$, then

$$\sup_{0\leq t\leq T}\Big|\sum_{i=1}^{\infty}h_\varpi(t_{i-1})\Big[\int_{\mathbb{R}_*^m\times]t\wedge t_{i-1},t\wedge t_i]}\gamma_1(\zeta,s)\tilde{p}_1(\mathrm{d}\zeta,\mathrm{d}s)\Big]\times \\ \times\Big[\int_{\mathbb{R}_*^m\times]t\wedge t_{i-1},t\wedge t_i]}\gamma_2(\zeta,s)\tilde{p}_2(\mathrm{d}\zeta,\mathrm{d}s)\Big]\Big|\to 0,$$

in probability, as $|\varpi|\to 0$.

Proof. Indeed, the argument is the same as in the proof of Corollary 4.6 with two independent Wiener processes, first by localization, we may assume that

$$\int_0^T\big[1+|h(s)|+|h_\varpi(s)|\big]\Big[1+\int_{\mathbb{R}_*^m}|\gamma_1(\zeta,s)|^2\pi_1(\mathrm{d}\zeta)+ \\ +\int_{\mathbb{R}_*^m}|\gamma_2(\zeta,s)|^2\pi_2(\mathrm{d}\zeta)\Big]\leq C_T,$$

for a deterministic constant C_T, as $|\varpi|\to 0$. Next, by continuity, it suffices to consider processes $\gamma_{1,\varpi}$ and $\gamma_{2,\varpi}$ that are piecewise constant relative to the partition ϖ, and in this case, each of the stochastic integrals can be considered 'inside' the other to have

$$B_\varpi(t,h_\varpi,\gamma_{1,\varpi}\mathrm{d}\tilde{p}_1,\gamma_{2,\varpi}\mathrm{d}\tilde{p}_2)=\int_{\mathbb{R}_*^m\times]0,t]}\gamma_\varpi(\zeta,s)\tilde{p}_2(\zeta,\mathrm{d}s),$$
$$\gamma_\varpi(\zeta,s)=h_\varpi(s)\gamma_{2,\varpi}(\zeta,s)\int_{\mathbb{R}_*^m\times]t_{i-1},t_i]}\gamma_{1,\varpi}(\zeta,s)\tilde{p}_1(\zeta,\mathrm{d}s),\ t_{i-1}<s\leq t_i.$$

Hence,

$$\mathbb{E}\Big\{\Big|\sup_{0\leq t\leq T}B_\varpi(t,h_\varpi,\gamma_{1,\varpi}\mathrm{d}\tilde{p}_1,\gamma_{2,\varpi}\mathrm{d}\tilde{p}_2)\Big|^2\Big\}\leq \\ \leq 4\mathbb{E}\Big\{\int_0^T\mathrm{d}s\int_{\mathbb{R}_*^m}|\gamma_\varpi(\zeta,s)|^2\pi_2(\zeta)\Big\},\quad(4.27)$$

and

$$\mathbb{E}\big\{|\gamma_\varpi(\zeta,s)|\big\}=\mathbb{E}\Big\{|h_\varpi(s)|^2|\gamma_{2,\varpi}(\zeta,s)|^2(t_i-t_{i-1})\int_{\mathbb{R}_*^m}|\gamma_{1,\varpi}(\zeta,s)|^2\pi_1(\zeta)\Big\},$$

for any $t_{i-1} < s \le t_i$. The term in the right of (4.27) can be written as

$$4\sum_{i=1}^{n}\mathbb{E}\Big\{|h_\varpi(t_{i-1})|^2|(t_i-t_{i-1})^2\int_{\mathbb{R}_*^m}|\gamma_{1,\varpi}(\zeta,t_{i-1})|^2\pi_1(\mathrm{d}\zeta)\times \\ \times\int_{\mathbb{R}_*^m}\gamma_{2,\varpi}(\zeta,t_{i-1})|^2\pi_2(\mathrm{d}\zeta)\Big\},$$

and its integrand is bounded by

$$4\Big(\sup_i\int_{t_{i-1}}^{t_i}|h_\varpi(s)|^2|\mathrm{d}s\int_{\mathbb{R}_*^m}|\gamma_{1,\varpi}(\zeta,s)|^2\pi_1(\mathrm{d}\zeta)\Big)\times \\ \times\Big(\int_0^T|h_\varpi(s)|^2|\mathrm{d}s\int_{\mathbb{R}_*^m}\gamma_{2,\varpi}(\zeta,s)|^2\pi_2(\mathrm{d}\zeta)\Big)\le C_T,$$

which vanishes as $|\varpi|\to 0$. □

QV Definition for Poisson Integrals (part 2)

Now, if γ is a function in $L^2(\mathbb{R}_*^m,\pi)$ and h_ϖ is a cag-lad predictable real-valued process associated with a partition ϖ then define the 'Riemann-Stieltjes' sums

$$K_\varpi(t,h_\varpi,\mathrm{d}^2\tilde p(\gamma))=\sum_{i=1}^{\infty}h_\varpi(t_{i-1})\big[\tilde p(\gamma,t\wedge t_i)-\tilde p(\gamma,t\wedge t_{i-1})\big]^2, \\ I_\varpi(t,h_\varpi,\mathrm{d}p(\gamma^2))=\sum_{i=1}^{\infty}h_\varpi(t_{i-1})\big[p(\gamma^2,t\wedge t_i)-p(\gamma^2,t\wedge t_{i-1})\big], \tag{4.28}$$

as a real-valued (random) finite sum. Note that since h_ϖ is piecewise constant,

$$\int_0^t h_\varpi(s)p(\gamma,\mathrm{d}s)=I_\varpi(t,h_\varpi,\mathrm{d}p(\gamma^2))=I(t,h_\varpi\gamma^2,\mathrm{d}p),$$

as defined early. Also remark that the property

$$\mathbb{E}\big\{\big[\tilde p(\gamma,t\wedge t_i)-\tilde p(\gamma,t\wedge t_{i-1})\big]^2\,\big|\,\mathcal{F}(t_{i-1})\big\}=[t_i-t_{i-1}]\int_{\mathbb{R}_*^m}[\gamma(\zeta)]^2\pi(\mathrm{d}\zeta),$$

yields the isometric equality

$$\mathbb{E}\big\{K_\varpi(t,h_\varpi,\mathrm{d}^2\tilde p(\gamma))\big\}=\mathbb{E}\big\{I_\varpi(t,h_\varpi,\mathrm{d}p(\gamma^2))\big\}= \\ =\mathbb{E}\Big\{\int_0^t h_\varpi(t_{i-1})\mathrm{d}t\Big\}\Big(\int_{\mathbb{R}_*^m}[\gamma(\zeta)]^2\pi(\mathrm{d}\zeta)\Big),$$

valid for every $t\ge 0$.

Theorem 4.12. *Under the same assumptions of Theorem 4.3 and the above notation with a function γ in $L^2(\mathbb{R}^m_*, \pi)$ we have*

$$\sup_{0\le t\le T} \left| K_\varpi(t, h_\varpi, \mathrm{d}^2\tilde{p}(\gamma)) - I(t, h\gamma^2, \mathrm{d}p) \right| \to 0, \tag{4.29}$$

in probability. Moreover, if f and g are two predictable processes as in Corollary 4.6 then for any nonnegative numbers a, b and c, with $a+b+c>1$, $a+b>0$, $t_n = T$,

$$\sum_{i=1}^n |h_\varpi(t_{i-1})| \Big| \int_{t_{i-1}}^{t_i} f(s)\mathrm{d}w(s) \Big|^{2a} \Big(\int_{t_{i-1}}^{t_i} |g(s)|^2 \mathrm{d}s \Big)^b \times \\ \times |\tilde{p}(\gamma, t_i) - \tilde{p}(\gamma, t_{i-1})|^{2c} \to 0, \tag{4.30}$$

in probability.

Proof. First, note that

$$I(t, h\gamma^2, \mathrm{d}p) = \sum_{0<s\le t} h(s)\gamma\big(\delta p(s)\big)$$

as defined by (4.25), is a pathwise integral with respect to the counting measure p, and thus

$$\sup_{0\le t\le T} \left| I_\varpi(t, h_\varpi, \mathrm{d}p(\gamma^2)) - I(t, h\gamma^2, \mathrm{d}p) \right| \le I(T, |h_\varpi - h|\gamma^2, \mathrm{d}p) \to 0,$$

as the mesh $|\varpi| \to 0$. Moreover, by localization, as in the proof of Theorem 4.3, we can assume that the processes are also bounded, i.e.,

$$\int_0^T \big[|h(s)| + |h_\varpi(s)| \big] \mathrm{d}s \le C < \infty, \quad \text{a.s.},$$

for some constant $C > 0$.

Now to establish the convergence (4.30), we proceed as in the proof of Corollary 4.6, but for the Poisson integral, the estimate

$$\mathbb{E}\big\{ |\tilde{p}(\gamma, t_i) - \tilde{p}(\gamma, t_{i-1})|^{2c} \big\} \le M_c \Big[\Big((t_i - t_{i-1}) \int_{\mathbb{R}^m_*} [\gamma(\zeta)]^2 \pi(\mathrm{d}\zeta) \Big)^c + \\ + \mathbb{1}_{c>1} \Big((t_i - t_{i-1}) \int_{\mathbb{R}^m_*} [\gamma(\zeta)]^{2c} \pi(\mathrm{d}\zeta) \Big) \Big],$$

for a constant $M_c > 0$, imposes $a + b > 0$, i.e., the contribution of the Poisson integral has order $c \wedge 1$ by itself.

To check (4.29), first assume that the function γ in $L^2(\mathbb{R}^m_*, \pi)$ satisfies $\pi(\{|\gamma| \ne 0\}) < \infty$, which implies that γ also belongs $L^1(\mathbb{R}^m_*, \pi)$. Hence the Poisson measure p becomes a compound Poisson process (with finite second

moment) when regarded on $\{\zeta \in \mathbb{R}^m_* : \gamma(\zeta) \neq 0\}$ instead of the whole space $\mathbb{R}^m_*$, and

$$\big[\tilde{p}(\gamma,t) - \tilde{p}(\gamma,s)\big]^2 = \Big(\sum_{s<r\le t} \gamma\big(\delta p(r)\big) - (t-s) \int_{\mathbb{R}^m} \gamma(\zeta)\pi(\mathrm{d}\zeta)\Big)^2,$$

and developing the square, this is equal to

$$\Big(\sum_{s<r\le t} \gamma\big(\delta p(r)\big)\Big)^2 - 2(t-s) \sum_{s<r\le t} \gamma\big(\delta p(r)\big) \int_{\mathbb{R}^m} \gamma(\zeta)\pi(\mathrm{d}\zeta) + \\ + \Big((t-s) \int_{\mathbb{R}^m} \gamma(\zeta)\pi(\mathrm{d}\zeta)\Big)^2,$$

and the argument in (4.30) shows that the expectation of last two term vanish as the mesh $|\varpi| \to 0$, i.e., for the case of a compound Poisson process $p(\gamma,t)$

$$\mathbb{E}\big\{\big|K_\varpi(T,h_\varpi,\mathrm{d}^2\tilde{p}(\gamma)) - K_\varpi(T,h_\varpi,\mathrm{d}^2p(\gamma))\big|\big\} \to 0, \quad \text{as} \quad |\varpi| \to 0,$$

where K_ϖ without $\mathrm{d}^2p(\gamma)$ instead of $\mathrm{d}^2\tilde{p}(\gamma)$ is given by

$$K_\varpi(t,h_\varpi,\mathrm{d}^2p(\gamma)) = \sum_{i=1}^{\infty} h_\varpi(t_{i-1})\big[p(\gamma,t\wedge t_i) - p(\gamma,t\wedge t_{i-1})\big]^2,$$

and the expression

$$\big[p(\gamma,t\wedge t_i) - p(\gamma,t\wedge t_{i-1})\big]^2 = \Big(\sum_{t\wedge t_{i-1}<r\le t\wedge t_i} \gamma\big(\delta p(r)\big)\Big)^2,$$

is only meaningful when $\pi(\{\gamma \neq 0\}) < \infty$. Moreover, because the process $K_\varpi(T,h_\varpi,\mathrm{d}^2\tilde{p}(\gamma)) - K_\varpi(T,h_\varpi,\mathrm{d}^2p(\gamma))$ is a martingale, Doob's maximal inequality yields

$$P\big\{ \sup_{0\le t\le T} |K_\varpi(T,h_\varpi,\mathrm{d}^2\tilde{p}(\gamma)) - K_\varpi(T,h_\varpi,\mathrm{d}^2p(\gamma))| \ge \varepsilon\big\} \to 0,$$

as $|\varpi| \to 0$, for every $\varepsilon > 0$.

Now, since the jumps (i.e., when $\gamma(\delta p(r)) \neq 0$) can be ordered as $\theta_1 < \theta_2 < \cdots < \theta_n < \cdots$ with θ_i random variables, for every fixed ω and as $|\varpi| \to 0$, the interval $]t\wedge t_{i-1}, t\wedge t_i]$ contains at most one jumps, which means that $K_\varpi(t,h_\varpi,\mathrm{d}^2p(\gamma)) = I_\varpi(t,h_\varpi\gamma^2,\mathrm{d}p)$, for $|\varpi| \le \inf_{i\le n}\{\theta_i - \theta_{i-1}\}$, $\theta_n < t \le \theta_{n+1}$. This proves

$$P\big\{ \sup_{0\le t\le T} |K_\varpi(t,h_\varpi,\mathrm{d}^2p(\gamma)) - I_\varpi(t,h_\varpi,\mathrm{d}p(\gamma^2))| \ge \varepsilon\big\} \to 0 \tag{4.31}$$

as $|\varpi| \to 0$. Recall that $I_\varpi(t,h_\varpi,\mathrm{d}p(\gamma^2)) = I(t,h_\varpi\gamma^2,\mathrm{d}p)$. Therefore, the convergence (4.29) has been established for a compound Poisson process with finite second moment, i.e., under the extra assumption that $\pi(\{|\gamma| \neq 0\}) < \infty$.

Next, consider $\gamma_n = \gamma \mathbb{1}_{\{1 \le n|\gamma|\}}$, with $n \to \infty$, which satisfies

$$\pi(\{|\gamma_n| \neq 0\}) \le \frac{1}{n^2} \int_{\mathbb{R}_*^m} \gamma^2(\zeta)\pi(\mathrm{d}\zeta) < \infty.$$

The equality $a^2 - b^2 = (a-b)(a+b)$ implies

$$\begin{aligned}[\tilde{p}(\gamma,t) - \tilde{p}(\gamma,s)]^2 - [\tilde{p}(\gamma_n,t) - \tilde{p}(\gamma_n,s)]^2 = \\ = [\tilde{p}(\gamma-\gamma_n,t) - \tilde{p}(\gamma-\gamma_n,s)][\tilde{p}(\gamma+\gamma_n,t) - \tilde{p}(\gamma+\gamma_n,s)],\end{aligned}$$

which yields

$$\begin{aligned}&\mathbb{E}\big\{\big|[\tilde{p}(\gamma,t) - \tilde{p}(\gamma,s)]^2 - [\tilde{p}(\gamma_n,t) - \tilde{p}(\gamma_n,s)]^2\big|\big\} \le \\ &\quad \le \Big(\mathbb{E}\big\{\big|\tilde{p}(\gamma-\gamma_n,t) - \tilde{p}(\gamma-\gamma_n,s)\big|^2\big\}\Big)^{\frac{1}{2}} \Big(\mathbb{E}\big\{\big|\tilde{p}(\gamma+\gamma_n,t) - \tilde{p}(\gamma+\gamma_n,s)\big|^2\big\}\Big)^{\frac{1}{2}}\end{aligned}$$

and

$$\mathbb{E}\big\{\big|\tilde{p}(\gamma-\gamma_n,t) - \tilde{p}(\gamma-\gamma_n,s)\big|^2\big\} = (t-s)\int_{\mathbb{R}_*^m} \gamma^2(\zeta)\mathbb{1}_{\{n|\gamma(\zeta)<1\}}\pi(\mathrm{d}\zeta),$$

$$\mathbb{E}\big\{\big|\tilde{p}(\gamma+\gamma_n,t) - \tilde{p}(\gamma+\gamma_n,s)\big|^2\big\} \le 4(t-s)\int_{\mathbb{R}_*^m} \gamma^2(\zeta)\pi(\mathrm{d}\zeta).$$

Hence

$$\begin{aligned}&\mathbb{E}\big\{\big|K_\varpi(t,h_\varpi,\mathrm{d}^2\tilde{p}(\gamma)) - K_\varpi(t,h_\varpi,\mathrm{d}^2\tilde{p}(\gamma_n))\big|\big\} \le \\ &\quad \le 2\Big(\int_0^t \mathbb{E}\{|h_\varpi(s)|\}\mathrm{d}s\Big)\Big(\int_{\mathbb{R}_*^m} \gamma^2(\zeta)\pi(\mathrm{d}\zeta)\Big)^{\frac{1}{2}}\Big(\int_{\{n|\gamma(\zeta)<1\}} \gamma^2(\zeta)\pi(\mathrm{d}\zeta)\Big)^{\frac{1}{2}}.\end{aligned}$$

Also

$$\begin{aligned}\mathbb{E}\big\{\big|I(t,h_\varpi\gamma^2,\mathrm{d}p) - I(t,h_\varpi\gamma_n^2,\mathrm{d}p)\big|\big\} \le \mathbb{E}\big\{I(t,|h_\varpi\gamma^2 - h_\varpi\gamma_n^2|,\mathrm{d}p)\big|\big\} = \\ = \Big(\int_0^t \mathbb{E}\{|h_\varpi(s)|\}\mathrm{d}s\Big)\Big(\int_{\{n|\gamma(\zeta)<1\}} \gamma^2(\zeta)\pi(\mathrm{d}\zeta)\Big),\end{aligned}$$

after using linearity and the isometric L^2-equality.

Since the processes

$$K_\varpi(t,h_\varpi,\mathrm{d}^2\tilde{p}(\gamma)) - \int_0^t h_\varpi(s)\mathrm{d}s\int_{\mathbb{R}_*^m}[\gamma(\zeta)]^2\pi(\mathrm{d}\zeta)),$$

$$I(t,h_\varpi\gamma^2,\mathrm{d}p) - \int_0^t h_\varpi(s)\mathrm{d}s\int_{\mathbb{R}_*^m}[\gamma(\zeta)]^2\pi(\mathrm{d}\zeta),$$

are martingales, and similarly for γ_n instead of γ, the difference is also a martingale and Doob's maximal inequality implies

$$\begin{aligned}&\varepsilon P\big\{\sup_{0\le t\le T}\big|K_\varpi(t,h_\varpi,\mathrm{d}^2\tilde{p}(\gamma)) - K_\varpi(t,h_\varpi,\mathrm{d}^2\tilde{p}(\gamma_n))\big|\big\} \le \\ &\quad \le 2\Big(\int_0^T |h_\varpi(s)|\mathrm{d}s\Big)\Big(\int_{\mathbb{R}_*^m} \gamma^2(\zeta)\pi(\mathrm{d}\zeta)\Big)^{\frac{1}{2}}\Big(\int_{\{n|\gamma(\zeta)<1\}} \gamma^2(\zeta)\pi(\mathrm{d}\zeta)\Big)^{\frac{1}{2}}\end{aligned}$$

and

$$\varepsilon P\Big\{\sup_{0\le t\le T}\big|I(t,h_\varpi\gamma^2,\mathrm{d}p)-I(t,h_\varpi\gamma_n^2,\mathrm{d}p)\big|\Big\}\le \\ \le\Big(\int_0^T|h_\varpi(s)|\mathrm{d}s\Big)\Big(\int_{\{n|\gamma(\zeta)<1\}}\gamma^2(\zeta)\pi(\mathrm{d}\zeta)\Big).$$

for any $\varepsilon>0$.

Finally, combine convergence (4.31), (4.29) valid for any fixed γ_n with the last two estimates to show that indeed, (4.29) holds true for any function γ in $L^2(\mathbb{R}^m_*,\pi)$. □

• *Remark* 4.13. The argument is Theorem 4.12 can be adapted to include the case of a general compound Poisson process p_γ, with second moment non necessarily finite provided in the definition (4.28) of K_ϖ, the stochastic integral (i.e., its compensated Lévy process may not defined) $\tilde p(\gamma,\cdot)$ is replaced by $p_\gamma(\cdot)$ and only the argument to deduce (4.31) is used. Therefore, if ℓ is a Lévy process with only jumps (i.e., no continuous part, for simplicity) then

$$\sum_{i=1}^\infty h_\varpi(t_{i-1})\big[\ell(t\wedge t_i)-\ell(t\wedge t_{i-1})\big]^2\to\sum_{0<s\le t}h(s)\big[\delta\ell(s)\big]^2,$$

in probability, uniformly for t within any bounded time interval $[0,T]$. If the Lévy measure π satisfies

$$\int_{\mathbb{R}^m_*}|\zeta|^2\pi(\mathrm{d}\zeta)<\infty,$$

then Theorem 4.12 can applied with the function $\gamma(\zeta)=\zeta_i\zeta_j$, i.e., this is the case of a Lévy process with finite second moment. □

QV Definition for Poisson Integrals (part 3)

A function in $L^2(\mathbb{R}^m_*\times[0,T],\pi\times\mathrm{d}s)$ or a predictable locally square-integrable process γ with values in $L^2(\mathbb{R}^m_*,\pi)$ could be used instead of a function γ in $L^2(\mathbb{R}^m_*,\pi)$ to define the optional (purely jump) process $t\mapsto I(t,\gamma^2,\mathrm{d}p)$ given by (4.25), which could be denoted by $t\mapsto p(\gamma^2,t)$, but we prefer to keep the notation $p(\gamma^2,t)$ for a time-independent γ, so that $p(\gamma^2,t)$ is itself a Poisson measure when γ is deterministic. Recall that contrary to the stochastic integral $I(t,\gamma,\mathrm{d}\tilde p)$, the expression is the fact that $I(t,\gamma^2,\mathrm{d}p)$ is always (or is only defined as) a pathwise integral. The expression K_ϖ and I_ϖ of (4.28) should be reconsidered for two predictable processes γ and γ_1 as a bilinear form, i.e.,

$$K_\varpi(t,h_\varpi,\gamma\mathrm{d}\tilde p,\gamma_1\mathrm{d}\tilde p)=\sum_{i=1}^\infty h_\varpi(t_{i-1})\Big[\int_{\mathbb{R}^m_*\times]t\wedge t_{i-1},t\wedge t_i]}\gamma(\zeta,s)\tilde p(\mathrm{d}\zeta,\mathrm{d}s)\Big]\times \\ \times\Big[\int_{\mathbb{R}^m_*\times]t\wedge t_{i-1},t\wedge t_i]}\gamma_1(\zeta,s)\tilde p(\mathrm{d}\zeta,\mathrm{d}s)\Big]$$

and

$$I_\varpi(t,h_\varpi,\gamma\gamma_1\mathrm{d}p)=\sum_{i=1}^{\infty}h_\varpi(t_{i-1})\big[I(t\wedge t_i,\gamma\gamma_1,\mathrm{d}p)-I(t\wedge t_{i-1},\gamma\gamma_1,\mathrm{d}p)\big],$$

$$I(t,h,\mathrm{d}p)=\int_{\mathbb{R}^m_*\times]0,t]}h(\zeta,s)p(\mathrm{d}\zeta,\mathrm{d}s)=\sum_{0<s\le t}h(\delta p(s),s)$$

In particular, if γ and γ_1 are piecewise constant on each sub-interval $]t_{i-1},t_i]$ then

$$I(t_i,\gamma\gamma_1,\mathrm{d}p)-I(t_{i-1},\gamma\gamma_1,\mathrm{d}p)=\\ =p(\gamma(t_{i-1})\gamma_1(t_{i-1}),t_i)-p(\gamma(t_{i-1})\gamma_1(t_{i-1}),t_{i-1}),$$

with the previous notation, i.e., all the jumps within the sub-interval $]t_{i-1},t_i]$ are counted with the same weight $\gamma(\delta p(s),t_{i-1})\gamma_1(\delta p(s),t_{i-1})$.

Corollary 4.14. *Let h be a predictable locally integrable real-valued process, and let γ and γ_1 be predictable locally square integrable processes with values in $L^2(\mathbb{R}^m_*,\pi)$. With the above notation and as in Theorem 4.3, suppose h_ϖ as piece-constant predictable process corresponding to a sequence of partition ϖ with mesh $|\varpi|\to 0$. If*

$$P\Big\{\int_0^T|h_\varpi(s)-h(s)|\Big[1+\int_{\mathbb{R}^m_*}\big(|\gamma(\zeta,s)|^2+|\gamma_1(\zeta,s)|^2\big)\pi(\zeta)\mathrm{d}s\Big]\ge\varepsilon\Big\}\to 0$$

for every $\varepsilon>0$, then

$$\sup_{0\le t\le T}\big|K_\varpi(t,h_\varpi,\gamma\mathrm{d}\tilde p,\gamma_1\mathrm{d}\tilde p)-I(t,h\gamma\gamma_1,\mathrm{d}p)\big|\to 0, \tag{4.32}$$

in probability. Moreover, if f and g are two predictable processes as in Corollary 4.6 then for any nonnegative numbers a, b and c, with $a+b+c>1$, $a+b>0$, $t_n=T$,

$$\sum_{i=1}^n|h_\varpi(t_{i-1})|\Big|\int_{t_{i-1}}^{t_i}f(s)\mathrm{d}w(s)\Big|^{2a}\Big(\int_{t_{i-1}}^{t_i}|g(s)|^2\mathrm{d}s\Big)^b\times\\ \times\Big|\int_{\mathbb{R}^m_*\times]t\wedge t_{i-1},t\wedge t_i]}\gamma(\zeta,s)\tilde p(\mathrm{d}\zeta,\mathrm{d}s)\Big|^{2c}\to 0, \tag{4.33}$$

in probability.

Proof. Proceed as in Theorem 4.12, first by localization, we may assume that

$$\int_0^T\big[|h_\varpi(s)|+|h(s)|\big]\Big[1+\int_{\mathbb{R}^m_*}\big(|\gamma(\zeta,s)|^2+|\gamma_1(\zeta,s)|^2\big)\pi(\zeta)\mathrm{d}s\Big]\le C$$

for a deterministic constant $C=C_T$, as $|\varpi|\to 0$.

Second, the parallelogram inequality

$$4K_\varpi(t,h_\varpi,\gamma\mathrm{d}\tilde{p},\gamma_1\mathrm{d}\tilde{p}) = K_\varpi(t,h_\varpi,(\gamma+\gamma_1)\mathrm{d}\tilde{p},(\gamma+\gamma_1)\mathrm{d}\tilde{p}) - \\ - K_\varpi(t,h_\varpi,(\gamma-\gamma_1)\mathrm{d}\tilde{p},(\gamma-\gamma_1)\mathrm{d}\tilde{p})$$

and similarly for $I(t,h\gamma\gamma_1,\mathrm{d}p)$, allows us to consider only the case where $\gamma=\gamma_1$, without any loss of generality, and in this case we use the notation

$$K_\varpi(t,h_\varpi,\gamma\mathrm{d}\tilde{p},\gamma\mathrm{d}\tilde{p}) = K_\varpi(t,h_\varpi,\gamma\mathrm{d}^2\tilde{p})$$

to simplify.

At this point, as in the calculation in the proof of Theorem 4.12 with γ and $\gamma_n=\gamma\mathbb{1}_{\{1\le n\gamma\}}$, write

$$K_\varpi(t,h_\varpi,\gamma\mathrm{d}^2\tilde{p}) - K_\varpi(t,h_\varpi,\gamma_n\mathrm{d}^2\tilde{p}) = \\ = K_\varpi(t,h_\varpi,(\gamma-\gamma_n)\mathrm{d}\tilde{p},(\gamma+\gamma_n)\mathrm{d}\tilde{p}),$$

to deduce

$$\mathbb{E}\big\{K_\varpi(t,h_\varpi,\gamma\mathrm{d}^2\tilde{p}) - K_\varpi(t,h_\varpi,\gamma_n\mathrm{d}^2\tilde{p})\big\} \le \\ \le \big(\mathbb{E}\{K_\varpi(t,h_\varpi,(\gamma-\gamma_n)\mathrm{d}^2\tilde{p}\}\big)^{1/2}\big(\mathbb{E}\big\{K_\varpi(t,h_\varpi,(\gamma+\gamma_n)\mathrm{d}^2\tilde{p}\}\big)^{1/2},$$

after using Hölder inequality. Therefore, we are reduced to study the limit as the mesh $\varpi|\to 0$, for each γ_n, which is the case of a process having paths with bounded variation.

To prove (4.33), proceed as in Corollary 4.6, first by localization, we may assume that

$$\int_0^T \big[|h(s)|+|h_\varpi(s)|\big]\Big[1+|f(s)|^2+|g(s)|^2 + \\ + \int_{\mathbb{R}_*^m}|\gamma(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big]\mathrm{d}s \le C<\infty,$$

for some constant $C=C_T>0$, as $|\varpi|\to 0$. Next, the arguments are the same, except that the inequality

$$\mathbb{E}\Big\{\Big|\int_{\mathbb{R}_*^m\times]t_{i-1},t_i]}\gamma(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)\Big|^{2c}\Big|\mathcal{F}(t_{i-1})\Big\} \le \\ \le M_c\mathbb{E}\Big\{\Big(\int_{t_{i-1}}^{t_i}\mathrm{d}s\int_{\mathbb{R}_*^m}|\gamma(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big)^c + \\ + \mathbb{1}_{c>1}\int_{t_{i-1}}^{t_i}\mathrm{d}s\int_{\mathbb{R}_*^m}|\gamma(\zeta,s)|^{2c}\pi(\mathrm{d}\zeta)\Big|\mathcal{F}(t_{i-1})\Big\},$$

for some constant $M_c>0$, forces the condition $a+b>0$. □

• *Remark* 4.15. If two independent Poisson measures p_1 and p_2 with Lévy measure π_1 and π_2 on R_*^m are given then $p = p_1 + p_2$ is a Poisson measure with Lévy measure π on $\mathbb{R}_*^m$, where π is the restriction to $\mathbb{R}_*^{2m}$ (i.e., integrands are extended by zero to $\mathbb{R}^{2m}$ before integration) of the sum of the product measures $(\pi_1 \times \delta_0) + (\delta_0 \times \pi_2)$ on $\mathbb{R}^{2m}$, with δ_0 being the Dirac measure in $\mathbb{R}^m$. This means that

$$\gamma(\zeta_1, \zeta_2)\pi(\mathrm{d}\zeta_1, \mathrm{d}\zeta_2) = \gamma(\zeta_1, 0)\pi_1(\mathrm{d}\zeta_1) + \gamma(0, \zeta_2)\pi_2(\mathrm{d}\zeta_2).$$

In other words, for two independent Poisson measures p_1 and p_2, the stochastic integrals becomes

$$\int_{\mathbb{R}_*^m \times]0,t]} \gamma_1(\zeta_1, s)\tilde{p}_1(\mathrm{d}\zeta_1, \mathrm{d}s) = \int_{\mathbb{R}_*^m \times]0,t]} \gamma_1(\zeta_1, s)\mathbb{1}_{\{\zeta_2=0\}}\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s)$$

and there is no simultaneous jumps, i.e., if $\delta p_1(s_1) \neq 0$ and $\delta p_2(s_2) \neq 0$ then $s_1 \neq s_2$. Therefore, in term of a unique Poisson measure p, the Poisson measures obtained from the stochastic integrals $I(t, \gamma_1, \mathrm{d}\tilde{p})$ and $I(t, \gamma_2, \mathrm{d}\tilde{p})$ are independent iff $\gamma_1\gamma_2 = 0$ on $\mathbb{R}_*^{2m}$, and applying Corollary 4.14, we deduce that

$$K_\varpi(t, h_\varpi, \gamma_1 \mathrm{d}\tilde{p}_1, \gamma_2 \mathrm{d}\tilde{p}_2) \to I(t, h\gamma_1\gamma_2, \mathrm{d}p) = 0,$$

when p_1 and p_2 are independent of each other, see also Corollary4.11. □

• *Remark* 4.16. An integer random measure ν with a predictable compensator $\nu^p(\mathrm{d}\zeta, \mathrm{d}t)$ which is absolutely continuous with respect to the Lebesgue measure $\mathrm{d}t$ could be used instead of the Poisson measure p, and both convergences remain valid. □

QV for Local-Martingales

Going back to a general case where M is a martingale and without applying Doob-Meyer Decomposition (Theorem 2.7) the quadratic variation can be obtained but contrary to the stochastic integrals, a priori, some more work is necessary. For instance, in the construction of the stochastic integral with respect to a Wiener process w or a Poisson martingale measure $\tilde{p}$, the fact that the expressions

$$t \mapsto \Big(\int_0^t f(s)\mathrm{d}w(s)\Big)^2 - \int_0^t |f(s)|^2\mathrm{d}s \quad \text{and}$$
$$t \mapsto \Big(\int_{\mathbb{R}_*^m \times]0,t]} h(\zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s)\Big)^2 - \int_0^t \mathrm{d}s \int_{\mathbb{R}_*^m} |h(\zeta, s)|^2\pi(\mathrm{d}\zeta)$$

are local-martingales is directly deduced from properties of the Wiener and Poisson processes when the integrand f and h are approximated by piecewise constant processes. However, this same assertion for a martingale M requires the predictable quadratic variation $\langle M \rangle$ (and optional quadratic variation $[M]$ when M is discontinuous).

For a sequence of partitions ϖ with mesh $|\varpi| \to 0$ and real-valued processes X and Y define

$$[X,Y]_t^{\varpi} = \sum_{i=1}^{n} \big(X_{t\wedge t_i} - X_{t\wedge t_{i-1}}\big)\big(Y_{t\wedge t_i} - Y_{t\wedge t_{i-1}}\big), \quad 0 < t \le T \le t_n,$$

with $[X,Y]_0^{\varpi} = 0$. Clearly, the sum stops after $i = k$ with $t_k \le t < t_{k+1}$. As mentioned early, for two optional processes this is refer to as the (optimal) quadratic co-variation of X and Y relative to the partition ϖ. If $X = Y$ then $[X,X]^{\varpi}$ is called (optimal) quadratic variation of X and it is written as $[X]^{\varpi}$. The limit as the mesh $|\varpi| \to 0$ is denoted by $[X,Y]$ or $[X]$, which is referred to as the (optional) quadratic co-variation (or variation) process.

If the processes X and Y are continuous then the notation $\langle X,Y\rangle^{\varpi} = [X,Y]^{\varpi}$, $\langle X\rangle^{\varpi} = [X]^{\varpi}$, $\langle X,Y\rangle = [X,Y]$, and $\langle X\rangle^{\varpi} = [X]$ could be used, referring to the predictable (co-)quadratic variation.

Several steps are necessary to show that the limit $\langle X,Y\rangle_t^{\varpi}$ as $|\varpi| \to 0$ exists when X and Y are continuous local-martingales, e.g., see Kunita [104, Section 2.2, pp. 46–56]. These is a follows:

(1) If a process X has continuous paths and a process Y has paths with bounded variation are on a given time interval $[0,T]$, then the estimate (4.21) shows that for any sequence of partition ϖ with mesh $|\varpi| \to 0$ we have $\langle X,Y\rangle_t^{\varpi} \to 0$ for every t in $[0,T]$, almost surely. Therefore, if X has continuous paths with bounded variation then $\langle X\rangle^{\varpi} = [X]^{\varpi} \to 0$ as $|\varpi| \to 0$.

(2) If the (optimal) quadratic variation process $\langle X\rangle$ exists for a vector space of processes (e.g., continuous local-martingales) then bi-linearity yields the parallelogram equality

$$4[X,Y] = [X+Y, X+Y] + [X-Y, X-Y],$$

can be used to obtain the co-variation process. Moreover, if

$$Y_t^{\varpi} = \sum_{i=1}^{n} X_{t\wedge t_{i-1}}(X_{t\wedge t_i} - X_{t\wedge t_{i-1}}) \quad \text{then} \quad X_t^2 - X_0^2 = [X]_t^{\varpi} + 2Y_t^{\varpi},$$

i.e., the convergence of $[X]^{\varpi}$ is reduced to the convergence of the process Y^{ϖ}. Furthermore, if stochastic integrals are used then

$$Y_t^{\varpi} \to \int_{]0,t]} X_{s-}\mathrm{d}X_s \quad \text{in } L^2, \text{ uniformly in } t, \text{ as } \quad |\varpi| \to 0,$$

which shows the existence of the (optional) quadratic variation process, and its connection with stochastic integrals.

(3) If X is a bounded martingale then Y^{ϖ} is a martingale with zero mean and

$$\sup_{0\le t\le T} \mathbb{E}\big\{(Y_t^{\varpi})^2\big\} \le 2C^2, \quad \text{where} \quad \sup_{0\le t\le T} |X_t| \le C.$$

Indeed, first for any $s < t$ choose j such that $t_{j-1} \le s < t_j$ to obtain

$$Y_t^{\varpi} - Y_s^{\varpi} = X_{t_{j-1}}(X_{t\wedge t_j} - X_s) + \sum_{i\ge j} X_{t\wedge t_{i-1}}(X_{t\wedge t_i} - X_{t\wedge t_{i-1}}),$$

which implies that Y^{ϖ} is a martingale satisfying $\mathbb{E}\{Y_t^{\varpi}\} = 0$, for every t in $[0,T]$. Moreover, note that

$$Y_{t_i}^{\varpi} - Y_{t_{i-1}}^{\varpi} = X_{t_{i-1}}(X_{t_i} - X_{t_{i-1}})$$

to calculate the quadratic variation

$$[Y^{\varpi}]_t^{\varpi} = \sum_{i=1}^{n} X^2_{t\wedge t_{i-1}}(X_{t\wedge t_i} - X_{t\wedge t_{i-1}})^2 \le C^2[X]_t^{\varpi}$$

and to deduce that

$$\mathbb{E}\big\{(Y_t^{\varpi})^2\big\} = \mathbb{E}\big\{[Y^{\varpi}]_t^{\varpi}\big\} \le C^2\mathbb{E}\big\{[X]_t^{\varpi}\big\} = C^2\mathbb{E}\big\{X_t^2 - X_0^2\big\},$$

as desired.

(4) If X is a cad-lag bounded martingale and ϖ is a sequence of partition with mesh $|\varpi| \to 0$ then Y^{ϖ} converges uniformly in L^2, i.e., for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|\varpi'|, |\varpi''| < \delta \quad \text{implies} \quad \mathbb{E}\big\{\sup_{0\le t\le T}(Y_t^{\varpi'} - Y_t^{\varpi''})^2\big\} < \varepsilon.$$

Indeed, any two partitions ϖ' and ϖ'' can be combined into one partition $\varpi = \varpi' \cup \varpi''$ so that

$$Y_t^{\varpi'} - Y_t^{\varpi''} = \sum_{i=1}^{n}(X_{t_{i-1}}^{\varpi'} - X_{t_{i-1}}^{\varpi''})(X_{t\wedge t_i} - X_{t\wedge t_{i-1}}),$$

where the cad-lag piecewise constant processes $X^{\varpi'}$ and $X^{\varpi''}$ are given by $X_t^{\varpi'} = X_{t'_{i-1}}$ for any $t'_{i-1} \le t < t'_i$ and similarly $X^{\varpi''}$. This implies

$$[Y^{\varpi'} - Y^{\varpi''}]_t^{\varpi} = \sum_{i=1}^{n}(X_{t_{i-1}}^{\varpi'} - X_{t_{i-1}}^{\varpi''})^2(X_{t\wedge t_i} - X_{t\wedge t_{i-1}})^2 \le$$

$$\le \Big(\sup_{0\le s\le t}(X_s^{\varpi'} - X_s^{\varpi''})^2\Big)\langle X\rangle_t^{\varpi},$$

and Doob's maximal inequality yields

$$\mathbb{E}\big\{\sup_{0\le t\le T}|Y_t^{\varpi'} - Y_t^{\varpi''}|^2\big\} \le 4\mathbb{E}\big\{|Y_T^{\varpi'} - Y_T^{\varpi''}|^2\big\} = 4\mathbb{E}\big\{[Y^{\varpi'} - Y^{\varpi''}]_T^{\varpi}\big\}$$

which is bounded by

$$\Big(\mathbb{E}\Big\{\Big(\sup_{0\le t\le T}|X_t^{\varpi'} - X_t^{\varpi''}|^2\Big)^2\Big\}\Big)^{1/2}\Big(\mathbb{E}\big\{\big([X]_T^{\varpi}\big)^2\big\}\Big)^{1/2},$$

after using Schwarz's inequality. Moreover

$$\mathbb{E}\big\{\big([X]_T^{\varpi}\big)^2\big\} \le 2\mathbb{E}\big\{\big(X_T^2 - X_0^2\big)^2\big\} + 4\mathbb{E}\big\{\big(Y_T^{\varpi}\big)^2\big\} \le 16C^2.$$

If X is continuous then the sup vanishes (almost surely) as mesh $|\varpi| \to 0$ otherwise, if X is only cad-lag, again by means of Doob's maximal inequality,

$$\mathbb{E}\Big\{\Big(\sup_{0\le t\le T}|X_t^{\varpi'} - X_t^{\varpi''}|^2\Big)^2\Big\} \le 4\mathbb{E}\big\{\big|X_T^{\varpi'} - X_T^{\varpi''}\big|^4\big\}$$

and the limit vanishes too.

(5) If X is a cad-lag bounded martingale and ϖ is a sequence of partition with mesh $|\varpi| \to 0$ then $[X]_t^{\varpi}$ converges to $[X]_t$ in L^2, uniformly in $0 \le t \le T$. The optional process $[X]$ is cad-lag and non-negative increasing. Moreover, the process $X^2 - [X]$ is a square-integrable martingale. Indeed, the fact that the limit is cad-lag follows form (4) above, and to check that $[X]$ is non-decreasing, note that any sequence of partitions ϖ with mesh $|\varpi| \to 0$ can be reorganized as an increasing sequence of partitions ϖ and the set D of all points belonging to some partition in the sequence is dense in $[0,T]$. Since the limit is cad-lag, it suffices to show that $[X]$ is non-decreasing only for points in D. Thus, for every $s < t$ in D, there exists $\delta > 0$ such that s, t belongs to ϖ whenever $|\varpi| < \delta$, so that from the definition $[X]_t^{\varpi} \ge [X]_s^{\varpi}$, and as $|\varpi| \to 0$ this becomes $[X]_t \ge [X]_s$ as desired. Finally, remark that Y^{ϖ} is a square-integrable martingale which converges to $t \mapsto X_t^2 - X_0^2 - [X]_t$, to deduce that $X^2 - [X]$ is a square-integrable martingale.

(6) If X is a cad-lag local-martingale with bounded jumps (i.e., $|X_t - X_{t-}| \le K$, for some deterministic constant K) then there exists an optional cad-lag increasing process $[X]$ such that $[X]_t^{\varpi}$ converges to $[X]_t$ in probability, uniformly in $0 \le t \le T$ as the mesh $|\varpi| \to 0$. Indeed, consider the sequence of stopping times $\tau_n = \inf\{t \in [0,T] : |X_t| > n\}$, which has the properties (a) $|X_t| \le n + K$, for every $0 < t \le \tau_n$, (b) the process $t \mapsto X_t^{\tau_n} = X_{t\wedge\tau_n}$ is a bounded martingale, (c) $P\{\tau_n < T\} \to 0$ as $n \to \infty$. Now, apply (5) above to the $X_t^{\tau_n}$ to obtain the optional cad-lag increasing process $[X^{\tau_n}]$ as the limit of $[X^{\tau_n}]^{\varpi}$ as $|\varpi| \to 0$. As $n \to \infty$, this defines a process $[X]$, because $[X^{\tau_n}]_t = [X^{\tau_k}]_t$, for any $0 \le t \le \tau_n$ and $k \ge n$. Moreover, the inequality

$$P\Big\{\sup_{0\le t\le T}\big|[X]_t - [X]_t^{\varpi}\big|\Big\} \le P\Big\{\sup_{0\le t\le T}\big|[X^{\tau_n}]_t - [X^{\tau_n}]_t^{\varpi}\big|\Big\} + P\{\tau_n < T\},$$

shows the desired convergence, by letting first $|\varpi| \to 0$ and later $n \to \infty$. Remark that if the jumps of X are unbounded then the stopped local-martingale X^{τ_n} may not be necessarily a bounded martingale, where all previous argument can be applied.

(7) Let X be a cad-lag local-martingale with bounded jumps. Then X is continuous if and only if its (optional) quadratic variation process $[X]$ is continuous.

It is clear that if X is continuous then $[X]$ is also continuous; to check the converse, let us show that $[X]_t - [X]_{t-} \to (X_t - Xt-)^2$. Indeed, take a sequence of time $s_k \uparrow t$ and choose n such that $t_n < s_k < t \leq t_{n+1}$ to obtain

$$[X]^{\varpi}_t - [X]^{\varpi}_{s_k} = (X_t - X_{t_n})^2 - (X_{s_k} - X_{t_n})^2.$$

As the mesh $|\varpi| \to 0$ and $k \to \infty$, this implies $[X]^{\varpi}_t - [X]^{\varpi}_{s_k} \to (X_t - Xt-)^2$. Due to the uniform convergence, $[X]^{\varpi}_{s_k} \to [X]^{\varpi}_{t-}$, and the jump-relation follows.

(8) If X is a cad-lag local-martingale with bounded jumps and A is optional cad-lag increasing locally-integrable process vanishing at the origin (i.e, $A_0 = 0$) such that the process $X^2 - A$ is a local-martingale and the square-jumps $(X_t - X_{t-})^2 = A_t - A_{t-}$ then $A = [X]$. Indeed, combining (5), (6) and (7) above, it is clear that the process $t \mapsto X_t^2 - [X]_t$ satisfies the same conditions as the process A. Now, to check the uniqueness, note the process $A - [X]$ is a difference of two cad-lag local-martingales with no jumps, i.e., a continuous local-martingale, and therefore, a constant process, namely $A = [X]$.

(9) To include a Lévy process X in the previous analysis we need to consider the part with large-jumps. The arguments for large-jumps is practically a deterministic analysis, and it suffices to remark that on any bounded time interval there is necessarily a finite number of jumps larger than a positive constant $\varepsilon > 0$. Alternatively, we may apply the previous points (1),...,(8) to any continuous local-martingale, in particular to the continuous part of a cad-lag locally square-integrable local-martingale (i.e., the orthogonal decomposition $X = X^c + X^d$ with X^c being a continuous local-martingale), and define $[X] = [X^c] + [X^d]$, where the process of the square-jumps

$$[X^d]_t = \sum_{0<s\leq t} \big(X_s - X_{s-}\big)^2, \quad \forall t \geq 0$$

is, by definition, the quadratic variation of the discontinuous martingale X^d. Note that X^d is a purely jump martingale (but strictly speaking, it may contain something more than jumps, it could be a compensated purely jump martingale) and that the (optional) quadratic variation process $[X^d]$ may be continuous in probability. Even another way, as mentioned early, the stochastic integral can be used to define the (optional) quadratic variation process

$$[X,Y]_t = X_tY_t - X_0Y_0 - \int_{]0,t]} X_{s-}\mathrm{d}Y(s) - \int_{]0,t]} Y_{s-}\mathrm{d}X(s), \quad \forall t \geq 0,$$

for two cad-lag local-martingales X and Y, and a posteriori show the convergences for partitions with mesh vanishing, e.g., Jacob and Shiryaev [84, Section 4c, pp. 51–58].

As mentioned early, for a continuous local-martingale M the (optional) quadratic variation process $[M]$ is continuous and therefore, denoted by $\langle M\rangle$ and referred to as the predictable quadratic variation. Thus, for a cad-lag locally square-integrable local-martingale M (which includes local-martingales

with bounded jumps) the predictable quadratic variation process $\langle M\rangle$ is defined as the predictable dual-projection (also called the predictable compensator) of the (optional) quadratic variation process $[M]$.

A simplify argument begins with a filtration $\mathbb{F}$ satisfying the usual conditions, and the predictable and optional σ-algebras $\mathcal{P}$ and $\mathcal{O}$) defined. Any cad-lag adapted process is optional and any cag-lad adapted process is predictable. Stopping (also called optional) times are defined using the filtration $\mathbb{F}$, and strictly increasing sequences of stopping times yield the so-called predictable times as their limits. For a predictable (optional) time τ, the expression $t \mapsto \mathbb{1}_{t\leq\tau}$, is a predictable (optional) process. Given an optional integrable and increasing process A, its compensator A^p is a predictable integrable and increasing process satisfying

$$\mathbb{E}\Big\{\int \mathbb{1}_{t\leq\tau}\mathrm{d}A(t)\Big\} = \mathbb{E}\Big\{\int \mathbb{1}_{t\leq\tau}\mathrm{d}A^p(t)\Big\},$$

for every almost surely bounded predictable time τ. Certainly, this equality remains valid for any predictable nonnegative process f replacing $\mathbb{1}_{t\leq\tau}$. This compensator is unique except for an evanescent set, and as seen early, in the case of a Poisson measure p with Lévy measure π, the predictable quadratic variation of the stochastic integral

$$\int_{\mathbb{R}^m_*\times]0,t]} \gamma(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s) \quad \text{is} \quad A^p_t = \int_0^t \mathrm{d}s \int_{\mathbb{R}^m_*} |\gamma(\zeta,s)|^2\pi(\mathrm{d}\zeta),$$

while that the optional quadratic variation is

$$A_t = \int_{\mathbb{R}^m_*\times]0,t]} |\gamma(\zeta,s)|^2 p(\zeta,\mathrm{d}s),$$

and A^p is the compensator of A. Remark that the stochastic integral is defined for L^2-type classes of equivalence, so that the difference between A and its compensator A^p is very subtle. For instance, the predictable quadratic variation process $\langle M\rangle$ of a cad-lag locally square-integrable local-martingale M is identified as the unique predictable locally integrable increasing process vanishing at the origin such that $M^2 - \langle M\rangle$ is a cad-lag martingale. However, $\langle M\rangle$ is a continuous process if and only if M is quasi-continuous, i.e., $\mathbb{E}\{M_{\tau_n}\} \to \mathbb{E}\{M_\tau\}$, for any increasing sequence $\{\tau_n\}$ of stopping times converging to a bounded predictable time τ. Actually, by requiring $\bigvee_n \mathcal{F}(\tau_n) = \mathcal{F}(\tau)$, the quasi-continuity property can be attached directly to the filtration $\mathbb{F}$. For instance, the interested reader may consult the books by He et al. [68, Chapters V and VI, pp. 135–190] or Liptser and Shiryayev [111, Chapter 1, pp. 1–84].

4.3 Random Fields of Martingales

As mentioned early, sometimes the interest is on random fields with parameter in $\mathbb{R}^d \times [0,\infty)$, with values in $\mathbb{R}^n$ or $\mathbb{R}^d \times L^2(\pi)$, where π is a Lévy measure

in $\mathbb{R}^m_*$. For instance, a $\mathbb{R}^d$-valued (vector) predictable locally square-integrable process $\sigma_k = \{\sigma_k(x,s) : x \in \mathbb{R}^d, s \geq 0\}$ (for any $k = 1, \dots, n$ and any x in $\mathbb{R}^d$), and the jump $\gamma = \{\gamma(x,\zeta,s) : x \in \mathbb{R}^d, \zeta \in \mathbb{R}^m_*, s \geq 0\}$ is a $\mathbb{R}^d$-valued (vector) predictable locally square-integrable process (for every x in $\mathbb{R}^d$) relative to $\pi(\mathrm{d}\zeta)\mathrm{d}s$.

Therefore, when a smooth function from $\mathbb{R}^d \times [0,T]$ into $\mathbb{R}^n$ is allowed to be random (i.e., a smooth random field with parameter in $\mathbb{R}^d \times [0,\infty)$ and values in $\mathbb{R}^n$), the degree of smoothness refer to continuity differentiability of a certain order for each fixed ω (almost surely) may not completely adequate for further analysis, and another type of differentiability could be necessary. Of particular important is the smoothness in the time variable, since our interest is on $\mathbb{R}^d$-valued square-integrable (local-) martingales processes that also depend on an spacial parameter, i.e., $M = \{M(x,t) : x \in \mathbb{R}^d, t \geq 0\}$ such that for every x in $\mathbb{R}^d$ the stochastic process $t \mapsto M(x,t)$ is a (local-) martingale in a given filtered space $(\Omega, \mathbb{F}, P)$ satisfying the usual conditions. The so-called quasi-continuous special square-integrable (local-)martingales $M(x,t)$ of the form

$$M(x,t) = \sum_k \int_0^t \sigma_k(x,s)\mathrm{d}w_k(s) + \int_{\mathbb{R}^m_* \times [0,t]} \gamma(x,\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s),$$

where $w = \{w(s) : s \geq 0\}$ is a standard Wiener process in $\mathbb{R}^n$, $p(\mathrm{d}\zeta,\mathrm{d}s)$ is a Poisson measure on $\mathbb{R}^m_*$ with Levy measure π, $\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s) = p(\mathrm{d}\zeta,\mathrm{d}s) - \pi(\mathrm{d}\zeta)\mathrm{d}s$, and for each x in $\mathbb{R}^d$, the diffusion term $\sigma_k = \{\sigma_k(x,s) : x \in \mathbb{R}^d, s \geq 0\}$, $k = 1, \dots, n$, and the jump term $\gamma = \{\gamma(x,\zeta,t) : x \in \mathbb{R}^d, \zeta \in \mathbb{R}^m_*, s \geq 0\}$ as above, and using components, $\sigma_k = (\sigma_{ik})$, $\gamma = (\gamma_i)$, $w = (w_1, \dots, w_n)$, $\zeta = (\zeta_1, \dots, \zeta_m)$, $x = (x_1, \dots, x_d)$, and clearly, $\sigma = (\sigma_{ik})$ may be regarded as a $d \times n$-matrix.

4.3.1 Preliminary Analysis

We try to avoid the treatment of general random fields, and only a quick and superficial discussion is reported below. Indeed, it takes several sections in book Kunita [104] to carefully treat continuous martingales, and even more delicate arguments in Kunita [105] to include the jumps.

Orthogonal Decomposition

The (local) square-integrable (local-) martingale M can be expressed as $M = M^c + M^d$, where M^c is a continuous (local) square-integrable (local-) martingale, indeed, M^c and M^d are orthogonal in the sense that the angle-bracket $\langle M^c, M^d \rangle = 0$. Therefore, the d-square matrices

$$a(x,s) = \sigma(x,s)\sigma^*(x,s) = \Big(\sum_{k=1}^n \sigma_{ik}(x,s)\sigma_{jk}(x,s) \Big) \quad \text{and}$$

$$\mathtt{M}(x,\mathrm{d}\zeta,s) = \Big(\int_{\mathbb{R}^m_*} \gamma_i(x,\zeta,s)\gamma_j(x,\zeta,s)\pi(\mathrm{d}\zeta) \Big)$$

yield the predictable quadratic variation density of $M^c(x,t)$ and $M^d(x,t)$, i.e., the real valued processes

$$t \mapsto M_i^c(x,t)M_j^c(x,t) - \int_0^t a_{ij}(x,s)\mathrm{d}s, \quad t \geq 0, \quad \text{and}$$
$$t \mapsto M_i^d(x,t)M_j^d(x,t) - \int_0^t \mathrm{d}s \int_{\mathbb{R}_*^m} \gamma_i(x,\zeta,s)\gamma_j(x,\zeta,s)\pi(\zeta), \quad t \geq 0,$$

are local-martingales, for any $i,j = 1,\ldots,d$. A drift can be added to the (local) square-integrable (local-) martingales (field) $M(x,t)$ to get the semi-martingale

$$X(x,t) = \int_0^t g(x,s)\mathrm{d}s + \sum_k \int_0^t \sigma_k(x,s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}_*^m \times [0,t]} \gamma(x,\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s),$$

for every x in $\mathbb{R}^d$ and $t \geq 0$. The expression $\mathtt{M}(x,\mathrm{d}\zeta,s)$ can be reconsidered as a random kernel on $\mathbb{R}_*^d$ (instead of $\mathbb{R}_*^m$) defined by

$$\mathtt{M}(x,B,s) = \pi\big(\{\zeta : \gamma(x,\zeta,s) \in B\}\big),$$

for every Borel set B in $\mathbb{R}_*^d$, x in $\mathbb{R}^d$ and $s \geq 0$, and sometimes, the notation

$$\mathtt{M}(x,h,t) = \int_{\mathbb{R}_*^d} h(z)\mathtt{M}(x,\mathrm{d}z,t)$$

for any Borel function h is used, in particular if $h(z) = z_i z_j$ then

$$\mathtt{M}(x,z_iz_j,s) = \int_{\mathbb{R}_*^d} z_i z_j \mathtt{M}(x,\mathrm{d}z,t).$$

Hence, the real valued process

$$t \mapsto M_i^d(x,t)M_j^d(x,t) - \int_0^t \mathtt{M}(x,z_iz_j,s)\mathrm{d}s, \quad t \geq 0,$$

is a local-martingale for any $i,j = 1,\ldots,d$, and x in $\mathbb{R}^d$.

Thus the predictable random fields $\{g(x,t), a(x,t), \mathtt{M}(x,\mathrm{d}z,t)\}$ are called the characteristic densities of the semi-martingale field $\{X(x,t) : x \in \mathbb{R}^d,\ t \geq 0\}$. Actually, even the reference to the Lebesgue measure $\mathrm{d}s$ can be omitted by taking a predictable locally integrable, continuous and increasing process $\alpha(x,t)$ depending on the parameter x ans setting the tern

$$g(x,t)\mathrm{d}\alpha(x,t), \quad a(x,t)\mathrm{d}\alpha(x,t), \quad \mathtt{M}(x,\mathrm{d}z,t)\mathrm{d}\alpha(x,t)$$

as the characteristic of the random field $X(x,t)$ of semi-martingales.

Smoothness of the RF

It is clear that some smoothness on the characteristic should yield smoothness on the random field of semi-martingales, but some difficulties appear. For instance, the d-square matrix-valued process random field $a(x,t)$ could be Lipschitz continuous but the square-root matrix $\sigma(x,t)$ may not be so; and even more complicate is the relation between the random kernel $\mathtt{M}(x,\mathrm{d}z,t)$ and the random field $\gamma(x,\zeta,t)$ and the Lévy measure π.

Conversely, if a random field M of $\mathbb{R}^n$-valued square-integrable local-martingales is given then the predictable field of characteristic $\{a,\mathtt{M}\}$ is also defined, but the dependency on the parameter x is not easy to track.

Suppose that on a filtered space $(\Omega,\mathbb{F},P)$, there given a standard Wiener process $w=\{w(s):s\geq 0\}$ in $\mathbb{R}^n$, and a Poisson measure $p(\mathrm{d}\zeta,\mathrm{d}s)$ on $\mathbb{R}^m_*$ with Levy measure π and martingale measure on $\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)=p(\mathrm{d}\zeta,\mathrm{d}s)-\pi(\mathrm{d}\zeta)\mathrm{d}s$ and for each x in $\mathbb{R}^d$. Moreover, if predictable fields are given, the drift $g(x,t)=(g_i:i=1,\dots,d)$, the diffusion $\sigma(x,t)=(\sigma_{ik}:i=1,\dots,d,\ k=1,\dots,n)$, and the jump $\gamma(x,\zeta,t)=(\gamma_i:i=1,\dots,d)$, then the semi-martingale field

$$X(x,t)=\int_0^t g(x,s)\mathrm{d}s+\sum_k\int_0^t\sigma_k(x,s)\mathrm{d}w_k(s)+ \\ +\int_{\mathbb{R}^m_*\times[0,t]}\gamma(x,\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s), \quad (4.34)$$

or equivalently,

$$X(x,t)=\int_0^t g(x,s-)\mathrm{d}s+\sum_k\int_0^t\sigma_k(x,s-)\mathrm{d}w_k(s)+ \\ +\int_{\mathbb{R}^m_*\times[0,t]}\gamma(x,\zeta,s-)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s),$$

is defined under some typical conditions on g, σ and γ. The characteristic of $X(x,t)$ is the $\mathbb{R}^d$-valued field $g(x,t)$, the $\mathbb{R}^{2d}$-valued field $a(x,y,t)=\sigma(x,t)\sigma^*(y,t)$ and the Lévy kernel $\mathtt{M}(x,\mathrm{d}z,s)=\pi\big(\{\zeta:\gamma(x,\zeta,s)\in\mathrm{d}z\}\big)$ in $\mathbb{R}^d_*$. These random fields are regarded as stochastic processes with values in some space of functions defined on $\mathbb{R}^d\times[0,\infty)$ which are continuous (or of class $C^{m,\alpha}$, i.e., continuously differentiable of order m with α-Hölder continuous m-derivative) in x and cad-lag in t, but some suitable bounds are also necessary to make sense, e.g.,

$$\sup_{x,y\in K}\int_0^T\big[|g(x,t)|+|a(x,y,t)|+\mathtt{M}(x,|z|^2,t)\big]\mathrm{d}t<\infty,$$

almost surely, for every compact subset K of $\mathbb{R}^d$ and any real number $T>0$. Note that regularity the predictable joint quadratic variation (of the continuous part) $a(x,y,t)$ is needed to ensure regularity of its square-root $\sigma(x,t)$ and a

specific expression of the kernel $\mathtt{M}(x, \mathrm{d}z, t)$ is actually required to obtain regularity of the jump term $\gamma(x, \zeta, s)$. For instance, a semi-martingale random field of class $C^2(\mathbb{R}^d)$ is a $\mathbb{R}^d$-valued semi-martingale $X(x,t)$ expressed by (4.34) with coefficients g, σ and γ satisfying

$$\sup_{x\in K}\int_0^T \big[|\partial_x^\alpha g(x,t)| + \sum_k |\partial_x^\alpha \sigma_k(x,t)|^2\big]\mathrm{d}t + \\ + \int_0^T \mathrm{d}t \int_{\mathbb{R}^m_*} |\partial_x^\alpha \gamma(x,\zeta,t)|^2 \pi(\mathrm{d}\zeta) < \infty, \quad (4.35)$$

almost surely, for every compact subset K of $\mathbb{R}^d$, any real number $T > 0$ and any multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$ of order $|\alpha| = \alpha_1 + \cdots + \alpha_d \le 2$.

Composition and Stochastic Integral wrt RF

Therefore, if $y = y(t)$ is a predictable $\mathbb{R}^d$-valued piecewise constant process, i.e., $y(t) = \sum_{i=1}^n c_i \mathbb{1}_{t_{i-1}<t\le t_i}$ for some number c_i and $0 = t_0 < t_1 < \cdots < t_n$ then the stochastic integral with respect to a random fields $X(x,t)$ of semi-martingales is defined by $\sum_{i=1}^n [X(c_i, t_i) - X(c_i, t_{i-1})]$. Assuming continuity of X and its characteristic in the parameter x, this stochastic integral goes to the limit to be defined for every predictable $\mathbb{R}^d$-valued process. Certainly this agrees with the expression

$$\int_0^t X(y(s), \mathrm{d}s) = \int_0^t g(y(s), s)\mathrm{d}s + \sum_k \int_0^t \sigma_k(y(s), s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}^m_* \times [0,t]} \gamma(y(s), \zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s),$$

the stochastic integral defined previously, i.e., same technique with a different viewpoint. If $y = y(t)$ is a cad-lag process then its predictable version $t \mapsto y(t-)$ can be used to obtain the semi-martingale

$$t \mapsto \int_0^t X(y(s-), \mathrm{d}s).$$

Hence, the stochastic integral with respect to a random field of semi-martingales can be decomposed into a composition (with a predictable process, the integrand) and then the integration (with the integrator), but the resulting stochastic integral is linear with respect to the integrand only for linear coefficients (i.e., when $g(z,t)$, $\sigma(x,t)$ and $\gamma(x,\zeta,t)$ are linear in x).

In contract, even if the representation (4.34) holds, the composition process $t \mapsto Y(t) = X(y(t), t)$ is certainly not represented in term of

$$\sum_k \int_0^t \sigma_k(y(t), s)\mathrm{d}w_k(s) \quad \text{and} \quad \int_{\mathbb{R}^m_* \times [0,t]} \gamma(y(t), \zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s),$$

since the integrands are not predictable and therefore the stochastic integrals are really defined for a fixed x and then substitution with $y(t)$ takes place. This procedure is of particular interest when the process y is itself a semi-martingale, for instance, if

$$y(t) = \int_0^t f(s)\mathrm{d}s + \sum_k \int_0^t \varsigma_k(s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}_*^m \times [0,t]} \vartheta(\zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad (4.36)$$

where the predictable processes f, ς and ϑ satisfy

$$\int_0^T \left[|f(t)| + \sum_k |\varsigma_k(t)|^2\right]\mathrm{d}t + \int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} |\vartheta(\zeta, t)|^2 \pi(\mathrm{d}\zeta) < \infty, \quad (4.37)$$

almost surely, for every real number $T > 0$, then we may expect an stochastic differential for the composition of semi-martingales, i.e., a stochastic differential rule the semi-martingale $Y = \{X(y(s), s) : s \geq 0\}$.

It is interesting to remark that if the random field X is time-independent (i.e., has parameter in $\mathbb{R}^d$) then the Itô rule we have proved for (deterministic) smooth functions $\varphi(x, t)$ could be used with $\varphi = X(x)$, for a fixed ω, but, as soon as X dependent also on the time (which is the case of interest for us!), the function $\varphi(x, s) = X(x, s)$ could never be smooth in time, indeed, even continuity is not granted when a Poisson integral term is present. Several questions are of interest, e.g., study how the composition of two smooth random fields of suitable dimensions preserve the form (4.34). Nevertheless, our interest is how the semi-martingale y changes when composed with a smooth random field, instead of a smooth (deterministic) function.

4.3.2 Itô Formula for RF

As mentioned early, firstly consider a (local square-integrable) semi-martingale (field) given by (4.34), i.e.,

$$X(x,t) = \int_0^t g(x,s)\mathrm{d}s + \sum_k \int_0^t \sigma_k(x,s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}_*^m \times [0,t]} \gamma(x,\zeta,s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s),$$

for every x in $\mathbb{R}^d$ and $t \geq 0$, where $g(x,s)$, $\sigma_k(x,s)$ and $\gamma(x,\zeta,s)$ are themselves smooth predictable random fields satisfying (4.35), i.e.,

$$\sup_{x\in K} \int_0^T \left(|\partial_x^\alpha g(x,t)| + \sum_k |\partial_x^\alpha \sigma_k(x,t)|^2\right)\mathrm{d}t + \\ + \int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} |\partial_x^\alpha \gamma(x,\zeta,t)|^2 \pi(\mathrm{d}\zeta) < \infty,$$

almost surely, for every compact subset K of $\mathbb{R}^d$, any real number $T > 0$ and any multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$ of order $|\alpha| = \alpha_1 + \cdots + \alpha_d \leq 2$. Secondly, consider another (local square-integrable) semi-martingale given by (4.36), i.e.,

$$y(t) = \int_0^t f(s)\mathrm{d}s + \sum_k \int_0^t \varsigma_k(s)\mathrm{d}w_k(s) + \int_{\mathbb{R}_*^m \times [0,t]} \vartheta(\zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s),$$

where the predictable processes f, ς and ϑ satisfy

$$\int_0^T \big(|f(t)| + \sum_k |\varsigma_k(t)|^2\big)\mathrm{d}t + \int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} |\vartheta(\zeta, t)|^2 \pi(\mathrm{d}\zeta) < \infty,$$

almost surely, for every real number $T > 0$. Remark that by enlarging the dimensions of driving processes (i.e., of the standard Wiener process w and the standard Poisson measure p), this setting includes the case when the field of semi-martingales X and the semi-martingale y use independent driving processes. The dimension of the random field X is usually d, the same are the parameter x, however, this is not necessary.

Continuous Case IF

First take a look at the case without jumps, i.e., $\gamma = 0$ and $\vartheta = 0$, namely,

$$Y(t) = X(y(t), t) = \int_0^t g(y(t), s)\mathrm{d}s + \sum_k \int_0^t \sigma_k(y(t), s)\mathrm{d}w_k(s),$$

where the stochastic integral is necessarily calculated with a fix value x and then x becomes $y(t)$,

$$y(t) = \int_0^t f(s)\mathrm{d}s + \sum_k \int_0^t \varsigma_k(s)\mathrm{d}w_k(s),$$

and the (predictable) quadratic co-variation processes are given by

$$\mathrm{d}\langle X_i(x, \cdot), X_j(x, \cdot)\rangle(t) = \sum_k \sigma_{ik}(x, t)\sigma_{jk}(x, t)\mathrm{d}t,$$
$$\mathrm{d}\langle y_i(\cdot), y_j(\cdot)\rangle(t) = \sum_k \varsigma_{ik}(t)\varsigma_{jk}(t)\mathrm{d}t.$$

In contract, our interest is to express $\mathrm{d}Y(t) = a(t)\mathrm{d}t + \sum_k b_k(t)\mathrm{d}w_k(t)$.

Theorem 4.17. *With the previous setting, including assumptions (4.35) and (4.37), if there is not jumps, i.e., $\gamma = 0$ and $\vartheta = 0$, then the composition process $t \mapsto Y(t) = X(y(t), t)$ is a semi-martingale with Itô differential* $\mathrm{d}Y(t) =$

$a(t)\mathrm{d}t + \sum_k b_k(t)\mathrm{d}w_k(t)$, *where the predictable processes* a *and* b_k *are given by*

$$a(t) = g(y(t),t) + \sum_i \partial_i X(y(t),t) f_i(t) + \frac{1}{2}\sum_{ijk} \partial^2_{ij} X(y(t),t)\varsigma_{ik}(t)\varsigma_{jk}(t) + \\ + \sum_{jk} \partial_j \sigma_k(y(t),t)\varsigma_{jk}(t),$$

$$b_k(t) = \sigma_k(y(t),t) + \sum_i \partial_i X(y(t),t)\varsigma_{ik}(t),$$

with

$$\partial^2_{ij} X(x,t) = \int_0^t \partial^2_{ij} g(x,s)\mathrm{d}s + \sum_k \int_0^t \partial^2_{ij}\sigma_k(x,s)\mathrm{d}w_k(s),$$

and similarly for the random field $\partial_i X$.

Proof. Without any loss of generality (i.e., use a localization argument) assume that the coefficients f and g are integrable, while σ_k and ς_k are L^2-integrable, and begin with a partition $\varpi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$, to write

$$Y(t) = \sum_{h=1}^n [X(y(t_h),t_h) - X(y(t_{h-1}),t_h)] + \\ + \sum_{h=1}^n [X(y(t_{h-1}),t_h) - X(y(t_{h-1}),t_{h-1})].$$

Note that the second sum approximate a stochastic integral, and on the first sum, use Taylor second-order approximation to write

$$X(y(t_h),t_h) - X(y(t_{h-1}),t_h) \approx \sum_i \partial_i X(y(t_{h-1}),t_h)[y_i(t_h) - y_i(t_{h-1})] + \\ + \frac{1}{2}\sum_{ij} \partial^2_{ij} X(y(t_{h-1}),t_h)[y_i(t_h) - y_i(t_{h-1})][y_j(t_h) - y_j(t_{h-1})].$$

Next,

$$\partial_i X(y(t_{h-1}),t_h) \approx \partial_i X(y(t_{h-1}),t_{h-1}) + \partial_i g(y(t_{h-1}),t_{h-1})[t_h - t_{h-1}] + \\ + \sum_k \partial_i \sigma_k(y(t_{h-1}),t_{h-1})[w_k(t_h) - w_k(t_{h-1})].$$

and similarly with the second derivative $\partial^2_{ij} X(y(t_{h-1}),t_h)$, where the approximation means that the sum (in h) of differences of both terms (i.e., on the right and on the left of sign $\approx$) vanishes as the mesh $|\varpi| \to 0$.

Substitute $\partial_i X(y(t_{h-1}),t_h)$ and $\partial^2_{ij} X(y(t_{h-1}),t_h)$ into the first equality, and note that some combinations contain factors that make the variation vanishes

as the mesh $|\varpi| \to 0$ (i.e., apply Theorem 4.3 and Corollary 4.6) to deduce

$$\begin{aligned} X(y(t_h),t_h)-X(y(t_{h-1}),t_k) \approx \sum_i \partial_i X(y(t_{h-1}),t_{h-1})[y_i(t_h)-y_i(t_{h-1})]+ \\ +\sum_{ik} \partial_i\sigma_k(y(t_{h-1}),t_{h-1})[y_i(t_h)-y_i(t_{h-1})][w_k(t_h)-w_k(t_{h-1})]+ \\ +\frac{1}{2}\sum_{ij} \partial^2_{ij} X(y(t_{h-1}),t_{h-1})[y_i(t_h)-y_i(t_{h-1})][y_j(t_h)-y_j(t_{h-1})]. \end{aligned}$$

Collect all pieces and let the mesh of the partition vanishes to obtain

$$\begin{aligned} \mathrm{d}X(y(t),t) = \partial_i X(y(t),t)\mathrm{d}y_i(t) + \Big(\sum_{ik} \partial_i\sigma_k(y(t),t)\varsigma_{ik}(t)\Big)\mathrm{d}t + \\ +\Big(\frac{1}{2}\sum_{ijk} \partial^2_{ij} X(y(t),t)\varsigma_{ik}(t)\varsigma_{jk}(t)\Big)\mathrm{d}t + X(y(t),\mathrm{d}t), \end{aligned}$$

where the relations with the differentials

$$\begin{aligned} &\partial_i X(y(t),t)\mathrm{d}y_i(t) = \partial_i X(y(t),t)f_i(t)\mathrm{d}t + \sum_k \partial_i X(y(t),t)\varsigma_k(t)\mathrm{d}w_k(t), \\ &\frac{1}{2}\sum_{ij} \partial^2_{ij} X(y(t),t)\mathrm{d}\langle y_i,y_j\rangle(t) = \frac{1}{2}\sum_{ijk} \partial^2_{ij} X(y(t),t)\varsigma_{ik}(t)\varsigma_{jk}(t)\mathrm{d}t, \\ &\mathrm{d}\langle X(y(\cdot),\cdot),y_j\rangle(t) = \sum_{ik} \partial_i\sigma_k(y(t),t)\varsigma_{ik}(t)\mathrm{d}t, \\ &X(y(t),\mathrm{d}t) = g(y(t),t)\mathrm{d}t + \sum_k \sigma_k(y(t),t)\mathrm{d}w_k(t), \end{aligned}$$

and the (derivative) random fields

$$\begin{aligned} &\partial_i X(x,t) = \int_0^t \partial_i g(x,s)\mathrm{d}s + \sum_k \int_0^t \partial_i\sigma_k(x,s)\mathrm{d}w_k(s), \\ &\partial^2_{ij} X(x,t) = \int_0^t \partial^2_{ij} g(x,s)\mathrm{d}s + \sum_k \int_0^t \partial^2_{ij}\sigma_k(x,s)\mathrm{d}w_k(s), \end{aligned}$$

are clearly valid. Remark the extra term involving the product of $\partial_i\sigma_k$ and ς_{ik}, which does not appear for smooth deterministic fields. □

Discrete Jumps Case IF

Only simple jumps are added, i.e., γ and ϑ vanish in a small neighborhood of the origin in $\mathbb{R}^m_*$, and the paths are piecewise continuous. Particularly, the jumps can be ordered, i.e., at a time $\tau_i > 0$ there is a jump of size ζ_i, and the next jump occurs at $\tau_{i+1} > \tau_i$ and there is no jump if $\tau_i = \infty$, where $\{\tau_i\}$ is a strictly (while finite) increasing sequence of stopping times and $\{\zeta_i\}$ is a

sequence of adapted random variables, i.e., ζ_i is $\mathcal{F}(\tau_i)$-measurable. If $p(\mathrm{d}\zeta, \mathrm{d}t)$ denotes corresponding the Poisson measure then

$$\sum_{i=1}^{\infty} c(\zeta_i, \tau_i) \mathbb{1}_{\tau_i \le t} = \int_{\mathbb{R}^m_* \times]0,t]} c(\zeta, s) p(\mathrm{d}\zeta, \mathrm{d}s),$$

for every measurable function c.

Let us begin with $y(t)$ driven by a Wiener process, i.e.,

$$y(t) = \int_0^t f(s)\mathrm{d}s + \sum_k \int_0^t \varsigma_k(s)\mathrm{d}w_k(s), \quad 0 \le t < \tau_1,$$

the first jump occurs at time $t = \tau_1$,

$$y(\tau_1) = y(\tau_1 -) + \vartheta(\zeta_1, \tau_1),$$

and

$$y(t) = y(\tau_1) + \int_{\tau_1}^t f(s)\mathrm{d}s + \sum_k \int_{\tau_1}^t \varsigma_k(s)\mathrm{d}w_k(s), \quad \tau_1 \le t < \tau_2,$$

the second jump occurs at time $t = \tau_2$,

$$y(\tau_2) = y(\tau_2 -) + \vartheta(\zeta_2, \tau_2),$$

and so on, alternating continuous evolution and jumps. Similarly, define the smooth random field $X(x,t)$ driven by a Wiener process and jumps, i.e.,

$$X(x,t) = \int_0^t g(x,s)\mathrm{d}s + \sum_k \int_0^t \sigma_k(x,s)\mathrm{d}w_k(s), \quad 0 \le t < \tau_1,$$

the first jump occurs at time $t = \tau_1$,

$$X(x,\tau_1) = X(x,\tau_1 -) + \gamma(x, \zeta_1, \tau_1),$$

and

$$X(x,t) = X(x,\tau_1) + \int_{\tau_1}^t g(x,s)\mathrm{d}s + \sum_k \int_{\tau_1}^t \sigma_k(x,s)\mathrm{d}w_k(s), \quad \tau_1 \le t < \tau_2,$$

the second jump occurs at time $t = \tau_2$,

$$X(x,\tau_2) = X(x,\tau_2 -) + \gamma(x, \zeta_2, \tau_2),$$

and so on, alternating continuous evolution and jumps, and certainly, keeping x fixed.

Therefore, the stochastic definition for $X(y(t),t)$ is

$$X(y(t),t) = \int_0^t g(y(t),s)\mathrm{d}s + \sum_k \int_0^t \sigma_k(y(t),s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}^m_* \times]0,t]} \gamma(y(t),\zeta,s)p(\mathrm{d}\zeta,\mathrm{d}s),$$

and

$$y(t) = \int_0^t f(s)\mathrm{d}s + \sum_k \int_0^t \varsigma_k(s)\mathrm{d}w_k(s) + \int_{\mathbb{R}^m_* \times]0,t]} \vartheta(\zeta,s)p(\mathrm{d}\zeta,\mathrm{d}s),$$

both equations are valid for $t \geq 0$.

Theorem 4.18. *With the previous setting, including assumptions (4.35) and (4.37), if the jumps are described above then the composition process $t \mapsto Y(t) = X(y(t),t)$ is a semi-martingale with Itô differential with jumps*

$$\mathrm{d}Y(t) = a(t)\mathrm{d}t + \sum_k b_k(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}^m_*} c(\zeta,t)p(\mathrm{d}\zeta,\mathrm{d}t),$$

where the optional processes a and b_k, and predictable process c are given by

$$a(t) = g(y(t),t) + \sum_i \partial_i X(y(t),t)f_i(t) + \frac{1}{2}\sum_{ijk} \partial^2_{ij}X(y(t),t)\varsigma_{ik}(t)\varsigma_{jk}(t) + \\ + \sum_{jk} \partial_j \sigma_k(y(t),t)\varsigma_{jk}(t),$$

$$b_k(t) = \sigma_k(y(t),t) + \sum_i \partial_i X(y(t),t)\varsigma_{ik}(t),$$

$$c(\zeta,t) = \big(X(y(t-)+\vartheta(\zeta,t),t-) - X(y(t-),t-)\big) + \\ + \gamma(y(t-)+\vartheta(\zeta,t),\zeta,t),$$

with

$$\partial^2_{ij}X(x,t) = \int_0^t \partial^2_{ij}g(x,s)\mathrm{d}s + \sum_k \int_0^t \partial^2_{ij}\sigma_k(x,s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}^m_* \times]0,t]} \partial^2_{ij}\gamma(x,\zeta,s)p(\mathrm{d}\zeta,\mathrm{d}s),$$

and similarly for the random field $\partial_i X$.

Proof. Since the expression $X(y(t),t)$ has two steps, (a) get $y(t)$ and $X(x,t)$ separately and (b) compose them to obtain $X(y(t),t)$. It is clear that vector $(X(x,t),y(t))$ makes a jump at time τ_i of size $(\gamma(x,\zeta_i,\tau_i),\vartheta(\zeta_i,\tau_i))$, so if $t = \tau_i$ then the jump at $X(y(t),t)$, i.e.,

$$\delta X(y(t),t) = X(y(t),t) - X(y(t-),t-),$$

can be calculated at $t = \tau_i$ as

$$X(y(t),t) - X(y(t-),t-) = \big(X(y(t-) + \vartheta(\zeta_i,t),t-) + \\ + \gamma(y(t+\vartheta(\zeta_i,t)),\zeta_i,t)\big) - \big(X(y(t-),t-)\big).$$

Therefore, there is an alternative stochastic expression for $X(y(t),t)$, namely,

$$X(y(t),t) = \int_0^t g(y(t),s)\mathrm{d}s + \sum_k \int_0^t \sigma_k(y(t),s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}^m_* \times]0,t]} \big[X(y(s-) + \vartheta(\zeta,s),s-) - X(y(s-),s-)\big]p(\mathrm{d}\zeta,\mathrm{d}s) + \\ + \int_{\mathbb{R}^m_* \times]0,t]} \gamma(y(s-) + \vartheta(\zeta,s),\zeta,s)p(\mathrm{d}\zeta,\mathrm{d}s),$$

which provides the definition of the predictable process $c(\zeta,t)$.

Now, remark that between two consecutive jumps, the pathwise integral with respect to the Poisson measure p behaves like a constant (in t) drift (i.e., added to g) and use the arguments of Theorem 4.17 to deduce the expressions of the processes a and b_k and to complete the proof. □

• *Remark* 4.19. The formula obtained in Theorem 4.18 remains valid when the Poisson measure p is not necessarily a compound Poisson process, but a process with bounded variation, i.e., as long as the coefficients γ and ϑ are integrable with respect to p almost surely, i.e.,

$$\sup_{x\in K} \int_0^T \mathrm{d}t \int_{\mathbb{R}^m_*} \big(|\gamma(x,\zeta,t)| + |\vartheta(\zeta,t)|\big)\pi(\mathrm{d}t) < \infty$$

where π is the corresponding Lévy measure, compare with assumptions (4.35) and (4.37). □

Corollary 4.20. *Under the same assumptions of Theorem 4.18, if the Poisson measure p is replaced by the compensated Poisson measure $\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t) = p(\mathrm{d}\zeta,\mathrm{d}t) - \pi(\mathrm{d}\zeta)\mathrm{d}t$ in the expressions of the random field $X(x,t)$ and the semimartingale $y(t)$, then the Itô differential with jumps is*

$$\mathrm{d}Y(t) = a(t)\mathrm{d}t + \sum_k b_k(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}^m_*} c(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t),$$

where the optional processes a and b_k, and predictable process c are given by

$$a(t) = g(y(t),t) + \sum_{jk} \partial_j\sigma_k(y(t),t)\varsigma_{jk}(t) + A(t)X(y(t-),t) + \\ + I(t)\gamma(y(t-),t),$$

$$b_k(t) = \sigma_k(y(t),t) + \sum_i \partial_i X(y(t),t)\varsigma_{ik}(t),$$

$$c(\zeta,t) = \big[X(y(t-) + \vartheta(\zeta,t),t-) - X(y(t-),t-)\big] + \\ + \gamma(y(t-) + \vartheta(\zeta,t),\zeta,t),$$

where the integro-differential operators

$$I(t)\gamma(x,t) = \int_{\mathbb{R}^m_*} \big[\gamma(x+\vartheta(\zeta,t),\zeta,t) - \gamma(x,\zeta,t) - \vartheta(\zeta,t)\cdot\nabla\gamma(x,\zeta,t)\big]\pi(\mathrm{d}\zeta),$$

$$A(t)X(x,t) = \frac{1}{2}\sum_{ijk}\partial^2_{ij}X(x,t)\varsigma_{ik}(t)\varsigma_{jk}(t) + \sum_i \partial_i X(x,t) f_i(t) + \\ + \int_{\mathbb{R}^m_*} \big[X(x+\vartheta(\zeta,t),t) - X(x,t) - \vartheta(\zeta,t)\cdot\nabla X(x,t)\big]\pi(\mathrm{d}\zeta),$$

and the random fields

$$\partial^2_{ij}X(x,t) = \int_0^t \partial^2_{ij}g(x,s)\mathrm{d}s + \sum_k \int_0^t \partial^2_{ij}\sigma_k(x,s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}^m_*\times]0,t]} \partial^2_{ij}\gamma(x,\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s),$$

and similarly for the random field $\partial_i X$.

Proof. It suffices to note that under these assumptions, the stochastic integral with respect to the compensated Poisson measure $\tilde{p}$ can be separated and expressed as two pieces, so that

$$f(t) \quad \text{becomes} \quad f(t) - \int_{\mathbb{R}^m_*} \vartheta(\zeta,t)\pi(\mathrm{d}\zeta), \quad \text{and}$$

$$g(x,t) \quad \text{becomes} \quad g(x,t) - \int_{\mathbb{R}^m_*} \gamma(x,\zeta,t)\pi(\mathrm{d}\zeta).$$

At this point, the Itô formula with jumps in Theorem 4.18 can be used to deduce the desired result.

It should be clear that the integro-differential operators A and I act on the variable x and depend (beside the Lévy measure π) on the coefficients $\vartheta(\zeta,t)$, $\varsigma(t)$ and $f(t)$ which are attached only to the semi-martingale y. In particular, the notations $I(t)$ and $A(t)$ make evident the (possible) dependency on the variable t fo the coefficients.

Also remark that the integrand c in the (compensated) Poisson integral is a predictable process, while the integrands a and b_k in the pathwise integral (in $\mathrm{d}t$) and the stochastic integral (in $\mathrm{d}w_k$) are optional processes. However, the predictable process

$$a_-(t) = g(y(t-),t) + \sum_{jk}\partial_j\sigma_k(y(t-),t)\varsigma_{jk}(t) + \\ + A(t)X(y(t-),t-) + I(t)\gamma(y(t-),t)$$

yields the same the pathwise integral, and the predictable process

$$b_k(t) = \sigma_k(y(t-),t) + \sum_i \partial_i X(y(t-),t)\varsigma_{ik}(t)$$

is actually used (by definition) for the stochastic integral relative to the Wiener process. This actually can be applied to the stochastic Poisson integral, and replace the predictable coefficient c with its optional version, but this may cause confusion, since in this case, the pathwise integral (if defined) does not necessarily agree with the stochastic integral.

Note that the definition of the integro-differential operator I (as well as the purely integro-differential part of A) requires $y(t-)$, due to the Poisson integral (in $\mathrm{d}\zeta$). Clearly, this is not necessary in Theorem 4.18, where only the pathwise jump-integral (in $p(\mathrm{d}\zeta, \mathrm{d}t)$) requires $y(t-)$ and $X(x, t-)$. □

General Jumps Case IF

Now, going back to the general type of jumps, i.e., as in the beginning of this subsection with the definitions (4.34) and (4.36), and the assumptions (4.35) and (4.37). Note that if the Poisson measure p with Lévy measure π in $\mathbb{R}^m_*$ is approximated by the Poisson measure p_ε corresponding to the Lévy measure $\pi_\varepsilon(\mathrm{d}\zeta) = \mathbb{1}_{\varepsilon<|\zeta|<1/\varepsilon}\pi(\mathrm{d}\zeta)$, then the previous construction applies to p_ε.

Theorem 4.21. *Under the previous general setting (4.34) and (4.36) on random field $X(x,t)$ and the semi-martingale $y(t)$, and under the assumptions (4.35) and (4.37), the Itô differential with jumps is*

$$\mathrm{d}Y(t) = a(t)\mathrm{d}t + \sum_k b_k(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}^m_*} c(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t),$$

where the predictable processes a, b_k and c are given by

$$\begin{aligned}
a(t) &= g(y(t-),t) + \sum_{jk}\partial_j\sigma_k(y(t-),t)\varsigma_{jk}(t) + A(t)X(y(t-),t) + \\
&\qquad + I(t)\gamma(y(t-),t),\\
b_k(t) &= \sigma_k(y(t-),t) + \sum_i \partial_i X(y(t-),t)\varsigma_{ik}(t),\\
c(\zeta,t) &= \big[X(y(t-)+\vartheta(\zeta,t),t-) - X(y(t-),t-)\big] + \\
&\qquad + \gamma(y(t-)+\vartheta(\zeta,t),\zeta,t),
\end{aligned}$$

where the integro-differential operators

$$\begin{aligned}
I(t)\gamma(x,t) &= \int_{\mathbb{R}^m_*}\big[\gamma(x+\vartheta(\zeta,t),\zeta,t) - \gamma(x,\zeta,t) - \vartheta(\zeta,t)\cdot\nabla\gamma(x,\zeta,t)\big]\pi(\mathrm{d}\zeta),\\
A(t)X(x,t) &= \frac{1}{2}\sum_{ijk}\partial^2_{ij}X(x,t)\varsigma_{ik}(t)\varsigma_{jk}(t) + \sum_i \partial_i X(x,t) f_i(t) + \\
&\quad + \int_{\mathbb{R}^m_*}\big[X(x+\vartheta(\zeta,t),t) - X(x,t) - \vartheta(\zeta,t)\cdot\nabla X(x,t)\big]\pi(\mathrm{d}\zeta),
\end{aligned}$$

and the random fields

$$\partial^2_{ij}X(x,t)=\int_0^t\partial^2_{ij}g(x,s)\mathrm{d}s+\sum_k\int_0^t\partial^2_{ij}\sigma_k(x,s)\mathrm{d}w_k(s)+\\+\int_{\mathbb{R}^m_*\times]0,t]}\partial^2_{ij}\gamma(x,\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s),$$

and similarly for the random field $\partial_i X$.

Proof. Indeed, all what it takes now is to justify the limit of the Poisson measure p_ε as $\varepsilon\to 0$. To simplify the arguments, we may think that the jumps in the random field (i.e., γ) and in the semi-martingale (i.e., ϑ). In any way, by localization, the coefficient become square-integrable so that the L^2-arguments can be applied. The details are reasonable simple (but perhaps tedious) and so, not included here. □

It interesting to remark the two new terms

$$\sum_{jk}\partial_j\sigma_k(y(t-),t)\varsigma_{jk}(t)\quad\text{and}\\\int_{\mathbb{R}^m_*}\big[\gamma(y(t-)+\vartheta(\zeta,t),\zeta,t)-\gamma(y(t-),\zeta,t)-\\-\vartheta(\zeta,t)\cdot\nabla\gamma(y(t-),\zeta,t)\big]\pi(\mathrm{d}\zeta)$$

that appear due to the present of a smooth random field, instead of a simple smooth (deterministic) function. These are the quadratic co-variation of the continuous parts in the random field X and the semi-martingale y. As expected, these two terms vanish when the driving processes of X are independent of the driving processes of y, and clearly, there is not interaction between the continuous part and the discontinuous (or jump) part. Actually, the second term (i.e., $I\gamma$) does not vanish when the jumps of $t\mapsto X(x,t)$ and $t\mapsto y(t)$ are independent, i.e., they do not occur simultaneously (i.e., $\gamma\vartheta=0$ π-almost everywhere), the expression under the integral over $\mathbb{R}^m_*$ vanishes except when $\vartheta\neq 0$, i.e., it is just the contribution of the jumps relative to X.

As mentioned early, if two Wiener processes and two Poisson measures are used, one for the expression of the smooth random field $\{X(x,t):x\in\mathbb{R}^d,\,t\geq 0\}$ and one for the semi-martingale $\{y(t):t\geq 0\}$ then a description on the quadratic co-variation is necessary. All this is simplify by assuming only one source of Wiener process and Poisson measure of a suitable dimension to accommodate the former setting, as a particular case.

4.3.3 Stochastic Flows

Stochastic differential equations is the key tool needed to study stochastic flows, which are an improvement (in the pathwise view) over Markov processes. Therefore, this section is necessarily only a brief introduction, as mentioned early, the

interested reader may check, for instance, the book Kunita [104] for a carefully treatment beginning with Brownian flows, and later, even more delicate arguments can be found in Kunita [105] to include the jumps, as well as several references there.

Recall that a function f from the Euclidean space $\mathbb{R}^d$ into itself is called a homeomorphism (or bicontinuous) if (a) f is one-to-one and onto (i.e., bijection), (b) f is continuous, and (c) its inverse f^{-1} is continuous (f is open, maps open sets into open sets). A homeomorphism f is called a diffeomorphism of class C^k if f and its inverse f^{-1} are continuously differentiable up to the order k. The topology associated with homeomorphisms (diffeomorphisms of class C^k) is the local (over compact sets of $\mathbb{R}^d$) uniform convergence of itself and its inverse (including derivatives up to the order k). Certainly, this topological space of homeomorphisms or diffeomorphisms of class C^k is a Polish space (i.e., a separable complete metrizable space) denoted by $H^k = H^k(\mathbb{R}^d)$, with $H = H^0$ corresponding to diffeomorphisms of class C^0, i.e., homeomorphisms.

Definition 4.22. A continuous (or cad-lag) random field $\{\varphi_{s,t}(x) : s,t \in [0,T],\, x \in \mathbb{R}^d\}$ on a probability space $(\Omega, \mathcal{F}, P)$ is called a *stochastic flow of homeomorphism* (or stochastic flow of diffeomorphisms of class C^k, $k \geq 0$) if $\{\varphi_{s,t}\}$ is a two-parameter continuous (cad-lag) process in (s,t) taking values into the Polish space of homeomorphisms of diffeomorphisms of class C^k) from $\mathbb{R}^d$ into itself, i.e. $H^k(\mathbb{R}^d)$, such that there exists a null set N such that for every ω in $\Omega \setminus N$ we have: (a) the function $(s,t,x) \mapsto \varphi_{s,t}(x,\omega)$ is continuous (or cad-lag in (s,t) and continuous in x), (b) for every (s,t) the function $x \mapsto \varphi_{s,t}(x,\omega)$ is a homeomorphism (or a diffeomorphism of class C^k), (c) for every s,t and x the equality $\varphi_{s,t}(x,\omega) = x$ holds, (d) if $\circ$ denotes the composition of maps then $\varphi_{s,r} \circ \varphi_{r,t} = \varphi_{s,t}$, for every s,t,r in $[0,T]$. The *forward flow* is $\{\varphi_{s,t}(x) : 0 \leq s \leq t \leq T,\, x \in \mathbb{R}^d\}$, while the *backward flow* is its inverse $\{\varphi_{t,s}(x) = \varphi_{s,t}^{-1}(x) : 0 \leq s \leq t \leq T,\, x \in \mathbb{R}^d\}$. □

In general, if $\{\varphi_t : t \geq 0\}$ is a cad-lag process with values in H^k, $k \geq 0$, such that its inverse φ_t^{-1} is a cad-lag process (with values in H^k) and φ_0 is the identity (in $\mathbb{R}^d$) then $\varphi_{s,t} = \varphi_t \circ \varphi_s^{-1}$ is a stochastic flow of diffeomorphisms of class C^k. It is clear that the analysis of stochastic flows is reduced to random fields with values in H^k, for the forward flow. Indeed, given a stochastic forward flow $\{\varphi_{s,t}(x) : 0 \leq s \leq t \leq T,\, x \in \mathbb{R}^d\}$, the expression $\varphi_{t,s}(x) = \varphi_{s,t}^{-1}(x)$ shows that there exists a unique stochastic flow $\{\bar{\varphi}_{s,t}(x) : s,t \in [0,T],\, x \in \mathbb{R}^d\}$ such that its restriction to the forward time parameters $0 \leq s \leq t \leq T$ coincides with $\varphi_{s,t}(x)$.

Typical Examples

Two typical examples could be used as prototypes. First, if w is a Wiener process in $\mathbb{R}^d$ with zero mean then

$$\varphi_{s,t} = x + w_{t-s} \text{ if } t \geq s \quad \text{and} \quad \varphi_{s,t} = x - w_{s-t} \text{ if } s \geq t.$$

Since the inverse flow is $\varphi_{s,t}^{-1} = x - w_{t-s}$ when $t \geq s \geq 0$, the flow condition (d) needs to be verified only for the forward flow, namely, $\varphi_{s,r} \circ \varphi_{r,t} = \varphi_{s,t}$, for every $s \leq r \leq t$, i.e., the equality $w_{r-s} + w_{t-r} = w_{t-s}$, which is satisfied from the construction. Indeed, with the notation of Chapter 2, if $\{e_{i,n} : i = 1, 2, \dots, 4^n,\ n \geq 1\}$ is a sequence of independent standard normally distributed random variables, then

$$\varphi_{s,t} = x + \sum_n 2^{-n} \sum_{i=1}^{4^n} e_{i,n} \mathbb{1}_{i2^{-n} \leq t} - \sum_n 2^{-n} \sum_{i=1}^{4^n} e_{i,n} \mathbb{1}_{i2^{-n} \leq s}, \quad t, s \geq 0,$$

is a realization of the stochastic flow for a standard Wiener process, which accepts a continuous version in s and t. In particular,

$$\varphi_{s,t} = x + \sum_n 2^{-n} \sum_{i=1}^{4^n} e_{i,n} \mathbb{1}_{s < i2^{-n} \leq t}, \quad t \geq s \geq 0,$$

becomes the forward flow.

The second example is a compound Poisson process, i.e., begin with a given (non-zero) finite measure $\mathtt{m}$ in $(\mathbb{R}^d_*, \mathcal{B}_*)$, to construct a sequence $\dot{q} = \{(z_n, \tau_n) : n \geq 1\}$ of independent random variables such that each τ_n is exponentially distributed with parameter $\mathtt{m}(\mathbb{R}^d_*)$ and z_n has the distribution law $A \mapsto \mathtt{m}(A)/\mathtt{m}(\mathbb{R}^d_*)$, thus, the random variables $\theta_n = \tau_1 + \cdots + \tau_n$ have $\Gamma(\mathtt{m}(\mathbb{R}^d_*), n)$ distribution. The series $\eta_t = \sum_n \mathbb{1}_{t \geq \theta_n}$ is almost surely a finite sum and defines a Poisson process with parameter $\mathtt{m}(\mathbb{R}^d_*)$, satisfying $\mathbb{E}\{\eta_t\} = t\mathtt{m}(\mathbb{R}^d_*)$ and $\mathbb{E}\{|\eta_t - t\mathtt{m}(\mathbb{R}^d_*)|^2\} = t\mathtt{m}(\mathbb{R}^d_*)$. In short, given a $\mathbb{R}^d$-valued compound Poisson process $\{N_t : t \geq 0\}$ with parameter $\lambda = \mathtt{m}(\mathbb{R}^d_*)$ and $\mathtt{m}/\lambda$, or simply $\mathtt{m}$, i.e., with the following characteristic function

$$\mathbb{E}\{\mathrm{e}^{\mathrm{i}\zeta \cdot N_t}\} = \exp\Big\{t \int_{\mathbb{R}^d_*} \big(\mathrm{e}^{\mathrm{i}\zeta \cdot z} - 1\big)\mathtt{m}(\mathrm{d}z)\Big\}, \qquad \forall \xi \in \mathbb{R}^d,$$

as a Lévy process, with $N_t = \sum_n z_n \mathbb{1}_{t \geq \theta_n}$. With all this, the expression $\varphi_{s,t} = x + N_t - N_s$, or equivalently,

$$\varphi_{s,t} = x + \sum_n z_n \mathbb{1}_{\theta_n \leq t} - \sum_n z_n \mathbb{1}_{\theta_n \leq s}, \quad t, s \geq 0$$

is a realization of the stochastic flow for a compound process, corresponding to Lévy measure $\mathtt{m}$. while

$$\varphi_{s,t} = x + \sum_n z_n \mathbb{1}_{s < \theta_n \leq t}, \quad t \geq s \geq 0,$$

provides the forward flow.

The solution to Ordinary Differential Equations provides typical deterministic examples, i.e., if $g(y,t)$ is a Lipschitz continuous maps from $\mathbb{R}^d$ into itself (of class C^k) then the unique solution $\varphi_{s,t}(x)$ of the initial valued problem

$\dot{y}(t) = g(y(t), t)$, for $t > s$, with $y(s) = x$, is a deterministic homeomorphism (diffeomorphism of class C^k). Indeed, this is easily checked by noting that the ODE can be solved forward and backward in time, i.e., $\varphi_{s,t}^{-1}(x)$ is the unique solution of the IVP $\dot{y}(t) = g(y(t), t)$, for $t < s$, with $y(s) = \varphi_{s,t}(x)$.

Going back to the second example of a compound Poisson process with Lévy measure $\mathtt{m}$ on $\mathbb{R}_*^d$, we may consider a sequence $\{\gamma(x, z, t)\}$ of jumps rules to be superimposed $x \mapsto x + \gamma(x, z, t)$ by induction as follows: $x_0 = x$, $z_0 = 0$, $\theta_0 = 0$, $\gamma(x, 0, t) = 0$, and

$$x_{n+1} = x_n + \gamma(x_n, z_n, \theta_n) \quad \text{and} \quad \varphi_{s,t}(x) = x_n \quad \text{if} \quad \theta_{n-1} \le t < \theta_n,$$

for any $n = 0, 1, 2, \dots$ This generalizes the compound Poisson process example, but for this jump-mechanism to be a homeomorphism we need to require that the mapping $x \mapsto x + \gamma(x, z, t)$ be a homeomorphism in $\mathbb{R}^d$, for each fixed z, t, plus Borel measurable in z and cad-lag in t. Indeed, if $y = h_{z,t}(x) = x + \gamma(x, z, t)$ and $x = h_{z,t}^{-1}(y)$ its inverse then define $\gamma^{-1}(y, z, t) = h_{z,t}^{-1}(y) - y$, i.e.,

$$\gamma^{-1}(x, z, t) = y - x \quad \text{iff} \quad y + \gamma(y, z, t) = x, \tag{4.38}$$

which satisfies $\gamma^{-1}(x, z, t) = -\gamma(y, z, t)$ when $y + \gamma(y, z, t) = x$. The inverse stochastic flow $\varphi_{s,t}^{-1}(x)$ is obtained by (reverse) induction as follows: for t find k such that $\theta_k \le t < \theta_{k+1}$ to define $x_k = \varphi_{s,t}(x)$, $\theta_0 = 0$, and $\varphi_{s,t}^{-1}(x) = x_k$ if $\theta_k \le t < \theta_{k+1}$, and then

$$x_{n-1} = x_n + \gamma^{-1}(x_n, z_n, \theta_n) \quad \text{and} \quad \varphi_{s,t}^{-1}(x) = x_{n-1} \quad \text{if} \quad \theta_{n-1} \le t < \theta_n,$$

for any $n = k, k-1, \dots, 1$. It should be also clear that the sequence $\{(z_n, \theta_n) : n \ge\}$ are the jumps of the piecewise constant process $\varphi_{s,t}(x)$. Also note that for "small jumps", e.g., $|\gamma(x, z, t)| \le \gamma_0 < 1$, the mapping $x \mapsto x + \gamma(x, z, t)$ is necessarily one-to-one.

Moreover, if a sequence $\{g_n(x, t) : n \ge 1\}$ of Lipschitz continuous drifts is given then the previous example can be modified to include the so-called piecewise deterministic process, e.g., see Davis [30, Sections 25 and 26]. Indeed, replace x_n with the unique solution $x_n(t)$ of the IVP

$$\begin{aligned} &\dot{x}_{n+1}(t) = g_n(x_{n+1}(t), t), \quad t > \theta_n, \\ &x_{n+1}(\theta_n) = x_n(\theta_n) + \gamma(x_n(\theta_n), z_n, \theta_n), \end{aligned}$$

and define $\varphi_{s,t}(x) = x_n(t)$ if $\theta_{n-1} \le t < \theta_n$.

Stochastic Differential Equations

It is not the objective of this section to study stochastic ordinary differential equations (SODE), but merely mention them as motor of stochastic flows.

Perhaps the simplest SODE is with constant coefficients, i.e.,

$$\varphi_{s,t}(x) = x + \int_s^t g(r)\mathrm{d}r + \sum_k \int_s^t \sigma_k(r)\mathrm{d}w_k(r-s) + \\ + \int_{\mathbb{R}^m_* \times]s,t]} \gamma(\zeta, r)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}r), \quad \forall\, t \geq s, \quad (4.39)$$

where the standard Wiener process w in $\mathbb{R}^n$ and the Poisson measure p and its corresponding martingale (or compensated) measure $\tilde{p}$ are all given. In this case, the composition-flow (or co-cycle) property is automatically satisfied, and inverse flow is defined by subtracting the integrals. However, for variables coefficients the situation is very different.

Another simple case is for instance the one-dimensional linear SODE, i.e., apply Itô formula to the one-dimensional stochastic integral

$$\varphi_{s,t}(x) = x\exp\Big(\int_s^t \big(g(r) - \frac{1}{2}\sigma^2(r)\big)\mathrm{d}r + \int_s^t \sigma(r)\mathrm{d}w_{r-s}\Big), \quad t \geq s \geq 0,$$

to deduce that

$$\mathrm{d}\varphi_{s,t}(x) = \varphi_{s,t}(x)g(t)\mathrm{d}t + \varphi_{s,t}(x)\sigma(t)\mathrm{d}w_{t-s}, \quad t \geq s \geq 0,$$

which is a linear stochastic ordinary differential equation for $y(t) = \varphi_{s,t}(x)$ with initial condition $y(s) = x$. Also note that the forward flow satisfies

$$\varphi_{s,r}(x) \circ \varphi_{r,t}(x) = x\exp\Big(\int_s^r \big(g(\tau) - \frac{1}{2}\sigma^2(\tau)\big)\mathrm{d}\tau + \int_s^t \sigma(\tau)\mathrm{d}w_{\tau-s}\Big)\times \\ \times \exp\Big(\int_r^t \big(g(\tau) - \frac{1}{2}\sigma^2(\tau)\big)\mathrm{d}\tau + \int_r^t \sigma(\tau)\mathrm{d}w_{\tau-t}\Big) = \\ = x\exp\Big(\int_s^t \big(g(\tau) - \frac{1}{2}\sigma^2(\tau)\big)\mathrm{d}\tau + \int_s^t \sigma(\tau)\mathrm{d}w_{\tau-s}\Big),$$

for any $t \geq r \geq s \geq 0$, as expected. The inverse flow $\varphi_{t,s}(x) = \varphi^{-1}_{s,t}(x)$, for $t \geq s \geq 0$, is given by

$$\varphi_{t,s}(x) = x\exp\Big(-\int_s^t \big(g(r) - \frac{1}{2}\sigma^2(r)\big)\mathrm{d}r - \int_s^t \sigma(r)\mathrm{d}w_{r-s}\Big).$$

Clearly, the multidimensional case is more delicate and requires the use of the fundamental matrix-solution for a linear (deterministic) ODE.

A way of setting-up stochastic ordinary differential equations (SODE) is to begin with a probability space $(\Omega, \mathcal{F}, P)$ with a standard Wiener process $w = (w_1, \ldots, w_n)$ in $\mathbb{R}^n$ and a Poisson measure p with Lévy measure π on $\mathbb{R}^m_*$, independent of each other. This allow us to consider w as a continuous martingale and the compensated Poisson measure $\tilde{p}(\mathrm{d}\zeta, \mathrm{d}t) = p(\mathrm{d}\zeta, \mathrm{d}t) - \pi(\mathrm{d}\zeta)\mathrm{d}t$ as a purely discontinuous martingales, both defined on the filtered probability

space $(\Omega, \mathbb{F}, P)$ generated by w and p. Note that once the Lévy measure π has been chosen, a (canonical) realization of w and p can be constructed to establish this setting. As seen later, it is convenient to assume that all moment of the 'large jumps' are finite, i.e.,

$$\int_{\mathbb{R}^m_*} |\zeta|^q \pi(\mathrm{d}\zeta) < \infty, \quad \forall q \geq 2, \tag{4.40}$$

so that they are incorporated into the stochastic integral. Recall that the integrals against the 'large jumps' are actually pathwise integrals.

Thus, beside the Lévy measure π and the time horizon $T > 0$, the coefficients of the SODE are part of the data, i.e., the drift $g\colon \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$, the diffusion $\sigma = (\sigma_1, \dots, \sigma_n)$ with $\sigma_k\colon \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$, and the jumps $\gamma\colon \mathbb{R}^d \times \mathbb{R}^m_* \times [0,T] \to \mathbb{R}^d$, and all coefficients are at least Borel measurable functions.

Therefore, the stochastic ordinary differential equation takes the form

$$y(t) = x + \int_s^t g(y(r), r)\mathrm{d}r + \sum_k \int_s^t \sigma_k(y(r), r)\mathrm{d}w_k(r) + \\ + \int_{\mathbb{R}^m_* \times]s,t]} \gamma(y(r), \zeta, r)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}r), \quad \forall\, t \geq s, \tag{4.41}$$

or in differential form as

$$\mathrm{d}y(t) = g(y(t), t)\mathrm{d}t + \sum_k \sigma_k(y(t), t)\mathrm{d}w_k(t) + \\ + \int_{\mathbb{R}^m_*} \gamma(y(t), \zeta, t)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}t), \tag{4.42}$$

plus the initial condition $y(s) = x$. Usually $y(t) = y_{xs}(t)$ to emphasize the dependency on the initial condition. A solution to the SODE is an optional process y such that the equality (4.41) holds true. Because the stochastic integrals are defined initially as an element in the space L^2 with the product measure $P \times \mathrm{d}t$, a solution is an adapted measurable process of which an optional cad-lag version is taken, and for the stochastic integrals, a predictable version is obtained by replacing $y(t)$ with $y(t-)$ inside the integrand. In this sense, the uniqueness is modulo $P \times \mathrm{d}t$ for adapted square-integrable processes or modulo an evanescence set for cad-lag processes.

To develop a neat existence and uniqueness theory, the coefficients have linear growth and are locally Lipschitz continuous, namely, there exists a constant $C > 0$ such that

$$|g(x,t)|^2 + \sum_k |\sigma_k(x,t)|^2 + \int_{\mathbb{R}^m_*} |\gamma(x,\zeta,t)|^2 \pi(\mathrm{d}\zeta) \leq C(1 + |x|^2), \tag{4.43}$$

for every (x,t) in $\mathbb{R}^d \times [0,T]$, and for any $r > 0$ there exists a positive constant

$M = M(r)$ such that

$$|g(x,t) - g(x',t)|^2 + \sum_k |\sigma_k(x,t) - \sigma_k(x',t)|^2 + \\ + \int_{\mathbb{R}^m_*} |\gamma(x,\zeta,t) - \gamma(x',\zeta,t)|^2 \pi(\mathrm{d}\zeta) \le M|x-x'|^2, \quad (4.44)$$

for every (x,t), (x',t) in $\mathbb{R}^d \times [0,T]$ with $|x| \le r$ and $|x'| \le r$.

Theorem 4.23 (existence and uniqueness)**.** *Under the above setting, including* (4.40), (4.43), (4.44)*, and for any given s in $[0,T]$ and x in $\mathbb{R}^d$, there exists one and only one solution $(t,\omega) \mapsto y(t,\omega)$ of the d-dimensional stochastic ordinary differential equation* (4.41) *on the time interval $[s,T]$.*

Proof. On an idea of a proof is given. Indeed, first the coefficients are assume globally Lipschitz (i.e., the constant $M = M(r)$ in (4.44) can be chosen independently of any $r > 0$), and a fixed point in found in a convenient space as in the deterministic ODE. For instance, without using martingale inequalities, a unique fixed point is found in the Hilbert space L^2 of all adapted square-integrable processes. Then Doob's maximal estimate is used to get a cad-lag (or continuous, when $\gamma = 0$) optional version. Alternatively, martingales estimates can be used to obtain a fixed point directly in the Banach space of cad-lag square-integrable processes with a $\sup - L^2$ type norm.

Next, approximate the coefficients with globally Lipschitz functions, the existence of a solution is established. Finally, the uniqueness follows from a convenient Gronwall-type inequality.

Furthermore, the initial condition could be stochastic, e.g., the initial time s could be a stopping time and x a $\mathcal{F}(s)$-measurable random variable. □

• *Remark* 4.24. It should be clear that only (4.40) with $q = 2$ is used in the above arguments, and even this is not necessary, the L^2 integrability of γ with respect to π in assumptions (4.43) and (4.44) is all what is required of the Radom measure π on $\mathbb{R}^m_*$. However, if the coefficients are globally Lipschitz (i.e., the constant $M(r)$ in (4.44) can be chosen independent of r) and the jump coefficient γ satisfies for every $q \ge 2$ and $r > 0$ there exists constants $C = C_q$ and M_q such that

$$\int_{\mathbb{R}^m_*} |\gamma(x,\zeta,t)|^q \pi(\mathrm{d}\zeta) \le C_q(1+|x|^q), \quad \forall x,t, \\ \int_{\mathbb{R}^m_*} |\gamma(x,\zeta,t) - \gamma(x',\zeta,t)|^q) \pi(\mathrm{d}\zeta) \le M_q|x-x'|^q, \quad \forall x,x',t, \quad (4.45)$$

then the solution $y(t) = y_{xs}(t)$ belongs to L^q and

$$\mathbb{E}\Big\{ \sup_{s\le t\le T} (1+|y_{xs}(t)|)^q \Big\} \le C'_q(1+|x|)^q, \quad \forall x,s, \\ \mathbb{E}\Big\{ \sup_{s\le t\le T} |y_{xs}(t) - y_{x's}(t)|^q \Big\} \le M'_q|x-x'|^q, \quad \forall x,x',s,$$

for some suitable constants C_q' and M_q' depending also on T. Moreover, the derivative with respect to the initial conditions can be calculated if the coefficient are smooth. Certainly, this requires a carefully treatment which is not described here. □

Homeomorphisms via SODE

Assuming the previous setting, i.e., a filtered probability space $(\Omega, \mathbb{F}, P)$ with a standard Wiener process $w = (w_1, \dots, w_n)$ on $\mathbb{R}^n$, a Poisson measure p with intensity π on $\mathbb{R}^m_*$ and compensated martingale measure $\tilde{p} = p(\mathrm{d}\zeta, \mathrm{d}t) - \pi(\mathrm{d}\zeta)\mathrm{d}t$, and with coefficients g, σ, γ satisfying (4.43) and (4.44), define the random field $\varphi_{s,t}(x) = y_{xs}(t)$, with $0 \le s \le t \le T$, x in $\mathbb{R}^d$, where $y_{xs}(t)$ is the unique cad-lad optional solution of the SODE (4.41).

Assume that $x \mapsto x + \gamma(x, \zeta, t)$ is an homeomorphism in $\mathbb{R}^d$ and if the inverse jump γ^{-1} is given by (4.38) then for every $q \ge 2$ there exists constants $C = C_q$ and M_q such that

$$\begin{aligned} &\int_{\mathbb{R}^m_*} |\gamma^{-1}(x, \zeta, t)|^q \pi(\mathrm{d}\zeta) \le C_q(1 + |x|^q), \quad \forall x, t, \\ &\int_{\mathbb{R}^m_*} |\gamma^{-1}(x, \zeta, t) - \gamma^{-1}(x', \zeta, t)|^q \pi(\mathrm{d}\zeta) \le M_q |x - x'|^q, \quad \forall x, x', t \end{aligned} \tag{4.46}$$

Remark that if the jumps are small, i.e., $|\gamma(x, \zeta, t)| \le c$ for every x, ζ, t and some constant $0 < c < 1$, then the required condition on the inverse jump γ^{-1} is a direct consequence of the that on γ.

Theorem 4.25 (homeomorphism)**.** *Under the assumptions of Theorem 4.23 and (4.45), (4.46), and for any given $0 \le s \le t \le T$, the mapping $x \mapsto y_{xs}(t)$ is an homeomorphism in $\mathbb{R}^d$, almost surely, i.e., $\varphi_{s,t}(x) = y_{xs}(t)$, $0 \le s \le s \le T$ is a forward flow of homeomorphism in $\mathbb{R}^d$.*

Proof. Actually, this is beyond the scope of these lectures, and only the key ideas are presented. Essentially, the key arguments is bases on convenient application of Itô formula.

A first step is to reinforce the assumption on the inverse jumps (4.46) with

$$\begin{aligned} &|\gamma^{-1}(x, \zeta, t)| \le C(1 + |x|), \quad \forall x, \text{ a.e. } P \times \pi \times \mathrm{d}t, \\ &|\gamma^{-1}(x, \zeta, t) - \gamma^{-1}(x', \zeta, t)| \le M|x - x'|, \quad \forall x, x', \text{ a.e. } P \times \pi \times \mathrm{d}t, \end{aligned} \tag{4.47}$$

and to obtain the estimates

$$\begin{aligned} &\big(1 + |x + \gamma(x, \zeta, t)|^2\big)^{-1} \le C\big(1 + |x|^2\big)^{-1}, \quad \forall x, \text{ a.e. } P \times \pi \times \mathrm{d}t, \\ &\int_{\mathbb{R}^m_*} \big(1 + |x + \gamma(x, \zeta, t)|^2\big)^{-q} \pi(\mathrm{d}\zeta) \le C_q\big(1 + |x|^2\big)^{-q}, \quad \forall x, \text{ a.e. } P \times \mathrm{d}t, \end{aligned}$$

and

$$|I|_{\gamma,\pi}(x, t)\big(1 + |\cdot|^2\big) \le C\big(1 + |x|^2\big), \quad \forall x, \text{ a.e. } P \times \mathrm{d}t,$$

for suitable constants C, C_q, any $q \geq 2$, where $I_{\gamma,\pi}(x,t)$ denotes the (purely) integro-differential operator

$$I_{\gamma,\pi}(x,t)\phi = \int_{\mathbb{R}^m_*} \big(\phi(x+\gamma(x,\zeta,t)) - \phi(x) - \gamma(x,\zeta,t)\cdot\nabla\phi(x)\big)\pi(\mathrm{d}\zeta),$$

and $|I|_{\gamma,\pi}(x,t)$ with the absolute value inside the integral (i.e., the first and the last parentheses are replaced by | inside the integral).

In a second step, Itô formula is used to deduce the estimate

$$\mathbb{E}\big\{\sup_{s\leq t\leq T}(1+|y_{xs}(t)|^2)^{-q}\big\} \leq C_q(1+|x|^2)^{-q}, \quad \forall x,s,$$
$$\mathbb{E}\big\{\sup_{s\leq t\leq T}|y_{xs}(t)-y_{x's}(t)|^{-2q}\big\} \leq M_q|x-x'|^{-2q}, \quad \forall x,x',s,$$

for some suitable constants C_q and M_q depending also on T.

Now, to show that the mapping $x \mapsto y_{xs}(t)$ is one-to-one, consider the random field $Y_{st}(x,y) = 1/|y_{xs}(t) - y_{xs}(t)|$, where s is fixed. Apply previous estimate to deduce

$$\mathbb{E}\big\{\sup_{s\leq t\leq T}|Y_{st}(x,y)-Y_{st}(x',y')|^{-2q}\big\} \leq C\delta^{-4q}\big(|x-x'|^{-2q}+|y-y'|^{-2q}\big),$$

holds true, for some constant $C>0$ and for any s, x, y, x', y' with $|x-y|\geq\delta$ and $|x'-y'|\geq\delta$. Hence, by Kolmogorov's criterion, the random field $Y_{st}(x,y)$ is continuous in the domain $D_\delta = \{(x,y) : |x-y|\geq\delta\}$, for every $\delta>0$, which show that $x\mapsto y_{xs}(t)$ is one-to-one from $\mathbb{R}^d$ into itself.

Next, to check that $x\mapsto y_{xs}(t)$ is onto $\mathbb{R}^d$, take $x\neq 0$ and set $\bar{x} = x|x|^{-2}$ to define $\bar{Y}_{st}(\bar{x}) = 1/(1+|y_{xs}(t)|)$ and $\bar{Y}_{st}(0)=0$. Again, combine previous estimates to obtain the estimate

$$\mathbb{E}\big\{\sup_{s\leq t\leq T}|Y_{st}(\bar{x})-Y_{st}(\bar{y})|^{2q}\big\} \leq C_q|\bar{x}-\bar{y}|^{2q},$$

for some constant $C>0$, any $q\geq 1$. Hence, by Kolmogorov's criterion, $\bar{Y}_{st}(\bar{x})$ can be extended continuously as $\bar{x}\to 0$, i.e., $\bar{Y}_{st}(\bar{x})\to 0$ as $\bar{x}\to 0$, and this implies $y_{st}(x)\to\infty$ as $|x|\to\infty$, which establishes the onto property of the maps.

To check the composition-flow (or co-cycle) property, i.e., $\varphi_{s,r}\circ\varphi_{r,t} = \varphi_{s,t}$ it suffices to remark that $\varphi_{s,t}(x)$ and $y_{rx'}(t)$ with $x' = \varphi_{s,r}(x)$ are both solutions of the same SODE, and uniqueness of solution yields the desired property. □

Certainly, more work is necessary to show that $\varphi_{s,t}(x) = y_{xs}(t)$, $0\leq s\leq s\leq T$ is a forward flow of diffeomorphism of class C^k or $C^{k,\alpha}$. Essentially, if the coefficients g, σ_k and γ are of class C^k or $C^{k,\alpha}$ with suitable estimates then solution of the SODE $y_{xs}(t)$ is differentiable with respect to the initial data, which implies the desired property on the stochastic flows.

It perhaps important to mention that to express the backward flow of homeomorphism the SODE should be considered backward, i.e., first the stochastic

integral should be defined for (cad-lag) backward adapted processes, which are the cad-lag version of processes like (4.39) in the variable s. This procedure take care of the continuous part, but also the inverse jumps γ^{-1} intervene in the backward SODE. This was nicely seen in the simple example of discrete jumps.

For instance, the reader is referred to the book Kunita [104] for a carefully treatment of continuous stochastic flows of homeomorphism, where a comprehensive study can be found. If jumps are added then some extra conditions are necessary, like (4.45), (4.46) or even (4.47), the arguments are more delicate, see Kunita [105].

4.4 Convergence of Integral Processes

A crucial point is to find a convergent (in various topologies) subsequence from a given sequence of stochastic processes, e.g., the reader may take a look at Ethier and Kurtz [45, Chapter 3, pp. 95–154]. In the following we collect various sufficient (and necessary in some cases) conditions to this end.

4.4.1 Standard Convergences

An important result related with stochastically (left or right) continuous processes can be found in Skorokhod [161, Section 1.6, pp. 9–14]

Theorem 4.26 (Skorokhod). *Let $X_n = (X_n(t) : t \geq 0)$, $n = 1, 2, \dots$ be a sequence of stochastically continuous processes with values in $\mathbb{R}^d$ in the probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$. Assume that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n, t, s satisfying $0 \leq t \leq 1/\varepsilon$, $0 \leq s \leq 1/\varepsilon$, $|t - s| < \delta$ we have*

$$P_n\{|X_n(t)| \geq 1/\delta\} + P_n\{|X_n(t) - X_n(s)| \geq \varepsilon\} \leq \varepsilon. \tag{4.48}$$

Then there exist a stochastically continuous process $\tilde{X} = (\tilde{X}(t) : t \geq 0)$ and a subsequence, indexed by N, of stochastic processes $\tilde{X}_n = (\tilde{X}_n(t) : t \geq 0)$, n in N, all with values in $\mathbb{R}^d$ and defined in another probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, such that X_n and $\tilde{X}_n$ have the same finite-dimensional distributions for every n in N and

$$\lim_{n \in N} \sup_{0 \leq t \leq 1/\varepsilon} \tilde{P}\{|\tilde{X}_n(t) - \tilde{X}(t)| \geq \varepsilon\} = 0, \tag{4.49}$$

for every $\varepsilon > 0$. □

Certainly, the construction uses the canonical probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, where $\tilde{\Omega} = [0, 1]$ with the Lebesgue measure $\tilde{P}$ on the Borel σ-algebra $\tilde{\mathcal{F}} = \mathcal{B}([0, 1])$. It is clear that each process X_n or $\tilde{X}_n$ may be only left (or right) stochastically continuous and the result remain valid. Moreover, if the processes $\{X_n : n \in N\}$ are continuous or cad-lag then there are continuous or cad-lag version of the processes $\{\tilde{X}_n : n \in N\}$. Indeed, denote by $\tilde{P}_n^*$ the outer measure

on the product space $(\mathbb{R}^d)^{[0,\infty)}$ associated with the process $\tilde{X}_n$, or equivalently to X_n. Since X_n is cad-lag, $\tilde{P}_n^*\{D([0,\infty),\mathbb{R}^d)\} = 1$, and therefore the set

$$\{\tilde{\omega} \,:\, \tilde{X}_n(\cdot,\tilde{\omega}) \notin D([0,\infty),\mathbb{R}^d)\}$$

has $\tilde{P}_n^*$-measure zero. However, the limit process $\tilde{X}$ may not be continuous nor cad-lag, since in (4.48) the sup is outside of the probability.

The Skorokhod Representation Theorem can be generalized to a metric space $(\mathbb{X},\rho)$ we have the following result, where on $[0,1]$ is considered with the usual Borel σ-algebra and the standard Lebesgue measure (sometimes referred to as the universal probability space).

Theorem 4.27. *Given a weak convergent sequence of probability measures on metric space* $\mathbb{X}$, $\mu_n \to \mu_0$, *assume that either* $\mathbb{X}_0$ *is* separable *or each* μ_n, $n = 0,1,\dots\}$ *is* tight. *Then there exist a sequence of random variables* $X_n\colon [0,1] \to \mathbb{X}_0$, $n = 0,1,\dots\}$, *such that* (1) *the image measures of* X_n *are the* μ_n *and* (2) $X_n(\theta) \to X(\theta)$, *for any* θ *in* $[0,1]$. □

Lévy processes are characterized by being stochastically continuous processes with a specific Lévy characteristic function (or Fourier transform), with drift vector b, covariance matrix a and Lévy measure (or jump intensity) π. Hence, if X_n are Lévy processes then so are the processes $\tilde{X}_n$, $\tilde{X}$, after choosing appropriate versions. In particular this applies to Wiener processes and Poisson measures.

Another point of view in this direction is to consider a $\mathbb{R}^d$-valued stochastic process as a probability measure in a canonical space such as $C([0,\infty),\mathbb{R}^d)$ or the space $D([0,\infty),\mathbb{R}^d)$, of continuous or cad-lag functions from $[0,\infty)$ into $\mathbb{R}^d$, which are Polish (i.e., separable, complete and metric) spaces. Thus, a continuous or cad-lag stochastic process is a random variable with values in either $C([0,\infty),\mathbb{R}^d)$ or $D([0,\infty),\mathbb{R}^d)$. The modulus of continuity and its equivalent for cad-lag process can be estimated as follows:

(1) if X is a separable process on $[0,T]$ such that there exist positive constants p, q, C such that

$$\mathbb{E}\big\{|X(t) - X(s)|^p\big\} \le C|t-s|^{1+q}, \quad \forall t,s \in [0,T],$$

then for every $0 < \alpha < q/p$ we have

$$\lim_{\varepsilon\to 0} \varepsilon^{-\alpha}\, \rho_{\scriptscriptstyle C}(\varepsilon,X,T) = 0,$$
$$\rho_{\scriptscriptstyle C}(\varepsilon,X,T) := \sup_{0\le t\le s\le t+\varepsilon\le T} \big\{|X(t) - X(s)|\big\},$$

almost surely.

(2) if X is a separable process on $[0,T]$ such that there exist positive constants p, q, C such that

$$\mathbb{E}\big\{\big[|X(t+\delta) - X(s)| \wedge |X(s) - X(t)|\big]^p\big\} \le C\delta^{1+q}, \quad \forall \delta > 0,$$

for every $0 \le t \le s \le t+\delta \le T$ then for every $0 < \alpha < q/p$ we have

$$\lim_{\varepsilon\to 0} \varepsilon^{-\alpha}\, \rho_{\scriptscriptstyle D}(\varepsilon, X, T) = 0,$$
$$\rho_{\scriptscriptstyle D}(\varepsilon, X, T) := \sup_{0\le t\le s\le t+\varepsilon\le T} \big\{|X(t+\varepsilon) - X(s)| \wedge |X(s) - X(t)|\big\},$$

almost surely.

To check these statements, we consider the diadic numbers $\mathbb{D} = \{k2^{-n} : k = 0, 1, \dots, 2^n,\ n \ge 0\}$ on the time interval $[0, T] = [0, 1]$, for simplicity. For each path, define

$$Z_n := \sum_{k=0}^{2^n-1} Z_n(k), \qquad Z_n(k) := |X((k+1)2^{-n}) - X(k2^{-n})|,$$

to get

$$|X(t) - X(s)| \le Z_n, \quad \forall t, s \in \mathbb{D},\ |t-s| = 2^{-n}.$$

Since $\mathbb{D}$ is a separant subset of X, this shows that

$$\rho_{\scriptscriptstyle C}(2^{-n}, X, T) \le \sup_{m\ge n} Z_m \le \sum_{m\ge n} Z_m.$$

The assumption on the process X in (1) yields

$$P\big\{|X(t) - X(s)| \ge |t-s|^{\alpha}\big\} \le C|t-s|^{1+\beta},$$

for every t, s in $[0, 1]$ and with $\beta := p - q\alpha$. Therefore

$$P\big\{Z_m \ge 2^{-m\alpha}\big\} \le \sum_{k=0}^{2^m-1} P\big\{Z_m(k) \ge 2^{-m\alpha}\big\} \le 2^{-m\beta},$$

and

$$P\Big\{\sum_{m\ge n} Z_m \ge \sum_{m\ge n} 2^{-m\alpha}\Big\} \le \sum_{m\ge n} 2^{-m\beta} = \frac{2^{-n\beta}}{1-2^{-\beta}}.$$

Hence

$$P\Big\{2^{n\alpha}\, \rho_{\scriptscriptstyle C}(2^{-n}, X, T) \ge \frac{1}{1-2^{-\alpha}}\Big\} \le \frac{2^{-n\beta}}{1-2^{-\beta}},$$

and by means of the Borel-Cantelli lemma, we deduce that

$$\limsup_{\varepsilon\to 0} \varepsilon^{-\alpha}\, \rho_{\scriptscriptstyle C}(\varepsilon, X, T) \le \frac{1}{1-2^{-\alpha}},$$

almost surely, i.e, statement (1) for any $0 < \alpha' < \alpha$. To show assertion (2), we may redefine

$$Z_n(k,\ell) := |X((k+1)2^{-n}) - X(k2^{-n})| \wedge \\ \wedge |X(\ell 2^{-n}) - X((\ell-1)2^{-n})|,$$
$$Z_n := \sum_{0<\ell\le k<2^m} Z_n(k,\ell)$$

to get

$$\rho_{_D}(2^{-n}, X, T) \le \sum_{m\ge n} Z_m,$$

and then to conclude similarly as above.

Going back to previous Theorem 4.26, if the processes X_n are cad-lag and the condition (4.48) is replaced by the following assumption: for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n

$$P_n\{w(X_n,\delta,1/\delta) \ge \varepsilon\} + P_n\{\sup_{0\le t\le 1/\varepsilon} |X_n(t)| \ge 1/\delta\} \le \varepsilon, \\ w(X_n,r,T) = \inf_{t_i} \max_i \sup_{t_{i-1}\le s,t<t_i} |X_n(t) - X_n(s)| \tag{4.50}$$

where $0 = t_0 < t_1 < \cdots < t_{n-1} < T \le t_n$, $t_i - t_{i-1} \ge r$, $i = 1,\dots,n$, then the limit $\tilde X$ is a cad-lag process and the sequence of laws $\tilde P_n$ (of X_n or equivalently of $\tilde X_n$) on the canonical space $D([0,\infty),\mathbb{R}^d)$ converge weakly to the law of $\tilde X$. Similarly, if the processes X_n are continuous and the condition (4.48) is replaced by: for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n

$$P_n\{\sup_{0\le t\le 1/\varepsilon} |X_n(t)| \ge 1/\delta\} + P_n\{\sup_{T(\varepsilon,\delta)} |X_n(t) - X_n(s)| \ge \varepsilon\} \le \varepsilon, \tag{4.51}$$

where now $T(\varepsilon,\delta)$ is the subset of t, s satisfying $0 \le t \le 1/\varepsilon - \delta$, $0 \le s \le 1/\varepsilon$, $|t-s| \le \delta$, then the limit $\tilde X$ is a continuous process and the sequence of law $\tilde P_n$ on the canonical space $C([0,\infty),\mathbb{R}^d)$ converges weakly to the law of $\tilde X$.

Sometime the above criteria (of tightness) of a sequence X_n could be not usable or hard to meet, specially the condition relative the uniform sup-bound on the increments in either (4.50) or (4.51). The so-called Aldous' criterion for tightness is a suitable tool. This reads as follows

Theorem 4.28. *Let $X_n = (X_n(t) : t \ge 0)$, $n = 1,2,\dots$ be a sequence of adapted cad-lag processes with values in $\mathbb{R}^d$ in the filtered probability spaces $(\Omega_n, \mathcal{F}_n, P_n, \mathcal{F}_n(t) : t \ge 0)$. Assume that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n and stopping times τ, θ satisfying $\theta \le \tau \le 1/\varepsilon$, $\tau - \theta \le \delta$, we have*

$$P_n\{\sup_{0\le t\le 1/\varepsilon} |X_n(t)| \ge 1/\delta\} + P_n\{|X_n(\tau) - X_n(\theta)| \ge \varepsilon\} \le \varepsilon. \tag{4.52}$$

Denote by $\tilde{P}_n$ the probability law of the process X_n in the canonical space $D([0,\infty),\mathbb{R}^d)$ of cad-lag functions. Then there exist a probability measure $\tilde{P}$ in $D([0,\infty),\mathbb{R}^d)$ and a subsequence, indexed by N, of $\{\tilde{P}_n : n \geq 1\}$ such that

$$\lim_{n\in N} \tilde{P}_n(F) \leq \tilde{P}(F), \quad \forall \text{ closed } F \in D([0,\infty),\mathbb{R}^d), \tag{4.53}$$

and we also have $\tilde{P}_n(h) \to \tilde{P}(h)$, for every bounded h which is $\tilde{P}$-almost surely continuous, i.e., $\tilde{P}_n$ converge weakly to $\tilde{P}$. Moreover, in some probability space $(\Omega,\mathcal{F},P)$ there are random variables $\tilde{X}_n$ and $\tilde{X}$ with values in $D([0,\infty),\mathbb{R}^d)$ and distributions $\tilde{P}_n$ and $\tilde{P}$, respectively, such that $\tilde{X}_n$ converges in probability $\tilde{X}$. Furthermore, if we assume that for every $\varepsilon > 0$ there exists an index n_ε such that

$$P_n\{\sup_{0\leq t\leq 1/\varepsilon} |X_n(t) - X_n(t-)| \geq \varepsilon\} \leq \varepsilon \quad \forall n \geq n_\varepsilon \tag{4.54}$$

then the limiting probability measure $\tilde{P}$ satisfies $\tilde{P}\big(C([0,\infty),\mathbb{R}^d)\big) = 1$, i.e., $\tilde{P}$ defines a probability measure on the canonical space $C([0,\infty),\mathbb{R}^d)$ of continuous functions. □

It is clear that the statement regarding the $D([0,\infty),\mathbb{R}^d)$-valued random variables comes from Skorokhod theorem. Recall that, if $\rho_D(\cdot,\cdot)$ denotes the metric in the Polish space $D([0,\infty),\mathbb{R}^d)$, then $\tilde{X}_n$ converges in probability $\tilde{X}$ if and only if for every $\varepsilon > 0$ we have

$$\lim_{n\in N} \tilde{P}\{\rho_D(\tilde{X}_n,\tilde{X}) \geq \varepsilon\} = 0,$$

in particular

$$\lim_{n\in N} \tilde{P}\{\sup_{T(\varepsilon)} |\tilde{X}_n(t+\varepsilon) - \tilde{X}_n(s)| \wedge |\tilde{X}_n(s) - \tilde{X}_n(t)| \geq \varepsilon\} = 0,$$

where $T(\varepsilon)$ is the subset of t, s satisfying $0 \leq s,t \leq 1/\varepsilon$, $0 \leq t \leq s \leq t+\varepsilon$.

Note that the filtration $\{\mathcal{F}_n(t) : t \geq 0\}$ is always right-continuous (in this case, not necessarily completed). It is customary to identify a cad-lag process X_n defined on the probability spaces $(\Omega_n,\mathcal{F}_n,P_n)$ with its probability law $\tilde{P}_n$ on $D([0,\infty),\mathbb{R}^d)$. Elements in the canonical space $D([0,\infty),\mathbb{R}^d)$ are denoted by ω and the canonical process $\omega \mapsto \omega(t)$, which is interpreted as the projection from $D([0,\infty),\mathbb{R}^d)$ into $\mathbb{R}^d$ or as the identity mapping from $D([0,\infty),\mathbb{R}^d)$ into itself is denoted by $x : (t,\omega) \mapsto \omega(t)$ or $x_t = x_t(\omega) = \omega(t)$ or $x(t) = x(t,\omega) = \omega(t)$ as long as no confusion may arrive. Recalling that $\omega_n \to \omega$ in the Skorokhod topology if and only if there exists a sequence λ_n of continuous and strictly increasing functions with $\lambda_n(0) = 0$ and $\lambda_n(\infty) = \infty$ such that

$$\sup_s |\lambda_n(s) - s| \to 0, \qquad \sup_{s\leq T} |\omega_n(\lambda_n(s)) - \omega(s)| \to 0,$$

for every $T > 0$, we can prove that, besides the projection function x, the functions

$$\omega \mapsto \sup_{t\leq T} |\omega(t)|, \qquad \omega \mapsto \sup_{t\leq T} |\omega(t) - \omega(t-)|$$

are continuous from $D([0,\infty),\mathbb{R}^d)$ into $\mathbb{R}$ for any ω such that $\omega(T)=\omega(T-)$. Moreover,

$$\omega \mapsto \sum_{t\le T} h\big(\omega(t)-\omega(t-)\big),$$

with h a continuous function vanishing near zero, is also continuous. It is clear that if $\tilde{P}$ is quasi-left continuous then $\tilde{P}\{\omega(T)=\omega(T-)\}=0$ for every $T>0$, and the above functionals are $\tilde{P}$-almost surely continuous.

Also on $D([0,\infty),\mathbb{R}^d)$, there is another canonical (integer random measure) process $\omega\mapsto\nu(\mathrm{d}z,\mathrm{d}t,\omega)$, defined as

$$\nu(B,]a,b],\omega) := \sum_{a<t\le b} \mathbb{1}_{\{\omega(t)-\omega(t-)\in B\}}, \quad \forall B\in\mathcal{B}(\mathbb{R}^d_*),\ b>a\ge 0,$$

which is interpreted as the counting jumps measure. Once a probability P is given so that the canonical process x is a local martingale, then its continuous martingale part x^c, the predictable jump compensator measure ν^p and the local martingale measure $\tilde{\nu}:=\nu-\nu^p$ are defined.

Let $h(t,x,v)$ be a real valued Borel measurable function which is bounded and locally uniform continuous in x. For every $T>0$ consider the expression

$$\omega\mapsto\int_0^T h(t,x(t,\omega))\mathrm{d}t,$$

which is a continuous and bounded function from $D([0,\infty),\mathbb{R}^d)$ into $\mathbb{R}$. Then, with the notation of the previous Theorem 4.28 we have

$$\mathbb{E}_n\Big\{\int_0^T h(t,X_n(t))\mathrm{d}t\Big\}=\tilde{\mathbb{E}}_n\Big\{\int_0^T h(t,x(t))\mathrm{d}t\Big\}=\mathbb{E}\Big\{\int_0^T h(t,\tilde{X}_n(t))\mathrm{d}t\Big\},$$

Hence, either if $\tilde{P}_n$ is weakly convergent to $\tilde{P}$ or if $\tilde{X}_n$ converge in probability to $\tilde{X}$ we deduce that the above expression converges to

$$\tilde{\mathbb{E}}\Big\{\int_0^T h(t,x(t))\mathrm{d}t\Big\}=\mathbb{E}\Big\{\int_0^T h(t,\tilde{X}(t))\mathrm{d}t\Big\},$$

where $\mathbb{E}_n$ and $\mathbb{E}$ are the mathematical expectation in the probability spaces $(\Omega_n,\mathcal{F}_n,P_n)$ and $(\Omega,\mathcal{F},P)$, respectively, and $\tilde{\mathbb{E}}_n$ and $\tilde{\mathbb{E}}$ are the integral with respect to the probability laws $\tilde{P}_n$ and $\tilde{P}$, respectively. Moreover, the convergence holds true if we have a sequence $\{h_n(t,x)\}$ of measurable functions, which are equi-bounded in (t,x) and equi-locally uniform continuous in x, and pointwise convergent to some function $h(t,x)$.

There is a key class of discontinuous functions in $D([0,\infty),\mathbb{R}^d)$, namely, the so-called *counting* functions of the form

$$\sum_{t_i\le t}\mathbb{1}_{t_i\le s},\quad \forall t\ge 0,$$

for some strictly increasing sequence $0 < t_i < t_{i+1}$, $t_i \to \infty$. Recall that a *point process* is a cad-lag process with counting functions as sample paths. The following result is sometime useful

Theorem 4.29. *Let $X_n = (X_n(t) : t \geq 0)$, $n = 1, 2, \ldots$ be a sequence of* increasing *cad-lag processes with values in $\mathbb{R}^d$ in the probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$. Suppose that the distributions of $\big(X_n(t_1), \ldots, X_n(t_m)\big)$ in $(\mathbb{R}^d)^m$ converges to $\big(X(t_1), \ldots, X(t_m)\big)$ for every $t_1, \ldots, t_m$ in some dense set of $[0, \infty)$, where $X = (X(t) : t \geq 0)$ is an increasing cad-lag process with values in $\mathbb{R}^d$. If either X is continuous or all X_n and X are point processes then the law of X_n converges weakly to the probability law of X in the canonical space $D([0, \infty), \mathbb{R}^d)$.* □

The reader is referred to Proposition VI.3.26 and Theorems VI.3.37, VI.4.5 in the book Jacod and Shiryaev [84, Chapter VI, pp. 312–322].

Again, with the notation of Theorem 4.28, if the canonical process x is a local martingale relative to $\tilde{P}_n$ then its continuous part x_n^c and its local martingale measure $\tilde{\nu}_n$ are defined and the expressions

$$\omega \mapsto \int_0^T h(t, x(t)) \mathrm{d}x_n^c(t) \quad \text{and} \quad \omega \mapsto \int_{]0,T]\times\mathbb{R}_*^d} \tilde{h}(t, x(t), z)\tilde{\nu}_n(\mathrm{d}z, \mathrm{d}t)$$

are $\tilde{P}_n$-almost surely continuous, as long as $\tilde{h}(t, x, z)$ is locally uniform continuous in x and uniformly integrable in z with respect to ν_n^p, the compensator of ν under $\tilde{P}_n$. However, to pass to the limit we will need the $\tilde{P}$-almost surely continuity. If X_n is a specific Lévy process then its characteristic function (or Fourier transform) is determined by the drift vector b, the covariance matrix a and the Lévy measure (or jump intensity) π (all independent of n or conveniently convergent as n tends to infinite). Hence, $x_n^c = X_n^c$ is its continuous local martingale part with predictable quadratic variation at and $\tilde{\nu}_n$ its local martingale measure with predictable jump compensator (Lévy measure) π. The limiting probability law $\tilde{P}$ has the same properties, so that the mathematical expectations

$$\tilde{\mathbb{E}}_n\Big\{ \int_0^T h(t, x(t)) \mathrm{d}x_n^c(t) \Big\} \quad \text{and} \quad \tilde{\mathbb{E}}_n\Big\{ \int_{]0,T]\times\mathbb{R}_*^d} \tilde{h}(t, x(t), z)\tilde{\nu}_n(\mathrm{d}z, \mathrm{d}t) \Big\}$$

converge to

$$\tilde{\mathbb{E}}\Big\{ \int_0^T h(t, x(t)) \mathrm{d}x^c(t) \Big\} \quad \text{and} \quad \tilde{\mathbb{E}}\Big\{ \int_{]0,T]\times\mathbb{R}_*^d} \tilde{h}(t, x(t), z)\tilde{\nu}(\mathrm{d}z, \mathrm{d}t) \Big\}.$$

Moreover, the processes $\tilde{X}_n$ and $\tilde{X}$ on the probability space $(\Omega, \mathcal{F}, P)$ are Lévy processes with the same characteristic function. Since the distributions of $\tilde{X}_n$ and $\tilde{X}$ coincide with those of the canonical process x under $\tilde{P}_n$ and $\tilde{P}$, respectively, and the stochastic integrals are limits (in L^2) of finite sums, we deduce that the previous stochastic integrals can be considered in the probability space

$(\Omega, \mathcal{F}, P)$. Thus

$$\mathbb{E}\Big\{ \int_0^T h(t, \tilde{X}_n(t)) \mathrm{d}\tilde{X}_n^c(t) \Big\} \quad \text{and} \quad \mathbb{E}\Big\{ \int_{]0,T]\times\mathbb{R}_*^d} \tilde{h}(t, \tilde{X}_n(t), z) \tilde{\nu}_{\tilde{X}_n}(\mathrm{d}z, \mathrm{d}t) \Big\}$$

converge to

$$\mathbb{E}\Big\{ \int_0^T h(t, \tilde{X}(t)) \mathrm{d}\tilde{X}^c(t) \Big\} \quad \text{and} \quad \mathbb{E}\Big\{ \int_{]0,T]\times\mathbb{R}_*^d} \tilde{h}(t, \tilde{X}(t), z) \tilde{\nu}_{\tilde{X}}(\mathrm{d}z, \mathrm{d}t) \Big\},$$

under the previous conditions, i.e., $\tilde{X}_n$ is a Lévy process with $\tilde{P}_n$ (its probability law) that converges weakly to $\tilde{P}$, the probability law of $\tilde{X}$. More delicate arguments apply if X_n are local martingale with characteristics determined as continuous predictable functionals on the paths of X_n, see Jacod and Shiryaev [84, Chapter VII, pp. 348–387].

However, because the processes $\tilde{X}_n$ converge in probability to $\tilde{X}$, we can establish the above convergence independently. We rephrase the result as follows:

Theorem 4.30. *Let f_n, g_n and w_n, ν_n, $n = 1, 2, \dots$ be sequences of real-valued predictable processes in $[0, \infty)$ and $[0, \infty) \times \mathbb{R}_*^m$, d-dimensional Wiener processes and Poisson measures with Lévy measure π on $\mathbb{R}_*^m$, all defined in a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$. Suppose that for some processes f, g, w and ν we have*

$$\int_0^T |f_n(t) - f(t)|^2 \mathrm{d}t \to 0 \quad \textit{and}$$
$$\int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} |g_n(z,t) - g(z,t)|^2 \pi(\mathrm{d}z) \to 0$$

and

$$w_n(t) \to w(t), \qquad \nu_n(K\times]0,t]) \to \nu(K\times]0,t]),$$

in probability, for every t in $[0, T]$ and any compact subset K of $\mathbb{R}_*^m$, where it is implicitly assumed that

$$\int_0^T \big[|f_n(t)|^2 + |f(t)|^2\big] \mathrm{d}t < \infty,$$
$$\int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} \big[|g_n(z,t)|^2 + |g(z,t)|^2\big] \pi(\mathrm{d}z) < \infty,$$

almost surely. Then the stochastic integrals

$$\int_0^T f_n(t) \mathrm{d}w_n(t) \to \int_0^T f(t) \mathrm{d}w(t),$$
$$\int_{\mathbb{R}_*^m\times]0,T]} g_n(z,t) \tilde{\nu}_n(\mathrm{d}z, \mathrm{d}t) \to \int_{\mathbb{R}_*^m\times]0,T]} g(z,t) \tilde{\nu}(\mathrm{d}z, \mathrm{d}t)$$

in probability, where $\tilde{\nu}_n := \nu_n - \pi \mathrm{d}t$ and $\tilde{\nu} := \nu - \pi \mathrm{d}t$ are the Poisson (local) martingale measures associated with Poisson measures ν_n and ν.

Proof. We follows the arguments in Skorokhod [161, Section 2.3, pp. 29–34]. First, recall that elementary predictable processes have the form $h(t,\omega) = h_{i-1}(\omega)$ if $t_{i-1} < t \le t_i$ with some $i = 1,\dots,n$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers and h_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $h(t,\omega) = 0$ otherwise, or $h(z,t,\omega) = h_{i-1,j}(\omega)$ if $t_{i-1} < t \le t_i$ and z belongs to K_j with some $i = 1,\dots,n$, and $j = 1,\dots,m$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers, K_j are disjoint compact subsets of $\mathbb{R}_*^m$ and $h_{i-1,j}$ is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $h(t,\omega) = 0$ otherwise. Then, we find sequences of elementary predictable processes $f_{n,k}$, $g_{n,k}$, f_k and g_k, such that

$$\int_0^T |f_{n,k}(t) - f_n(t)|^2 \mathrm{d}t \to 0, \quad \int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} |g_{n,k}(z,t) - g_n(z,t)|^2 \pi(\mathrm{d}z) \to 0$$
$$\int_0^T |f_k(t) - f(t)|^2 \mathrm{d}t \to 0 \quad \text{and} \quad \int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} |g_k(z,t) - g(z,t)|^2 \pi(\mathrm{d}z) \to 0$$

in probability as $k \to \infty$, for every n. It is clear that

$$\int_0^T f_{n,k}(t)\mathrm{d}w_n(t) \to \int_0^T f_k(t)\mathrm{d}w(t),$$
$$\int_{\mathbb{R}_*^m \times]0,T]} g_{n,k}(z,t)\tilde{\nu}_n(\mathrm{d}z,\mathrm{d}t) \to \int_{\mathbb{R}_*^m \times]0,T]} g_k(z,t)\tilde{\nu}(\mathrm{d}z,\mathrm{d}t)$$

in probability for each k. Now, based on the inequalities

$$P\Big\{ \sup_{0\le t\le T} \Big| \int_0^t h(s)\mathrm{d}w(s)\Big| \ge \varepsilon \Big\} \le \frac{\delta}{\varepsilon^2} + P\Big\{ \int_0^T |h(s)|^2 \mathrm{d}s \ge \delta \Big\},$$

and

$$P\Big\{ \sup_{0\le t\le T} \Big| \int_{\mathbb{R}_*^m \times (0,t]} h(z,s)\, \tilde{\nu}(\mathrm{d}z,\mathrm{d}s)\Big| \ge \varepsilon \Big\} \le \frac{\delta}{\varepsilon^2} + \\ + P\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}_*^m} |h(z,s)|^2 \pi(\mathrm{d}z) \ge \delta \Big\},$$

valid for every positive constant T, δ and ε, we deduce that

$$\int_0^T f_{n,k}(t)\mathrm{d}w_n(t) \to \int_0^T f_n(t)\mathrm{d}w_n(t),$$
$$\int_{\mathbb{R}_*^m \times]0,T]} g_{n,k}(z,t)\tilde{\nu}_n(\mathrm{d}z,\mathrm{d}t) \to \int_{\mathbb{R}_*^m \times]0,T]} g_n(z,t)\tilde{\nu}_n(\mathrm{d}z,\mathrm{d}t)$$

in probability as $k \to \infty$, uniformly in n, which complete the proof. □

Notice that in the context of the previous Theorem 4.30 the conditions $\nu_n \to \nu$ and $\tilde{\nu}_n \to \tilde{\nu}$ are equivalents. On the other hand, if $w_n(t)$ and $\nu_n(K\times]0,t])$

converge in probability uniformly for t in $[0,T]$ then the same is true for the stochastic integrals.

Let w_n and ν_n, $n = 1, 2, \ldots$ be sequences of d-dimensional (standard) Wiener processes and Poisson measures with Lévy measure π on $\mathbb{R}^m_*$, all defined in a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_n(t) : t \geq 0)$, where $\mathcal{F}_n(t)$ is the σ-algebra generated by $\{x_n(s) : s \leq t\}$, where x_n is a cad-lag process for each n. It is clear that w_n and $\tilde{\nu}_n(\mathrm{d}z, \mathrm{d}t) := \nu(\mathrm{d}z, \mathrm{d}t) - \pi(\mathrm{d}z)\,\mathrm{d}t$ are martingale and martingale measures relative to $(\mathcal{F}_n(t) : t \geq 0)$ and $(\mathcal{F}_n(t+) : t \geq 0)$. If $w_n(t)$, $\nu_n(K,]0,t])$ and $x_n(t)$ converge in probability to $w(t)$, $\nu(K,]0,t])$ and $x(t)$, for every $t \geq 0$ and any compact subset K of $\mathbb{R}^m_*$, then w and ν are too, a d-dimensional (standard) Wiener process and a Poisson measure with Lévy measure π on $\mathbb{R}^m_*$, but with respect to the limiting filtration either $(\mathcal{F}(t) : t \geq 0)$ or $(\mathcal{F}(t+) : t \geq 0)$, where $\mathcal{F}(t)$ is the σ-algebra generated by $\{x(s) : s \leq t\}$. The above remark can be generalized with a random change of time, i.e., if $\ell(t)$ is a continuous nondecreasing (adapted) process and such that $\ell(0) = 0$ and $\ell(t)$ is a stopping time relative to each filtration $(\mathcal{F}_n(t) : t \geq 0)$ the we can change t into $\ell(t)$ for each process w_n, ν_n and x_n. This means that if w_n a square integrable martingale with ℓ as its predictable quadratic covariation (just one dimension to simplify notation) and ν_n has a jump compensator ν_n^p given by

$$\nu_n^p(B,]a,b]) = \pi(B)\,[\ell(b) - \ell(a)], \quad \forall B \in \mathcal{B}(\mathbb{R}^m_*),\ b > a \geq 0,$$

relative to $(\mathcal{F}_n(t) : t \geq 0)$, then the same is valid for the limiting process w, ν and x. Therefore, the previous Theorem 4.30 can be modified for this case, replacing $\mathrm{d}t$ with $\ell(t)$. However, if $\mathrm{d}\ell_n$ changes with n then the situation requires more details.

Recall the locally uniform and the Skorokhod's topologies given by the family of functions $\rho(\omega, \delta,]a,b])$ and $w(\omega, \delta,]a,b])$, which are defined for ω in the space of cad-lag functions $D([0,\infty), \mathbb{R}^d)$, by the expressions

$$\begin{aligned}
\rho(\omega, \delta,]a,b]) &:= \sup\{|\omega(t) - \omega(s)| \ :\ a < s,\, t \leq b,\ |t-s| \leq \delta\},\\
w(\omega, \delta,]a,b]) &:= \inf_{\{t_i\}} \sup_i \sup\{|\omega(t) - \omega(s)| \ :\ t_{i-1} \leq s < t < t_i\},
\end{aligned}$$

where $\{t_i\}$ ranges over all partitions of the form $a = t_0 < t_1 < \cdots < t_{n-1} < b \leq t_n$, with $t_i - t_{i-1} \geq \delta$ and $n \geq 1$. Both family of functions $\rho(\omega, \delta,]a,b])$ and $w(\omega, \delta,]a,b])$, define the same topology on the space of continuous functions $C([0,\infty), \mathbb{R}^d)$. It is clear ρ and w is the oscillation (or variation) for continuous and cad-lag functions.

If $\ell(t)$ is a nondecreasing element in $C([0,\infty), \mathbb{R}^d)$ and $f(t)$ is another element in $D([0,\infty), \mathbb{R}^d)$, then the Riemann-Stieltjes integral

$$\int_0^T f(t)\mathrm{d}\ell(t), \quad \forall T \geq 0,$$

is defined as the limit of the Riemann sums

$$R(f,\ell,\pi,[0,T[) := \sum_{i=1}^{n} f(t_i^*)[\ell(t_i)-\ell(t_{i-1})],$$

$$\varpi = \{t_i,\ t_i^*\},\quad 0=t_0<t_1<\cdots<t_n=T,\quad t_{i-1}\le t_i^*<t_i,$$

when the mesh of the partition $|\varpi| := \max_i\{t_i - t_{i-1}\}$ vanishes. Moreover, if we define

$$f^{\varpi}(t) := f(t_i^*)\quad \text{if}\quad t_{i-1}<t\le t_i,$$

then

$$\lim_{|\varpi|\to 0}\int_0^T |f(t)-f^{\varpi}(t)|\,\mathrm{d}\ell(t)=0,$$

for every $T>0$. Indeed, since $f(t)-f^{\varpi}(t)=f(t)-f(t_i^*)$ for some t in $]t_{i-1},t_i]$ and t_i^* in $[t_{i-1},t_i[$, we see that for every $\varepsilon>0$, $i\ge 1$ and any t in $]t_{i-1},t_i]$,

$$|f(t)-f^{\varpi}(t)| \le w(f,|\varpi|,]0,T]) + \varepsilon + \\ + \max_{t_{i+1}<s\le t_i} \mathbb{1}_{\{|f(s)-f(s-)|>\varepsilon\}}\,|f(s)-f(s-)|,$$

i.e., the variation (or oscillation) is bounded by its continuous variation, plus ε, plus the maximum jumps bigger than ε. Hence

$$\int_0^T |f(t)-f^{\varpi}(t)|\,\mathrm{d}\ell(t) = \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} |f(t)-f_{\varpi}(t)|\,\mathrm{d}\ell(t) \le \\ \le \sum_{i=1}^{n}\Big(\sup_{t_{i-1}<t\le t_i}|f(t)-f_{\varpi}(t)|\Big)[\ell(t_i)-\ell(t_{i-1})]$$

which yields

$$\int_0^T |f(t)-f^{\varpi}(t)|\,\mathrm{d}\ell(t) \le \big[w(f,|\varpi|,]0,T]) + \varepsilon\big]\big[\ell(T)-\ell(0)\big] + \\ + \rho(\ell,|\varpi|,]0,T])\sum_{0<s\le T}\mathbb{1}_{\{|f(s)-f(s-)|>\varepsilon\}}\,|f(s)-f(s-)|.$$

From the definition of the cad-lag modulus of continuity w we have

$$\sum_{a<t\le b}\mathbb{1}_{\{|\omega(t)-\omega(t-)|\ge w(\omega,\delta,]a,b])\}} \le \frac{b-a}{\delta},$$

for every ω, $\delta>0$, and $b>a\ge 0$. Therefore, for $\varepsilon = w(f,\delta,]0,T])$ we obtain

$$\int_0^T |f(t)-f^{\varpi}(t)|\,\mathrm{d}\ell(t) \le \frac{T}{\delta}\,\rho(\ell,|\varpi|,]0,T]) + \\ + \big[w(f,|\varpi|,]0,T]) + w(f,\delta,]0,T])\big]\big[\ell(T)-\ell(0)\big]. \quad (4.55)$$

Actually, this estimate implies the following result.

Lemma 4.31. *Let $\{f_n\}$ be a family of cad-lag processes and $\{\ell_n\}$ be another family of continuous and nondecreasing processes, defined in a probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Assume that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n*

$$P_n\{w(f_n, \delta,]0, 1/\varepsilon]) \geq \varepsilon\} + P_n\{\sup_{0\leq t\leq 1/\varepsilon} |f_n(t)| \geq 1/\delta\} \leq \varepsilon$$

and

$$P_n\{|\ell_n(1/\varepsilon) - \ell_n(0)| > 1/\delta\} + P_n\{\rho(\ell_n, \delta,]0, 1/\varepsilon]) \geq \varepsilon\} \leq \varepsilon.$$

Now, for any partition $\varpi = \{t_i,\ t_i^*\}$, $0 = t_0 < t_1 < \cdots < t_n = T$, $t_{i-1} \leq t_i^* < t_i$, define $f_n^\varpi(t) := f_n(t_i^*)$ if $t_{i-1} < t \leq t_i$ as above. Then for every $\varepsilon > 0$ there is $\delta > 0$ such that for every n

$$P_n\Big\{\int_0^T |f_n(t) - f_n^\varpi(t)|\, \mathrm{d}\ell_n(t) \geq \varepsilon\Big\} \leq \varepsilon,$$

for every ϖ with $|\varpi| \leq \delta$.

Proof. Notice that the assumptions means that $\{f_n\}$ is tight (or pre-compact) in the space $D([0,\infty), \mathbb{R}^d)$ and $\{\ell_n\}$ is tight in $C([0,\infty), \mathbb{R}^d)$. The conclusion is the uniform convergence in probability of the integral processes, which is a direct consequence of the a priori estimate (4.55). □

If we are looking at processes $g_n(z,t)$ instead of just $f_n(t)$, with $t \geq 0$ and z in $\mathbb{R}_*^d$, we may consider g_n as having values in the function space $L^2_{\pi_n}(\mathbb{R}_*^m)$, i.e., we use the following definition of the cad-lag modulo

$$w^{\pi_n}(x, \delta,]a, b]) := \\ := \inf_{\{t_i\}} \sup_i \sup_{t_{i-1}\leq s<t<t_i} \Big\{\Big(\int_{\mathbb{R}_*^m} |x(z,t) - x(z,s)|^2 \pi_n(\mathrm{d}z)\Big)^{1/2}\Big\},$$

where $\{t_i\}$ ranges over all partitions of the form $a = t_0 < t_1 < \cdots < t_{n-1} < b \leq t_n$, with $t_i - t_{i-1} \geq \delta$ and $n \geq 1$. Estimate (4.55) becomes

$$\int_0^T \Big(\int_{R_*^d} |g(z,t) - g^\varpi(z,t)|^2\, \pi_n(\mathrm{d}z)\Big)^{1/2} \mathrm{d}\ell(t) \leq \\ \leq \frac{T}{\delta}\, \rho(\ell, |\varpi|,]0, T]) + \big[w^{\pi_n}(g, |\varpi|,]0, T]) + \\ + w^{\pi_n}(g, \delta,]0, T])\big]\big[\ell(T) - \ell(0)\big], \quad (4.56)$$

and the previous Lemma 4.31 remain valid under the assumption that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n

$$P_n\{w^{\pi_n}(g_n, \delta,]0, 1/\varepsilon]) \geq \varepsilon\} + P_n\{\sup_{0\leq t\leq 1/\varepsilon} |g_n(\cdot, t)|_{\pi_n} \geq 1/\delta\} \leq \varepsilon,$$

where

$$|g_n(\cdot,t)|_{\pi_n} := \Big(\int_{\mathbb{R}^m_*} |g_n(z,t)|^2 \pi_n(\mathrm{d}z) \Big)^{1/2}.$$

The a priori estimate obtained is written as for every $\varepsilon > 0$ there is $\delta > 0$ such that for every n

$$P_n \Big\{ \int_0^T \mathrm{d}\ell_n(t) \int_{\mathbb{R}^m_*} |g_n(z,t) - g_n^{\varpi}(z,t)|^2 \, \pi_n(\mathrm{d}z) \geq \varepsilon \Big\} \leq \varepsilon, \tag{4.57}$$

for every ϖ with $|\varpi| \leq \delta$.

Now based on above Lemma 4.31 we are able to generalize Theorem 4.30 as follows

Theorem 4.32. *Let ℓ_n, w_n, ν_n, and x_n, $n = 1, 2, \ldots$ be sequences of processes defined in a probability space $(\Omega, \mathcal{F}, P)$ and let $(\mathcal{F}_n(t) : t \geq 0)$ be the filtration generated by x_n. Assume that (1) ℓ_n are continuous nondecreasing adapted processes, (2) w_n are a d-dimensional square integrable martingales with predictable quadratic covariation $\langle w_{n,i}, w_{n,j} \rangle = \ell_n$ if $i = j$ and $\langle w_{n,i}, w_{n,j} \rangle = 0$ if $i \neq j$. (3) ν_n are integer measures with jump compensator $\nu_n^p(\mathrm{d}z, \mathrm{d}t) = \pi(\mathrm{d}z)\, \mathrm{d}\ell_n(t)$, where π is a given Lévy measure in $\mathbb{R}^m_*$. Suppose that ℓ_n converges to ℓ, i.e., for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that*

$$P\{ \sup_{0 \leq t \leq 1/\varepsilon} |\ell_n(t) - \ell(t)| \geq \varepsilon \} \leq \varepsilon, \quad \forall n \geq N(\varepsilon),$$

$\ell_n(0) = 0$ and that $x_n(t) \to x(t)$, $w_n(t) \to w(t)$ and $\nu_n(K \times]0,t]) \to \nu(K \times]0,t])$ in probability, for every $t \geq 0$ and any compact subset K of $\mathbb{R}^m_*$. Then (a) w is also a square integrable martingale with predictable quadratic covariation $\langle w_i, w_j \rangle = \ell$ if $i = j$ and $\langle w_i, w_j \rangle = 0$ if $i \neq j$, (b) ν is also an integer measure with jump compensator $\nu^p(\mathrm{d}z, \mathrm{d}t) = \pi(\mathrm{d}z)\, \mathrm{d}\ell(t)$, both relative to the limiting filtration $(\mathcal{F}(t) : t \geq 0)$ generated by x. Furthermore, if f_n and g_n are cad-lag adapted processes pointwise (on a dense set of time) convergent to f and g in probability and for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ satisfying

$$P\{ w(f_n, \delta,]0, 1/\varepsilon]) + w^{\pi}(g_n, \delta,]0, 1/\varepsilon]) \geq \varepsilon \} \leq \varepsilon, \quad \forall n \geq 1,$$

the limiting processes f and g are certainly cad-lag, and there exist sequences of partitions $\{\varpi_k = \varpi_k^f : k \geq 1\}$ and $\{\varpi_k = \varpi_k^g : k \geq 1\}$ with mesh $|\varpi_k^f| \to 0$ and $|\varpi_k^g| \to 0$ such that in probability we have

$$\int_0^T f_n^{\varpi_k}(t)\, \mathrm{d}\ell_n(t) \to \int_0^T f^{\varpi_k}(t)\, \mathrm{d}\ell(t),$$
$$\int_0^T \mathrm{d}\ell_n(t) \int_{\mathbb{R}^d_*} g_n^{\varpi_k}(z,t)\, \pi(\mathrm{d}z) \to \int_0^T \mathrm{d}\ell(t) \int_{\mathbb{R}^d_*} g^{\varpi_k}(z,t)\, \pi(\mathrm{d}z),$$

for every k and T, then the Riemann-Stieltjes integrals converge in probability, i.e.,

$$\lim_n P\Big\{\Big|\int_0^t f_n(s)\,\mathrm{d}\ell_n(s) - \int_0^t f(s)\,\mathrm{d}\ell(s)\Big| \geq \varepsilon\Big\} = 0,$$

and

$$\lim_n P\Big\{\Big|\int_0^t \mathrm{d}\ell_n(s)\int_{\mathbb{R}^d_*} g_n(z,s)\pi(\mathrm{d}z) - \int_0^t \mathrm{d}\ell(s)\int_{\mathbb{R}^d_*} g(z,s)\pi(\mathrm{d}z)\Big| \geq \varepsilon\Big\} = 0,$$

for every $t, \varepsilon > 0$. Also the stochastic integrals

$$M_n(t) := \int_{]0,t]} f_n(s)\,\mathrm{d}w_n(s), \qquad J_n(t) := \int_{\mathbb{R}^m_* \times]0,t]} g_n(z,s)\,\tilde{\nu}_n(\mathrm{d}z,\mathrm{d}s),$$

converge also in probability to

$$M(t) := \int_{]0,t]} f(s)\,\mathrm{d}w(s), \qquad J(t) := \int_{\mathbb{R}^m_* \times]0,t]} g(z,s)\,\tilde{\nu}(\mathrm{d}z,\mathrm{d}s),$$

for every $t > 0$, where $\tilde{\nu}_n := \nu_n - \pi\mathrm{d}\ell_n$ and $\tilde{\nu} := \nu - \pi\mathrm{d}\ell$ are the (local) martingale measures associated with integer measures ν_n and ν. □

It is also clear that under the conditions of the above Theorem and the assumption that f_n and g_n converge to f and g in probability as random variable with values in the Polish space $D([0,\infty),\mathbb{R}^d)$ and $D([0,\infty),L^2_\pi(\mathbb{R}^m_*))$, the stochastic integrals M_n and J_n converge to M and J in probability as random variable with values in the Polish space $D([0,\infty),\mathbb{R}^d)$ and $D([0,\infty),\mathbb{R})$. Moreover, if M_n are continuous then we can replace the cad-lag space $D([0,\infty),\mathbb{R}^d)$ with $C([0,\infty),\mathbb{R}^d)$. In any case, the Riemann-Stieltjes integral processes converge as random variables with values in the Polish space $C([0,\infty),\mathbb{R}),$, i.e., for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that

$$P\Big\{\sup_{0\leq t\leq 1/\varepsilon}\Big|\int_0^t f_n(s)\,\mathrm{d}\ell_n(s) - \int_0^t f(s)\,\mathrm{d}\ell(s)\Big| \geq \varepsilon\Big\} \leq \varepsilon,$$

for every $n \geq N(\varepsilon)$, see estimate (4.55).

It is possible to consider the Lévy measure π in Theorem 4.32 depending on n, i.e. π_n, provided some uniform integrability at the origin is imposed, e.g.,

$$\lim_{\varepsilon\to 0}\sup_n \int_{|z|\geq\varepsilon} |z|^2\,\pi_n(\mathrm{d}z) = 0,$$

or replacing the function $|z|^2$ with either $|z|^2 \wedge |z|$ or $|z|^2 \wedge 1$, depending on the integrability condition imposed on each π_n.

4.4.2 Other Convergence of Probabilities

Mainly, we discuss here Jakubowski convergence of probability measures. The canonical spaces $C([0,\infty),\mathbb{R}^d)$ and $\mathbb{D} = D([0,\infty),\mathbb{R}^d)$, of continuous and cad-lag functions, are Polish (complete separable metric) spaces, with the local uniformly convergence and the Skorokhod topology (usually referred to as the J_1-topology. Clearly, the addition and multiplication are continuous operation on $C([0,\infty),\mathbb{R}^d)$, but not on $\mathbb{D}$, i.e., $C([0,\infty),\mathbb{R}^d)$ is a topological vector space but $\mathbb{D}$ is not. Moreover, the topology in $D([0,\infty),\mathbb{R}^d)$ is strictly stronger that the product topology in $D([0,\infty),\mathbb{R}^{d_1}) \times D([0,\infty),\mathbb{R}^{d_2})$, $d = d_1 + d_2$.

Now, the spaces of probability measures on $C([0,\infty),\mathbb{R}^d)$ and $\mathbb{D}$, denoted respectively $\wp(C([0,\infty),\mathbb{R}^d))$ and $\wp(\mathbb{D})$, are Polish spaces, with the weak convergence topology, i.e. $\mu_n \to \nu$ if $\mu_n(f) \to \mu(f)$ for every bounded continuous function f from $C([0,\infty),\mathbb{R}^d)$ (or $\mathbb{D}$) into $\mathbb{R}$; moreover any probability measure is tight. The reader is referred to the book by Ethier and Kurtz [45, Chapter 3, pp. 95–154] or Jacod and Shiryaev [84, Chapter VI, pp. 288–347] for a comprehensive discussion.

The *operation stochastic integral* can be regarded as a functional on either $C([0,\infty),\mathbb{R}^d)$ or $\mathbb{D}$, i.e., given a probability measure in $\mathbb{D}$ with a certain number of properties (relative to some integrands and integrators), the law of the stochastic integral process defines another probability measure. Loosely speaking, if we have a sequence of integrands and integrators then we actually have a sequence of probability measures on $C([0,\infty),\mathbb{R}^d)$ or $\mathbb{D}$. Specifically, we are interested in the functional defined by an stochastic differential equation. When dealing with cad-lag processes of (local) bounded variation, the Skorokhod topology seems too strong for some cases of reflected stochastic differential equation, and a weak topology is necessary. One of the key difficulties is that we exit the framework of Polish spaces and we need to recall or review certain points of general topology.

Sequential Convergence

First, it is a necessary some basic terminology on sequential convergence. In a given topological space $(\mathbb{X},\tau)$ the closure of any subset of $\mathbb{X}$ could be defined as a map $A \to \bar{A}\colon 2^{\mathbb{X}} \to 2^{\mathbb{X}}$ with the following properties: (a) $\bar{\emptyset} = \emptyset$, (b) $A \subset \bar{A}$, (c)$\overline{A \cup B} = \bar{A} \cup \bar{B}$ and (d) $\bar{\bar{A}} = \bar{A}$. This previous four properties are called the Kuratowski axioms.

Suppose now to have defined on a set $\mathbb{X}$ (without a topology) a map on the subsets of $\mathbb{X}$, say $\kappa\colon 2^{\mathbb{X}} \to 2^{\mathbb{X}}$, such that: (1) $\kappa(\emptyset) = \emptyset$, (2) $A \subset \kappa(A)$ and (3) $\kappa(A \cup B) = \kappa(A) \cup \kappa(B)$. Then, we can endow $\mathbb{X}$ with a topology τ_κ by defining as "closed sets" those subsets F such that $F = \kappa(F)$. We can easily check that the properties (1), (2) and (3) imply that the family of the complements of "closed sets", just defined, is a topology. The closure operator with respect to this topology has the property $A \subset \kappa(A) \subset \bar{A}$. Thus, if for any subset A we have that $\kappa(A) = \bar{A}$, then we have also the property (4) $\kappa(\kappa(A)) = \kappa(A)$. Hence, we can shows that if the map κ verifies (1), (2), (3) and (4) as above, then the

above topology τ_κ is the unique topology such that $\bar{A} = \kappa(A)$, for any subset $A \subset \mathbb{X}$.

Convergent Sequences in a Given Topology

Now given a topological space $(\mathbb{X}, \tau)$, the family of converging sequences $x_n \to x$ is determined. We can define the map $\kappa(A) = [A]_{\mathsf{seq}}$ as the set of all limits points of τ-converging sequences of points of A. It is easy to check that κ satisfies (1) $[\emptyset]_{\mathsf{seq}} = \emptyset$, (2) $A \subset [A]_{\mathsf{seq}} \subset \bar{A}^\tau$ and (3) $[A \cup B]_{\mathsf{seq}} = [A]_{\mathsf{seq}} \cup [B]_{\mathsf{seq}}$, but in general the point (4) is not true, i.e., we may have $[A]_{\mathsf{seq}} \subsetneq [[A]_{\mathsf{seq}}]_{\mathsf{seq}}$.

Thus we can introduce, as before, a topology $\tau_{\mathsf{seq}} (= \tau_\kappa)$, by defining the new closed sets as $F = [F]_{\mathsf{seq}}$ and we have that $\tau \subset \tau_{\mathsf{seq}}$. Obviously that the two topologies have the same converging sequences, moreover, there is also the weakest topology τ' with the same converging sequences of τ, and

$$A \subset [A]_{\mathsf{seq}} \subset \bar{A}^{\tau_{\mathsf{seq}}} \subset \bar{A}^{\tau} \subset \bar{A}^{\tau'}.$$

Hence, a topological space $(\mathbb{X}, \tau)$ is called *sequential space* if $\tau = \tau_{\mathsf{seq}}$, with $\kappa(A) = [A]_{\mathsf{seq}}$. Since it could happen that $[[A]_{\mathsf{seq}}]_{\mathsf{seq}} \neq [A]_{\mathsf{seq}}$, a topological space $(\mathbb{X}, \tau)$ is called *Fréchet-Urysohn space* if it a sequential space such that $[[A]_{\mathsf{seq}}]_{\mathsf{seq}} = [A]_{\mathsf{seq}}$ or equivalently $\bar{A}^{\tau_{\mathsf{seq}}} = \bar{A}^{\tau}$. Note that any metric space $\mathbb{X}$ is a Fréchet-Urysohn space.

Topology After Convergent Sequences

We can define the convergence of a sequence without introducing necessarily a topology, in other words we can define a convergence of a sequence not in terms of a given topology, as, for example, in the usual weak convergence of probability measures on topological spaces. Now, if we assume that a notion of convergence of sequences on a set (arbitrary) $\mathbb{X}$ is given, then to find a topology τ on the space $\mathbb{X}$ such that all the converging sequences converge also in this topology, we need to impose the the following properties to the family of converging sequences:

(i) The uniqueness of the limit holds.

(ii) For every $x \in \mathbb{X}$, the constant sequence $\{x, x, x, \ldots\}$ is convergent to x.

(iii) Given a convergent sequence $\{x_1, x_2, x_3, \ldots\}$ $(x_n \to x)$, then every subsequence is convergent to the same limit x.

These hypotheses imply that the sequential closure map $\kappa(A) = [A]_{\mathsf{seq}}$ (as the set of all limits points of converging sequences of points of A) verifies the properties (1), (2) and (3) above. Hence we can introduce the topology τ_{seq}, and all converging sequences are also convergent in this topology. But in general, there are more τ_{seq}-converging sequences than converging sequences (in the initial sense).

Since a sequence $\{x_n\}$ is τ_{seq}-converging to x_0 if and only if from any subsequence it is possible to extract a further subsequence convergent to the same

x_0 (in the initial sense). This motivates the following further property, after the properties (i), (ii) and (iii),

(iv) a sequence $\{x_n\}$ is converge to x_0 if from any subsequence of $\{x_n\}$ it is possible to extract a further subsequence convergent to the same x_0.

Therefore, if (iv) holds then all converging sequences in the topology τ_{seq} are just the given converging sequences.

If in a set $\mathbb{X}$ we have defined (initially) the meaning of converging sequences satisfying (i), (ii) and (iii), then we say that we have space of type $\mathcal{L}$ or *sequential convergence of type* $\mathcal{L}$. Moreover, if also the property (iv) is satisfied then we called it a space of type $\mathcal{L}^*$ or *sequential convergence of type* $\mathcal{L}^*$.

Now, starting from a space $\mathbb{X}$ with sequential convergence of type $\mathcal{L}$, we can endow $\mathbb{X}$ of the corresponding τ_{seq} topology. Next, if we take *all* the convergent sequences in this τ_{seq} topology, which is called the $*$-convergence (relative to the initial convergence), then we have a sequential convergence of type $\mathcal{L}^*$. Clearly, if a sequence converges in the initial convergence then it also converges in the $*$-convergence, but not necessarily the converse. On the other hand, if we start from a space $\mathbb{X}$ of type $\mathcal{L}^*$ and we endow $\mathbb{X}$ with the corresponding τ_{seq} topology, then this time *all* the convergent sequences in this τ_{seq} topology are exactly the same convergent sequences given initially, i.e., the initial convergence and the $*$-convergence are the same.

The simplest example is perhaps the space of real-valued Borel measurable functions $B([0,\infty),\mathbb{R})$ with the pointwise convergence (which yields a sequential topology), where all four properties are satisfied, i.e., a sequential convergence of type $\mathcal{L}^*$. However, it is clear that $B([0,\infty),\mathbb{R})$ is not a Fréchet-Urysohn space. Indeed, if continuous functions are called Baire function of class 0 or of at most class 0, then pointwise limit of Baire functions of at most class n are called Baire function of at most class $(n+1)$, and Baire functions of class $(n+1)$ those that are of at most class $(n+1)$ without being of at most class n. Clearly, all Baire functions are Borel measurable functions. Thus denoting by $[\cdot]_{\mathsf{seq}}$ the sequential closure, the Baire functions of at most class 1 is the closure $[C([0,\infty),\mathbb{R})]_{\mathsf{seq}}$ while the Baire functions of at most class 2 is the double closure $[[C([0,\infty),\mathbb{R})]_{\mathsf{seq}}]_{\mathsf{seq}}$. Their difference is the Baire functions of class 2, e.g., the Dirichlet function ($=1$ for all rational and $=0$ for all irrational) $\lim_{n\to\infty}\left(\lim_{k\to\infty}(\cos n!\,\pi x)^{2k}\right)$ is a Baire function of class 2. It is clear that similar remarks apply to the pointwise and bounded convergence. Actually, if T is an interval and X is a $\mathcal{L}^*$ space, so is $B(T,X)$ and $C(T,X)$. Clearly, the pointwise convergence makes $B(T,X)$ a Hausdorff topological spaces, which is neither a countable separated space nor a separable space.

Another interesting example is the space $L^0(\Omega,\mathcal{F},P)$ of the equivalence classes of real-valued random variables with the almost surely pointwise convergence. This space is of type $\mathcal{L}$, but is not of type $\mathcal{L}^*$. Moreover, the convergence (iv), i.e., the τ_{seq} convergence or $*$-convergence (due to the topology induced by the almost surely pointwise convergence) is actually the convergence in probability, i.e., in this case, $L^0(\Omega,\mathcal{F},P)$ with the $*$-convergence becomes a complete metric space.

Sequence of Probability Measures

The notions of *tightness* (or *boundedness in probability*), *regularity*, and of *weak convergence* (or *convergence in law*) of measures need that the underlying measure space be a topological space $\mathbb{X}$ with the corresponding Borel σ-algebra $\mathcal{B}(\mathbb{X})$.

Definition 4.33. Given a topological space $\mathbb{X}$ and its Borel σ-algebra $\mathcal{B}(\mathbb{X})$, a family of probability measures $P_i, i \in \mathcal{I}$ is *uniformly tight* if for any $\varepsilon > 0$ there exists a compact K_ε such that, for every $i \in \mathcal{I}$, $P_i(K_\varepsilon) > 1 - \varepsilon$. □

Let us mention two key results on Borel (measures defined on the Borel σ-algebra) and Radon measures (measures finite on any compact set):

(a) Any probability measure P on a *metric space* $\mathbb{X}$ is *regular*, i.e., for any Borel set A and every $\varepsilon > 0$ there exist a closed set F and a open set G such that $F \subset A \subset G$ and $P(G \setminus F) < \varepsilon$.

(b) Any probability measure P on a *Polish space* (i.e., complete separable metric space) $\mathbb{X}$ is tight, i.e., for any $\varepsilon > 0$ there exists a compact K_ε such that $P(K_\varepsilon) > 1 - \varepsilon$.

In particular, any probability measure on a Polish space is regular and tight.

Weak convergence

Given a topological space $(\mathbb{X}, \tau_0)$ and its Borel σ-algebra $\mathcal{B}(\mathbb{X})$, we can consider the space of probability measures on $\mathbb{X}$, which is denoted by $\wp(\mathbb{X})$ and endowed with the weakest topology such that every linear functional of the form $\mu \to \mu(f)$ is continuous, when f is any bounded and continuous function on $\mathbb{X}$. This topological space is denoted by $\mathfrak{W}(\wp(\mathbb{X}))$ or simply $\mathfrak{W}$. Note that if $\mathbb{X}$ is a metric (or Polish) space then $\wp(\mathbb{X})$ results also a metric (or Polish) space with the Prohorov's distance.

Note that $\mu_n \xrightarrow{\mathfrak{W}} \mu$ implies that $\mu_n(f) \to \mu(f)$ for every f bounded and continuous. But the converse may be false, i.e. we can have that $\mu_n(f) \to \mu(f)$ for every f bounded and continuous, but not converging in the $\mathfrak{W}$ topology.

Recall that usually we have the weak convergence defined by $\mu_n(f) \to \mu(f)$ for every f bounded and continuous. This convergent yields a space of type $\mathcal{L}$, which is not necessarily $\mathfrak{W}$. Thus, it makes sense to introduce the sequential weak topology (the previous τ_{seq} topology) $\mathfrak{W}_{\mathsf{seq}}$, the weakest topology with respect to which we have $\mu_n(f) \to \mu(f)$ for every f bounded and continuous. Certainly, $\mathfrak{W} \subset \mathfrak{W}_{\mathsf{seq}}$. We have

Theorem 4.34. *The space of probability measures on a* Polish *space with the weak convergence is a space of type* $\mathcal{L}^*$ *with* $\mathfrak{W}_{\mathsf{seq}} = \mathfrak{W}$. □

On the other hand, starting with $(\mathbb{X}, \tau_0)$, let us suppose that there is another (weaker) topology τ_1 on $\mathbb{X}$, such that $\tau_1 \subset \tau_0$ and that the Borel σ-algebra generated by τ_1 is the same as that generated by τ_0. In such a case the space

$\wp(\mathbb{X})$ is uniquely defined, with either τ_0 or τ_1. We have that $\mathfrak{W}(\tau_1) \subset \mathfrak{W}(\tau_0)$ and $\mathfrak{W}_{\mathsf{seq}}(\tau_1) \subset \mathfrak{W}_{\mathsf{seq}}(\tau_0)$. If $(\mathbb{X}, \tau_0)$ is a Polish space then

$$\mathfrak{W}(\tau_1) \subset \mathfrak{W}_{\mathsf{seq}}(\tau_1) \subset \mathfrak{W}(\tau_0) = \mathfrak{W}_{\mathsf{seq}}(\tau_0).$$

In fact, $\mu_n \to \mu$ in $\mathfrak{W}_{\mathsf{seq}}(\tau_0)$ means that $\mu_n(f) \to \mu(f)$ for every bounded and τ_0-continuous function. Since a τ_1-continuous function is also τ_0-continuous, we deduce also that $\mu_n \to \mu$ in $\mathfrak{W}_{\mathsf{seq}}(\tau_1)$, i.e., $\mathfrak{W}_{\mathsf{seq}}(\tau_1) \subset \mathfrak{W}_{\mathsf{seq}}(\tau_0)$.

Prohorov Theorem

There are two implication, the direct and the converse:

(1) Given a metric space $\mathbb{X}$, a family of probability measures $\{P_i, i \in \mathcal{I}\}$ on $\mathbb{X}$, is uniformly tight if it is relatively compact with respect to weak convergence.

(2) Given a separable, complete metric space $\mathbb{X}$ (i.e. a Polish space), a family of probability measures $\{P_i, i \in \mathcal{I}\}$ on $\mathbb{X}$, is relatively compact with respect to weak convergence if it is uniformly tight.

For instance, see for example Dudley [37, Section 11.5, pp. 402–405].

Non Metric Case

Now, let $(\mathbb{X}, \tau)$ be a topological space countably separated, i.e., such that the following hypothesis holds:

> there exists a countable family $\{f_i \colon \mathbb{X} \to [-1, 1], i = 1, 2, \ldots\}$ of τ-continuous functions which separates the points of $\mathbb{X}$, that is for any two distinct points $x_1, x_2 \in \mathbb{X}$ there exists a function f_κ such that $f_\kappa(x_1) \neq f_\kappa(x_2)$.

Consider $[0, 1]$ is considered with the usual Borel σ-algebra and the standard Lebesgue measure (sometimes referred to as the universal probability space), see Theorem 4.27 to compare assumptions.

Theorem 4.35. *Let $\{\mu_n\}$ be a sequence of* tight *probability measures on a topological space $\mathbb{X}$, with also the previous hypothesis. Then there exist a subsequence $\{\mu_{n_k}\}$ and a sequence of random variables $X_k \colon [0, 1] \to \mathbb{X}$ and a further random variable $X \colon [0, 1] \to \mathbb{X}$, such that* (1) *the image measures of X_k are the μ_{n_k} and* (2) *$X_n(\theta) \to X(\theta)$, for any θ in $[0, 1]$.* □

Note that given any compact $K \subset \mathbb{X}$ the set

$$C_K = \{\theta \in [0, 1] : X_k(\theta) \to X(\theta)\} \cap \bigcap_{k=1}^{\infty} \{\theta \in [0, 1] : X_k(\theta) \in K\}$$

is Borel measurable in $[0, 1]$. Moreover, for any ε there exists a compact K_ε such that the Lebesgue measure of C_{K_ε} is greater or equal to $1 - \varepsilon$.

Star-convergence of Tight Probability on $\mathbb{X}$

Let us continue with

Definition 4.36. Given a sequence of *tight* probability measures $\{\mu_n\}$ on $\mathbb{X}$, we say that $\mu_n \overset{*}{\Longrightarrow} \mu$ if from every subsequence $\{\mu_{n_k}\}$ there exist a further subsequence $\{\mu_{n_{k_i}}\}$ and a sequence of "random" variables $X_i\colon [0,1] \to \mathbb{X}$, whose image measures are just the $\mu_{n_{k_i}}$'s and a further "random" variable $X\colon [0,1] \to \mathbb{X}$, whose image measure is μ such that, for each $\theta \in [0,1]$, $X_i(\theta) \to X(\theta)$ and for each $\varepsilon > 0$ there exists a compact $K_\varepsilon \subset \mathbb{X}$ such that $\mathsf{Leb}\big(\bigcap_{i=1}^{\infty}\{\theta : X_i(\theta) \in K_\varepsilon\}\big) > 1-\varepsilon$. □

This definition gives to the space of tight probability measures (denoted by $\wp^o(\mathbb{X}) \subset \wp(\mathbb{X})$) the structure of space of type $\mathcal{L}^*$. Hence we have the corresponding $\mathfrak{W}_{\mathsf{Jak}}$ sequential topology. Moreover, referring to $\wp^o(\mathbb{X})$ instead of the whole $\wp(\mathbb{X})$, this topology $\mathfrak{W}_{\mathsf{Jak}}$ is stronger than the sequential topology $\mathfrak{W}_{\mathsf{seq}}$, i.e. $\mathfrak{W}_{\mathsf{seq}} \subset \mathfrak{W}_{\mathsf{Jak}}$.

Theorem 4.37. *This $\mathfrak{W}_{\mathsf{Jak}}$ topology has the property that the family of relatively compact sets coincides with the family of relatively uniformly tight sets.* □

If $\mathbb{X}$ is a metric space then the weak topology and the $\mathfrak{W}_{\mathsf{Jak}}$ topology coincide.

4.4.3 Back to the Canonical Space

Consider the dual space of $C([0,T])$, $T > 0$, which is the space of functions $v(\cdot)$ with bounded variation with the duality pairing

$$\langle \varphi, v\rangle_T = \int_{[0,T]} \varphi(t)\mathrm{d}v(t).$$

Since each function $v(\cdot)$ with bounded variation can be modified (without changing the dual pairing) so that $v(\cdot)$ is also cad-lag, we denote by $D_{\mathsf{BV}}([0,T])$, the set of cad-lag functions with bounded variation, which is considered as a subspace of the canonical space $D([0,T])$. Because $D_{\mathsf{BV}}([0,T])$ is the dual of the Banach space $C([0,T])$, we can use the weak* topology on $D_{\mathsf{BV}}([0,T])$, where balls are weakly* compact. Thus, we are interested in a topology on the space $D([0,T])$ such that relatively to the subspace $D_{\mathsf{BV}}([0,T])$ the convergence is similar to the weak* convergence and any set of equi-bounded variation functions is compact. The topology introduced by Jakubowski [85] has this property. Clearly, what is done for $D([0,T])$ can be extended to $D([0,\infty),\mathbb{R}^d)$.

Again, let $D_{\mathsf{BV}}([0,\infty),\mathbb{R}^d)$ denote the space of functions x in $D([0,\infty),\mathbb{R}^d)$ that locally are of bounded variation, with the sup-norm

$$\|x\|_{T,\infty} = \sup\big\{|x(t)| \,:\, 0 \le t \le T\big\}$$

and the variation-norm

$$\|x\|_{T,\mathsf{BV}} = \sup\Big\{\sum_{i=0}^{n} |x(t_{i+1}) - x(t_i)| \,:\, t_i < t_{i+1}\Big\}.$$

where the supremum is taken with respect to all partitions with $t_0 = 0$, $t_n = T$ and t_i belonging only to a dense subset of $(0, T)$. We consider the following convergence in $D([0,\infty), \mathbb{R}^d)$, as introduced by Jakubowski [85].

Definition 4.38. We say that the sequence $\{x_n\}$ in $D([0,\infty), \mathbb{R}^d)$ is convergent in the sense of Jakubowski, denoting by $x_n \xrightarrow{\text{Jak}} x$, if and only if, for any $\epsilon > 0$ there exist a sequence $\{v_n^\epsilon\}$ and v^ϵ in $D_{\text{BV}}([0,\infty), \mathbb{R}^d)$ such that

$$\|v_n^\epsilon - x_n\|_{1/\varepsilon,\infty} \le \epsilon, \quad \forall n \ge 1, \qquad \|v^\epsilon - x\|_{1/\varepsilon,\infty} \le \epsilon$$

and

$$\int_0^{1/\varepsilon} \varphi(t)\, \mathrm{d}v_n^\epsilon \to \int_0^{1/\varepsilon} \varphi(t)\, \mathrm{d}v^\epsilon,$$

for any $\varphi \in C([0, 1/\varepsilon])$. □

Star-convergence of Probability on $\mathbb{D}_S$

Now we look at $\mathbb{D}_S$, the canonical space $\mathbb{D} = D([0,\infty), \mathbb{R}^d)$ space of cad-lag functions with the S-topology defined below (i.e., the $*$-convergence, denoted by "$\xrightarrow{\text{Jak}^*}$", and derived from Definition 4.38).

It can be proved that the S-topology on $\mathbb{D}$ generates the same Borel sets that we have with the metric (J_1) topology, thus the probability measures are the same. Since the compact sets in the metric topology are also compact in the S-topology, we have that every probability measure is also tight in the S-topology. Hence, the sequential convergence is defined on the entire space of probability measures $\wp(\mathbb{D}_S) = \wp^o(\mathbb{D}_S)$, because all probability measures are tight in $\mathbb{D}_S$.This new topology $\mathfrak{W}_{\mathsf{Jak}}(\mathbb{D}_S)$ is stronger than the topology of weak convergence $\mathfrak{W}_{\mathsf{seq}}(\mathbb{D}_S)$ (where we consider the S-topology on $\mathbb{D}$), in other words if we have the $\mathfrak{W}_{\mathsf{Jak}}$ convergence then we have also the weak S-convergence. On the other hand the $\mathfrak{W}_{\mathsf{seq}}(\mathbb{D}_S)$-topology is weaker than the classical (metric) topology of weak convergence (that is with J_1 as topology on $\mathbb{D}$). However, in general, we cannot say anything (from only this information) on the classical weak convergence, hence every case needs a specific study: for example, the Łaukajtys-Słomiński paper [109] shows that we don't have the classical convergence, nevertheless they prove the weak S-convergence (proving namely the stronger $\mathfrak{W}_{\mathsf{Jak}}$ convergence).

In other words, rephrasing Definition 4.38), $x_n \xrightarrow{\text{Jak}} x$ if and only if there exists a double sequence $\{v_{n,k}\}$ in $D_{\text{BV}}([0,\infty), \mathbb{R}^d)$ such that (1) for every n, $v_{n,k} \to x_n$ locally uniform as $k \to \infty$, (2) for every k and any continuous function φ, $\langle \varphi, v_{n,k} \rangle \to \langle \varphi, v_k \rangle$ as $n \to \infty$, and (3) $v_k \to x$ locally uniform as $k \to \infty$.

Actually we can endow $D([0,\infty), \mathbb{R}^n)$ of the topology τ_{Jak} defined by the following family of open sets

> G is *open* if and only if for any sequence $\{x_n\}$, converging to a element $x \in G$ in the previous sense of Jakubowski, a *tail* of the

sequence belongs to G, i.e. there exists an integer N such that $x_n \in G$ for any $n \geq \mathbb{N}$.

We now remark that the converging sequences in the sense of Jakubowski are *not the only sequences* that converge in the above topology! There are sequences that are convergent in the sense of τ_{Jak} topology without being convergent in the sense of Jakubowski. Thus, we will indicate this weaker convergence in the topology τ_{Jak} as $x_n \xrightarrow{\mathsf{Jak}^*} x$.

We can also endow $D([0,\infty),\mathbb{R}^n)$ of the usual Skorokhod topology (the so called J_1 topology), under which $D([0,\infty),\mathbb{R}^n)$ is a separable, complete metric space (Polish space).

Actually the topology τ_{Jak} is weaker of the usual Skorokhod topology, but the Borel sets with respect to τ_{Jak} are just the same Borel sets with respect to Skorokhod topology, both coinciding with the σ-algebra generated by the cylindrical sets.

Note the contrast, $D_{\mathsf{BV}}([0,\infty),\mathbb{R}^d)$ is dense in $\mathbb{D}_S$, but $C([0,\infty),\mathbb{R}^d)$ is closed in $D([0,\infty),\mathbb{R}^d)$, and $C([0,\infty),\mathbb{R}^d)$ is dense in $\mathbb{D}_S$. To check this, first recall that for any x in $D([0,\infty),\mathbb{R}^d)$ and any $\varepsilon > 0$ there exist $0 = t_0 < t_1 < \cdots < t_r = 1/\varepsilon$ such that for any $i = 1,\dots,r$ and for any s,t in $[t_{i-1},t_i)$ we have $|x(t)-x(s)| < \varepsilon$. Indeed, by means of the right continuity property, we can define inductively $t_0 = \inf\{t > 0 : |x(t) - x(0)| \geq \varepsilon/2\}$ and $t_i = \inf\{t > t_{i-1} : |x(t) - x(t_{i-1})| \geq \varepsilon/2\}$ for $i \geq 1$. This sequence $\{t_k\}$ is divergent, namely if $t_k \to \tilde{t}$ we would have also $\varepsilon/2 \leq |x(t_k) - x(t_{k-1}| \to |x(\tilde{t}-) - x(\tilde{t}-)| = 0$, in view of the existence of left-hand limits, which is a contradiction. Thus we can define

$$v^\varepsilon(t) = x(t_{i-1}) \quad \text{if} \quad t \in [t_{i-1},t_i), \quad i = 1,\dots,r,$$

which is a piecewise constant function (so cad-lag with bounded variation) satisfying $\|v^\varepsilon - x\|_{1/\varepsilon,\infty} < \varepsilon$. Now defined

$$x_n(t) = n\int_t^{(t+1/n)} x(s)\mathrm{d}s \quad \text{and} \quad v_n^\varepsilon(t) = n\int_t^{(t+1/n)} v^\varepsilon(s)\mathrm{d}s,$$

which are absolutely continuous (and so continuous with bounded variation), $\|v_n^\varepsilon - x_n\|_{1/\varepsilon,\infty} < \varepsilon$, and as $n \to \infty$, converge pointwise (i.e., for each t fixed) to x and v^ε. This proves that the space of absolutely continuous functions is dense $D([0,\infty,\mathbb{R}^d)$ with the Jakubowski topology.

4.4.4 Uniform Tightness or UT Condition

The following statements described this topology:

(1) The space $D([0,\infty),\mathbb{R}^n)$ equipped with the sequential topology τ_{Jak} is a Hausdorff topological space which is not a metric space. Recall that with the Skorokhod topology, it is a complete separable metrizable space.

(2) There exists a countable family of τ_{Jak}-continuous functions which separate points in $D([0,\infty),\mathbb{R}^n)$.

(3) The addition is sequentially continuous with respect to convergence in the sense of Jakubowski. In particular, $x_n \xrightarrow{\text{Jak}} x$ if and only if $x_n - x \xrightarrow{\text{Jak}} 0$. Recall, this holds with the Skorokhod topology only if x is continuous.

(4) Compact subsets $K \subset D([0,\infty),\mathbb{R}^n)$ are metrizable spaces.

(5) A subset $K \subset D([0,\infty),\mathbb{R}^n)$ is relatively τ_{Jak}-compact if for any each $\varepsilon > 0$ there exists a constant C_ε such that for each $x \in K$ there exists $v_{x,\varepsilon}$ in $D_{\text{BV}}([0,\infty),\mathbb{R}^n)$ such that

$$\|x - v_{x,\varepsilon}\|_{1/\varepsilon,\infty} \le \varepsilon \quad \text{and} \quad \|v_{x,\varepsilon}\|_{1/\varepsilon,\text{BV}} \le C_\varepsilon \tag{4.58}$$

is satisfied.

(6) The evaluation or projection operators $x \mapsto x(t)$ from $D([0,\infty),\mathbb{R}^n)$ into $\mathbb{R}^d$ are nowhere continuous with the τ_{Jak} topology. However, the functionals

$$x \mapsto \frac{1}{\varepsilon}\int_t^{t+\varepsilon} x(s)\mathrm{d}s \quad \text{and} \quad x \mapsto \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} x(s)\mathrm{d}s$$

are continuous and converges to $x(t)$ and $x(t-)$ as $\varepsilon \to 0$. Thus, τ_{Jak}-Borel subsets $\mathcal{B}_{\text{Jak}}$ coincide with the standard σ-algebra generated by evaluations (projections). This fact implies that any probability measure on $(D([0,\infty),\mathbb{R}^n),\mathcal{B}_{\text{Jak}})$ is tight. Recall that with the Skorokhod topology, the evaluation operators are continuous at any continuity time of the limit point.

(7) τ_{Jak} is coarser (weaker) than the usual Skorokhod topology. Thus the cadlag space $D([0,\infty),\mathbb{R}^n)$ endowed with the τ_{Jak} Jakubowski topology is a Lusin space, i.e., it is a one-to-one continuous image of a Polish space. However, it is unknown if the space is completely regular (i.e., the topology is given by a family of pseudo-metric which is Hausdorff separated.)

(8) The subspace $D_{\text{BV}}([0,\infty),\mathbb{R}^n)$ is dense, indeed, if x is any point in in the space $D([0,\infty),\mathbb{R}^n)$ the singleton $\{x\}$ is compact, hence we can find a sequence of functions in $D_{\text{BV}}([0,\infty),\mathbb{R}^d)$ converging to x in the sense of Jakubowski, in view of (4.58). Actually, we proved above that the space of absolutely continuous is dense. We may consider $D_{\text{BV}}([0,\infty),\mathbb{R}^n)$ with the relative τ_{Jak} topology on $D([0,\infty),\mathbb{R}^n)$, but it is weaker than the weak* topology.

(9) Let $N_T^{a,b}(x)$ be the number of up-crossing (of x) of the interval $[a,b]$ in the time interval $[0,T]$, i.e., $N_T^{a,b}(x) \ge k$ if there exist $0 \le t_1 < t_2 < \ldots < t_{2k-1} < t_{2k}$ such that $x(t_{2i-1}) < a < b < x(t_{2i})$ for any $i = 1,2,\ldots k$. On the other hand, let $N_{T,\eta}(x)$ be the number of oscillations (of x) greater than η in the time interval $[0,T]$, i.e., $N_{T,\eta}(x) \ge k$ if there exist $0 \le t_1 < t_2 < \ldots < t_{2k-1} < t_{2k}$ such that $|x(t_{2i}) - x(t_{2i-1})| > \eta$ for any $i = 1,2,\ldots k$. Then, it can be proved that any uniformly equi-bounded subset K of $D([0,\infty),\mathbb{R}^n)$ (i.e., for any $T > 0$ there exists a constant $C > 0$ such that $\|x\|_{T,\infty} \le C$ for every x in K) is compact (i.e., condition (4.58) holds) if and only if one of the following two conditions, for every $T > 0$,

$$\sup_{x\in K} N_T^{a,b}(x) < +\infty \qquad \text{or} \qquad \sup_{x\in K} N_{T,\eta}(x) < +\infty,$$

for each $a < b$, or for each $\eta > 0$, is satisfied.

The following result is useful to check the τ_{Jak} convergence. Let $\{x_n\}$ is relatively τ_{Jak}-compact. Then $x_n \xrightarrow{\mathsf{Jak}^*} x$ if there exists a (countable) dense set Q in $[0,\infty)$ such that $x_n(q) \to x(q)$, as $n \to +\infty$, for every q in Q.

On the other hand, if $\varphi(t,x)$ is a locally bounded Charathéodory function (measurable in t and continuous in x) defined on $[0,\infty) \times \mathbb{R}^d$, and ℓ is a continuous bounded variation function, then the integral functionals

$$x \mapsto \int_{]0,T]} \varphi(t, x(t-)) \mathrm{d}\ell(t) \quad \text{and} \quad x \mapsto \int_{]0,T]} \varphi(t, x(t)) \mathrm{d}\ell(t)$$

are continuous the Jakubowski topology.

However, if we consider the sequence $k_n(t) = \mathbb{1}_{[1/2-1/n,1]}(t)$, the function $k(t) = 1_{[1/2,1]}(t)$ and the sequence $x_n(t) = k(t)$, constant for any n, then $k_n \to k$ and $x_n \to k$ in the Skorokhod space $D([0,1],\mathbb{R})$ but

$$k(t) = \int_{]0,t]} k_n(s-) \mathrm{d}x_n(s) \not\to \int_{]0,t]} k(s-) \mathrm{d}k(s) = 0.$$

Another way to get a convergence of integrals is to consider convergence in $D([0,1],\mathbb{R}^2)$. There is a general result from Jakubowski-Mémin-Pagès [86]: we have (note that in the following theorem the topology in $D([0,T],\mathbb{R}^n)$ is the Skorokhod topology J_1)

Now we take a look at

Definition 4.39 (UT Condition)**.** Given a sequence of stochastic processes (semi-martingales) X_n with respect to the stochastic basis $(\Omega^n, \mathcal{F}^n, \{\mathcal{F}_t^n\}, P^n)\}$, the UT condition means that the family of all random variables of the form

$$\sum_{i=1}^{N} H_{t_{i-1}}^n (X_{t_i}^n - X_{t_{i-1}}^n)$$

is uniformly tight, where N is any integer, $0 = t_0 < t_1 < \cdots < t_N = T$ and $|H_{t_i}^n| \le 1$ with $H_{t_i}^n$ is $\mathcal{F}_{t_i}^n$-measurable for any i. □

Theorem 4.40. *Given a sequence of semi-martingales M_n with respect to the stochastic basis $(\Omega^n, \mathcal{F}^n, \{\mathcal{F}_t^n\}, P^n)\}$ satisfying the UT condition and a sequence of stochastic processes K_n adapted to $\{\mathcal{F}_t^n\}$, with trajectories in $D([0,\infty),\mathbb{R}^d)$, let us suppose that*

$$(K_n, M_n) \to (K, M)$$

weakly in $D([0,\infty),\mathbb{R}^{2d})$. Then M is a semi-martingale with respect to the natural filtration generated by (K,M) and

$$\int_{]0,\cdot]} K_n(t-) \mathrm{d}M_n(t) \to \int_{]0,\cdot]} K(t-) \mathrm{d}M(t)$$

weakly in $D([0,\infty),\mathbb{R}^d)$*, and*

$$\Big(K_n, X_n, \int_{]0,\cdot]} K_n(t-)\mathrm{d}M_n(t)\Big) \to \Big(K, M, \int_{]0,\cdot]} K(t-)\mathrm{d}M(t)\Big)$$

weakly in $D([0,\infty),\mathbb{R}^{3d})$. □

Note that the convergence $(K_n, M_n) \to (K, M)$ in $D([0,\infty),\mathbb{R}^{2d})$ is strictly stronger than the convergence $(K_n, M_n) \to (K, M)$ in the product topology of $D([0,\infty),\mathbb{R}^d) \times D([0,\infty),\mathbb{R}^d)$. We have from Jakubowski [85]

Theorem 4.41. *The UT condition implies the tightness in the space of probability measures on* $D([0,\infty),\mathbb{R}^n)$ *with respect to* $*$*-convergence.* □

The above statements are valid sometimes also in some not Polish space. If Ω is topological space having a countable family of continuous functions separating points then Skorokhod representation and Prohorov's theorem hold, i.e., for any uniformly tight sequence $\{P_n : n \geq 1\}$ of probability measures on Ω there exist a subsequence of indexes $\{n_k\}$ and random variables $\{X_k : k \in \mathbb{N}\}$ and X on the universal (Lebesgue) probability space $([0,1],\mathcal{B}([0,1]),\ell)$ with values in Ω such that P_{n_k} is the distribution of X_k for any $k \in \mathbb{N}$ and $X_k(t)$ converges to $X(t)$ for every t in $[0,1]$. In particular this applies when Ω is the canonical space $D([0,\infty),\mathbb{R}^d)$ with the τ_{Jak} Jakubowski topology (which is then only a Lusin space).

Because this is based on Prohorov's theorem, the above result holds for the $D([0,\infty),\mathbb{R}^d)$ with τ_{Jak} Jakubowski topology (see Definition 4.38) if the criterium of compactness (a') and (b') are modified accordingly.

Theorem 4.42 (tight)**.** *Let* $X_1, X_2, \ldots$ *be a sequence of random variables with values in* $D([0,\infty),\mathbb{R}^d)$*, and* $P_1, P_2, \ldots$ *be its associated probability law on* $D([0,\infty),\mathbb{R}^d) = \mathbb{D}_S$*, endowed with the Jakubowski topology. Then the sequence* $P_1, P_2, \ldots$ *is tight (hence relatively compact) in* $\mathbb{D}_S$ *if and only if the following two conditions hold:*

(a) *almost equi-bounded, i.e., for any* $\varepsilon > 0$ *there exists* $C > 0$ *such that for any index* n *we have*

$$P_n\{\sup_{0\leq t\leq 1/\varepsilon} |X_n(t)| \leq C\} \geq 1-\varepsilon,$$

(b) *equi-UT-condition, i.e., for any* $\varepsilon > 0$ *and for each* $T, \eta > 0$*, there exists* $K > 0$ *such that for any index* n *we have*

$$P_n\{N_{T,\eta}(X_n) \leq K\} \geq 1-\varepsilon.$$

Moreover, if the sequence is tight, then it is weakly convergent if and only its finite-dimensional distributions converge. □

Certainly we can replace (b) with

(b') for any $\varepsilon > 0$ and for each $T > 0$, $b > a$, there exists $K > 0$ such that for any index n we have

$$P_n\{N_T^{a,b}(X_n) \le K\} \ge 1 - \varepsilon,$$

where $N_T^{a,b}$ and $N_{T,\eta}(X_n)$ are as in (9) above.

For instance, if the processes (P_n, X_n) has local bounded variation, i.e., $X_n = X_n^+ - X_n^-$, with X_n^+ and X_n^- being increasing monotone, then the condition: for any $\varepsilon > 0$ there exists $C > 0$ such that for any index n we have

$$\sup_{0\le t\le 1/\varepsilon} P_n\{|X_n^+(t)| + |X_n^-(t)| > C\} \le \varepsilon,$$

implies both (a) and (b) above, since $N_{T,\eta}$ is controlled by the variation $|X_n^+| + |X_n^-|$ process.

Similarly, if the processes (P_n, X_n) is a local continuous martingale with predictable variation process $\langle X_n\rangle$, then the condition: for any $\varepsilon > 0$ there exists $C > 0$ such that for any index n we have

$$\sup_{0\le t\le 1/\varepsilon} P_n\big\{\big|\langle X_n\rangle(t)\big| > C\big\} \le \varepsilon,$$

implies both (a) and (b), since $N_T^{a,b}$ is controlled by the predictable variation process. Similarly, if the processes (P_n, X_n) is a local purely discontinuous (square-integrable) martingale with integer measure ν_n and predictable jumps compensator ν_n^p, then the condition: for any $\varepsilon > 0$ there exists $C > 0$ such that for any index n we have

$$\sup_{0\le t\le 1/\varepsilon} P_n\Big\{ \int_{\mathbb{R}^d_*\times]0,t]} \big(|z|^2 \wedge 1\big)\nu_n^p(\mathrm{d}z, \mathrm{d}s) > C\Big\} \le \varepsilon,$$

implies both (a) and (b), since $N_T^{a,b}$ is controlled by the predictable jumps compensator process. Note that

$$\int_{\mathbb{R}^d_*\times]0,t]} \big(|z|^2 \wedge 1\big)\nu_n^p(\mathrm{d}z, \mathrm{d}s) = \\ = \sum_{0<s\le t} \big[\mathbb{1}_{|X_n(s)-X_n(s-)|\ge 1} + |X_n(s)-X_n(s-)|^2 \mathbb{1}_{|X_n(s)-X_n(s-)|<1}\big],$$

i.e., adding the number of jumps greater than 1 and the square of the small jumps. Actually, these martingale cases can be treated directly with the classic Skorokhod topology, since

$$P_n\{\sup_{a\le t,s\le b} |X_n(t) - X_n(s)| \ge \varepsilon\} \le \frac{\delta}{\varepsilon^2} + P_n\big\{\big|\langle X_n\rangle(b) - \langle X_n\rangle(a)\big| \ge \delta\big\},$$

$$P_n\{\sup_{a\le t,s\le b} |X_n(t) - X_n(s)| \ge \varepsilon\} \le \frac{\delta}{\varepsilon^2} + \\ + P_n\Big\{ \int_{\mathbb{R}^d_*\times]a,b]} \big(|z|^2 \wedge 1\big)\nu_n^p(\mathrm{d}z, \mathrm{d}s) \ge \delta\Big\},$$

for every $\varepsilon, \delta > 0$, in view of Lenglart dominate property, e.g., see Jacod and Shiryaev [84, Section I.3c, pp. 35–36]. Essentially, the local bounded variation processes are of main interest for the Jakubowski topology.

Another situation is the following, see Section 3.3.3. Let ℓ_n, υ_n, w_n and ν_n, $n = 1, 2, \dots$ be sequences of processes defined in a probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Assume that:

(1) ℓ_n and υ_n are cad-lag processes with values in $\mathbb{R}^d$ and non-anticipative relative to (w_n, ν_n), and ℓ_n are nondecreasing,

(2) w_n are a d-dimensional continuous square integrable martingales with predictable quadratic covariation $\langle w_{n,i}, w_{n,j} \rangle = \varsigma_{n,j}$ if $i = j$, and $\langle w_{n,i}, w_{n,j} \rangle = 0$ if $i \neq j$,

(3) ν_n are integer measures with jump compensator $\nu_n^p(\mathrm{d}z, \mathrm{d}t) = \pi_n(\mathrm{d}z)\, \mathrm{d}\varrho_n(t)$, where π_n is a given Lévy measure in $\mathbb{R}_*^m$, and q_n denotes the corresponding purely discontinuous square-integrable martingale, i.e.,

$$q_n(t) = \int_{\mathbb{R}_*^m \times]0,t]} |z|^2 \tilde{\nu}_n(\mathrm{d}z, \mathrm{d}s), \quad \forall t \geq 0,$$

with $\tilde{\nu}(\mathrm{d}z, \mathrm{d}t) = \nu(\mathrm{d}z, \mathrm{d}t) - \pi_n(\mathrm{d}z)\varrho_n(\mathrm{d}t)$. Suppose that ℓ_n, υ, ς_n and ϱ_n are equi-bounded in probability, i.e., for every $\varepsilon > 0$ there exists $C = C(\varepsilon)$ such that

$$\sup_{0 \leq t \leq 1/\varepsilon} P_n\big\{ \big[|\ell_n(t)| + |\upsilon_n(t)| + |\varsigma_n(t)| + |\varrho_n(t)| \big] \geq C \big\} \leq \varepsilon, \quad \forall n,$$

and ℓ_n, ς_n and ϱ_n vanish at time $t = 0$.

(4) Also assume that υ_n satisfies the UT-condition, in term of the number of up-crossing $N_T^{a,b}$ or the number of oscillations $N_{T,\eta}$, e.g., for any $\varepsilon > 0$ and for each $T, \eta > 0$, there exists $K > 0$ such that for any index n we have

$$P_n\{N_{T,\eta}(\upsilon_n) \geq K\} \leq \varepsilon;$$

also that ς_n and ϱ_n are equi-continuous in probability, i.e., for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$P_n\Big\{ \sup_{0 \leq s,t \leq 1/\varepsilon,\ |t-s| < \delta} \big[|\varsigma_n(t) - \varsigma_n(s)| + |\varrho_n(t) - \varrho_n(s)| \big] \geq \varepsilon \Big\} \leq \varepsilon, \quad \forall n;$$

and that $\{\pi_n\}$ is a uniformly integrable Lévy sequence, i.e., there is a constant $C >$ such that

$$\int_{\mathbb{R}_*^m} |z|^2 \pi_n(\mathrm{d}z) \leq C, \quad \forall n,$$

and for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\int_{\{z : |z| < \delta\} \cup \{z : |z| > 1/\delta\}} |z|^2 \pi_n(\mathrm{d}z) < \varepsilon, \quad \forall n.$$

Now, consider (a) the probability law Q_n defined by $(P_n, \ell_n, \upsilon_n, w_n, \varsigma_n, q_n, \varrho_n)$ in the canonical space $D([0,\infty), \mathbb{R}^{d_0})$, with $d_0 = 4d + m + 1$, (b) the canonical processes ℓ, υ, w, ς, q, ϱ, and (c) endowed with the Jakubowski topology in the first $2d$ variables (relative to ℓ and υ) and with the Skorokhod topology in the remaining variables. Actually, for the variables w, ς and ϱ, we could use the sample space $C([0,\infty), \mathbb{R}^{d_1})$, $d_1 = 2d + 1$, with the usual locally uniform convergence. Then we can extract a subsequence, still denoted by $\{Q_n, \pi_n\}$, which is weak convergent to Q, π.

Clearly, all limiting processes are cad-lag. Moreover w, ς and ϱ are also continuous. Then, relative to Q on $D([0,\infty), \mathbb{R}^{d_0})$, we have:

(a) ℓ and υ are non-anticipating processes relative to (w, ν), and ℓ is a nondecreasing,

(b) w is also a continuous square integrable martingale with predictable quadratic covariation $\langle w_i, w_j\rangle = \varsigma_i$ if $i = j$ and $\langle w_i, w_j\rangle = 0$ if $i \neq j$,

(c) the integer measure ν associated with q has $\nu^p(\mathrm{d}z, \mathrm{d}t) = \pi(\mathrm{d}z)\mathrm{d}\varrho(t)$ as it predictable jump compensator,

(d) on the universal probability space $([0,1], \mathcal{B}([0,1]), l)$, where l is the Lebesgue measure, there exist random variables x, x_n with values in the canonical space $D([0,\infty), \mathbb{R}^{d_0})$ such that first (i) x and x_n have the same finite distributions as $(\ell, \upsilon, w, \varsigma, q, \varrho)$ and $(\ell_n, \upsilon_n, w_n, \varsigma_n, q_n, \varrho_n)$, respectively, and secondly (ii) $x_n(\theta) \to x(\theta)$, for every θ in $[0,1]$.

At this point, we can take limit on any continuous functional defined on the space $\mathcal{D}([0,\infty), \mathbb{R}^{d_0})$, e.g.,

$$\int_{]0,T]} f(t, x_n)\mathrm{d}\ell_n(t), \qquad \int_0^T f(t, x_n)\mathrm{d}w_n(t), \quad \text{and}$$

$$\int_{\mathbb{R}^m_* \times (0,T]} g(z, t, x_n)\tilde{\nu}_n(\mathrm{d}z, \mathrm{d}t),$$

preserved almost surely through finite-dimensional distributions. It is also clear that if the processes υ_n are equi-continuous in probability, we may use the Skorokhod topology in the variable υ instead of the weaker Jakubowski topology. Moreover, the cad-lag modulus of continuity can also be used. Note that $f(t,x)$ and $g(z,t,x)$ are regarded as deterministic (real or vector-valued) random fields with $t \geq 0$ and x in $\mathbb{D} = D([0,\infty), \mathbb{R}^{d_0})$, i.e., the mappings $(t,x) \mapsto f(t,x)$ and $(z,t,x) \mapsto g(z,t,x)$ are measurable with respect to the product Borel σ-algebras $\mathcal{B}([0,\infty)))\times\mathcal{B}(\mathbb{D})$ and $\mathcal{B}(\mathbb{R}^m_0)\times\mathcal{B}([0,\infty)))\times\mathcal{B}(\mathbb{D})$, plus some appropriate regularity conditions, e.g., continuity in x and causality, i.e., if $x(s) = y(s)$ for any $0 \leq s < t$ then $f(t,x) = f(t,y)$. Typical $f(t,x)$ has the form $f(t, x(t))$ or $f(t, x(t-))$. It is perhaps important to recall that under the Skorokhod topology, the evaluation or projection functional $x \mapsto x(t)$ are continuous only at any point of continuity of x, however, under the Jakubowski topology, they are nowhere continuous. Thus, when the measures $\mathrm{d}\ell_n$ have atoms (i.e., the processes ℓ_n are discontinuous) some extra special care should be taken to ensure the passage

to the limit in the Lebesgue-Stieltjes integral. Certainly all this applies to our case of interest, i.e., for a Wiener process or a Poisson measure.

Chapter 5

Transition Functions and PDEs

In this chapter is an introduction to Markov processes as a modelling tool, with a focus on the analytic aspect (i.e., the transition probability function) and partial differential equations (PDE). The style is to state the key results, give some comments on the proofs, but not complete proofs (but the reader can find them in the references quoted). Certainly, there are many excellent books on Markov processes with various orientations, e.g., Applebaum [1], Bertoldi and Lorenzi [8], Blumenthal and Getoor [15], Dellacherie and Meyer [32, 31], van Casteren [176], Dynkin [42], Ethier and Kurtz [45], Fukushima et al. [57], Jacob [80], Liggett [110], Rogers and Williams [153], Sharpe [158], Taira [172], among many others.

The first Section is a simple discussion on one-dimensional Markov processes, where most of the key characters are presented. Section 2 takes a more general view within cad-lag Markov processes. Next, Section 3 presents the semi-group approach, in a very general way, which covers processes on infinite dimension and semi-group non necessarily continuous (and therefore, main proofs are given). The remaining sections are mainly PDE stuff. Section 4 studies integro-differential equation corresponding to Lévy processes or in general, diffusion processes with jumps in a domain of the Euclidean space, with special attention to a priori estimates in Hölder and Sobolev spaces. Next, Section 5 is dedicated to the Green and Poisson functions or the transition probability density functions associated with diffusion process with jumps. Finally, Section 6 gives a list of examples.

5.1 Markov Processes

As a modelling tool, Markov processes have proved to be very efficient, as being the key element necessary to apply (in a simple way) the so-called *dynamic programming*. In its simplest form, this is know as Markov chain in discrete

time with a finite number of states, as typically used in many applications of operation research (e.g., see the textbooks by Bellman [4], Bertsekas [9], or Hillier and Lieberman [73], among many other excellent books). The passage to continuous time is (as usual) very technical, but again, numerous applications are found under the many of *queue theory*, which essentially deals with Markov chains in continuous time with a finite (sometimes countable) number of states. This Section begins with a simple introduction to Markov processes (which deals with one-dimensional Markov processes in continuous time) and ends with general considerations.

5.1.1 Processes without after-effect

Markov processes, or stochastic processes without after-effect, or non-hereditary (or memory less) stochastic processes refer always to the same property: any additional knowledge concerning the states of a process $\{X(\cdot)\}$ at previous times $s < t_0$ relative to the present time t_0 does not affect the statistics of the process at future times $s > t_0$. As will be noticed later, this means that a stochastic processes without after-effect is completed determined by its transition function, i.e., a function $P(s, x, t, \mathrm{d}y)$, which is the condition (transition) probability of $X(t)$ knowing that $X(s) = x$. For real valued process, we may use the transition distribution function $F(s, x, t, y)$, for $s < t$ and x, y in $\mathbb{R}$, associated with $\{X(\cdot)\}$, which is the probability that $X(t) < y$ assuming that $X(s) = x$. Clearly the function F should satisfies

$$\lim_{y\to-\infty} F(s, x, t, y) = 0, \qquad \lim_{y\to\infty} F(s, x, t, y) = 1,$$

and continuity from the left in y

$$\lim_{z\uparrow y} F(s, x, t, z) = F(s, x, t, y)$$

as any distribution function. To be able to use Bayes' formula for conditional distribution it is convenient to assume that F is continuous with respect to the variables s, x and t. This yields the Chapman-Kolmogorov equation (or identity)

$$F(s, x, t, y) = \int_{\mathbb{R}} F(s, x, \tau, \mathrm{d}\zeta)\, F(\tau, \zeta, t, y),$$

valid for any $s < \tau < t$ and x, y in $\mathbb{R}$. So, we may complete the definition of $F(s, x, t, y)$ for $s = t$ by a limit continuation as follows

$$\lim_{t\to s+} F(s, x, t, y) = \lim_{s\to t-} F(s, x, t, y) = F_0(x, y), \tag{5.1}$$

where $F_0(x, y) = 1$ if $y > x$ and $F_0(x, y) = 0$ otherwise. In what follows, this condition (5.1) is always assumed.

If a density function $f(s, x, t, y) = \partial_y F(s, x, t, y)$ exists then

$$F(s, x, t, y) = \int_{-\infty}^{y} f(s, x, t, \zeta)\mathrm{d}\zeta$$

and clearly

$$f(s,x,t,y) = \int_{\mathbb{R}} f(s,x,\tau,\zeta) f(\tau,\zeta,t,y) \mathrm{d}\zeta,$$

for any $s < \tau < t$ and x, y in $\mathbb{R}$.

For instance, the interested reader should consult the classic books Feller [48, Vol II, Sections X.3–5, pp. 316–331] and Gnedenko [64, Sections 53–55, pp. 358–376], for a more detailed discussion and proofs of most the results in this section.

For a *continuous* stochastic process we assume

$$\begin{aligned} \lim_{t\to s+} \frac{1}{t-s} \int_{|y-x|\geq\delta} F(s,x,t,\mathrm{d}y) = 0, \\ \lim_{s\to t-} \frac{1}{t-s} \int_{|y-x|\geq\delta} F(s,x,t,\mathrm{d}y) = 0, \end{aligned} \tag{5.2}$$

for every s, x and t, plus the following two conditions: (a) the first and second partial derivatives $\partial_x F(s,x,t,y)$ and $\partial_x^2 F(s,x,t,y)$ exit and are continuous for any $s < t$, x and y, and (b) for every $\delta > 0$ the limits

$$\begin{aligned} \lim_{t\to s+} \frac{1}{t-s} \int_{|y-x|<\delta} (y-x)^2 F(s,x,t,\mathrm{d}y) = 2a(s,x), \\ \lim_{s\to t-} \frac{1}{t-s} \int_{|y-x|<\delta} (y-x)^2 F(s,x,t,\mathrm{d}y) = 2a(t,x), \end{aligned} \tag{5.3}$$

and

$$\begin{aligned} \lim_{t\to s+} \frac{1}{t-s} \int_{|y-x|<\delta} (y-x) F(s,x,t,\mathrm{d}y) = b(s,x), \\ \lim_{s\to t-} \frac{1}{t-s} \int_{|y-x|<\delta} (y-x) F(s,x,t,\mathrm{d}y) = b(t,x), \end{aligned} \tag{5.4}$$

exit uniformly in x for every $s < t$ fixed. Note that the limiting coefficients a and b in (5.3) and (5.4) seem to depend on the value δ, but in view of the continuity condition (5.2) is merely apparent.

Under the above assumptions (5.2), (5.3) and (5.4) we can easily prove the validity of the *backward* Kolmogorov equation, namely

$$\partial_s F(s,x,t,y) + a(s,x)\partial_x^2 F(s,x,t,y) + b(s,x)\partial_x F(s,x,t,y) = 0, \tag{5.5}$$

for any $s < t$, x and y. Actually, only the first part of conditions (5.3) and (5.4) play any role here.

However, to deduce the *forward* Kolmogorov (also called Fokker-Planck) equation satisfied by the (probability) density function, namely

$$\partial_t f(s,x,t,y) + \partial_y \big[b(t,y) f(s,x,t,y)\big] = \partial_y^2 \big[a(t,y) f(s,x,t,y)\big], \tag{5.6}$$

for any $s < t$, x and y, we do need the second part of conditions (5.3) and (5.4), as well as the existence and continuity of the derivatives $\partial_t f(s,x,t,y)$, $\partial_y\big[b(t,y)f(s,x,t,y)\big]$ and $\partial_y^2\big[a(t,y)f(s,x,t,y)\big]$.

If the continuity condition (5.2) is slightly strengthen into

$$\lim_{t\to s+}\frac{1}{t-s}\int_{|y-x|\geq\delta}(y-x)^2F(s,x,t,\mathrm{d}y)=0,$$
$$\lim_{s\to t-}\frac{1}{t-s}\int_{|y-x|\geq\delta}(y-x)^2F(s,x,t,\mathrm{d}y)=0,$$

then conditions (5.3) and (5.4) are valid also for $\delta=0$, which mean

$$\mathbb{E}\{[X(t)-X(s)]-(t-s)b(s,X(s))\}=(t-s)\varepsilon_1,$$
$$\mathbb{E}\{[X(t)-X(s)]^2-2(t-s)a(s,X(s))\}=(t-s)\varepsilon_2,$$

where $\varepsilon_1,\varepsilon_2\to 0$ as $t-s\to 0+$, in term of the first and second moment of the Markov process $x(\cdot)$. As seem later, the actual construction of the stochastic process $\{X(t):t\geq 0\}$ is quite delicate, in particular if $a>0$ then almost surely the path $t\mapsto X(t,\omega)$ are continuous, but with infinite variation.

The transition distribution F of a *purely jump* Markov process $\{X(\cdot)\}$ can be expressed as follows:

$$F(s,x,t,y)=[1-(t-s)\lambda(s,x)]F_0(x,y)+ \\ +(t-s)\lambda(s,x)\rho(s,x,y)+(t-s)\varepsilon, \quad (5.7)$$

where $\varepsilon\to 0$ as $(t-s)\to 0^+$, uniformly in x and y. Thus, on any time interval (s,t) the value $X(\cdot)$ remains constant and is equal to $X(s)=x$ with probability $1-(t-s)\lambda(s,x)+(t-s)\varepsilon$, with $\varepsilon\to 0$ as $t-s\to 0^+$ (so that it may only have a jump with a probability $(t-s)\lambda(s,x)+(t-s)\varepsilon$). The coefficient $\rho(s,x,y)$ denotes the conditional distribution function of $X(s)$ under the condition that a jump has indeed occurred at time s and that immediately before the jump $X(\cdot)$ was equal to $X(s-)=x$.

Certainly, $\lambda(s,x)$ and $\rho(s,x,y)$ are nonnegative, $y\mapsto\rho(s,x,y)$ is a nondecreasing function continuous from the left, $\rho(s,x,-\infty)=0$, $\rho(s,x,\infty)=1$, and we assume that $\rho(s,x)$ is bounded, and that both $\rho(s,x)$ and $\rho(s,x,y)$ are continuous in s and Borel measurable in x. Under these conditions, (5.7) and the continuity (5.1), the following *backward* and *forward* Kolmogorov *integro-differential* equations

$$\begin{aligned}\partial_s F(s,x,t,y)&=\lambda(s,x)\int_{\mathbb{R}}\big[F(s,x,t,y)-F(s,\zeta,t,y)\big]\rho(s,x,\mathrm{d}\zeta),\\ \partial_t F(s,x,t,y)&=\int_{\mathbb{R}}\big[\lambda(s,\zeta)-\lambda(t,\zeta)\rho(t,\zeta,y)\big]F(s,x,t,\mathrm{d}\zeta),\end{aligned} \quad (5.8)$$

for any $s<t$, x and y. In the case of purely jump, the construction of the process $\{X(t):\geq 0\}$ is relatively simple and the path $t\mapsto X(t,\omega)$ are piecewise

constant (and normalized to be left-hand continuous) almost surely, however, the average or mean $t \mapsto \mathbb{E}\{X(t)\}$ is a continuous function.

Since Gaussian and Poisson processes are the prototypes of continuous and jump processes, we would like to quote some results related to the convergence of a sum of independent random variables to the Gaussian and Poisson distributions, e.g., see Gnedenko [64, Sections 49, pp. 336–339].

Let $\{\xi_{n,1}, \xi_{n,2}, \dots, \xi_{n,m_n}\}$ be a set of m_n independent random variables for $n = 1, 2, \dots$, and set $\zeta_n = \xi_{n,1} + \xi_{n,2} + \cdots \xi_{n,m_n}$ and denote by $F_{n,m}(x)$ the distribution function of the random variable $\xi_{n,m}$.

Suppose that we normalize the sequence so that

$$\lim_{n\to\infty} \sum_{m=1}^{m_n} \mathbb{E}\{\xi_{n,m}\} = 0, \qquad \lim_{n\to\infty} \sum_{m=1}^{m_n} \mathbb{E}\big\{\big[\xi_{n,m} - \mathbb{E}\{\xi_{n,m}\}\big]^2\big\} = a > 0.$$

Then the sequence of distribution functions $G_n(x)$ of the sum of random variables ζ_n converge to the Gaussian (or normal) distribution

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\big(-\frac{y^2}{2a}\big)\mathrm{d}y$$

if and only if

$$\lim_{n\to\infty} \sum_{m=1}^{m_n} \int_{|x|>\delta} x^2 F_{n,m}(\mathrm{d}x) = 0,$$

for every $\delta > 0$.

Alternately, suppose that we normalize the sequence so that

$$\lim_{n\to\infty} \sum_{m=1}^{m_n} \mathbb{E}\{\xi_{n,m}\} = \lambda > 0, \qquad \lim_{n\to\infty} \sum_{m=1}^{m_n} \mathbb{E}\big\{\big[\xi_{n,m} - \mathbb{E}\{\xi_{n,m}\}\big]^2\big\} = \lambda.$$

Then the sequence of distribution functions $G_n(x)$ of the sum of random variables ζ_n converge to the Poisson distribution

$$P(x) = \begin{cases} 0 & \text{for} \quad x \le 0, \\ \mathrm{e}^{-\lambda} \displaystyle\sum_{0\le k<x} \frac{\lambda^k}{k!} & \text{for} \quad x > 0 \end{cases}$$

if and only if

$$\lim_{n\to\infty} \sum_{m=1}^{m_n} \int_{|x-1|>\delta} x^2 F_{n,m}(\mathrm{d}x + \lambda_{n,m}) = 0,$$

for every $\delta > 0$, where $\lambda_{n,m} = \mathbb{E}\{\xi_{n,m}\}$.

Clearly, plenty of general results existent in the current literature regarding limits of sum of random variables, the point we remark is the fact that a normalization, the sum converges to a Gaussian distribution if the mean and the

variance have limits of different order, one zero and the other non-zero. However, if the mean and the variance have limits of the same order (non-zero) the sum converges to a Poisson distribution. This gives the two flavors, continuous and discontinuous (jumps) processes.

Now, let us discuss (Markov) jump processes from sample space point of view as in the classic book Doob [33, Section VI.2, pp. 255–273]. Similarly to the transition distribution function, an homogeneous transition probability function $P(x,t,A)$, x in a E (usually an open or closed subset of $\mathbb{R}^d$), $t>0$ and A in $\mathcal{B}(E)$, the Borel σ-algebra in E, satisfies: (a) for each $t>0$ and x in E the function $A \mapsto P(x,t,A)$ is a probability measure on $(E,\mathcal{B}(E))$, (b) for each $t>0$ and A in $\mathcal{B}(E)$ the function $x \mapsto P(x,t,A)$ is a (Borel) measurable, (c) for any x in E and A in $\mathcal{B}(E)$ we have

$$\lim_{t\to 0} P(x,t,A) = \delta_x(A),$$

i.e., the limit is equal to 1 if x belongs to A, and 0 otherwise, (d) for each $t,s>0$, x in E and A in $\mathcal{B}(E)$ we have

$$P(t+s,x,A) = \int_E P(t,x,\mathrm{d}y)\, P(s,y,A)$$

which is referred to as the Chapman-Kolmogorov identity.

If the continuity condition (c) above is strengthen into (or replace by)

$$\lim_{t\to 0} \inf_{x\in E} P(x,t,\{x\}) = 1, \quad \forall x \in E, \tag{5.9}$$

then Doeblin's result proves that there is a homogeneous piecewise constant Markov process $\{X(t) : t \geq 0\}$ with a cad-lag path and transition probability function $P(x,t,A)$. By piecewise constant (cad-lag) paths we means that if $X(t,\omega) = x$ then there exists a positive constant $\delta = \delta(t,\omega)$ such that $X(s) = x$ for every s in the stochastic interval $[t, t+\delta)$.

By means of the Chapman-Kolmogorov identity and under the continuity assumption (5.9), where the uniform limit in x is essential, we may use the transition probability function $P(x,t,A)$ to define the pair of functions $\lambda(x)$ and $\Lambda(x,A)$, for x in E and A in $\mathcal{B}(E)$ as follows:

$$\begin{aligned} \lim_{t\to 0^+} \frac{1-P(x,t,\{x\})}{t} &= \lambda(x), \\ \lim_{t\to 0^+} \frac{P(x,t,A)}{t} &= \Lambda(x,A), \end{aligned} \tag{5.10}$$

for any x in E, A in $\mathcal{B}(E \smallsetminus \{x\})$, where the definition is completed by setting $\Lambda(x,\{x\}) = 0$. Thus the function $x \mapsto \lambda(x)$ results bounded in E, the function $x \mapsto \Lambda(x,A)$ is Borel measurable for every fixed A in $\mathcal{B}(E)$, the function $A \mapsto \Lambda(x,A)$ is finite Borel measure on E (or in $E \smallsetminus \{x\}$) for every fixed x in E, and $\lambda(x) = \Lambda(x,E) = \Lambda(x, E\smallsetminus\{x\})$, so that $0 \leq \Lambda(x,A) \leq \lambda(x)$, for every x in E and $\mathcal{B}(E)$. Moreover, both convergence in (5.10) is uniform in x and A in $\mathcal{B}(E\smallsetminus\{x\})$,

x in E. Usually, we normalize the coefficients so that $\bar\Lambda(x,A)=\Lambda(x,A)/\Lambda(x,E)$ is a probability measure (with a distribution $\rho(x,A)$ if $E=\mathbb{R}$ as in the previous discussion) and $\bar\lambda(x)=\lambda(x)\Lambda(x,E)$.

Note that as in the initial discussion with the transition distribution function, for every x in E and A in $\mathcal{B}(E)$ we do have

$$P(x,t,A)=\big[1-t\lambda(x)\big]\delta(x,A)+t\Lambda(x,A)+t\varepsilon,$$

with $\varepsilon\to 0$ as $t\to 0^+$, uniformly in x and A in $\mathcal{B}(E)$. The *Backward* and *Forward* Kolmogorov *integro-differential* equations have the form

$$\begin{aligned}\partial_t P(x,t,A)&=\lambda(x)\int_E\big[P(x,t,A)-P(\zeta,t,A)\big]\Lambda(x,\mathrm{d}\zeta),\\ \partial_t P(x,t,A)&=\int_E\lambda(\zeta)\big[\Lambda(\zeta,E)-\Lambda(\zeta,A)\big]P(x,t,\mathrm{d}\zeta),\end{aligned}\tag{5.11}$$

for any $s<t$, x in E and A in $\mathcal{B}(E)$. The last equation takes also the form

$$\partial_t P(x,t,A)=\int_{E\smallsetminus A}\Lambda(\zeta,A)P(x,t,\mathrm{d}\zeta)-\int_A\Lambda(\zeta,E\smallsetminus A)P(x,t,\mathrm{d}\zeta).$$

Moreover, with suitable assumptions all this extends to non-homogeneous transition probability functions on Polish (separable, complete and metric) spaces.

Density and Intensity

The exponential distribution with intensity $\lambda>0$ of a random variable T is given by the expression $P\{T\geq t\}=\mathrm{e}^{-\lambda t}$ for any $t\geq 0$, and to accommodate the case λ, it is customary to allow $T=\infty$, i.e., if $\lambda=0$ then $T(\omega)=\infty$ for every ω; while if $\lambda\to\infty$ then $T(\omega)\to 0$ for every ω. Such a random variables are used to model 'waiting time' (of a certain event), i.e., the event arrives at the random time $T<\infty$ and it never arrives when $T=\infty$. A key property of this random variables T (or this distribution) is the so-called memory-less, which can be expressed as 'the conditional distribution $P\{T\,|\,T\geq t\}$ is the same as the full distribution $P\{T\geq t\}$'. When the typical Poisson process is viewed as a jumps process, we recognised jumps arrive at the times $T_1, T_1+T_2,\dots$ with $\{T_n\}$ being an IID sequence having exponential distribution.

Recall that all distribution in $[0,\infty]$ (or $\overline{\mathbb{R}}_+=[0,\infty[\cup\{\infty\}$) are given by a cadlag increasing function $\pi\colon[0,\infty[\to[0,1]$ such that its corresponding Stieltjes-Lebesgues measure $\mathrm{d}\pi$ is a probability measure on $[0,\infty]$, i.e., abusing notation, we can write $\pi(\{0\})=\pi(0)$, $\pi(]a,b])=\pi(b)-\pi(a)$, with $a<b<\infty$, and $\pi(\{\infty\})=1-\pi(\infty):=1-\lim_{t\to\infty}\pi(t)$. This, if T is a random variable with distribution π on $[0,\infty]$ then $P\{T\leq t\}=\pi([0,t])=\pi(t)$ and $P\{T>t\}=\pi(]t,\infty])=1-\pi(t)$, for any $0\leq t<\infty$, i.e., assuming T represents the arrival time of an 'event' then besides the obvious interpretations, $\pi(\{0\})$ is the probability of 'missing' the event (waiting since time zero), while $\pi(\{\infty\})$ is the probability that the event never arrive at a finite time. Also, a random variable

T with distribution π on $[0,\infty]$, which is absolutely continuous (with respect to the Lebesgue measure), has a density given by the limit

$$\lim_{h\to 0}\frac{P\{t\le T<t+h\}}{h}=\dot\pi(t),\quad \text{a.e. } y\ge 0$$

or equivalently, there exists an integrable non-negative function $\dot\pi$ satisfying

$$\begin{aligned}&\pi(]a,b[)=\int_a^b\dot\pi(t)\mathrm{d}t\le 1,\quad \forall b>a\ge 0,\quad\text{and}\\&\pi(\{\infty\})+\pi(\{0\})=1-\int_0^\infty\dot\pi(t)\mathrm{d}t,\end{aligned}\tag{5.12}$$

usually $\pi(\{0\})=0$, and most of the time $\pi(\{\infty\})=0$, which is the correct or strict meaning of 'density'. Its corresponding 'intensity' is defined by the limit

$$\lambda(t)=\lim_{h\to 0}\frac{P\{t<T<t+h\,|\,T>t,\}}{h}=\frac{\dot\pi(t)}{\pi(]t,\infty])}$$

or equivalently

$$\lambda(t)=\Big[-\ln\Big(\pi(\infty)+\int_t^\infty\dot\pi(s)\mathrm{d}s\Big)\Big]'=\big[-\ln\big(1-\pi(t)\big)\big]'=$$
$$=\frac{\dot\pi(t)}{1-\pi(t)}=\frac{[1-\pi(t)]'}{1-\pi(t)}$$

provided $\pi(t)<1$. Since

$$\int_0^t\lambda(s)\mathrm{d}s=\int_0^t\big[-\ln\big(1-\pi(s)\big)\big]'\mathrm{d}s=\ln\big(1-\pi(0)\big)-\ln\big(1-\pi(t)\big)$$

yields

$$\exp\Big(-\int_0^t\lambda(s)\mathrm{d}s\Big)=\frac{1-\pi(t)}{1-\pi(0)}$$

and so

$$-\Big[\exp\Big(-\int_0^t\lambda(s)\mathrm{d}s\Big)\Big]'=\lambda(t)\exp\Big(-\int_0^t\lambda(s)\mathrm{d}s\Big)=\frac{\dot\pi(t)}{1-\pi(0)}$$

which is valid as long as $\pi(0)<1$, any of these relations can be used to define the intensity when the density is known, and conversely, to obtain the density when the intensity is known. However if an intensity with a compact support (not simply such that $\int_0^\infty\lambda(t)\mathrm{d}t<\infty$) is to be allowed, then we should rethink the previous definition of density, since $\pi(t)<1$ may not be satisfied for every $t\ge 0$. Therefore, if an intensity vanishes after some value $t^\star<\infty$ then the same

happen with its density and really, we are dealing with a density in $[0, t^\star]$ and the role of ∞ is played by $t^\star$, i.e., instead of (5.12) we could require

$$\begin{aligned} &\pi(]a,b[) = \int_a^b \dot\pi(t)\mathrm{d}t \le 1, \quad \forall 0 \le a < b < t^\star, \\ &\pi(]t^\star,\infty]) = 0 \ \text{ and } \ \pi(\{t^\star\}) + \pi(\{0\}) = 1 - \int_0^{t^\star} \dot\pi(t)\mathrm{d}t, \end{aligned} \tag{5.13}$$

which means that $[t^\star, \infty]$ has been collapsed into the point $\{t^\star\}$, and thus, the point $t^\star$ could be considered as a 'coffin' state. Similarly, if the intensity vanishes prior to $t_\star > 0$, we could require

$$\pi([0,t_\star[) = \pi(]t^\star,\infty]) = 0 \ \text{ and } \ \pi(\{t^\star\}) + \pi(\{t_\star\}) = 1 - \int_{t_\star}^{t^\star} \dot\pi(t)\mathrm{d}t,$$

indeed, as mentioned below,

$$\int_t^{t^\star} \dot\pi(r)\mathrm{d}r > 0, \quad \forall t \in]t_\star, t^\star[,$$

is a convenient condition on the density $\dot\pi$ to assume.

There several points to remark:

• the values of a distribution π on $[0,\infty]$ at $\{0\}$ and $\{\infty\}$ are adjusted after the desired density/intensity has been chosen, i.e., if $[0, t^\star]$ (with $t^\star < \infty$) is the support of the intensity λ then (if necessary) the distribution π should have a mass at $t^\star$ instead of ∞, and still π can be considered as a distribution on $[0,\infty]$ with density $\dot\pi$ satisfying condition (5.13);

• if T is a random variable with distribution π on $[0,\infty]$, density/intensity in $\dot\pi/\lambda$, and such that $\pi(\{0\}) = 0$ (for simplicity), and

$$\pi(\{\infty\}) = P\{T = \infty\} = 1 - \int_0^{t^\star} \dot\pi(t)\mathrm{d}t = \exp\Big(-\int_0^{t^\star} \lambda(s)\mathrm{d}s\Big) > 0,$$

then $\mathbb{E}\{T\} = \infty$ and our interest turns on the moment, for any $n > 0$,

$$\mathbb{E}\big\{T^n \mathbb{1}_{\{T<\infty\}}\big\} = \int_0^{t^\star} t^n \dot\pi(t)\mathrm{d}t = \int_0^{t^\star} t^n \lambda(t) \exp\Big(-\int_0^t \lambda(s)\mathrm{d}s\Big)\mathrm{d}t,$$

which is finite under some (reasonable) conditions on the density/intensity, e.g., integrability of

$$t \mapsto t^n \dot\pi(t) = t^n \lambda(t) \exp\Big(-\int_0^t \lambda(s)\mathrm{d}s\Big)$$

on the interval $[0, t^\star[$ suffices, and in particular, when $t^\star < \infty$ this is automatically satisfied (or implicitly assumed);

• the intensity is very useful to calculate conditional probabilities for a random variable T with distribution π on $[0,\infty]$, i.e.,

$$P\{a<T<b\,|\,T>s\}=\pi(s,]a,b])=\frac{\pi(]a,b[\cap]s,\infty])}{\pi(]s,\infty])}=$$
$$=\frac{\pi(b)-\pi(a)}{1-\pi(s)}=\frac{\displaystyle\int_a^b\lambda(t)\exp\big(-\int_0^t\lambda(r)\mathrm{d}r\big)\mathrm{d}t}{\big(1-\pi(0)\big)\,\exp\big(-\displaystyle\int_0^s\lambda(r)\mathrm{d}r\big)}=$$
$$=\frac{1}{1-\pi(0)}\int_a^b\lambda(s+t)\exp\big(-\int_0^t\lambda(s+r)\mathrm{d}r\big)\mathrm{d}t$$

for any real numbers $b>a\geq s\geq 0$, and the last expression make sense as long as $\pi(\{0\})<1$, without requiring $1-\pi(t)=P\{T>s\}>0$ for every $s\geq 0$, actually, if $t^\star<\infty$ then the condition $1-\pi(s)=P\{T>s\}>0$ for every $0<s<t^\star$ makes sense for an intensity with a compact support;

• a distribution π, with $\pi(\{0\})=\pi(0)=0$ and having a given intensity $\lambda(t)$ are used to model the 'waiting time' for the next jump of a semi-Markov jump process $\{z_t:t\geq 0\}$, i.e., all jumps occur at times $T_1, T_1+T_2,\dots$ with $\{T_n\}$ a IID sequence with intensity $\lambda(z)$ (i.e.,

$$P\{T_n<t\}=\pi(t)=1-\exp\Big(-\int_0^t\lambda(s)\mathrm{d}s\Big),\quad\forall t\geq 0$$

and usually, $t\mapsto\lambda(t)$ is a non-negative Borel bounded (a minimum requirement would be locally integrable) function defined on $[0,\infty[$, and $\lambda(t)$ is interpreted as the jump-rate (i.e., the jumps take longer time to arrive as $\lambda(t)$ becomes close to zero, and also the jumps take shorter time to arrive as $\lambda(t)$ becomes large), and the value $\exp\Big(-\int_0^\infty\lambda(t)\mathrm{d}t\Big)$ represents the probability that jumps never arrive (meaning no an actual jump);

• if T is a random variable representing the arrival time of an event (say, a jump) with distribution π on $[0,\infty]$ and intensity λ as above, then the survivor function is defined as $\Psi(t)=P\{T>t\}$, i.e., whenever $\Psi(t)>0$ an event may arrive and if $\Psi(t)=0$ the event already occurred, and an expression of the form (with an intensity λ not identically zero)

$$P\{T>t\}=\Psi(t)=\mathbb{1}_{t<t^\star}\exp\Big(-\int_0^t\lambda(r)\mathrm{d}r\Big),\quad\forall t\geq 0,$$
$$\text{with}\;\;P\{T=t^\star\}=\pi(\{t^\star\})=\lim_{t\to t^\star}\Psi(t)<1,$$

makes sense for any positive $t^\star$ (finite or infinite), with $P\{T=0\}=\pi(\{0\})=0$, and the conditional probability density $\lim_{h\to 0}P\{T\in]s+t,s+t+h[\,|\,T>s\}/h$,

with $t, s \geq 0$, $s + t < t^\star$, can be calculated as

$$\lim_{h\to 0} \frac{\Psi(s+t) - \Psi(s+t+h)}{h\Psi(s)} = -\frac{\Psi'(s+t)}{\Psi(s)} = \\ = \lambda(s+t) \exp\Big(- \int_0^t \lambda(s+r)\mathrm{d}r\Big),$$

provided $\Psi(s) > 0$, moreover the case $\lambda = 0$ corresponds to $\Psi(t) = 1$ for every $t > 0$ (i.e., $\pi(\{t^\star\}) = 1$), while $t^\star = 0$ (or $\lambda \geq 0$ only Borel measurable but non integrable close to 0) is also an extreme situation with $\Psi(t) = 0$ for every $t > 0$ (i.e., $P\{T = 0\} = 1$);

• the special case where the intensity $\lambda(t)$ has support $[0, t^\star]$, with $t^\star < \infty$, means that a 'maximum waiting time' equal to $t^\star$ is in effect, i.e., the time elapse between two consecutive jumps cannot exceed the quantity $t^\star$, i.e., the cadlag semi-Markov process $\{z_t : t \geq 0\}$ should such that $z_t = z_{t-}$ for every $\tau \leq t \leq \tau + t^\star$ has probability zero, for any stopping time τ, or in other words, a jumps must arrive in at most $t^\star$ unit of time;

• if the waiting time s is added to the state z of a cadlag semi-Markov process $\{z_t : t \geq 0\}$ on E then the couple $\{(z_t, s_t) : t \geq 0\}$ becomes a cadlag Markov process in $E \times [0, s^\star[$, $s^\star = t^\star$, with infinitesimal generator

$$Av(z,s) = \partial_s v(z,s) + \lambda(z,s)\Big[\int_{E\times[0,s^\star[} \big(v(\zeta,0) - v(z,s)\big) Q(z,s,\mathrm{d}\zeta)\Big],$$

where $Q(z, s, \mathrm{d}\zeta)$ is the conditional distribution of jumps given the state (z, s), which is constructed on the basis of a given kernel/distribution $Q(z, \mathrm{d}\zeta)$ of jumps occurring at times with jump-rate $\lambda(z, s)$ ($\lambda(z, s) = 0$ for any $s > s^\star$), usually $s^\star = \infty$ and $\lambda(z, s)$ is a given non-negative bounded and continuous function, for instance, see Davis [30, Appendix, pp. 256–279], Gikhman and Skorokhod [63, Section III.3, pp. 226–249], Jacod [82], and Robin [152], among others.

Let us consider a particular case of piecewise deterministic processes (PDP) as a (time-homogeneous) Markov process with an infinitesimal generator of the form

$$A_y v(y) = g(y)\partial_y + \lambda(y) \int_{I^\star} \big(v(z) - v(y)\big) Q(y, \mathrm{d}z), \quad y \in I^\star,$$

where either $I^\star = [0, \infty[$ or $I^\star = [0, y^\star]$ with $0 < y^\star < \infty$, and

(a) $y \mapsto \lambda(y)$ is a non-negative Borel function defined on $[0, \infty[$ with support (as a distribution) on $[0, y^\star]$ when $y^\star < \infty$, see also (d);

(b) g: $[0, \infty[\mapsto [0, \infty[$ a Lipschitz continuous function satisfying $g(0) \geq 0$ and also $g(y^\star) > 0$ if $y^\star < \infty$, so that $\mathtt{Y}(y, t)$ is a forward/backward flow on $I^\star$ (i.e., it solves $\dot{y}(t) = g(y(t))$, $y(0) = y$) and if $y^\star < \infty$ then the first exit time from $I^\star$,

$$\mathtt{T}(y) = \inf\big\{t > 0 : \mathtt{Y}(y, t) \notin I^\star\big\},$$

is defined for every y in $I^\star$, with the convention that $\mathtt{T}(y) = \infty$ if $\mathtt{Y}(y,t)$ belongs to $I^\star$ for every $t > 0$ (so that, there is no need to define $\mathtt{T}(y) = \infty$ for every y, when $y^\star = \infty$), and moreover, the map $y \mapsto \mathtt{T}(y)$ is continuous on the set $\{y \in I^\star : \mathtt{T}(y) < \infty\}$;

(c) $Q(y, \mathrm{d}z)$ a stochastic kernel on $I^\star$ (i.e., $Q(y,\cdot)$ is a probability measure on $I^\star$ for each fixed y in $I^\star$, and $Q(\cdot, B)$ is Borel measurable for any fixed B in $\mathcal{B}(I^\star)$, Borel σ-algebra of subsets of $I^\star$) such that $Q(y, \{y\}) = 0$;

(d) $t \mapsto \lambda\big(\mathtt{Y}(y,t)\big) \geq 0$ is an integrable function (intensity) on $[0, \varepsilon]$, for some $0 < \varepsilon < \mathtt{T}(y)$;

(e) to prevent accumulation of small jumps, e.g., assume that $\lambda(y)$ is bounded by a constant c in $I^\star$ and either $\mathtt{Y}(y,t)$ belongs to $I^\star$ for every $t > 0$ or there exists $\varepsilon > 0$ such that $Q(y, \{y \in I^\star : \mathtt{T}(y^\star) \geq \varepsilon\}) = 1$.

Consider the survivor function of jumps times

$$\Psi(y,r) = \mathbb{1}_{r<\mathtt{T}(y)} \exp\Big(- \int_0^r \lambda(\mathtt{Y}(y,s))\mathrm{d}s\Big), \quad \forall y, r \in I^\star$$

and its (generalized) inverse

$$\Psi^{-1}(y,u) = \inf\big\{r \geq 0 : \Psi(y,r) \leq u\big\}$$

with the convention that $\Psi^{-1}(y,u) = \infty$ if $\Psi(y,r) > u$ for every $r \geq 0$. This means that if U is a uniformly distributed random variable on $[0,1]$ then $P\{\Psi^{-1}(y,U) > r\} = \Psi(y,r)$, i.e., the random variable $T = \Psi^{-1}(y,U) \leq \mathtt{T}(y)$ has intensity $\lambda(\mathtt{Y}(y,s))$ on $[0, \mathtt{T}(y)]$, with

$$P\{T = \mathtt{T}(y)\} = \lim_{r \uparrow \mathtt{T}(y)} \Psi(y,r) = \exp\Big(- \int_0^{\mathtt{T}(y)} \lambda(\mathtt{Y}(y,s))\mathrm{d}s\Big),$$

and T represents the waiting time for the next jump, while in the path $t \mapsto \mathtt{Y}(y,t)$. Also, assumptions (c) and (e) imply that there exists a Borel function on the canonical space $([0,1], \ell)$ satisfying $\Upsilon : I^\star \times [0,1] \to \{y \in I^\star : \mathtt{T}(y) \geq \varepsilon\}$, and $\ell(\{u : \Upsilon(y,u) \in B\}) = Q(y,B)$, for every B in $\mathcal{B}(I^\star)$.

Now a realization of the Markov jump (or piecewise deterministic) process (with the characteristics (g, Q, λ) on $I^\star$ and the above infinitesimal generator A_y) can be construct by means of a IID sequence $\{U_k : k \geq 1\}$ of uniformly distributed random variable on $[0,1]$ as follows, with the jump-times $T_i = \Psi^{-1}(y, U_{2i})$ and the jumps $Y_i = \Upsilon(y, U_{2i+1})$, $i = 1, 2, \ldots$, and initialize the procedure by setting $i = 1$ to begin, for each given ω and initial value $Y_0 = y$ in $I^\star$, with:

1.- if $T_i = T_i(y) = \Psi^{-1}(y, U_{2i}) = \infty$ then $y_t = \mathtt{Y}(y,t)$, $t \geq 0$, and stop;

2.- if $T_i = T_i(y) = \Psi^{-1}(y, U_{2i}) < \infty$ then $y_t = \mathtt{Y}(y,t)$, $0 \leq t < T_i$ and $y_{T_i} = Y_i = \Upsilon(y_{T_i}, U_{2i+1})$;

3.- reset $y = y_{T_i}$ and increase i to $i+1$ to restart with 1 and 2 (the same recipe), repeating all over again.

Note that in each step if $\mathtt{T}(Y_{i-1}) < \infty$ then $T_i \le \mathtt{T}(Y_{i-1})$, so that y_t belongs always to $I^\star$ (i.e., y_t cannot leave the compact region $I^\star = [0, y^\star]$ when $y^\star < \infty$, in other words, if $y_t = y^\star$ then an immediate jump occurs, and the process is sent back to the $[0, y^\star[$). Therefore the process y_t constructed above should be considered as defined on $I^\star$, i.e., either $[0, \infty[$ or $[0, y^\star]$ when $y^\star < \infty$. This algorithm effectively define the Markov piecewise deterministic process y_t and the Markov jump process $z_t = \sum_i \mathbb{1}_{t \ge T_i}$, for any time $t \ge 0$. Making a good use of the 'Martingale Theory', it can be proved (see Davis [30, (24.6) Proposition, pp. 60–61]) that under assumption (d), $\mathbb{E}\{N_t\} \le (c + 1/\varepsilon)t + 1$ for any $t \ge 0$, so that there is not accumulation of small jumps anywhere.

If $y^\star < \infty$ then this (time-homogeneous) Markov piecewise deterministic process would have the desired infinitesimal generator A_y with some 'boundary' conditions at $y = 0$ and $y = y^\star$. Indeed, there is no boundary conditions at $y = 0$, because the drift points to the right and therefore, the process y_t cannot exit at $y = 0$ (i.e., $\mathtt{T}(0) > 0$), and in view of (d), we have $P\{T = 0\} = \lim_{r \downarrow 0} \Psi(y, r) = 0$, i.e., there is no an immediate jump at $y = 0$. On the contrary, the boundary condition at $y = y^\star$ is more delicate, as mentioned early, it could be an immediate jump at $y^\star = 0$ sending the process back to $[0, y^\star[$. Actually, depending on the forward/backward flow $\mathtt{Y}(y^\star, t)$ (or drift g at $y^\star$) at the boundary $y^\star$ it could be only two possibilities, either (a) $y^\star$ is, with probability 1, not hit by the process y_t from any (other) starting point, or else (b) $P\{T = \Psi^{-1}(y, U) = t\} \to 1$, for some y in $[0, y^\star[$ (see Davis [30, pp. 60–61], after (24.6) Proposition, definition of Γ). In case (a), the stochastic kernel $Q(y, \cdot)$ need not to be specify (i.e., no boundary condition is necessary), and in case (b), the extended generator requires the boundary condition

$$v(y^\star) = \int_{I^\star} v(z) Q(y^\star, \mathrm{d}z),$$

and some integrability conditions on the predictable jumps, indeed, case (b) is equivalent to the condition: $P\{T = \Psi^{-1}(y, U) = t\} > 0$ for some $t > 0$ with $y = \mathtt{Y}(-t, y^\star)$, with U an uniformly distributed random variable, see Davis [30, Section 26, pp. 66–74]. This difficulty at $y^\star$ can be avoided if the process y_t is stopped at the first hitting time of $y^\star$, which produces the boundary condition $v(y^\star) = 0$, and the process 'live' in $[0, y^\star[$. Certainly, this procedure works fine with an extra single variable following the equation $\dot{t} = 1$, or even in a region of $\mathbb{R}^d$ (instead of $I^\star = [0, y^\star]$), provided the boundary points are property discussed.

Actually, our interest is in even a particular case, with $g(y) = 1$ and $Q(y, \mathrm{d}z) = \delta$ (i.e., all jumps are concentrated at $y = 0$), which yields

$$A_y v(y) = \partial_y v(y) + \lambda(y)\big(v(0) - v(y)\big), \quad y \in I^\star,$$

as the infinitesimal generator. This (time-homogeneous) Markov process $\{y_t : t \ge 0\}$ represents the 'time elapsed since the last signal' (or the 'waiting time' for a signal), where 'signal' means a jump or in general, the event under consideration. To decide on the two cases (a) and (b) for this special situation, we

remark that $y = \mathtt{Y}(-t, y^\star) = y^\star - t$, and $\mathtt{T}(\mathtt{Y}(-t, y^\star)) = t > 0$, for any $t > 0$ sufficiently small, to deduce

$$\begin{aligned}
&P\{T = \Psi^{-1}(y, U) = t\} = \\
&= \lim_{\varepsilon \downarrow 0} \big[P\{\Psi^{-1}(y^\star - t, U) > t - \varepsilon\} - P\{\Psi^{-1}(y^\star - t, U) > t\}\big] = \\
&= \lim_{\varepsilon \downarrow 0} \big[\Psi(y^\star - t, t - \varepsilon) - \Psi(y^\star - t, t)\big] = \\
&\qquad\qquad = \lim_{\varepsilon \downarrow 0} \Psi(y^\star - t, t - \varepsilon) = \exp\Big(- \int_0^t \lambda(y^\star + s - t) \mathrm{d}s \Big),
\end{aligned}$$

and because λ is bounded (it suffices assumption (d), i.e., that λ is integrable close to 0), we obtain $P\{T = \Psi^{-1}(y, U) = t\} \to 1$, i.e., we are in the case (b), namely, the boundary condition

$$v(y^\star) = \int_{I^\star} v(z) Q(y^\star, \mathrm{d}z) = v(0),$$

should be used, namely, for the signal process y_t representing the 'time elapsed since last signal' a periodic boundary condition should be used.

Given a bounded intensity $0 \le \lambda(y) \le c$ on $[0, \infty[$ and a stochastic kernel $Q(y, \mathrm{d}z)$ on $\mathbb{R}^d$ satisfying $Q(y, \{y\}) = 0$ and represented by $\Upsilon(y, u)$ as $\ell(\{u : \Upsilon(y, u) \in B\}) = Q(y, B)$, for every B in $\mathcal{B}(\mathbb{R}^d)$, where ℓ is the Lebesgue measure on $[0, 1]$. To represent/combine $\lambda(y)$ and $Q(y, \mathrm{d}z)$ into a Lévy measure $\mathtt{M}(y, \mathrm{d}z)$, we can use the Lebesgue measure ℓ' on $[0, 1]$ to define $\gamma(y, u, u') = [\Upsilon(y, u) - y] \mathbb{1}_{\{cu' \le \lambda(y)\}}$, $\pi(\mathrm{d}u, \mathrm{d}u') = \ell(\mathrm{d}u)\, c\, \ell'(\mathrm{d}u')$ and to deduce

$$\begin{aligned}
&\int_{[0,1]} [\varphi(y + \gamma(y, u, u')) - \varphi(y)]\, c\, \ell'(\mathrm{d}u') = \\
&\qquad = c \int_0^{\lambda(y)/c} [\varphi(\Upsilon(y, u)) - \varphi(y)] \mathrm{d}u' = \lambda(y)[\varphi(\Upsilon(y, u)) - \varphi(y)],
\end{aligned}$$

which implies

$$\begin{aligned}
&\int_{[0,1]\times[0,1]} [\varphi(y + \gamma(y, u, u')) - \varphi(y)] \pi(\mathrm{d}u, \mathrm{d}u') = \\
&= \lambda(y) \int_0^1 [\varphi(\Upsilon(y, s, u)) - \varphi(y)] \ell(\mathrm{d}u) = \\
&\qquad\qquad = \lambda(y) \int_{\mathbb{R}^d_*} [\varphi(z) - \varphi(y)] Q(y, \mathrm{d}z),
\end{aligned}$$

so that

$$\mathtt{M}(y, \mathrm{d}y') = \pi\big\{(u, u') \in [0, 1] \times [0, 1] : y + \gamma(y, u, u') \in \mathrm{d}y'\big\}$$

is a (uniformly in y) bounded Lévy measure on $\mathbb{R}^d_*$. Certainly, if $\lambda(y)$ is finite but not necessarily bounded by the constant c then the above relation would be

valid for $c \wedge \lambda(y)$ instead of $\lambda(y)$, and final key relations hold true as $c \to \infty$, but now, the Lévy measure $\mathtt{M}(y, \mathrm{d}y')$ is not necessarily uniformly bounded in y; however, in this case it is better to use an unbounded measure on $[0,1]$, e.g., $\mathrm{d}u'/u'$ instead of ℓ', with $\gamma(y,u,u') = [\Upsilon(y,u) - y]\mathbb{1}_{\{\mathrm{e}^{-\lambda(y)} \le u' < 1\}}$.

Another key estimate is

$$\begin{aligned}\int_{[0,1]} \big|\gamma(y,u,u') - \gamma(\bar{y},u,u')\big|\, c\, \ell'(\mathrm{d}u') = \\ = c\Big|\int_0^{\lambda(y)/c} \Upsilon(y,u)\mathrm{d}u' - \int_0^{\lambda(\bar{y})/c} \Upsilon(\bar{y},u)\mathrm{d}u'\Big| = \\ = \big|\lambda(y)\Upsilon(y,u) - \lambda(\bar{y})\Upsilon(\bar{y},u)\big|,\end{aligned}$$

which implies

$$\begin{aligned}\int_{[0,1]\times[0,1]} \big|\gamma(y,u,u') - \gamma(\bar{y},u,u')\big|\pi(\mathrm{d}u,\mathrm{d}u') = \\ = \int_{[0,1]} \big|\lambda(y)\Upsilon(y,u) - \lambda(\bar{y})\Upsilon(\bar{y},u)\big|\mathrm{d}u\end{aligned}$$

and some conditions on the intensity $\lambda(y)$ and the jumps-size $\Upsilon(y,u)$ can be imposed to obtain regularity (e.g., locally Lipschitz in y, linear growth in y) so that the stochastic ordinary differential equations (SODEs) with jumps (Lévy measure) in $\mathbb{R}^d$ as above.

For instance, $Q(y,\mathrm{d}z) = \delta(z-y)\mathrm{d}z$, $\gamma(y,u) = -y\mathbb{1}_{\{cu \le \lambda(y)\}}$, and $\tilde{\pi}(\mathrm{d}u,\mathrm{d}t)$ the martingale measure corresponding with $\pi(\mathrm{d}u) = c\,\mathrm{d}u$. Actually, all this can be reviewed and the following assertions are true:

(a) Given a non-negative locally Lipschitz function $\lambda(y)$, bounded (by $c > 0$) and supported in $[0,b]$ (this implies $\lambda(y) = 0$ for any $y \ge b$) we can consider the intensity measure $\pi_c(\mathrm{d}\zeta) = c\,\mathbb{1}_{\{0<\zeta\le 1\}}\mathrm{d}\zeta$ and the jump-coefficient $\gamma(y,\zeta) = -y\mathbb{1}_{\{c\zeta \le \lambda(y)\}}$ which satisfies

$$\begin{aligned}&\int_{]0,1[} [\varphi(y+\gamma(y,\zeta)) - \varphi(y)]\pi_c(\mathrm{d}\zeta) = \lambda(y)[\varphi(0) - \varphi(y)],\\ &\int_{]0,1[} |\gamma(y,\zeta))|\pi_c(\mathrm{d}\zeta) = y\,\lambda(y),\\ &\int_{]0,1[} |\gamma(y,\zeta) - \gamma(y',\zeta)|\pi_c(\mathrm{d}\zeta) \le |y-y'|\lambda(y) + y'|\lambda(y') - \lambda(y)|,\end{aligned}$$

for every y, y' in $[0,\infty[$ (in particular $[0,b]$). Hence, the SDE

$$y(t) = y(0) + t + \int_{]0,t]\times[0,1]} \gamma(y(r),\zeta)\tilde{\pi}_c(\mathrm{d}r,\mathrm{d}\zeta), \quad t \ge 0$$

(where $\tilde{\pi}_c(\mathrm{d}\zeta)$ is the martingale measure corresponding to the Lévy/intensity measure π_c) defines a Markov-Feller process on either $[0,\infty[$ or $[0,b[$ with

$$A_y v(y) = \partial_y v(y) + \lambda(y)[v(0) - v(y)]$$

as its infinitesimal generator. If $b = \infty$ then the process is considered in $[0, \infty[$ and the boundary condition on $y = b$ seems 'non-existent' or 'natural' (i.e., it never reaches $y = \infty$). However, if the process is considered in $[0, b]$ (compact) and $y = b < \infty$ then there is a jump (to zero) immediately, and therefore, this is a jump-discontinuity with full probability, i.e., it could not be a Feller process, unless the jump at $y = b$ is not seen, by assuming periodic conditions $v(0) = v(b)$. Alternatively, we may decide to stop (immediately) the process at $y = b$, and it becomes a Markov-Feller process on the compact $[0, b]$, which yields the boundary condition $v(b) = 0$ for the infinitesimal generator.

(b) The above Lévy measure $\pi_c(\mathrm{d}\zeta)$ [which yields the martingale measure $\tilde{\pi}_c(\mathrm{d}\zeta)$] defines a compound Poisson process which can be construct as follows: begin with a Poisson process N_t (which, itself, can be construct from a sequence of IID random variables with exponential distribution) with rate c (constant), i.e., $P\{N_t = n\} = \mathrm{e}^{-ct}(ct)^n/n!$, with mean $\mathbb{E}\{N_t\} = ct$ and variance $\mathbb{E}\{(N_t - ct)^2\} = ct$, and if $s_1 < s_2 < \cdots$ are the time of jumps (all jumps have size 1) then $P\{s_{k+1} - s_k \geq t\} = \mathrm{e}^{-ct}$; and then find an another sequence $\{z_k\}$ of IID (and independent of N_t) with values in $]0, b]$ and distribution $\pi_c(\mathrm{d}\zeta)/c$ (remark that we assume that the distribution does not charge zero, i.e., $z_k > 0$); thus the $[0, \infty[$-valued process

$$t \mapsto Z_t = \sum_{k=1}^{N_t} z_k = \sum_k z_k \mathbb{1}_{s_k \leq t}$$

is the desired compound Poisson process, and the stochastic integral can be (also) written as

$$\begin{aligned}
&\int_{]0,t]\times]0,1[} \gamma(y(r), \zeta)\tilde{\pi}_c(\mathrm{d}r, \mathrm{d}\zeta) = \\
&= \sum_{k=1}^{\infty} \gamma(y(s_k-), z_k)\mathbb{1}_{s_k \leq t} - \int_0^t \mathrm{d}r \int_{]0,1[} \gamma(y(r-), \zeta)\pi_c(\mathrm{d}\zeta) = \\
&\quad = -\sum_{k=1}^{\infty} y(s_k-)\mathbb{1}_{z_k \leq \lambda(y(s_k-))}\mathbb{1}_{s_k \leq t} - \int_0^t \lambda(y(r-))\, y(r-)\mathrm{d}r, \quad t \geq 0.
\end{aligned}$$

which is a finite stochastic sum and a regular integral.

(c) Moreover, as deduced in (b), the SDE can (also) be written

$$\begin{aligned}
y(t) = y(0) + \int_0^t \big(1 - y(r)\lambda(y(r-))\big)\mathrm{d}r - \\
- \sum_{k=1}^{\infty} y(s_k-)\mathbb{1}_{z_k \leq \lambda(y(s_k-))}\mathbb{1}_{s_k \leq t}, \quad t \geq 0,
\end{aligned}$$

and it is clear that the process $y(t)$ remains in $[0, b]$, i.e., it cannot exit at $y = 0$ and it has an immediate jump (back to 0) at $y = b$.

5.1.2 General Discussion

There is an important class of processes known as Markov processes which are used to model dynamical systems under disturbances. They are based on the principle that the *future* is independent of the *past* when we know the *present*. Similar to the *state variable* model for deterministic dynamical systems. Essentially, it is a matter of what is called *state* so that any dynamical process can be view a Markov process with a *larger* state. However, the price of the Markov character is the lack of differentiability in time of the process as we will see later. It is convenient to assume that state-space is a complete metric space (i.e. a Polish space) and that the index set T has a natural order e.g., T is a subset of $\mathbb{R}$. In most of our cases $T = [0, \infty)$ and E is a either closed or open subset of $\mathbb{R}^d$, but more general situations are also very interesting.

From the analysis viewpoint, 'Markov processes' are particular (or special) cases of the semigroups theory on Banach or Hilbert spaces, in particular (positive and contraction semigroups in $B(E)$, the Banach space of Borel and bounded real-valued functions on E. Several aspects of the theory of Markov process are necessary to discuss, and their connections are complicate, essentially, a key element of a Markov process is its transition probability function $P(s, x, t, A)$, which also define a semigroup (in one or two parameters) $\Phi(s, t)$ acting on some $B(\mathcal{O})$ under the relation $\Phi(s, t)f(x) = P(s, x, t, f)$, with the common notation the $P(s, x, t, f)$ means the integral of $f(\xi)$ with respect to $P(s, x, t, \mathrm{d}\xi)$. All this will be make clear later, after some discussion and definitions, but it suffices to mention that only 'normal' Markov processes are usable, and normal means that $\Phi(s, t)$ maps $B(E)$ into itself.

A stochastic process X on a (complete) probability space $(\Omega, \mathcal{F}, P)$ and values in a Polish space E satisfies the *Markov property* if for any $n = 1, 2 \ldots$, any bounded measurable (actually continuous suffices, because E is a complete metric space) functions $f_1, \ldots, f_n$, $g_1, \ldots, g_n$, h, and times $s_1 \leq \cdots \leq s_n \leq t \leq t_1 \leq \cdots \leq t_n$ we have

$$\mathbb{E}\big\{h(X_t)\big(\prod_{i=1}^{n} f(X_{s_i})\big)\big(\prod_{i=1}^{n} g(X_{t_i})\big)\big\} = \\ = \mathbb{E}\big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n} f(X_{s_i}) \,|\, X_t)\big\} \prod_{i=1}^{n} g(X_{t_i})\big\}, \quad (5.14)$$

where $\mathbb{E}\{\prod_{i=1}^{n} f(X_{s_i}) \,|\, X_t\}$ is X_t-measurable functions satisfying

$$\mathbb{E}\big\{h(X_t) \prod_{i=1}^{n} f(X_{s_i})\big\} = \mathbb{E}\big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n} f(X_{s_i}) \,|\, X_t)\big\}\big\},$$

i.e., it is the conditional expectations with respect to the σ-algebra generated by the random variable X_t. This is briefly expressed by saying that the future is independent of the past given the present. Clearly, this condition involves only the finite-dimensional distributions of the process, and (5.14) is equivalent to

(e.g., see Blumenthal and Getoor [15, Thm 1.3, pp. 12-14]) either

$$P(X_t \in B \mid X_{s_1}, \dots, X_{s_n}, X_s) = P(X_t \in B \mid X_s), \quad \text{a.s.}$$

for every $t > s \geq s_n > \cdots > s_1$, B in $\mathcal{B}(E)$, or

$$\mathbb{E}\{f(X_t) \mid X_{s_1}, \dots, X_{s_n}, X_s\} = \mathbb{E}\{f(X_t) \mid X_s\}, \quad \text{a.s.},$$

for every $t > s \geq s_n > \cdots > s_1$, and for any arbitrary bounded and continuous (actually, with compact support when E is locally compact) function f from E into $\mathbb{R}$.

Definition 5.1 (history)**.** Given a stochastic process X on a (complete) probability space $(\Omega, \mathcal{F}, P)$ we can define the *history* (or internal history or strict history) of the process as the increasing family of σ-algebras $\{\mathcal{H}(t) : t \in T\}$, where each $\mathcal{H}(t)$ is generated by the random variables $\{X(s) : s \leq t\}$ and the null sets. Similarly, the *innovation* $\{\mathcal{H}^{\perp}(t) : t \in T\}$ is the decreasing family of σ-algebras, where each $\mathcal{H}^{\perp}(t)$ is generated by all sets in some $\mathcal{H}(s)$ for $s > t$ which are independent of $\mathcal{H}(t)$. □

The internal history $\{\mathcal{H}(t) : t \in T\}$ of a process X is also denoted by $\{\mathcal{F}^X(t) : t \in T\}$ (or $\{\mathcal{F}_X(t) : t \in T\}$ or with $\mathcal{H}$ replacing $\mathcal{F}$) and contains (or records) all events linked to the process X, up to (and including) the time t, i.e., past and present. From the system-science point of view, the history $\{\mathcal{F}^X(t) : t \in T\}$ is best thought as an *increasing information pattern*. On the other hand the innovation $\{\mathcal{H}^{\perp}(t) : t \in T\}$ records all events linked to the process X, after time t and is unrelated to (independent of) the past.

Based on the observation of a stochastic process up to the present time we can know whether a *causal* phenomenon has (or has not) already taken place. If *causally* is understood in this way, a random variable τ with values in $[0, \infty]$ can be interpreted as a random time of occurrence of some phenomenon depending causally upon the process X when the event $\{\omega : \tau(\omega) \leq t\}$ is $\mathcal{F}^X(t)$-measurable, which correspond to the notion of optional or stopping times previously mentioned.

Most of the processes that we are going to discuss will be cad-lag, and the history $\{\mathcal{H}(t) : t \in T\}$ will be right-continuous and therefore be equal to the canonical filtration (associated with the given process), after being augmented with all zero-measure sets. By construction $\mathcal{H}(t)$ is independent of $\mathcal{H}^{\perp}(t)$ for any $t \in T$, $\mathcal{H}(t)$ represents the *past* and *present* information at time t and $\mathcal{H}^{\perp}(t)$ is the *future* new information to come.

Thus, another process Y is said to be *adapted* to X if $Y(t)$ is measurable with respect to $\mathcal{H}(t)$ for any $t \in T$. Similarly, the process Y is *non-anticipative* with respect to X if the random variable $Y(t)$ is independent of $\mathcal{H}^{\perp}(t)$ for any $t \in T$. It is clear that if Y is adapted to X then Y is non-anticipative with respect to X, but the converse does not hold in general.

Actually, we do not need a process X to define the *innovation*, if we start from a filtration $\{\mathcal{F}(t) : t \in T\}$ we can define its innovation or *independent*

complement $\{\mathcal{F}^{\perp}(t) : t \in T\}$, and then we can say that a process X is either adapted or non-anticipative with respect to the filtration $\{\mathcal{F}(t) : t \in T\}$.

At this point, the Markov property (5.14) can be re-phrased as

$$P(X_t \in B \mid \mathcal{H}_s) = P(X_t \in B \mid X_s), \quad \text{a.s.} \quad \forall t > s,\ B \in \mathcal{B}(E),$$

where $\mathcal{H}_t = \mathcal{H}(t) = \mathcal{H}_X(t)$. The Markov property identifies only how the finite-dimensional distributions of the process interact themselves or evolve in time.

Definition 5.2 (Markov)**.** A Markov process with states in $E \subset \mathbb{R}^d$ is a (complete) probability measure P on $(\Omega, \mathcal{F})$, together with a measurable mapping X (P-equivalence class) from $(\Omega, \mathcal{F})$ into $(E^T, \mathcal{B}^T(E))$ and an increasing family of completed σ-algebras $(\mathcal{F}_t : t \in T)$ on $(\Omega, \mathcal{F})$ satisfying the *Markov property*

$$P(X_t \in B \mid \mathcal{F}_s) = P(X_t \in B \mid X_s), \quad \text{a.s.} \quad \forall t > s,\ B \in \mathcal{B}(E).$$

If the family of σ-algebras $(\mathcal{F}_t : t \in T)$ is not mentioned, then it is assumed that $(\mathcal{F}_t : t \in T)$ is the history $(\mathcal{H}(t) : t \in T)$ of the process X. Moreover, if $(\mathcal{F}_t : t \in T)$ is a filtration satisfying the usual conditions and the paths of $(X_t : t \in T)$ are cad-lag, except in a set of P-probability zero, then $(P, X_t : t \in T)$ is called a *cad-lag* Markov process. □

As mentioned early, we are concerned with E-valued Markov processes where $E \subset \mathbb{R}^d$, and because cad-lag is usually assumed, the sample space Ω will be a Polish (separable, complete and metric) space as discussed later. However, the above definition is meaningful when E is a Polish, and even when E is only a Lusin space (homeomorphic to a Borel subset of a compact metric space).

In the above Markov property, assuming we have taken a regular conditional probability, the equality is true except on a set of probability zero which may depend on t, s. Thus some regularity is necessary on path of the process to completely identify the process in term of its finite-dimensional distributions. In order to avoid extra difficulties, we consider only cad-lag Markov processes, where the Markov property is satisfied with a complete and right-continuous increasing family of σ-algebras and the path of the process may have only discontinuities of first kind, which are normalized to be cad-lag. The larger the σ-algebras of the filtration $(\mathcal{F}_t : t \in T)$ are, the more significant is the assertion that $(P, X_t, \mathcal{F}_t : t \in T)$ has the Markov property. Thus, the process $(X_t : t \in T)$ is adapted to $(\mathcal{F}_t : t \in T)$ and the filtration $(\mathcal{F}_t : t \in T)$ is non-anticipative i.e., $\mathcal{F}_t$ is independent of $\mathcal{H}^{\perp}(t)$ for any t in T. Note that the Markov property can be re-phased as follows: for every time t the σ-algebra $\mathcal{F}_t$ is independent of $\sigma(X_s : s \geq t)$ given X_t.

In most of the literature, the word *standard* Markov processes refer to cad-lag Markov processes which are also *quasi-left-continuous* and satisfy the strong Markov property (as discussed later). It will become clear that the strong Markov property is highly desired, however, some applications involving *deterministic impulses* yield cad-lag Markov processes which are not quasi-left-continuous.

Usually, when talking about a Markov process we do not refer to a single process, we really mean a family of processes satisfying the Markov property and some given initial distribution. The following concept of transition function is then relevant if we can explicitly write

$$P\{X_t \in A \,|\, X_s = x\} = P(s, x, t, A), \quad \forall s < t,\ x \in E,\ A \in \mathcal{B}(E),$$

for some function $P(s, x, t, A)$. Note that

$$P\{X_t \in A \,|\, X_s = x\} = \frac{P(\{X_t \in A,\ X_s = x\})}{P(\{X_s = x\})} = \\ = \frac{1}{P(\{X_s = x\})} \int_{\{X_s = x\}} P\{X_t \in A \,|\, X_s\}(\omega) P(\mathrm{d}\omega),$$

whenever $P(\{X_s = x\}) > 0$ and $P\{X_t \in A \,|\, X_s = x\} = 0$ when $P(\{X_s = x\}) = 0$, under the condition that a regular conditional probability exists.

Definition 5.3 (transition)**.** A *transition probability function* on a given measurable space $(E, \mathcal{F})$, is a function $P(s, x, t, A)$ defined for $s < t$ in T (T is equal to $[0, +\infty)$ or $(-\infty, +\infty)$ in most of our cases), x in E and A in $\mathcal{F}$ such that

(a) for each $s < t$ in T and x in E the function $A \mapsto P(s, x, t, A)$ is a probability measure on $(E, \mathcal{F})$,

(b) for each $s < t$ in T and A in $\mathcal{F}$ the function $x \mapsto P(s, x, t, A)$ is a measurable,

(c) for any s in T, x in E and A in $\mathcal{F}$ we have

$$\lim_{t \to s} P(s, x, t, A) = \delta_x(A),$$

i.e., the limit is equal to 1 if x belongs to A, and 0 otherwise,

(d) for each $s < r < t$ in T, x in E and A in $\mathcal{F}$ we have

$$P(s, x, t, A) = \int_E P(s, x, r, \mathrm{d}y)\, P(r, y, t, A),$$

which is referred to as the Chapman-Kolmogorov identity. It is called *homogeneous* if $P(s, x, t, A) = P(0, x, t - s, A)$ for any $t > s$ in $T = [0, +\infty)$ (or $T = \{0, 1, 2, \dots\}$), x in E and any Borel measurable subset A of E, in this case we will denote $P(0, x, r, A)$ by $P(r, x, A)$. In most of the cases, the space E is a Polish space and $\mathcal{F} = \mathcal{B}(E)$, its Borel σ-algebra. We say that $P(s, x, t, A)$ is a *Feller* transition probability function

(e) if the function $(s, x) \mapsto P(s, x, t, f)$, with

$$P(s, x, t, f) = \int_E f(y) P(s, x, t, \mathrm{d}y),$$

is continuous from $[0, t] \times E$ into $\mathbb{R}$, for any fixed t in $(0, \infty)$ and any bounded continuous function f from E into $\mathbb{R}$. □

Note that conditions (a) and (b) are natural weak regularity assumptions, the limit in (c) is a more restrictive (but necessary) initial condition, and the Chapman-Kolmogorov identity follows from the Markov property in Definition 5.2. Usually, when the space E is locally compact Polish space and $T = [0, \infty)$, we replace the condition (c) by a stronger assumption, namely, for any compact subset K of E, any s in $[0, \infty)$ and any $\varepsilon > 0$ we have

$$\begin{aligned} &\text{(a)} \ \lim_{t \to s} \sup_{x \in K} [1 - P(s, x, t, B(x, \varepsilon))] = 0, \\ &\text{(b)} \ \lim_{x \to \infty} \sup_{0 \le s < t \le 1/\varepsilon} P(s, x, t, K) = 0, \end{aligned} \tag{5.15}$$

where $B(x, \varepsilon)$ is the ball of radius ε and center x, and neighborhood of ∞ are of the form $E \setminus K$ for some compact K of E. In (5.15), the first condition (a) is referred to as local *uniformly stochastic continuity* property, while condition (b) is only necessary when E is not compact. Note that by adding one dimension to the space E, e.g., replacing E by $E \times T$, we can always assume that the transition is homogeneous.

Theorem 5.4 (strong Markov)**.** *Let $(P, X_t, \mathcal{F}_t : t \in T)$ be a Markov process on a Polish space E with cad-lag paths and homogeneous transition function $P(t, x, A)$. If either $P(t, x, A)$ is a Feller transition, i.e., condition (e) holds, or at least the process $s \mapsto P(t, X_s, f)$ is cad-lag for every $t > 0$ and any bounded continuous function f, then (1) $(P, X_t, \mathcal{F}_t^+ : t \in T)$ is a Markov process with transition $P(t, x, A)$, where $\mathcal{F}_t^+ = \cap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}$, and (2) $\mathcal{F}_t^+ = \mathcal{F}_t$, for every $t \ge 0$, when ever $\mathcal{F}_t$ is the σ-algebra generated by the null sets and $\{X_s : s \le t\}$. Moreover, if the cad-lag Markov process $(P, X_t, \mathcal{F}_t : t \in T)$ exists for every initial condition $X_0 = x$, any x in E, and the transition function is Feller or at least the process $s \mapsto P(t, X_s, f)$ is cad-lag for every $t > 0$ and any initial condition $X_0 = x$, then $(P, X_t, \mathcal{F}_t^+ : t \in T)$ is a strong Markov process, i.e., for any τ optional (or stopping) time, any $t > 0$, and every Borel measurable set A in E,*

$$P(X(\tau + t) \in A \,|\, \mathcal{F}_\tau^+) = P(t, X(\tau), A), \quad \text{a.s. on } \{\tau < \infty\},$$

where $\mathcal{F}_\tau^+$ is the σ-algebra generated by the optional time τ, relative to $\{\mathcal{F}_t^+ : t \ge 0\}$. □

A proof of the above *strong Markov* property can be founded in Doob [34, Theorems 8 and 9, pp. 556-560] or in Blumenthal and Getoor [15, Chapter 1, Theorem 8.1, pp. 41-42], where almost surely right continuous (instead of cad-lag) processes is only assumed. Moreover, adding another coordinate to include time as a new state variable, this result is extended to non-homogenous Markov processes with almost no changes. Indeed, if $P(s, x, t, \mathrm{d}\xi)$ is a non-homogeneous transition probability function then $\dot{P}(\dot{x}, t, \mathrm{d}\dot{\xi}) = P(s, x, \tau, \mathrm{d}\xi)\delta(t - \tau)\mathrm{d}\tau$ is a homogeneous transition probability function associated to the Markov process $\dot{X}(t) = (t - s, X(t - s))$ with initial condition $\dot{X}(s) = (s, X(0))$, where $\delta(t-\tau)\mathrm{d}\tau$ is the Dirac measure at $\{t\}$, $\dot{x} = (s, x)$, $\dot{\xi} = (\tau, \xi)$ and $\dot{E} = [0, \infty) \times E$, and

the associated Markov process $\dot{X}(t) = (t - s, X(t - s))$ with initial condition $\dot{X}(s) = (s, X(0))$.

In most of the cases, the Markov process takes values in a locally compact metric space E endowed with its Borel σ-algebra. Using the fact that Radon measures can be regarded as linear continuous functions on the space of continuous functions with compact support, the properties in the Definition 5.3 of transition function including condition (e) and (5.15) can be rephrased as a family of linear operators $P(t, s) : C_0(E) \longrightarrow C_0(E)$, where $C_0(E)$ is the space of continuous functions vanishing at infinity (i.e., for any $\varepsilon > 0$ there exists a compact subset K of E such that $|\varphi(x)| \leq \varepsilon$ for every x in $E \setminus K$), such that

(a) $0 \leq P(t, s)\varphi \leq 1$, for every φ in $C_0(E)$ with $0 \leq \varphi \leq 1$

(b) $\lim_{t \to s} P(t, s)\varphi(x) = \varphi(x)$, for any x in E and φ in $C_0(E)$

(c) $P(t, s) = P(t, r)P(r, s)$, for any $s < r < t$.

Thus, if the transition function is homogeneous, i.e., $P(t, s) = P(t - s)$, we have a one-parameter semigroup in $C_0(E)$.

Sometimes, it is convenient to consider processes with values in a enlarged space $\bar{E} = E \cup \{\infty\}$, with an isolated point ∞ (usually, the one-point compactification), and even defined in the whole $[0, \infty]$. In this case, the *lifetime formalism* is used, i.e., define the *lifetime* of a process $X(\cdot)$ as $\varsigma(\omega) = \{t \geq 0 : X(t) = \infty\}$, and assume that $X(t) = \infty$ for every $t \geq \varsigma$. This allow to relax the condition (a) of the definition of transition function, only the fact that $P(s, x, t, \cdot)$ is a measure with total mass not larger than 1 (instead of a probability measure) is actually necessary.

Usually, the transition function $P(s, x, t, A)$ associated with a Markov process X_t is continuous in time (i.e., in s and t) and a standard realization makes X_t a cad-lag process. In this case, an extra property is desirable, the process $X_t = X(t)$ is *quasi-left* continuous, i.e., $X(T_n)$ converges to $X(T)$ on the set where $T < \infty$, for any increasing sequence of stopping time T_n converging to T, with $T_n < T$. It is clear that here the key is fact that T_n are stopping times. In this sense, the process X_t do not have any *deterministic* jumps.

If a stochastic process represents some kind of phenomenological process then it should characterized by its finite-dimensional distributions. Then, a mathematical model is a *realization* of such a process in a convenient sample space. However, a Markov process is also characterized by either its transition function or it infinitesimal generator (see next chapter). It is important to recognize that when only one stochastic process (or variable) is involved, it finite-dimensional distributions determine the process in an appropriate sample space (usually refer to as a *realization of the process*), however, when two or more stochastic processes (or variables) are discussed, it is important to know its *joint distribution*. Thus the concept of *version* or *modification* of a process is relevant, i.e., at the end we are always working with stochastic processes as random variables which are *almost surely* equals. Recall that two stochastic processes need not to be defined in the same probability space to have the same finite-dimensional distributions, but they do have the same law, once the

sample space has been chosen. However, to be a version (or a modification) one of each other, they do need to be defined in the same probability space. After a sample space has been chosen, the stochastic process are treated as random variable, with values in the sample space. The procedure of selecting a sample space on which a probability is constructed satisfying its characteristic properties (e.g., the finite-dimensional distributions are given, or in the case of a Markov process, its transition function or its infinitesimal generator is given) is called a *realization* of the stochastic process.

The reader may consult the classic books by Blumenthal and Getoor [15], Dynkin [41, 42] or more recent books, such as, Applebaum [1], Çınlar [26], Chung [24], Dellacherie and Meyer [32], Ethier and Kurtz [45], Liggett [110], Marcus and Rosen [118], Rogers and Williams [153], Taira [172], among many others.

5.1.3 Strong Markov Processes

Starting from a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$, we may consider stochastic processes X with values in some Polish space E (complete separable metric space, usually locally compact) as (1) a family of E-valued random variables $\{X(t) : t \geq 0\}$, (2) a function on a product space $X : [0, \infty) \times \Omega \to E$, (3) a function space valued random variable, i.e., either a random variable with values in some sub-space of $E^{[0,\infty)}$ or a mapping from $[0, \infty)$ into the space of E-valued random variables. Except when explicitly mentioned, we are looking at a stochastic process as a random variable with values in some function space, a Polish space non-locally compact which most of the cases is either $D([0, \infty), E)$ or $C([0, \infty), E)$, with E being an Borel (usually open or closed) subset of $\mathbb{R}^d$.

A stochastic process X with values in a Polish space E (even more general, E could be a Lusin space, i.e., a topological space homeomorphic to a Borel subset of a complete separable metric space) is called a *Markov process* in the filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ if the Markov property is satisfied, i.e.,

$$\mathbb{E}\{f(X(t) \mid \mathcal{F}(s)\} = \mathbb{E}\{f(X(t) \mid X(s)\}, \tag{5.16}$$

for every $t \geq s$ and any bounded Borel real function f on E. This is an almost surely equality due to the use of conditional probability. It means that the only information relevant for evaluating the behavior of the process beyond time s is the value of the current state $X(s)$. This implies in particular that X is adapted. Points x in E are called *states* and E is the *state space* of X.

A Markov process can be identified by its *transition function*, which is defined by taking a particular class of function f in (5.16), namely characteristic or indicator functions $f = \mathbb{1}_B$ of Borel subsets B of E, i.e., with B in $\mathcal{B}(E)$. The transition function $p(s, x, t, B)$ is defined with following (minimal) regularity conditions:

(1) for fixed $0 \leq s \leq t$, x in E, the function $B \mapsto p(s, x, t, B)$ is a probability measure on $\big(E, \mathcal{B}(E)\big)$,

(2) for fixed $0 \leq s \leq t$, B in $\mathcal{B}(E)$ the function $x \mapsto p(s, x, t, B)$ is Borel mea-

surable,

(3) for every $0 \le s \le t$, x in E, B in $\mathcal{B}(E)$ we have the identity $p(s, X(s), t, B) = \mathbb{E}\{\mathbb{1}_B(X(t)) \,|\, \mathcal{F}(s)\}$, almost surely.

Really, (1) and (2) are necessary conditions to make sense to the key condition (3). However, the Markov property alone is not sufficient to define the transition function. Condition (3) implies that for every $s \ge 0$, x in E, B in $\mathcal{B}(E)$ we have $p(s, x, s, B) = \mathbb{1}_B(x)$ and standard properties of the conditional probability yield the Chapman-Kolmogorov identity

$$p(s, x, t, B) = \int_E p(r, y, t, B) p(s, x, r, \mathrm{d}y), \tag{5.17}$$

valid for any $0 \le s < r < t$, x in E and B in $\mathcal{B}(E)$.

Markov processes are mathematical model for phenomena which evolve in time, in a random way and following some dynamic or evolution law. Most often, statistical experiments or physical considerations give only information about the so-called finite-dimensional distributions of a process. This means that for a given initial probability measure μ on $\big(E, \mathcal{B}(E)\big)$ and times $0 \le t_0 < t_1 < \cdots < t_n$ the probabilities $P_{t_0,t_1,\ldots,t_n}$ on E^{n+1} defined by

$$\begin{aligned} P_{t_0,t_1,\ldots,t_n}(B_0 \times B_1 \times \cdots \times B_n) = \\ = \int_{B_0} \mu(\mathrm{d}x_0) \int_{B_1} p(t_0, x_0, t_1, \mathrm{d}x_1) \int_{B_2} p(t_1, x_1, t_2, \mathrm{d}x_2) \cdots \\ \cdots \int_{B_{n-1}} p(t_{n-2}, x_{n-2}, t_{n-1}, \mathrm{d}x_{n-1})\, p(t_{n-1}, x_{n-1}, t_n, B_n) \end{aligned} \tag{5.18}$$

are the finite-dimensional distributions. Thus, starting from a function p satisfying the properties (1) and (2) of a transition function, and if the function p satisfies the Chapman-Kolmogorov identity (5.17), then the above relation (5.18) defines a consistent family of finite-dimensional distributions on the canonical product space $E^{[0,\infty)}$. Note that the Dirac measure $\delta(x_0)$, i.e., the unit mass concentrated at x_0, is the typical initial distribution at time t_0. For simplicity, let us discuss *homogeneous* Markov process, i.e., the case where the transition function is time invariant, i.e., $p(s, x, t, B) = p(0, x, t - s, B)$ for every $t \ge s$, x in E and B in $\mathcal{B}(E)$. Hence, the transition function can be taken as $p(x, t, B)$, with $t \ge 0$. Remark that by adding an extra variable (the time), we can always reduce to homogeneous case. Thus, Kolmogorov's existence theorem can be used to construct a Markov process with the given transition function p, for each initial probability measure μ at time $t_0 = 0$, and then we have a family of Markov processes. Therefore, by a *realization* of Markov process with transition probability function p we mean a collection $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P_x, t \ge 0, x \in E)$ where P_x is the probability measure constructed as above with initial probability $\mu = \delta_x$, the Dirac measure at x. In Kolmogorov's construction, the process $X_t(\omega) = X(t, \omega) = \omega(t)$ is the coordinate (or identity) mapping and $\mathcal{F}(t)$ is the natural filtration associated with the process $X(t)$, which is not always right-continuous. Some difficulties appear since $\mathcal{F}(t)$ should be completed with

respect to the probability measure P_x, given a completed filtration $\mathcal{F}_x(t)$ or $\mathcal{F}_\mu(t)$, which depend on initial parameter x or μ.

By means of the transition probability function, we may re-write the Markov property (5.16) as

$$P\{X(s+t) \in B \mid \mathcal{F}(s)\} = p(X(s), t, B), \tag{5.19}$$

for every $t \geq s \geq 0$ and any Borel subset B of E. Now, a *strong Markov process* is one for which the Markov property holds at stopping times of the filtration $\{\mathcal{F}(t) : t \geq 0\}$, i.e.,

$$P\{X(T+t) \in B \mid \mathcal{F}(T)\}\mathbb{1}_{T<\infty} = p(X(T), t, B)\mathbb{1}_{T<\infty}, \tag{5.20}$$

for every $t \geq 0$, any stopping time T and any Borel subset B of E. This says that the probabilistic evolution of the process after the stopping time T is just that of another process *restarted* at T, i.e., *the process restarts at stopping time*. The reader is referred to Doob [34, Theorems 8 and 9, pp. 556-560], see Theorem 5.4 in Chapter 1, for conditions ensuring the right-continuity of the filtration and the strong Markov property. In the statement (5.20), we remark the interest in using a filtration satisfying the usual condition, in particular the need of having a *completed* σ-algebra $\mathcal{F}(0)$. A useful definition in this context is the so-called *universally completed* filtration, which is constructed as follows. First, let $\{\mathcal{F}(t) : t \geq 0\}$ be the filtration (history) generated by the canonical process $X(t, \omega) = \omega(t)$, not necessarily satisfying the usual conditions. Denote by $\{\mathcal{F}^\mu(t) : t \geq 0\}$ the filtration which is obtained by completing $\mathcal{F}(0)$ with respect to the probability measure P_μ. Now the universally completed filtration is $\{\mathcal{F}^0(t) : t \geq 0\}$, where $\mathcal{F}^0(t) = \cap_\mu \mathcal{F}^\mu(t)$, for every $t \geq 0$. Note that the filtration $\{\mathcal{F}^0(t) : t \geq 0\}$, does not necessarily satisfies the usual conditions, but it is right-continuous if the initial filtration $\{\mathcal{F}(t) : t \geq 0\}$ is so.

As discussed earlier, the product space $E^{[0,\infty)}$ does not provide a suitable mathematical setting, we need to use the Polish sample space $D([0,\infty), E)$ or $C([0,\infty), E)$. This imposes more conditions on the transition function p, and eventually we are lead to the study of Markov-Feller processes and semigroups.

The reader may consult the classic references Blumenthal and Getoor [15], Dynkin [42] or more recent books, e.g., Davis [30], Rogers and Williams [153].

One of the most simple Markov processes in continuous time is the Poisson process. If $\{\tau_n : n, n = 1, 2, \ldots\}$ is a sequence of independent exponentially distributed (with parameter λ) random variables, then the random variable $\theta_n = \tau_1 + \cdots + \tau_n$ has a Γ-distribution with parameters λ and $n-1$, for $n = 1, 2, \ldots$, i.e.,

$$P\{\theta_n \leq t\} = \frac{\lambda^n}{(n-1)!}\int_0^t s^{n-1}\mathrm{e}^{-\lambda x}\mathrm{d}s, \quad \forall\, t \geq 0,$$

and the *counting process* defined by

$$p(t, \omega) = \sum_{n=1}^{\infty} \mathbb{1}_{\theta_n(\omega)\leq t}, \quad \forall\, t \geq 0 \tag{5.21}$$

is a Poisson process, i.e., $p(0) = 0$, $p(t) - p(s)$ is a Poisson variable with mean $\lambda(t-s)$, namely

$$P\{p(t) - p(s) = n\} = \big[\lambda(t-s)\big]^n \exp\big[-\lambda(t-s)\big],$$

for every $n = 0, 1, \dots$, and for any $0 \le t_0 < t_1 < \cdots < t_n$ the family $\{p(t_0), p(t_k) - p(t_{k-1}) : k = 1, 2, \dots, n\}$ is a set of independent random variables. The parameter λ is usually called *jump rate*.

In a *compound* Poisson process the construction (5.21) is modified as follows

$$p_c(t, \omega) = \sum_{k=1}^{\infty} \eta_n(\omega) \mathbb{1}_{\theta_n(\omega) \le t}, \quad \forall\, t \ge 0, \tag{5.22}$$

where $\{\eta_n : n = 1, 2, \dots\}$ is a sequence of independent identically distributed (with distribution law ν and independent of the $\{\tau_n\}$) $\mathbb{R}^d$-valued random variables. A integer-valued measure process can be associated, namely

$$\rho_c(t, B, \omega) = \sum_{k=1}^{\infty} \mathbb{1}_{\theta_k(\omega) \le t} \mathbb{1}_{\eta_k(\omega) \in B}, \quad \forall\, t \ge 0,\ B \in \mathcal{B}(\mathbb{R}^d), \tag{5.23}$$

which captures all features of the compound process and extends to the so-called *Poisson measures*. Note that $\mathbb{E}\{\rho_c(t, B, \omega)\} = t\, \lambda\, \nu(B)$. The parameters λ and ν yield the integral operator

$$Ih(x) = \lambda \int_{\mathbb{R}} \big[h(x+y) - h(x)\big]\, \nu(\mathrm{d}y), \quad \forall\, x \in \mathbb{R}^d, \tag{5.24}$$

which is a characteristic element of the compound Poisson process. This integral operator is the infinitesimal generator of the Markov process, which in turn is determined by its kernel, the Lévy measure $M(\mathrm{d}y) = \lambda\, \nu(\mathrm{d}y)$. Note that to make the expression (5.23) interesting, we assume $\nu(\{0\}) = 0$ and then the mass of the origin $M(\{0\})$ does not play any role in the definition of I, thus the Lévy measure is on $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$.

All these examples are time and spatially homogeneous Markov processes. To relax the homogeneity, we must allow the Lévy measure to depend on t and x. For instance, we take $M(x, \mathrm{d}y)$ in the expression (5.24) of the integral operator. The dependency on x of the kernel could be very general and in some cases hard to track. A typical assumption is the representation

$$M(x, B) = \lambda(x)\, \ell(\{\zeta \in [0,1] : x + j(x, \zeta) \in B\}), \tag{5.25}$$

for every x in $\mathbb{R}^d$ and B in $\mathcal{B}(\mathbb{R}^d_*)$, where $([0,1], \mathcal{L}, \ell)$ is the canonical Lebesgue probability measure space, $\lambda : \mathbb{R}^d \to [0, \infty)$ and $j : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d_*$ are measurable functions, on which some regularity (such as continuity) in x may be required.

If $\{Z_n, U_n : n = 1, 2, \dots\}$ are double sequence of independent uniformly distributed random variables in $([0,1], \mathcal{L}, \ell)$, then the transformation

$$\Theta(x, u) = \inf\big\{t \ge 0 : \exp[-t\, \lambda(x)] \le u\big\}, \tag{5.26}$$

with $\Theta(x,0) = +\infty$, yields the construction of the following Markov jump process by induction. Given θ_{k-1} and x_{k-1} we define

$$\begin{aligned} \theta_k &= \theta_{k-1} + \Theta(x_{k-1}, U_k), \\ x_k &= x_{k-1} + j(x_{k-1}, Z_k) \end{aligned} \tag{5.27}$$

and for any t in the stochastic interval $[\![\theta_{k-1}, \theta_k[\![$ set $x(t) = x_k$. Naturally, we can start from any initial time θ_0 and state x_0, but we use $\theta_0 = 0$ and any given x_0. Assuming that $\theta_n \to \infty$ (e.g., this hold if $\lambda(\cdot)$ is bounded) the process $x(t)$ is defined for every time $t \geq 0$. Its associated integer-valued measure process is given by

$$\rho(t, B, \omega) = \sum_{k=1}^{\infty} \mathbb{1}_{\theta_k(\omega) \leq t} \mathbb{1}_{x_k(\omega) \in B}, \quad \forall\, t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d). \tag{5.28}$$

The integral operator becomes

$$Ih(x) = \lambda(x) \int_{[0,1]} \big[h(x + j(x,\zeta)) - h(x)\big]\, \ell(\mathrm{d}\zeta), \quad \forall\, x \in \mathbb{R}^d, \tag{5.29}$$

which make sense for any bounded Borel measurable function h. The process $\{x(t) : t \geq 0\}$ a cad-lag realization (and piecewise constant) of a strong Markov process. Several other variations are possible.

As seen early, the waiting times between two consecutive jumps of a (compound) Poisson process is a sequence $\{\tau_k : k \geq 1\}$ of independent identically distributed (IID) with an exponential distribution. Moreover, even in the previous of a jump Markov process example $\{x(t) : t \geq 0\}$ with $x(t) = x_k$ for t in $[\![\theta_{k-1}, \theta_k[\![$ as in (5.27), the waiting times between two consecutive jumps $\tau_k = \Theta(x_{k-1}, U_k)$ follows an exponential distribution, when conditioned to the past. In general, a 'pure' Markov jump process is not suitable to include (or to describe) consecutive jumps given by a sequence $\{T_i : i \geq 1\}$ of IID random variables, with T_1 having a distribution π_0 (other than exponential). Instead, an homogeneous Markov process $\{y_t : t \geq 0\}$, representing the *time elapsed since the last jump*, could be constructed as follows. First, for the initial condition $y(0) = 0$ define $\theta_0 = 0$ and then by induction $\theta_n = \theta_{n-1} + T_n$, and

$$y(t) = t - \theta_{n-1} \quad \text{if} \quad \theta_{n-1} \leq t < \theta_n \quad \text{and} \quad y(\theta_n) = 0, \quad n \geq 1. \tag{5.30}$$

However, if $y(0) = y > 0$ then conditional probability must be used to define $y(t)$ as beginning at time '$-y$' conditional to 'having the first jump at sometime $t \geq 0$'. This means that if the initial IID sequence $\{T_1, T_2, \ldots\}$ of *waiting time between two consecutive jumps* has its common law π_0 supported on $[0, \infty[$, with $\pi_0(\{0\}) = 0$, and $y > 0$ is the initial condition at time $t = 0$, then first consider a non-negative random variable T^y independent of $\{T_1, T_2, \ldots\}$ with distribution

$$P\{T^y \in]a, b]\} = P\{T_1 \in]a + y, b + y] \,|\, T_1 \geq y\} = \frac{\pi_0(]a + y, b + y])}{\pi_0(]y, +\infty[)},$$

for any $b > a \geq 0$. Now define the sequence of time-jumps

$$\theta_0^y = 0, \quad \theta_1^y = T^y \quad \text{and} \quad \theta_{n+1}^y = \theta_n^y + T_n, \quad \forall n \geq 1,$$

and the process $\{y(t) : t \geq 0\}$ with $y(0) = y$ by the expressions

$$y(t) = y(\theta_{n-1}^y) + t - \theta_{n-1}^y \text{ if } \theta_{n-1}^y \leq t < \theta_n^y \quad \text{and} \quad y(\theta_n) = 0, \quad \forall n \geq 1,$$

which agree with (5.30) when $y = 0$, while the process $\xi(t) = \sum_{i=1}^{\infty} \mathbb{1}_{\theta_i \leq t}$ counts the jumps. In this case, either $\{y(t) : t \geq 0\}$ or $\{(y(t), \xi(t)) : t \geq 0\}$ is a Markov process.

If the common distribution π_0 has a density $\dot{\pi}_0$, i.e.,

$$\pi_0([0, y]) = \int_0^y \dot{\pi}_0(s)\mathrm{d}s, \quad \forall y \geq 0,$$

then the intensity (of jumps) is given by

$$\lambda(y) = \frac{\dot{\pi}_0(y)}{1 - \pi_0([0, y])} = \lim_{h \to 0} \frac{P\{y \leq T_1 \leq y + h \mid T_1 \geq y\}}{h}, \quad \forall y \geq 0,$$

or equivalently

$$\lambda(y) = \Big\{ -\ln\Big[1 - \int_0^t \dot{\pi}_0(s)\mathrm{d}s\Big]\Big\}', \quad \forall y \geq 0,$$

which yields the conditional distribution

$$\pi(y,]a, b]) = \frac{\pi_0(]a, b] \cap]y, \infty[)}{\pi_0(]y, \infty[)} = \int_a^b \exp\{-\int_0^t \lambda(y + s)\mathrm{d}s\}\lambda(t + y)\mathrm{d}t,$$

for any real numbers $b > a \geq y \geq 0$. Note that if the law π_0 is an exponential distribution then T^y has also the same exponential distribution π_0 (i.e., the jumps of y_t do not depend on the initial value $y(0)$, in other words, T^y can be regarded as one of T_i), and therefore, no need to introduce the Markov process $\{y(t) : t \geq 0\}$ in the model. Moreover, if the law π_0 satisfies $\pi_0(]y_{\max}, \infty[) = 0$ (with $0 < y_{\max} < \infty$) then the initial value y should be taken either $0 < y < y_{\max}$ (if $\pi_0(\{y_{\max}\}) = 0$) or $0 < y \leq y_{\max}$ (if $\pi_0(\{y_{\max}\}) > 0$).

The infinitesimal generator of the process $\{y(t) : t \geq 0\}$ is given by

$$A\varphi(y) = \partial_y \varphi(y) + \lambda(y)[\varphi(0) - \varphi(y)], \quad \forall y \geq 0.$$

More general, the counting process $\{\xi(t) : t \geq 0\}$ may become $\xi(t) = \sum_k \zeta_k \mathbb{1}_{t \geq \theta_k}$, with $\{\zeta_k\}$ (independent of $\{T_i\}$ and) with distribution π/c, $c = \pi(\mathbb{R}_*^m)$. In this case, $\{(y(t), \xi(t)) : t \geq 0\}$ is also a Markov process and

$$A\varphi(y, \xi) = \partial_y \varphi(y, \xi) + \lambda(y, \xi)\big[\mathbb{E}\{\varphi(0, \xi + \zeta_1)\} - \varphi(y, \xi)\big], \quad \forall y \geq 0,$$

and any $\xi \in \mathbb{R}^m$, is its infinitesimal generator.

5.1.4 Extended Generators

Let E be a Borel subset of Polish space, let $B(E)$ be the Banach space of bounded Borel measurable functions f from E into $\mathbb{R}$ with sup-norm $\|\cdot\|$, and let $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P_x, t \geq 0, x \in E)$ be a (strong) Markov process. For $t \geq 0$, define an operator $P(t) : B(E) \to B(E)$ by $P(t)f(x) = \mathbb{E}_x\{f(X(t)\}$, where $\mathbb{E}_x\{\cdot\}$ denotes the mathematical expectation relative to P_x. It is clear that $P(t)$ is a contraction, i.e. $\|P(t)f\| \leq \|f\|$, for every $t \geq 0$, and that the Chapman-Kolmogorov identity (5.17) are equivalent to the *semigroup property* $P(t)P(s) = P(s+t)$, for every $t, s \geq 0$.

Denote by B_0 the subset of $B(E)$ consisting of those functions f for which $\|P(t)f - f\|$ vanishes as t goes to zero. The contraction property shows that B_0 is a closed subspace of $B(E)$ and $\{P(t) : t \geq 0\}$ is called *strongly continuous* on B_0. Moreover, (1) B_0 is invariant under $P(t)$, for every $t \geq 0$, and (2) for every f in B_0 (which is itself a Banach space), the function $t \mapsto P(t)f$ is continuous from $[0, \infty)$ into B_0.

Now, let A be the *strong infinitesimal generator* of $\{P(t) : t \geq 0\}$ with domain $\mathcal{D}(A) \subset B_0 \subset B(E)$, i.e, f belong to $\mathcal{D}(A)$ and $Af = g$ if and only if

$$\lim_{t \to 0} \left\| \frac{P(t)f - f}{t} - g \right\| = 0.$$

Note that the domain $\mathcal{D}(A)$ is as important as the expression of A, there are examples of two different Markov process with the same expression for the infinitesimal generator A but with disjoint domains $\mathcal{D}(A)$, see Davis [30, Chapter 2].

Based on properties of derivatives and Riemann integrals of continuous functions with values in a Banach space, we can establish:

(1) if $f \in B_0$ and $t \geq 0$ then

$$\int_0^t P(s)f \mathrm{d}s \in \mathcal{D}(A) \quad \text{and} \quad A \int_0^t P(s)f \mathrm{d}s = P(t)f - f,$$

(2) if $f \in \mathcal{D}(A)$ and $t \geq 0$ then $P(t) \in \mathcal{D}(A)$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} P(t)f = AP(t)f = P(t)Af,$$
$$P(t)f - f = \int_0^t AP(s)f \mathrm{d}s = \int_0^t P(s)Af \mathrm{d}s.$$

In probabilistic terms, if $u(t) = u(x,t) = \mathbb{E}_x\{f(X(t))\} = P(t)f(x)$ with f in $\mathcal{D}(A)$ then u satisfies

$$\partial_t u(t) = Au(t), \quad u(0) = f, \tag{5.31}$$

which is an abstract version of the so-called *Kolmogorov backward* equation. The semigroup is determined by (5.31) and this determines the transition (probability) functions $p(x, t, B)$, which determines the finite-distributions and hence the

probability measure P_x, i.e., the Markov process itself. Certainly, some technical conditions are required to turn this calculation into reality. For practical calculations it is more important the second expression in (2) which can be written as

$$\mathbb{E}_x\{f(X(t)\} = f(x) + \mathbb{E}_x\Big\{\int_0^t Af(X(s)\mathrm{d}s\Big\}, \tag{5.32}$$

for every f in $\mathcal{D}(A)$, which is known as *Dynkin formula*.

Let f be in $\mathcal{D}(A)$ and define the real-valued process $\{M_f(t) : t \geq 0\}$ by

$$M_f(t) = f(X(t)) - f(X(0)) - \int_0^t Af(X(s))\mathrm{d}s. \tag{5.33}$$

By using the semigroup property and conditional expectation arguments, we can show that for every x in E the process $\{M_f(t) : t \geq 0\}$ is a martingale in $(\Omega, \mathcal{F}, P_x, \mathcal{F}(t), t \geq 0)$, i.e.,

$$\mathbb{E}_x\{M_f(t) \,|\, \mathcal{F}(s)\} = M_f(s), \quad \forall t \geq s \geq 0.$$

A natural extension of the domain $\mathcal{D}(A)$ of the (strong) infinitesimal generator is as follows.

Definition 5.5 (extended generator)**.** Let $B^*(E)$ be the space of all Borel measurable functions, not necessarily bounded, from E into $\mathbb{R}$. We say that a function f belongs to the domain of the *extended (infinitesimal) generator* if there exists another function g in $B^*(E)$ such that $t \mapsto g(X(t))$ is locally integrable P_x-almost surely and the process $\{M_f(t) : t \geq 0\}$ defined by

$$M_f(t) = f(X(t)) - f(X(0)) - \int_0^t g(X(s))\mathrm{d}s$$

is a local martingale, i.e., there exists an increasing sequence of stopping times $\{\tau_n : n = 1, 2, \ldots\}$, with $\tau_n \to +\infty$ such that the stopped process $M_f^n(t) = M_f(t \wedge \tau_n)$ is a uniformly integrable martingale for each n. We use the notation $\mathcal{D}(\bar{A})$ for the extended domain and $\bar{A}f = g$ for the extended generator. □

Note that $\mathcal{D}(A) \subset \mathcal{D}(\bar{A})$ and that $\bar{A}f$ is uniquely defined (module subset of potential zero). Indeed, if $f = 0$ then the process $\{M_f(t) : t \geq 0\}$ is a continuous martingale with locally bounded variation, therefore $M_f(t) = M_f(0)$ is the constant process zero. Hence, $Af = 0$ except possibly on some measurable set B of E such that

$$\int_0^\infty \mathbb{1}_B(X(t))\mathrm{d}t = 0, \quad P_x - \text{a.s.},$$

for every x in E. Such a set B is said to have *potential zero*. The process $\{X(t) : t \geq 0\}$ *spend no time* in B, regardless of the starting point, so the process $\{M_f(t) : t \geq 0\}$ does not depend on the values of Af for x in B, and Af is unique up to sets of zero potential.

When $\{M_f(t) : t \geq 0\}$ is a martingale, Dynkin formula (5.32) holds. Usually, it is quite difficult to characterize $\mathcal{D}(\bar{A})$ but in most of the cases, there are easily checked sufficient conditions for membership in the extended domain $\mathcal{D}(\bar{A})$. For instance, the reader is refereed to the books by Davis [30, Chapter 1], Ethier and Kurtz [45, Chapter 4] for more details.

Let us go back to the examples in the previous section. For the particular case of the Poisson process $(p(t) : t \geq 0)$ given by (5.21), the extended infinitesimal generator is

$$Af(x) = \lambda[f(x+1) - f(x)], \quad \forall\, x \in \mathbb{R}$$

while for the compound Poisson process $(p_c(t) : t \geq 0)$ (5.21), A is the integral operator I given by (5.24). What is perhaps more relevant is the extended domain $\mathcal{D}(\bar{A})$, which have not restriction at all (i.e., all real-valued function defined on $\mathbb{R}$) for the Poisson process, while a condition on local integrability, i.e., a measurable function $f : \mathbb{R} \to \mathbb{R}$ belongs to $\mathcal{D}(\bar{A})$, for the compound Poisson process with parameters λ and ν, if and only if

$$\mathbb{E}\{\sum_{i=1}^{\infty} |f(x+\eta_i) - f(x)|\, \mathbb{1}_{\theta_i < \sigma_n}\} < \infty, \quad \forall\, x, n,$$

where σ_n is a sequence of stopping times with $\sigma_n \to \infty$ almost surely.

For the class of Markov jump process constructed by induction, see (5.26) and (5.28), the full description of the extended domain $\mathcal{D}(\bar{A})$, with $A = I$ as in (5.29), is as follow. First, we say that a process $\{h(x,t,\omega) : t \geq 0,\, x \in \mathbb{R}^d\}$ belongs to $L^1(\rho)$, (where ρ is the integer-valued measure process) if

$$\mathbb{E}\Big\{\sum_{i=1}^{\infty} h(x_k, \theta_k, \omega)\Big\} < \infty.$$

Similarly, h belongs to $L^1_{\mathrm{loc}}(\rho)$, if there exists a sequence $\{\sigma_k : k \geq 0\}$ of stopping times with $\sigma_n \to \infty$ almost surely such that

$$\mathbb{E}\Big\{\sum_{i=1}^{\infty} h(x_k, \theta_k \wedge \sigma_n, \omega)\Big\} < \infty, \quad \forall\, n.$$

Now, a measurable function f belongs to $\mathcal{D}(\bar{A})$ if the process $h(x,t,\omega) = f(x) - f(x(t-,\omega))$ belongs to $L^1_{\mathrm{loc}}(\rho)$. This is particular case of Davis [30, Theorem 26.14, pp. 69–74].

5.1.5 Transition Functions

Now we focus on the transition functions of spatially homogeneous Markov processes or additive processes. There are several aspects of a Markov Process, depending on the specific emphasis given to the discussion, one of the following elements is first studied and then other elements are derived. A Markov process

with valued in $\mathbb{R}^d$ may be presented as

(a) a family of $\mathbb{R}^d$-valued stochastic processes $X = X_{sx}$ indexed by the initial distribution $X(s) = x$, $s \geq 0$,

(b) a probability transition function $P(s, x, t, A)$ with $t > s \geq 0$, $x \in \mathbb{R}^d$ and A a Borel subset of $\mathbb{R}^d$,

(c) a family of linear and bounded evolution operators $\Phi(t, s)$ from $B(\mathbb{R}^d)$, the Banach space of bounded Borel real-valued function on $\mathbb{R}^d$ into itself, indexed by $t \geq s \geq 0$,

(d) a family of linear and bounded operators $R(\lambda)$ from $B(\mathbb{R}^d)$ into itself, indexed by $\lambda > 0$,

(e) a family of linear possible unbounded (infinitesimal generator) operators $A(t)$ defined in a subspace $\mathcal{D}(A(t))$ of $B(\mathbb{R}^d)$ into $B(\mathbb{R}^d)$, indexed by $t \geq 0$.

Certainly, each of these (a),...,(e) elements should satisfy some specific conditions to yield a Markov process.

The elements $R(\lambda)$ in (d) are called *resolvent* operators and are mainly used with *time-homogeneous* Markov processes, i.e., when (a) $X_{sx} = X_{0x}$ for any $s > 0$ or (b) $P(s, x, t, A) = P(0, x, t - s, A)$ for any $t > s \geq 0$ or (c) the evolution operators $\Phi(t, s) = \Phi(t - s)$ for any $t > s \geq 0$ or (e) $A(t) = A$ for any $t \geq 0$. It is clear that by adding a new dimension to $\mathbb{R}^d$ we may always assume we are in the time-homogeneous, however, in most of the cases, we prefer to live the special time variable t with its preferential role and to work with non-time-homogeneous Markov processes. It is possible to use a Polish (separable complete metric space) $\mathcal{O}$ instead of the Euclidean space $\mathbb{R}^d$, usually $\mathcal{O}$ is locally compact since the infinite-dimensional case needs some special care.

The principle stating that *the future is independent of the past given the present* is called *Markov property* and formally is written as

$$P\{X(t) \in B \mid X(r),\, r \leq s\} = P\{X(t) \in B \mid X(s)\}, \tag{5.34}$$

for every $t > s \geq 0$ and $B \in \mathcal{B}(\mathbb{R}^d)$, which should be satisfied by the family of processes. This same property viewed by the transition function is called the *Chapman-Kolmogorov identity*,

$$P(s, x, t, B) = \int_{\mathbb{R}^d} P(s, x, r, \mathrm{d}y)\, P(r, y, t, B), \tag{5.35}$$

for every $t > r > s$, x in $\mathbb{R}^d$ and B in $\mathcal{B}(\mathbb{R}^d)$. For the evolution operators this is called the *semigroup property* are written as

$$\Phi(t, s) = \Phi(t, r)\Phi(r, s) \ \text{ in } \ B(\mathbb{R}^d), \quad \forall t > r > s > 0, \tag{5.36}$$

and in the case of time-homogeneous Markov processes, the resolvent operators satisfy the so-called *resolvent equation*, namely

$$R(\lambda) - R(\nu) = (\nu - \lambda)R(\lambda)R(\nu) \ \text{ in } \ B(\mathbb{R}^d), \quad \forall \lambda, \nu > 0. \tag{5.37}$$

The resolvent $\{R(\lambda) : \lambda > 0\}$ is mainly used in potential theory, the semi-group $\{\Phi(t) : t \geq 0\}$ and the infinitesimal generator A are well know in analysis, while the family of stochastic processes X and the transition function $P(s, x, t, B)$ are more probabilistic tools. At this general level, we ramark that the Markov property (5.34) is almost surely satisfied, i.e., only *version* of the stochastic processes are involved and therefore a property on the *sample path* should be added. The evolution and resolvent operators are defined on $B(\mathbb{R}^d)$, which is a non-separable Banach space, so that in general the theory is very delicate.

Out interest is in Markov-Feller or Feller-Dynkin processes, instead of the large space $B(\mathbb{R}^d)$ we use the separable Banach space $C_0(\mathbb{R}^d)$, of all continuous functions vanishing at infinity (i.e., for any $\varepsilon > 0$ there exists a compact subset K of $\mathbb{R}^d$ such that $|\varphi(x)| \leq \varepsilon$ for every x in $\mathbb{R}^d \setminus K$). Thus, after a one-point compactification method, we are reduced to $C(\bar{\mathbb{R}}^d)$, with $\bar{\mathbb{R}}^d = \mathbb{R}^d \cup \{\infty\}$ being a compact Polish space. For the family of stochastic processes X_x, this yields a *cad-lag condition* on the sample path. Regarding the Chapman-Kolmogorov identity (5.35) we have

Definition 5.6 (transition function)**.** A *(Markov) transition function* on the Borel space $(\mathbb{R}^d, \mathcal{B})$, $\mathcal{B} = \mathcal{B}(\mathbb{R}^d)$, is a function $P(s, x, t, B)$ defined for $t > s \geq 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$ such that

(a) for each $t > s \geq 0$ and x in $\mathbb{R}^d$ the function $B \mapsto P(s, x, t, B)$ is a positive measure on $(\mathbb{R}^d, \mathcal{B})$, with $P(s, x, t, \mathbb{R}^d) \leq 1$,

(b) for each $t > 0$ and B in $\mathcal{B}$ the function $(s, x) \mapsto P(s, x, t, B)$ is a measurable,

(c) for any $s \geq 0$, for any compact subset K of $\mathbb{R}^d$ and any $\varepsilon > 0$ we have

$$\lim_{t \to s} \sup_{x \in K} \left[1 - P(s, x, t, \{y \in \mathbb{R}^d : |y - x| \leq \varepsilon\})\right] = 0,$$

so-called uniformly stochastic continuous,

(d) for each $t > r > s \geq 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$ we have

$$P(s, x, t, B) = \int_{\mathbb{R}^d} P(s, x, r, \mathrm{d}y)\, P(r, y, t, B),$$

i.e., Chapman-Kolmogorov identity.

These properties can be rephrased in term of linear non-negative operators from $B(\mathbb{R}^d)$, the space of real-valued bounded and Borel functions on $\mathbb{R}^d$, into itself, defined by

$$P(t, s)\varphi(x) = \int_{\mathbb{R}^d} \varphi(y) P(s, x, t, \mathrm{d}y) = P(s, x, t, \varphi), \tag{5.38}$$

for every $t > s \geq 0$ and x in $\mathbb{R}^d$, which satisfies

(a') for each $t > s \geq 0$ and φ in $B(\mathbb{R}^d)$ with $0 \leq \varphi \leq 1$ we have $0 \leq P(t, s)\varphi \leq 1$,

(b') for each $t > s \geq 0$ and x in $\mathbb{R}^d$ the mapping $B \mapsto P(t, s)\mathbb{1}_B(x)$ is σ-additive on $\mathcal{B}(\mathbb{R}^d)$,

(c') for any $s \geq 0$ and φ in $C_0(\mathbb{R}^d)$, continuous functions on $\mathbb{R}^d$ vanishing at infinity, we have

$$\lim_{t \to s} P(t,s)\varphi(x) = \varphi(x), \quad \forall x \in \mathbb{R}^d,$$

i.e., the *stochastic continuity property* , a weaker version of *(c)*,

(d') for each $t > r > s \geq 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$ we have

$$P(t,s) = P(t,r)\, P(r,s), \quad \text{in} \quad B(\mathbb{R}^d),$$

usually referred to as the *semigroup property*, and the transition function is called a *Feller transition* if the following condition (e) , so-called *Feller property*, is satisfied

(e) for each $t > s \geq 0$ and φ in $C_0(\mathbb{R}^d)$ we have $P(t,s)\varphi$ in $C_0(\mathbb{R}^d)$, i.e., $P(t,s)$ can be considered as acting on $C_0(\mathbb{R}^d)$.

It is called *time-homogeneous* if $P(s,x,t,B) = P(0,x,t-s,B)$ and *spatially-homogeneous* if $P(s,x,t,B) = P(s,0,t,B-x)$, for any $t > s \geq 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$. It is called a transition *probability* function if $P(s,x,t,\mathbb{R}^d) = 1$, for any $t > s \geq 0$ and x in R^d. □

Certainly, to define a transition function we only need a measurable space $(E, \mathcal{E})$ and t belonging to some set T with a complete order, instead of the Euclidean space $\mathbb{R}^d$ and the real semi-line $[0,\infty)$. However, for time-homogeneous transition function, essentially we need the semi-line $[0,\infty)$ and for the spatially-homogeneous transition function E has to be a vector space, e.g., $\mathbb{R}^d$.

Condition (b') is satisfied when E is locally compact, i.e., $\mathbb{R}^d$, but it is mentioned above as a difficulty when considering the infinite-dimensional case. Instead of the transition function in the form $P(s,x,t,B)$ we may look at the family of linear non-negative operators $P(t,s)$ from $C_0(\mathbb{R}^d)$ into itself as a two-parameter C_0-semigroup, which satisfies $0 \leq P(t,s)\varphi \leq 1$ for any $0 \leq \varphi \leq 1$.

For instance, the reader is referred to Stroock and Varadhan [169, Chapter 9, pp. 208–247] for some useful estimates on the transition probability functions for diffusion processes in $\mathbb{R}^d$.

In either of these two equivalent forms of transition function we complete the definition by using the one-point compactification of E, say $\bar{E} = E \cup \{\infty\}$ with $P(s,x,t,\{\infty\}) = 1 - P(s,x,t,\mathbb{R}^d)$, so that $P(s,x,t,B)$ is a transition function in compact Polish space $\bar{E}$. Thus, time-homogeneous means $P(t,s) = P(t-s)$ while spatially-homogeneous means that $P(t,s)$ commutes with the translations operators $T_h\varphi(x) = \varphi(x-h)$, i.e., for any $t > s \geq 0$ and h in $\mathbb{R}^d$ we have $P(t,s)\, T_h = T_h\, P(t,s)$ in $C_0(\mathbb{R}^d)$.

Condition (c) or (c') means that the Markov process X is stochastically continuous, i.e., for any $\varepsilon > 0$ and $s \geq 0$ there is a $\delta > 0$ such that $P\{|X(t) - X(s)| \geq \varepsilon\} < \varepsilon$ for any t in $](s-\delta) \wedge 0, s+\delta[$. On a bounded interval, this is equivalent to a *uniformly stochastically continuous* property, namely for any $\varepsilon > 0$ there is a $\delta > 0$ such that $P\{|X(t) - X(s)| \geq \varepsilon\} < \varepsilon$ for any t, s in $[0, 1/\varepsilon]$ satisfying $|t - s| \leq \delta$. Actually, because the Polish space E is locally compact, both conditions (c) and (c') are equivalent under the Feller assumption (d).

The relation between a transition function and the evolution operators (or semigroup) is clearly (5.38) with $\Phi(t,s) = P(t,s)$. In the time-homogeneous case, this relates with the resolvent operators by

$$R(\lambda)\varphi(x) = \int_0^\infty \mathrm{e}^{-t\,\lambda}\,\Phi(t)\varphi(x)\mathrm{d}t = \\ = \int_0^\infty \mathrm{e}^{-t\,\lambda}\mathrm{d}t \int_{\mathbb{R}^d} \varphi(y)P(t,x,\mathrm{d}y),\ \ \forall x \in \mathbb{R}^d, \quad (5.39)$$

which may be generalized to the non-homogeneous case.

A crucial relation between the transition function $P(s,x,t,B)$ and the family of stochastic processes $X = X_{sx}$ is the equality

$$P\{X(t) \in B \mid X(r),\, r \le s\} = P(s, X(s), t, B), \quad (5.40)$$

for every $t > s \ge 0$ and B in $\mathcal{B}(\mathbb{R}^d)$, which is the Markov property itself. This is the primary building block, in the sense that when the family of stochastic processes X is given first, some property on their paths is necessary to construct the transition function, condition (5.34) is not sufficient. The general theory of Markov processes is rather delicate, so that we prefer to limit ourself to the case of standard Markov processes, i.e., cad-lag path and stochastically continuous in a filtered spaces (satisfying the usual conditions).

Generally, a Markov process is used for modeling the dynamic of a motion (e.g., of a particle). Intuitively, the Markov property expresses a *prediction* of subsequent motion (of a particle), knowing its position at time t, does not depend on what has been observed during the time interval $[0,t]$. In most of the cases, the above (simple) Markov property is not sufficient, this *starting afresh* property need to be used with stopping times. This is called the *strong Markov property* and written as

$$P\{X(t+\tau) \in B \mid X(r+\tau),\, r \le 0\} = P(\tau, t, X(\tau), B), \quad (5.41)$$

for every $t \ge 0$, B in $\mathcal{B}(\mathbb{R}^d)$, and every stopping time. It is clear that any Markov process with cad-lag paths and a Feller transition satisfies the strong Markov property (5.41).

Only in very particular cases the transition function is explicitly known, such as a Wiener or a Poisson process. In most of the cases, the transition function is constructed from a family of linear possible unbounded (infinitesimal generator) operators $A(t)$ defined in a domain $\mathcal{D}(A(t))$ and indexed in $t \ge 0$. Moreover, what is actually known is the expression to the operators $A(t)$ for *smooth* or *test* functions, e.g., $A(t)$ is a second order elliptic differential operator with given coefficients, or more general an integro-differential operator of a particular form. The semigroup theory or the theory of evolution operators address this question, i.e., (1) if a semigroup $\{\Phi(t) : t \ge 0\}$ is given then characteristic properties on its so-called infinitesimal generator A are listed and (2) if a given operator A satisfies the characteristic properties of an infinitesimal generator

then a semigroup $\{\Phi(t) : t \geq 0\}$ can be constructed. For a linear and bounded operator A the arguments go back to the exponential function, i.e.,

$$A\varphi = \lim_{t\to 0} \frac{\Phi(t)\varphi - \varphi}{t} \quad \text{and} \quad \Phi(t) = \sum_{n=0}^{\infty} \frac{(t\,A)^n}{n!} = \mathrm{e}^{tA}.$$

In general, a much more sophisticated argument is necessary, Conditions (a') and (e') of the Definition 5.6 are characteristic properties of the so-called Markov-Feller (or Feller-Dynkin) semigroups, which is the main tool we use to model stochastic dynamics. Clearly, assumption (e') imposes a certain type of regularity, while (a') translates into the so-called *maximum principle* satisfied by its infinitesimal generator, see Chapter 2 for an overview of the semigroup $\Phi(t)$ and its infinitesimal generator A

For a given transition probability function $P(s, x, t, B)$ as in Definition 5.6, since $P(s, x, t, B)$ and an initial distribution determine the finite-dimensional of the Markov process, we may use Kolmogorov's construction to define a family of $\mathbb{R}^d$-valued random variables $\{X_{sx}(t) : t \geq 0\}$ for each initial time $s \geq 0$ and initial distribution x in $\mathbb{R}^d$ such that the Markov property (5.40) is satisfied, i.e., for any real numbers $s < t_1 < \cdots < t_n$ and Borel subsets $B_1, \ldots, B_n$ of R^d the family of probability measures

$$P_{sx,t_1,\ldots,t_n}(B_1 \times \ldots \times B_n) = \int_{B_1} P(s, x, t_1, \mathrm{d}x_1) \times \\ \times \int_{B_2} P(t_1, x_1, t_2, \mathrm{d}x_2) \ldots \int_{B_n} P(t_{n-1}, x_{n-1}, t_n, \mathrm{d}x_n),$$

for any $s < t_1 < \cdots < t_n$, has the consistency property. Therefore there exists a unique probability measure P_{sx} on the space Ω of all functions from $[s, \infty)$ into $\mathbb{R}^d$ such that $P_{sx}\{X(t) \in B\} = P(s, x, t, B)$ for any $t > 0$ and B in $\mathcal{B}(\mathbb{R}^d)$, where X is the canonical (coordinate or projection) process, namely $X(t, \omega) = \omega(t)$ for any ω in Ω. Besides this, for any bounded and measurable function $f(x_1, \ldots, x_n)$ we have

$$\mathbb{E}_{sx}\{f(X(t_1), \ldots, X(t_n))\} = \int P(s, x, t_1, \mathrm{d}x_1) \times \\ \times \int P(t_1, x_1, t_2, \mathrm{d}x_2) \ldots \int f(x_1, \ldots, x_n)\, P(t_{n-1}, x_{n-1}, t_n, \mathrm{d}x_n).$$

Thus, the Markov property (5.40) holds true for this construction. Since no condition on the paths is assumed, this is referred to as a Markov process *in law*, where the crucial Markov property may be re-written as

$$\mathbb{E}_{sx}\{f(X(s_1), \ldots, X(s_m))\, g(X(r + t_1), \ldots, X(r + t_n))\} = \\ = \mathbb{E}_{sx}\{f(X(s_1), \ldots, X(s_m))\, h(X(r))\},$$

where $h(\xi) = \mathbb{E}_{r\xi}\{g(X(r + t_1), \ldots, X(r + t_n))\}$ and $s < s_1 < \ldots < s_m \leq r \leq t_1 < \ldots < t_n$. Note that only conditions (a), (b) and (d) in Definition 5.6 of

transition function are used to construct a Markov process in law. As mentioned previously, if the transition function $P(s,x,t,B)$ is not a full probability, i.e., $P(s,x,t,\mathbb{R}^d) \leq 1$ then we need to use the one-point compactification $\bar{\mathbb{R}}^d$ of R^d and define $P(s,x,t,\{\infty\}) = 1 - P(s,x,t,\mathbb{R}^d)$ and $P(s,\infty,t,\{\infty\}) = 1$. In this case, the above random variables $\{X_{sx}(t) : t \geq 0\}$ take values in $\bar{\mathbb{R}}^d$.

Given a transition function $P(s,x,t,B)$ we define the modulus of stochastic continuity by

$$\alpha(\varepsilon,T,\delta,K) = \sup\big\{1 - P\big(s,x,t,\{y : |y-x| \leq \varepsilon\}\big) : \\ : \forall x \in K,\ s,t \in [0,T],\ 0 < t-s \leq \delta\big\}, \quad (5.42)$$

where $K \subset \mathbb{R}^d$. Because of assumption (c) or (c') on a transition function we know that for any $\varepsilon, T > 0$ and any x in $\mathbb{R}^d$ we have $\alpha(\varepsilon,T,\delta,\{x\}) \to 0$ as $\delta \to 0$. However, we need to assume that

$$\lim_{\delta \to 0} \alpha(\varepsilon,T,\delta,\mathbb{R}^d) = 0, \quad \forall \varepsilon, T > 0, \quad (5.43)$$

This condition (5.43) is satisfied for a Feller transition.

The following result addresses the construction of standard Markov processes

Theorem 5.7. *Let $P(s,x,t,B)$ be a transition probability function satisfying (5.43). Then for any initial condition (s,x) there exists a probability measure P_{sx} on the canonical space $D([0,\infty),\mathbb{R}^d)$ such that the canonical process $X(t,\omega) = \omega(t)$ is a Markov process with transition function $P(s,x,t,B)$, which satisfies $P_{sx}\{X(t) = x,\ t \leq s\} = 1$. Moreover, if the transition function satisfies*

$$\lim_{\delta \to 0} \frac{\alpha(\varepsilon,T,\delta,\mathbb{R}^d)}{\delta} = 0, \quad \forall \varepsilon, T > 0, \quad (5.44)$$

then the support of the measure P_{sx} is the canonical space $C([0,\infty),\mathbb{R}^d)$. Furthermore, if $P(s,x,t,B)$ is a Feller transition function then the strong Markov property relative to the canonical filtration $(\mathcal{F}(t) : t \geq 0)$ (universally completed with respect to the family $\{P_{sx} : (s,x)\}$ and right-continuous), i.e.,

$$P_{sx}\{X(\theta) \in B \mid \mathcal{F}(\tau)\} = P(\tau, X(\tau), \theta, B),\ \ \forall B \in \mathcal{B}(\mathbb{R}^d), \quad (5.45)$$

for any finite stopping times $\theta \geq \tau \geq s$, and the filtration $(\mathcal{F}(t) : t \geq 0)$ is quasi-left continuous.

Proof. Since this is a classic result for the construction of Markov processes, only the key points will be discussed here, for instance, reader may consult the book by Dellacherie and Meyer [32, Section XIV.24, pp. 169–172] or Sato [157, Theorem 11.1, pp. 59–63] for details.

First, we need some notation. Let R be a subset of times in $[0,\infty)$ and $\varepsilon > 0$. We say that a family $X = \{X(t) : t \geq 0\}$ of $\mathbb{R}^d$-valued random variables (1) has ε-oscillations n-times in R for a fixed ω if there exist $t_0 < t_1 < \cdots < t_n$ in R such that $|X(t_i) - X(t_{i-1})| > \varepsilon$ for any $i = 1,\ldots,n$, or (2) has ε-oscillations

infinitely often in R for a fixed ω if for any n the family X has ε-oscillations n-times in R. Denote by $B^X(n,\varepsilon,R)$ and $B^X(\infty,\varepsilon,R)$ the set of ω where X has ε-oscillations n-times and infinitely often in R, respectively.

Most of the arguments is to find a modification of the Markov process in law constructed above. To that effect, denote by Ω_2 the set of ω such that the one-sided limits

$$\lim_{s\to t,\, s<t\, s\in\mathbb{Q}} X(s,\omega) \qquad \text{and} \qquad \lim_{s\to t,\, s>t\, s\in\mathbb{Q}} X(s,\omega)$$

exist in $\mathbb{R}^d$ for any $t\geq 0$. Note that for any strictly decreasing sequence $\{t_n\}$ to t, of rational numbers in $[0,\ell]$, there exists $N=N(\varepsilon,\ell)$ such that $|X(t_n,\omega)-X(t_N,\omega)|\leq\varepsilon$ for any $n\geq N$ and ω in $\Omega\setminus B^X(\infty,\varepsilon,[0,\ell]\cap\mathbb{Q})$. This shows that Ω_2 contains the set

$$\Omega_2^*=\Omega\setminus\bigcup_{\ell=1}^{\infty}\bigcup_{k=1}^{\infty}B^X(\infty,4/k,[0,\ell]\cap\mathbb{Q}),$$

which is measurable since $\mathbb{Q}$ is countable.

The following modification, $X^*(t,\omega)=0$ for every $\omega\in\Omega\setminus\Omega_2^*$, and

$$X^*(t,\omega)=\lim_{s\to t,\, s<t\, s\in\mathbb{Q}} X(s,\omega),\quad \forall\omega\in\Omega_2^*,$$

has cad-lag paths and because the stochastically continuity we obtain

$$P\{X(t,\omega)=X^*(t,\omega),\,\omega\in\Omega_2^*\}=1.$$

To complete this cad-lag modification we need to show that $P(\Omega_2^*)=1$.

The following estimate, proved by induction on the integer n, yields the result as discussed below. If $0\leq s_1<\cdots<s_m\leq r\leq t_1<\cdots<t_k<r+\delta\leq T$ and $R=\{t_1,\ldots,t_k\}$ then we have

$$\mathbb{E}\{Z\,\mathbb{1}_{B^X(n,4\varepsilon,R)}\}\leq\mathbb{E}\{Z\}\,[2\alpha(\varepsilon,T,\delta,\mathbb{R}^d)]^n, \tag{5.46}$$

for every $Z=f(X(s_1),\ldots,X(s_\ell))$ with a nonnegative measurable function f, and where $\alpha(\varepsilon,T,\delta,\mathbb{R}^d)$ is defined by (5.42). A key point is the fact that the right-hand side does not depend on k.

Thus, to show that $P(\Omega_2^*)=1$ we will prove that $P\{B^X(\infty,4/k,[0,\ell]\cap\mathbb{Q})\}=0$ for any integer k and ℓ. Indeed, by making a subdivision of $[0,\ell]$ into j equal intervals, we obtain

$$P\{B^X(\infty,4/k,[0,\ell]\cap\mathbb{Q})\}\leq \\ \leq\sum_{i=1}^{j}\lim_{n\to\infty}P\{B^X(n,4/k,[(i-1)\ell/j,i\ell/j]\cap\mathbb{Q}),$$

and from the above estimate (5.46) with $\{t_1,t_2,\ldots\}=[(i-1)\ell/j,i\ell/j]\cap\mathbb{Q}$ deduce

$$P\{B^X(n,4/k,[(i-1)\ell/j,i\ell/j]\cap\mathbb{Q})\}\leq[2\alpha(1/k,\ell,\ell/j,\mathbb{R}^d)]^n,$$

for every $n = geq1$. In view of condition (5.43), for a given ℓ we can select the integer j sufficiently large so that $2\alpha(1/k, \ell, \ell/j, \mathbb{R}^d) < 1$. Hence, as $n \to \infty$ we get $P\{B^X(n, 4/k, [(i-1)\ell/j, i\ell/j] \cap \mathbb{Q})\} = 0$, which implies $P(\Omega_2^*) = 1$.

When condition (5.44) is satisfied, we have to find a measurable set Ω_1^* with $P(\Omega_1^*) = 1$ and such that $X^*(t,\omega) = X^*(t-,\omega)$ for any $t > 0$ and ω in Ω_1^*. Indeed, for a given $\ell > 0$, consider the set $R(n, \varepsilon, \omega)$, with $n = 1, 2, \ldots$ and $\varepsilon > 0$, defined as the number of $i = 1, \ldots, n$ such that $|X^*(i\ell/n, \omega) - X^*((i-1)\ell/n, \omega)| > \varepsilon$. Then, $\omega \mapsto R(n, \varepsilon, \omega)$ is measurable and

$$\mathbb{E}\{R(n, \varepsilon, \cdot)\} \leq n\, \alpha(\varepsilon, \ell, \ell/n).$$

Hence, condition (5.44) and Fatou's lemma yield $\mathbb{E}\{\liminf_{n\to\infty} R(n, \varepsilon, \cdot)\} = 0$ and therefore the set

$$\Omega_1^*(\ell) = \bigcap_{k=1}^{\infty} \big\{\omega : \liminf_{n\to\infty} R(n, 1/k, \omega) = 0\big\}$$

is measurable with full measure, i.e., $P\{\Omega_1^*(\ell)\} = 1$. Moreover, if ω is in $\Omega_1^*(\ell)$ then for any t in $(0, \ell]$ we have $|X^*(t,\omega) - X^*(t-,\omega)| \leq \varepsilon$, for every $\varepsilon > 0$. Thus $\Omega_1^* = \bigcap_\ell \Omega_1^*(\ell)$ has the desired property.

It is clear that once a cad-lag version, namely X^*, has been found, we can take the image probability measure in the canonical space to produce P_{sx} as required. On the other hand, the stochastic continuity and the cad-lag regularity of the paths imply that $P\{X^*(t) = X^*(t-)\} = 1$ for any $t > s$.

The right-continuity of paths ensures that the process X^* is adapted to $\mathcal{F}(t) = \mathcal{F}^{sx}(t+) = \bigcap_{\varepsilon>0} \mathcal{F}^{sx}(t)$, where $\mathcal{F}^{sx}(t)$ is the σ-algebra generated by the canonical process and P-null sets. Thus (5.45) is satisfied after using the continuity of the transition probability function and approximating any finite stopping time.

Regarding the quasi-left continuity we proceed as follows. Let $\{\tau_n : n \geq 1\}$ be a sequence of stopping times convergence almost surely to τ, with $P(\tau_n < \tau < \infty, \tau > s) = 1$. For any two functions f and g in $C_0(\mathbb{R}^d)$ we have

$$\lim_{t\to 0} \lim_{n\to\infty} \mathbb{E}\{f(X^*(\tau_n))\, g(X^*(\tau_n + t))\} =$$
$$= \lim_{t\to 0} \mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau + t-))\} = \mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau))\},$$

because the right-continuity of the paths. On the other hand, the strong Markov property (5.41) and the Feller property imply

$$\lim_{n\to\infty} \mathbb{E}\{f(X^*(\tau_n))\, g(X^*(\tau_n + t))\} = \mathbb{E}\{f(X^*(\tau-))\, P(\tau, \tau + t, X^*(\tau-), g)\}$$

and

$$\lim_{t\to 0} \mathbb{E}\{f(X^*(\tau-))\, P(\tau, \tau + t, X^*(\tau-), g)\} = \mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau-))\}.$$

Hence,

$$\mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau))\} = \mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau-))\},$$

i.e., $P\{X^*(\tau) = X^*(\tau-)\} = 1$ and X^* is almost surely continuous at τ. □

Usually, condition (5.43) is replaced by

$$\begin{aligned} &(a) \quad \lim_{|x|\to\infty} \sup_{0\le s<t\le T} P(s,x,t,K) = 0, \\ &(b) \quad \lim_{\delta\to 0} \alpha(\varepsilon,T,\delta,K) = 0, \quad \forall \varepsilon, T > 0, \end{aligned} \tag{5.47}$$

for any compact subset K of $\mathbb{R}^d$, and assumption (5.44) can be substituted by

$$\lim_{\delta\to 0} \frac{\alpha(\varepsilon,T,\delta,K)}{\delta} = 0, \quad \forall \varepsilon, T > 0, \quad \text{any compact } K \subset \mathbb{R}^d, \tag{5.48}$$

and in general this construction ie valid for a transition function, without the probability condition $P(s,x,t,\mathbb{R}^d) = 1$, see Taira [171, Chapter 9 and 10, pp. 273–424].

To properly handle the strong Markov property, we need to use the *universally complete* σ-algebra, i.e., first we remark that the above construction can be used with any initial law μ at any time 0 and the corresponding filtration is $\{\mathcal{F}^{0\mu}(t) : t \ge 0\}$. Thus $\mathcal{F}^0(t) = \bigcap_\mu \mathcal{F}^{0\mu}(t)$, which is not necessarily complete with respect to $P^{0\mu}$, but it satisfies $\mathcal{F}^0(t+) = \mathcal{F}^0(t)$, i.e., it is right-continuous, and the so called Blumenthal's zero-one law, i.e., $P(A) = 0$ or $P(A) = 1$ for any A in $\bigcap_{t>0} \mathcal{F}^0(t)$.

Let us look at the particular case of additive processes, see Definition 2.1, which include the Lévy processes. The transition function of an additive process is spatially homogeneous, i.e., if $P(s,x,t,B)$ is the transition function of an additive process X then $P(s,x,t,B) = P(s,0,t,B-x)$ and we only have to consider transition functions of the form $P(s,t,B)$. Thus, any additive process X yields a transition function $P(s,t,B) = P\{X(t) - X(s) \in B\}$, for any $t > s \ge 0$ and B in $\mathcal{B}(\mathbb{R}^d)$ so that X is a (stochastically continuous) Markov process in $\mathbb{R}^d$ stating at 0. Its associated semigroup is called a *convolution* semigroup, i.e.,

$$P(t,s)\varphi(x) = \int_{\mathbb{R}^d} \varphi(x+y) P(s,t,\mathrm{d}y)$$

and Chapman-Kolmogorov identity is re-written as

$$P(s,t,B) = \int_{\mathbb{R}^d} P(s,r,\mathrm{d}x)\, P(r,t,B-x),$$

for every $t > r > s \ge 0$ and B in $\mathbb{R}^d$. It is also clear that the previous Theorem 5.7 applies to this case, to obtain a cad-lag of additive processes in law. Because the transition function $P(s,t,B)$ is spatially homogeneous, it satisfies the Feller conditions and the process is quasi-left continuous, i.e., $X(Tn)$ converges to $X(T)$ on the set where $T < \infty$, for any increasing sequence of stopping time T_n converging to T, with $T_n < T$..

Lévy processes X are also time-homogeneous and its semigroup is a true convolution and the infinitely divisible distribution $\mu = X(1)$ completely determines the process, see Section 2.2. Thus to each infinitely divisible distribution

μ there corresponds a Lévy process. For instance, Poisson and compound Poisson processes correspond to Poisson and compound Poisson distributions. The Lévy process on $\mathbb{R}^d$ corresponding to a Cauchy distribution with parameters γ in $\mathbb{R}^d$ and $c > 0$, namely, for any B in $\mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} &\mu(B) = \pi^{-(d+1)/2}\Gamma(\tfrac{d+1}{2})\, c \int_B \left(|x-\gamma|^2 + c^2\right)^{-(d+1)/2} \mathrm{d}x, \\ &\text{and} \quad \hat{\mu}(y) = \mathrm{e}^{-c|y| + \mathtt{i}\,\gamma\cdot y}, \quad \forall y \in \mathbb{R}^d, \end{aligned} \tag{5.49}$$

is called a *Cauchy process*. However, the Lévy process on $\mathbb{R}$ corresponding to an exponential distribution is called a Γ-process, since it has a Γ distribution at any $t > 0$.

If X is an additive process on $\mathbb{R}^d$ with a Gaussian distribution at each t, then X has continuous paths almost surely, see Sato [157, Theorem 11.7, pp. 63-64]. For instance, for dimension $d = 1$, the characteristic function is

$$\mathbb{E}\{\mathrm{e}^{\mathtt{i}\,y\cdot X(t)}\} = \mathrm{e}^{-t\,y^2/2}, \quad \forall t \geq 0,\ y \in \mathbb{R}^d,$$

and a simple calculation shows that condition (5.44) of Theorem 5.7 is satisfied. Actually, the only additive process with continuous paths are Wiener processes.

For a given additive process X we consider the σ-algebra $\mathcal{F}(t)$ generated by all null sets and the family of random variables $X(s)$ with $s \leq t$. Because of the independence of increments, an application of Kolmogorov's zero-one law to a *tail* σ-algebra shows that $\mathcal{F}(t)$ is already right-continuous, so that it is the filtration associated with X.

The reader is referred to the books by Blumenthal and Getoor [15], Dellacherie and Meyer [32, Chapters XI–XVI], Ethier and Kurtz [45], Sato [157, Chapter 1 and 2, pp. 1–68], among others.

5.2 Markov-Feller Semigroups

Let E be a locally compact Polish (i.e., complete separable metric) space and $\mathcal{B}(E)$ be the σ-algebra of Borel subsets of E. Usually, $E \subset \mathbb{R}^d$, is an smooth domain. Moreover, in this section we may even take E to be a locally compact Hausdorff space with countable basis. As before, $B(E)$ denotes the space of Borel measurable and bounded functions from E into $\mathbb{R}^d$, endowed with the supremum norm $\|\cdot\|$. Recall that a function f *vanishes at infinity* if for any $\varepsilon > 0$ there is a compact set K such that $|f(x)| < \varepsilon$ for any $x \in E \setminus K$. Also, $C_b(E)$, respectively $C_0(E)$, stands for the space of continuous functions which are bounded, respectively vanishing at infinity. It is clear that if the whole space $E = K$ is compact then $C_b(K) = C_0(K)$, in this case we use the notation $C(K)$. The Riesz representation theorem states that any bounded linear functional on $C(K)$ may be uniquely written in the form

$$\mu(f) = \int_K f(x)\, \mu(\mathrm{d}x), \quad \forall f \in C(K),$$

where μ is a regular bounded (signed) measure on $E = K$ (recall that regular means that for any measurable subset A of E and for every $\varepsilon > 0$ there exits an open set $\mathcal{O}$ and a closed set F, with $F \subset A \subset \mathcal{O}$ and $\mu(\mathcal{O} \setminus F) < \varepsilon$), see Dunford and Schwartz [38, p. 265, Theorem IV.6.3]. Moreover, if E is a locally compact Polish space, we may construct its one-point compactification $\bar{E} = E \cup \{\infty\}$ where open sets in $\bar{E}$ are of the open sets in E and the sets of the form $\{\infty\} \cup (E \setminus K))$ for any compact subsets K of E. Any regular bounded measure μ in $\bar{E}$ satisfying $\mu(\{\infty\} = 0)$, have the property that for any $\varepsilon > 0$ there exits a compact subset K of E such that $\mu(E \setminus K) < \varepsilon$, which usually refer to as μ being *tight* in E. The Banach space $C_0(E)$ of continuous functions on E vanishing at infinity can be identified with the Banach space of continuous function on $\bar{E}$ satisfying the zero boundary condition at infinity $f(\infty) = 0$. Thus, any bounded linear functional on $C_0(E)$ can be uniquely represented by a regular bounded (signed) measure (tight) in E, e.g., see Folland [52, Chapter 7] or Malliavin [115, Chapter II]. Therefore, we assume that the base space E is such that bounded linear nonnegative on $B(E)$ are uniquely given by (regular) bounded measures on E.

Definition 5.8 (Markov)**.** Let E be a locally compact Hausdorff space with countable basis and denote by $\mathcal{B}(E)$ its Borel σ-algebra.

(1) A one-parameter family $\{S(t) : t \geq 0\}$ of bounded linear operators from the Banach space $B(E)$ into itself is called a *(sub-)Markov semigroup* with *(sub-)Markovian kernels* $\{P(t, x, A) : t \geq 0,\ x \in E,\ A \in \mathcal{B}(E)\}$ given by

$$S(t)f(x) = \int_E f(y)\, P(t, x, \mathrm{d}y), \quad \forall f \in B(E),$$

if it satisfies

(a) $S(t + s) = S(t)S(s), \quad \forall t, s \geq 0,$
(b) $S(t)f(x) \geq 0, \quad \forall t \geq 0,\ x \in E \quad$ if $\quad f(x) \geq 0,\ \forall x \in E,$
(c) $S(t)\mathbb{1}_E(x) \leq 1, \quad \forall t \geq 0,\ x \in E.$

or equivalently

(a)′ for each $s, t \geq 0$, x in E and A in $\mathcal{B}(E)$ we have

$$P(s + t, x, A) = \int_E P(s, x, \mathrm{d}y)\, P(t, y, A),$$

which is referred to as the Chapman-Kolmogorov identity.

(b)′ for each t and x the function $A \mapsto P(t, x, A)$ is a (non-negative) measure on $\mathcal{B}(E)$ with $P(t, x, E) \leq 1$ and $P(0, x, \{x\}) = 1$,

(c)′ for each t and A in $\mathcal{B}(E)$ the function $x \mapsto P(t, x, A)$ is a Borel measurable,

(2) It is called a *transition function* if for every A in $\mathcal{B}(E)$ the mapping $(t, x) \mapsto P(t, x, A)$ is jointly Borel measurable in $[0, \infty) \times E$.

(3) It is called *stochastically continuous* if

$$\lim_{t \to 0} P(t, x, \mathcal{O}) = 1,$$

for every x in E and any open neighborhood $\mathcal{O}$ of x.

(4) It satisfies the (pointwise) *Feller property* (respectively, *strong Feller property*) if for every $t > 0$ the function $x \mapsto S(t)f(x)$ is continuous at each point of continuity of the function f (respectively, at each point x). □

Here, we have denoted by $\mathbb{1}_A$ the characteristic function of the subset A, i.e., $\mathbb{1}_A(x) = 1$ if x belongs to A and zero otherwise. It is clear that condition (a) is the usual semigroup property, condition (b) is the *weak maximum principle* and inequality (c) is a *normalization* condition. Actually, condition (c) can be replaced by the equality $S(t)\mathbb{1}_E(x) = 1$, for any $t \geq 0$, $x \in E$, without any lost of generality, by using the one-point compactification. This give rise to the distinction between Markov and sub-Markov semigroups.

If the base space E is not locally compact, then we normally add the condition

$$\text{(d)} \quad A \mapsto S(t)\mathbb{1}_A(x) \quad \text{is } \sigma\text{-additive on} \quad \mathcal{B}(E), \quad \forall t \geq 0,\ x \in E$$

to the definition of a Markov semigroup. This condition is automatically satisfied if E is a locally compact Polish space. We refer to a Markov semigroup $S(t)$ or to a Markov kernels $P(t, x, \cdot)$ indistinctly.

In general, a Markov semigroup is not strongly continuous in $B(E)$, even if it satisfies the above Feller property. Moreover, a joint measurability condition is needed to define the resolvent operators

$$R_\lambda f(x) = \int_0^\infty \mathrm{e}^{-\lambda t} S(t)f(x)\mathrm{d}t, \quad \forall x \in E,\ \lambda > 0,$$

as a mapping from $B(E)$ into itself. This is precisely the condition (2) in Definition 5.8, i.e., a Markov transition function.

If E is a Polish space, the Markov semigroup $\{S(t) : t \geq 0\}$ or its Markov kernels $P(t, x, \cdot)$ is stochastically continuous if and only if

$$\lim_{t \to 0} P(t, x, B(x, \delta)) = 1, \quad \forall x \in E,\ \delta > 0,$$

where $B(x, \delta)$ is the ball of center x and radius δ in E. Notice that the above Feller property refers to the *space* variable x, whilst stochastically continuous involves the *time* variable t. Even if the base space E is not locally compact, it is proved in Da Prato and Zabczyk [29, p. 13] that a Markov semigroup $\{S(t) : t \geq 0\}$ is stochastically continuous if and only if $S(t)f(x)$ converges to $f(x)$ as $t \to 0$, for any $x \in E$ and any function f which is either (a) bounded and continuous or (b) bounded and uniformly continuous or (c) bounded and Lipschitz continuous.

It is clear that a stochastically continuous Markov semigroup $\{S(t) : t \geq 0\}$ is (Borel) measurable, i.e, $(t, x) \mapsto S(t)f(x)$ is jointly Borel measurable in $[0, \infty) \times E$, i.e, a Markov transition function. Thus we can use the general results in Dellacherie and Meyer [32, Section XIV.24, pp. 169–172]) to construct a cadlag realization of the associated Markov (strong Markov, since it is stochastically

continuous) Markov process as described in Chapter 1. Note that a systematic study on analytic methods for Markov diffusion semigroup can be found in Bertoldi and Lorenzi [8].

5.2.1 Feller Semigroups

A good way is to consider the semigroup restricted to the space $C_b(E)$ or $C_0(E)$ and impose the Feller property.

Definition 5.9 (Feller)**.** Let E be a Polish space. Then a one-parameter family $\{S(t) : t \geq 0\}$ of bounded linear operators from a closed subspace C of the Banach space $C_b(E)$ (e.g., $C_0(E)$ or the whole space $C_b(E)$) into itself is called a *Feller semigroup* if it satisfies

(a) $S(t+s) = S(t)S(s), \quad \forall t, s \geq 0,$

(b) $0 \leq S(t)f \leq 1, \quad \forall t \geq 0 \quad \text{if} \quad 0 \leq f \leq 1,$

(c) $\lim_{t \downarrow 0} \|S(t)f - f\| = 0, \quad \forall f \in C.$

Actually, a Feller semigroup need not to be *strongly continuous* as indicated by condition (d), usually only a *weakly continuous* condition, namely

(e) $\lim_{t \downarrow 0} S(t)f(x) = f(x), \quad \forall x \in E, \quad \forall f \in C$

is required. However, since our base space E is locally compact (see Dellacherie and Meyer [32, Theorem XIII.19, pp. 98–99]) these two conditions are equivalent. □

Roughly speaking, a *Markov semigroup* is semigroup associated with some Markov process and a *Feller semigroup* (or Markov-Feller semigroup or Feller-Dynkin semigroup) is a Markov semigroup which satisfies the *Feller property*. Actually, most of the key results on Markov theory requires a stochastically continuous Markov semigroup.

The measure theory ensures that any Feller semigroup in either $C_b(E)$ or $C_0(E)$ can be extended to be a Markov semigroup in $B(E)$. It is clear that for a Feller semigroup, the condition (e) above on *weakly continuous* is the equivalent of *stochastically continuous* for a Markov semigroup. Clearly, in a locally compact Polish space, a stochastically continuous Markov semigroup is indeed a Feller semigroup as mentioned in the above definition.

The following result (on locally compact Polish space E) is taken from Taira [171, Chapter 9, Section 2, pp. 333–340],

Theorem 5.10 (Markov-Feller)**.** *Let $\{S(t) : t \geq 0\}$ be a Markov semigroup in $B(E)$, which leaves invariant the subspace $C_0(E)$ (i.e., $S(t)f \in C_0(E)$, $\forall t > 0$, $f \in C_0(E)$) and is uniformly stochastically continuous (i.e., the continuity condition in t at 0 holds uniformly on compact set in x). Then the restriction of $S(t)$ to $C_0(E)$ is a Feller semigroup if and only if the following property is satisfied*

(L) *For any $T, \varepsilon > 0$ and any compact $C \subset E$ there exists another compact set $K = K(T, \varepsilon, C)$ of E such that $P(t, x, C) < \varepsilon$, for any $t \in [0, T]$ and any $x \in E \setminus K$,*

where $P(t, x, \cdot)$ is the Markov transition function associated with $\{S(t) : t \geq 0\}$. □

Since a Feller semigroup $S(t)$ on $C_0(E)$ is strongly continuous, the general (contraction) semigroup theory applies to characterize the infinitesimal generator A of $S(t)$. The extra property involved in term of the resolvent operator $R(\lambda, A) = (\lambda I - A)^{-1}$ can be formulated as

$$f \in C_0(E),\ f \geq 0 \Longrightarrow R(\lambda, A) f \geq 0, \tag{5.50}$$

which is referred to as the *weak Maximum Principle*. In connection with this, we mention the following result, see Taira [171, Chapter 9, Section 3, pp. 340–349],

Theorem 5.11. *Let E be a compact Polish space and A be a linear operator (not necessarily bounded) in $C(E) = C_b(E) = C_0(E)$ densely defined on the domain $\mathcal{D}(A)$. Assume that for some $\lambda \geq 0$ the range $\mathcal{R}(\lambda I - A)$ of $\lambda I - A$ is dense in $C(E)$. If the weak Maximum Principle is satisfied in the following sense*

(wMP) *If u belongs to $\mathcal{D}(A)$ and it takes a positive maximum at the point x_0 in E then $Au(x_0) \leq 0$,*

then the closure $\bar{A}$ of the operator A is the infinitesimal generator of a Feller semigroup in the space $C(E)$.

Since a Feller semigroup $\{S(t) : t \geq 0\}$ has the property that $(t, x) \mapsto S(t)f(x)$ is Borel measurable, the Laplace transform

$$R(\lambda) f = \int_0^\infty \mathrm{e}^{\lambda t} S(t) \mathrm{d}t$$

defines the resolvent operators $\{R(\lambda) : \lambda > 0\}$ from $B(E)$ into itself. The specific properties of a Markov semigroup $\{S(t) : t \geq 0\}$ (on $B(E)$ Borel bounded functions on E) as in Definitions 5.8 or 5.9 become

$$\text{(a)}\ \ R(\lambda) f \geq 0,\ \forall f \geq 0, \quad \lambda > 0, \qquad \text{(b)}\ \ \lambda R(\lambda) \mathbb{1}_E(x) \leq 1, \quad \forall x \in E.$$

the converse is also valid.

For instance, the reader may consult the books Jacob [80, Vol I, Chapter 4] and Taira [171, Chapters 9 and 10] for a detailed presentation on Feller semigroups, among other topics.

5.2.2 Markov Process Realization

Until know, we have mentioned several results concerning the construction of a Markov process starting from a transition probability function. To understand better the relation between Markov processes and Markov semigroups,

we discuss the main steps and difficulties to construct a realization of a given Markov semigroup $S(t)$ with transition function $P(t,x,\mathrm{d}y)$ on a compact base space $\bar{E}$, the one-point compactification of E, assumed locally compact and $P(t,x,\{\infty\}) = 1 - P(t,x,E)$, so that $P(t,x,\bar{E}) = 1$. Intuitively, given any initial position x in $\bar{E}$ at time 0, the probability distribution for a location of the stochastic process X at a future time $t > 0$ is $P(t,x,\mathrm{d}y)$. Thus to construct a separable version (see Definition 1.13) of the stochastic process X, we need to construct a measure on the space of *paths* $\bar{E}^I$, where I is a countable dense set in $(0,\infty)$, for instance I is the set of positive rational numbers. Since $\bar{E}$ is compact and I countable, the product topology in $\bar{E}^I$ produces a compact metrizable space. In view of Riesz representation theorem, to construct a probability measure on $\bar{E}^I$, it suffices to construct a positive linear functional $\mathbb{E}^x$ from the space $C(\bar{E}^I)$ of real-valued continuous functions to $\mathbb{R}$. Naturally, first we define $\mathbb{E}^x$ on the subspace $C_c(\bar{E}^I)$ of cylindrical functions, consisting of continuous functions that depend on only finitely many factors of $\bar{E}^I$, i.e., functions on $C(\bar{E}^I)$ of the form $\psi(\omega) = F(\omega(t_1),\omega(t_2),\dots,\omega(t_n))$, $t_1 < t_2 < \cdots < t_n$, for some $n \geq 1$, where F is a continuous function in $\bar{E}^n$ and t_i belongs to I for $i = 1,2,\dots,n$. We set

$$\mathbb{E}^x(\psi) = \int P(t_1, x_1 - x, \mathrm{d}x_1) \int P(t_2 - t_1, x_2 - x_1, \mathrm{d}x_2) \dots \\ \dots \int F(x_1, x_2, \dots, x_n)\, P(t_n - t_{n-1}, x_n - x_{n-1}, \mathrm{d}x_n), \quad (5.51)$$

which is well defined in view of the semigroup or Chapman-Kolmogorov identity. Hence, $\mathbb{E}^x : C_c(\bar{E}^I) \to \mathbb{R}$ is a positive linear functional satisfying $\mathbb{E}^x(\mathbb{1}) = 1$. By the Stone-Weierstrass theorem on polynomial approximations, the subspace $C_c(\bar{E}^I)$ is dense and then $\mathbb{E}^x$ has a unique continuous extension to $C(\bar{E}^I)$, with the same properties. Then, we have a unique probability measure P_x on $C(\bar{E}^I)$ such that

$$\mathbb{E}^x(\psi) = \int_{\bar{E}} \psi(\omega) P_x(\mathrm{d}\omega)$$

and (5.51) holds on $C_c(\bar{E}^I)$. To go further in this construction, we need some regularity in the variable t, i.e., the function

$$\rho(\varepsilon,\delta) = \sup_x \int_{|y-x|\geq\varepsilon} P(\delta, x, \mathrm{d}y)$$

as δ goes to zero for any fixed ε. If we want to get a probability measure P_x on the space $C([0,\infty),\bar{E})$ of continuous functions from $[0,\infty)$ into $\bar{E}$, then we need to show first that the set $C_{\mathrm{lu}}(I,\bar{E})$, of uniformly locally continuous functions (which are extended uniquely to continuous functions from $[0,\infty)$ into $\bar{E}$) is a Borel subset of $\bar{E}^I$, which contains the support of P_x. Actually, this is the *hard* point in the construction of the measure P_x. To this purpose, the set $C_{\mathrm{lu}}(I,\bar{E})$ can be expressed as a countable intersection of a countable union closed sets,

namely

$$C_{\rm lu}(I,\bar E)=\bigcap_{k,\ell=1}^{\infty}\bigcup_{n=1}^{\infty}F'(k,1/\ell,1/n),$$

where $F'(k,\varepsilon,\delta)$ is the complement of the set

$$F(k,\varepsilon,\delta)=\bigcup\big\{E(a,b,\varepsilon)\ :\ 0\le a<b\le a+\delta\le k\big\},$$
$$E(a,b,\varepsilon)=\big\{\omega\in\bar E^I\ :\ \exists\, t,s\in I\cap[a,b],\ |\omega(t)-\omega(s)|>2\varepsilon\big\}.$$

Since the complement of $E(a,b,\varepsilon)$ is closed in $\bar E^I$, the set $F(k,\varepsilon,\delta)$ is open and therefore $C_{\rm lu}(I,\bar E)$ is a Borel set. Next, assuming

$$P_x\{E(a,b,\varepsilon)\}\le 2\rho(\varepsilon/2,b-a)), \tag{5.52}$$

we have

$$P_x\{F(k,\varepsilon,\delta)\}\le 2\frac{k}{\delta}\rho(\varepsilon,\delta))$$

and if $\rho(\varepsilon,\delta)/\delta$ goes to zero as δ goes to zero, we can deduce the equality $P_x\{C_{\rm lu}(I,\bar E)\}=1$. To obtain the estimate (5.52), we may express the set $E(a,b,\varepsilon)$ as an increasing limit of sets of the form $A=\{\omega\ :\ \exists\, i,j$ such that $|\omega(t_i)-\omega(t_j)|>2\varepsilon$, for some ε,δ,n and $t_1<t_2<\cdots<t_n$ in I, with $t_n-t_1\le\delta$. Each set A of the above form is contained in the union $B\cup(C_1\cap D_1)\dots(C_n\cap D_n)$, where $B=\{\omega\ :\ |\omega(t_n)-\omega(t_1)|>\varepsilon\}$, $C_i=\{\omega\ :\ |\omega(t_n)-\omega(t_i)|>\varepsilon\}$ and $D_i=\{\omega\ :\ |\omega(t_1)-\omega(t_i)|>2\varepsilon$ and $|\omega(t_1)-\omega(t_j)|\le 2\varepsilon,\ \forall j\le i-1\}$. It is clear that $P_x(B)\le\rho(\varepsilon,\delta)$ and $P_x(C_i)\le\rho(\varepsilon,\delta)$. Since C_i is independent of D_i, we have $P_x(C_i\cap D_i)=P_x(C_i)\,P_x(D_i)$, which allow us to conclude, see Taylor [173, Vol 2, Chapter 11, pp. 303–307].

The point here is that under the assumption $\rho(\varepsilon,\delta)/\delta\to 0$ as $\delta\to 0$, we are able to construct the probability measure in the sample space of continuous paths $C([0,\infty),\bar E)$, e.g. the Wiener measure. In some cases, e.g., the (compound) Poisson measure, we have only $[\rho(\varepsilon,\delta)]^2/\delta\to 0$ as $\delta\to 0$ so that the construction in the sample space $C([0,\infty),\bar E)$ fails. Then, the sample space of right-continuous (with left-hand limits) paths $D([0,\infty),\bar E)$ is used. In this case, the set $E(a,b,\varepsilon)$ is re-defined as

$$E(a,b,\varepsilon)=\big\{\omega\in\bar E^I\ :\ \exists\, t,s,r\in I,\ \text{ such that}$$
$$|\omega(t)-\omega(s)|\wedge|\omega(s)-\omega(r)|>2\varepsilon\quad a\le r<s<t\le b\big\}.$$

Using the fact that the two events $\{|\omega(t)-\omega(s)|>2\varepsilon\}$ and $\{|\omega(s)-\omega(r)|>2\varepsilon\}$ are actually of the previous form and independent of each other, we notice that estimate (5.52) is modified as follows

$$P_x\{E(a,b,\varepsilon)\}\le 2[\rho(\varepsilon/2,b-a))]^2. \tag{5.53}$$

Naturally, instead of the subspace $C_{\rm lu}(I,\bar{E})$ we use the subset $D_{\rm lu}(I,\bar{E})$ of $\bar{E}^I$ composed by the restriction to I of functions in $D([0,\infty),\bar{E})$. Some more detailed analysis is needed to effectively give a realization of the Markov process in the sample space $D([0,\infty),\bar{E})$. The interested reader may take a look at Jacob [80, Vol III, Chapter 3] for a more complete overview on Feller semigroups and processes.

5.2.3 Pointwise Continuous Semigroups

First, note that for a given strongly continuous semigroup in a Banach space, the weak infinitesimal generator denoted by $\bar{A}$ is not a genuine extension of the strong infinitesimal generator A, indeed their domain of definition are the same and they agree on it. Moreover, a weakly continuous semigroup in a Banach space is actually strongly continuous, see Pazy [142, Chapter 2, Section 2, pp. 42–44].

Thus, one way to proceed is to consider the weak-star topology in $B(E)$, i.e., boundedly pointwise convergence. Hence, the notion of *pointwise continuous* semigroup (also called weakly continuous) and *weak-star infinitesimal generator* (also called weak infinitesimal generator) are necessary, see Dynkin [42]. Given a stochastically continuous Markov semigroup $\{S(t) : t \geq 0\}$, we restrict our attention to the subspace $B^0(E)$ of real bounded Borel functions f on E such that the map $t \mapsto S(t)f(x)$ is continuous for any x in E. It is clear that $B^0(E)$ contains $C_b(E)$ and it is invariant under $S(t)$ for any $t \geq 0$. Thus $\{S(t) : t \geq 0\}$ is a pointwise continuous semigroup on $B^0(E)$, i.e., besides (a), (b) and (c) of Definition 5.8 it also satisfies the condition $S(t)f(x) \to f(x)$ for any x in E and any f in $B^0(E)$. Then, the *weak-star* infinitesimal generator $\bar{A}$ can be (densely) defined on $B^0(E)$ be means of the boundedly pointwise convergence, i.e., $\bar{A}f = g$ if and only if $[S(t)f - f]/t$ converges (boundedly pointwise) to g, this means

$$\sup_{t>0} \frac{|S(t)f(x) - f(x)|}{t} \leq C, \quad \forall x \in E,$$

for some constant $C = C_f > 0$ and

$$\lim_{t\to 0} \frac{S(t)f(x) - f(x)}{t} = g(x), \quad \forall x \in E,$$

where necessarily $g = \bar{A}f$ belongs to $B(E)$.

This approach is more relevant when the base space E is not locally compact Polish space, i.e., E may be an infinite dimensional Hilbert space endowed with the weak or strong topology. For instance, as in [121], suppose that a (strong) homogeneous Markov process $y(t,x)$ is know (e.g., via a stochastic partial differential equations) and then, a semigroup is define as follow

$$\Phi_\alpha(t)h(x) = E\{\mathrm{e}^{-\alpha t}h(y(t,x))\}, \tag{5.54}$$

for any $\alpha > 0$, on the space of $C_b(X)$ of real (uniformly) continuous and bounded functions, where X is an open subset (or the closure of an open subset) in a

separable Banach space. Sometimes, we are required to consider the semigroup $\{\Phi_\alpha(t) : t \geq 0\}$ on a space with unbounded functions, e.g., $C_p^0(X)$ be the space of real uniformly continuous functions on any ball and with a growth bounded by the norm to the $p \geq 0$ power, in another words, the space of real functions h on X such that $x \mapsto h(x)(1 + |x|^2)^{-p/2}$ is bounded and locally uniformly continuous, with the weighted sup-norm

$$\|h\| = \|h\|_{C_p^0} = \sup_{x \in X}\{|h(x)|(\lambda + |x|^2)^{-p/2}\}, \tag{5.55}$$

with $\lambda > 0$. Suppose that the Markov process $y(t, x)$, defined on some probability space $(\Omega, \mathcal{F}, P)$, satisfies the conditions:

(1) $x \mapsto y(t, x)$ is locally uniformly continuous (in x), locally uniformly continuous for t in $[0, \infty)$, i.e., for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any x, $\bar{x}$ in X satisfying $|x - \bar{x}| < \delta$, $|x| \leq 1/\varepsilon$ and $|\bar{x}| \leq 1/\varepsilon$ we have

$$P\Big\{ \sup_{0 \leq t \leq 1/\varepsilon} |y(t, x) - y(t, \bar{x})| \geq \varepsilon \Big\} < \varepsilon. \tag{5.56}$$

(2) $t \mapsto y(t, x)$ is locally uniformly continuous (in t), for any x in X, (actually in a dense subset suffices) i.e., for any x in X and for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$P\Big\{ \sup_{0 \leq t \leq \delta} \sup_{0 \leq s \leq 1/\varepsilon} |y(t + s, x) - y(s, x)| \geq \varepsilon \Big\} < \varepsilon. \tag{5.57}$$

(3) For any $p > 0$ there are positive constants α_0 and λ sufficiently large such that the following estimate

$$E\{\sup_{t \geq 0} \mathrm{e}^{-\alpha_0 t}(\lambda + |y(t, x)|^2)^{p/2}\} \leq C_p\,(\lambda + |x|^2)^{p/2}, \quad \forall x \in \mathcal{O} \tag{5.58}$$

holds, with some $C_p \geq 1$ and $C_p = 1$ if the sup is removed in the left-hand side. Here we are using the notation $|\cdot|$ for either the Euclidean norm or the norm in the Banach space containing X.

It is clear that (5.58) plays a role only when X is unbounded and that the closure of an open subset, say $\bar{X}$ could be used instead of X in all that follows. The associate semigroup $\Phi_\alpha(t)$ is not necessarily a *strongly continuous* semigroup on $C_b(X)$ nor on $C_p^0(X)$. Actually, we have in mind $X = \mathbb{R}^d$ (i.e., an stochastic ODE where the above conditions are easily verified and $\Phi_\alpha(t)$ is strongly continuous) but these conditions apply also for more general situations (stochastic PDE), such as the stochastic Navier-Stokes equation, e.g. Menaldi and Sritharan [124, 125].

It is clear that $C_b(X) \subset C_q^0(X) \subset C_p^0(X)$ for any $0 \leq q < p$. Then for any $\alpha \geq 0$, the (linear) semigroup $(\Phi_\alpha(t),\, t \geq 0)$ with an α-exponential factor is a weak-star continuous Markov semigroup in the space $C_p(X)$, i.e.,

$$\begin{aligned}
&\Phi_\alpha(t + s) = \Phi_\alpha(t)\Phi_\alpha(s), \quad \forall\, s, t \geq 0,\\
&\|\Phi_\alpha(t)h\| \leq \|h\|, \quad \forall\, h \in C_p^0(X),\\
&\Phi_\alpha(t)h(x) \to h(x) \ \text{ as } \ t \to 0, \quad \forall\, h \in C_p^0(X),\\
&\Phi_\alpha(t)h(x) \geq 0, \quad \forall\, h \geq 0, \quad h \in C_p^0(X).
\end{aligned} \tag{5.59}$$

This follows immediately from the conditions (5.56), (5.57) and (5.58) imposed on the Markov process $y(t,x)$.

Since the semigroup is not strongly continuous, we cannot consider the *strong* infinitesimal generator as acting on a dense domain in $C^0_p(X)$. However, this Markov semigroup $\{\Phi_\alpha(t) : t \geq 0\}$ may be considered as acting on real Borel functions with p-polynomial growth, which is a Banach space with the norm (5.55) and is denoted by $B_p(X)$. It is convenient to define the family of semi-norms on $B_p(X)$

$$p_0(h,x) = E\big\{ \sup_{s\geq 0} |h(y(s,x))|\, \mathrm{e}^{-\alpha_0 s}\big\}, \quad \forall x \in X, \tag{5.60}$$

where 2α, $2p$ and λ satisfy the estimate (5.58), and when $p = 0$ we may take $\alpha_0 = 0$. If a sequence $\{h_n\}$ of equi-bounded functions in $B_p(X)$ satisfies $p_0(h_n - h, x) \to 0$ for any x in X, we say that $h_n \to h$ boundedly pointwise relative to the above family of semi-norms. In view of (5.57), it is clear that $p_0(\Phi_\alpha(t)h - h, x) \to 0$ as $t \to 0$, for any function h in $C^0_p(X)$ and any x in X.

Let us defined $B^0_p(X)$ be the subspace of functions $\bar h$ in $B_p(X)$ such that the mapping $t \mapsto \bar h[y(t,x)]$ is almost surely continuous on $[0,\infty)$ for any x in X and satisfies

$$\lim_{t\to 0} p_0(\Phi_\alpha(t)\bar h - \bar h, x) = 0, \quad \forall\, x \in X. \tag{5.61}$$

where $p_0(\cdot,\cdot)$ is the semi-norm given by (5.60). This is the space of functions (uniformly) continuous over the random field $y(\cdot,x)$, relative to the family of semi-norms (5.60), and it is independent of α, as long as (5.58) holds. Hence, we may consider the semigroup on the Banach space $B^0_p(X)$, endowed with the norm (5.55). The *weak-star* infinitesimal generator $\bar A_\alpha$ with domain $\mathcal{D}_p(\bar A_\alpha)$ (as a subspace of $B^0_p(X)$) is defined by the boundedly pointwise limit $[\Phi_\alpha(t)h - h]/t \to \bar A_\alpha h$ as $t \to 0$, relative to the family of semi-norms (5.60). Also, it is clear that $p_0(\Phi_\alpha(t)\bar h, x) \leq p_0(\bar h, x)$ for any $t \geq 0$, $\bar h$ in $B^0_p(X)$ and x in X. We include the proof of the following results for the sake of completeness,

Proposition 5.12 (density)**.** *If assumptions* (5.56), (5.57) *and* (5.58) *hold, then* $C^0_p(X) \subset B^0_p(X)$*, the semigroup* $\{\Phi_\alpha(t) : t \geq 0\}$ *leaves invariant the space* $B^0_p(X)$*, and for any function* $\bar h$ *in* $B^0_p(X)$*, there is an equi-bounded sequence* $\{\bar h_n\}$ *of functions in* $\mathcal{D}_p(\bar A_\alpha)$ *satisfying* $p_0(\bar h_n - \bar h, x) \to 0$ *for any* x *in* X*.*

Proof. Indeed, since any function h in $C_p(X)$ is such that $x \mapsto h(x)\,(\lambda + |x|^2)^{-q/2}$, $q > p$, is uniformly continuous for x in X, we may use the estimate (5.58) to reduce the proof of the property (5.61) to the following condition

$$\lim_{t\to 0} P\big\{ \sup_{0\leq s\leq T} |y(t+s,x) - y(s,x)|\big\} = 0, \quad \forall\, x \in X,\ T > 0, \tag{5.62}$$

which follows from (5.57). This verifies the fact that $C^0_p(X) \subset \mathcal{B}^0_p(X)$.

Next, from the strong Markov property we deduce

$$p_0(\Phi_\alpha(t)\bar h, x) = E\{\sup_{s\geq 0} E\{|\bar h[y(t+s,x)]|\mathrm{e}^{-\alpha_0(t+s)} \mid y(t,x)\}\mathrm{e}^{-(\alpha-\alpha_0)t}\}$$
$$\leq E\{\sup_{s\geq 0} |\bar h[y(t+s,x)]|\mathrm{e}^{-\alpha_0(t+s)}\} = p_0(\bar h, x),$$

for any x in $\mathcal{O}$ and $t \geq 0$. Therefore,

$$p_0(\Phi_\alpha(r+t)\bar h - \Phi_\alpha(t)\bar h, x) = p_0(\Phi_\alpha(t)[\Phi_\alpha(r)\bar h - \bar h], x) \leq$$
$$\leq p_0(\Phi_\alpha(r)\bar h - \bar h, x),$$

which proves that the space $B^0_p(\mathcal{O})$ is invariant under the semigroup.

Finally, to approximate a function $\bar h$ in $B^0_p(\mathcal{O})$ by regular functions, we can define the sequence $\{\bar h_n \; n = 1, 2, \dots\}$ by

$$\bar h_n(x) = n\int_0^\infty \mathrm{e}^{-nt}\Phi_\alpha(t)\bar h(x)\mathrm{d}t = \int_0^\infty \mathrm{e}^{-t}E\{\bar h(y(\frac{t}{n}, x))\mathrm{e}^{-\alpha(\frac{t}{n})}\}\mathrm{d}t,$$

and apply the Markov property to get

$$\big|E\{\sup_{s\geq 0}[\bar h_n(y(s,x)) - \bar h(y(s,x))]\mathrm{e}^{-\alpha_0 s}\}\big| \leq$$
$$\leq \int_0^\infty \mathrm{e}^{-t}\big[E\{\sup_{s\geq 0}|\bar h(y(s+\frac{t}{n},x))\mathrm{e}^{-\alpha(\frac{t}{n})} - \bar h(y(s,x))|\mathrm{e}^{-\alpha_0 s}\}\big]\mathrm{d}t.$$

Thus, from the estimates (5.57) and (5.58) we deduce

$$\lim_{n\to\infty}\big|E\{\sup_{s\geq 0}[\bar h_n(y(s,x)) - \bar h(y(s,x))]\mathrm{e}^{-\alpha_0 s}\}\big| = 0,$$

for any fixed x in X. □

A clear consequence of the above results is that given $\alpha > 0$, $p \geq 0$, λ sufficiently large to ensure (5.58), and a function $\bar h$ in $B^0_p(\mathcal{O})$, there is another function $\bar u$ in $\mathcal{D}_p(\bar A_\alpha)$ such that $-\bar A_\alpha \bar u = \bar h$, where the solution admits the explicit representation

$$\bar u = \int_0^\infty \Phi_\alpha(t)\bar h\,\mathrm{d}t. \tag{5.63}$$

The right-hand side is called the *weak-star* resolvent operator and is denoted by either $R_\alpha = (-\bar A_\alpha)^{-1}$ or $R_\alpha = (\alpha I - \bar A_0)^{-1}$. For any $\alpha > \alpha_0$ we obtain

$$\|\Phi_\alpha(t)\bar h\| \leq \mathrm{e}^{-(\alpha-\alpha_0)t}\,\|\bar h\|, \qquad p_0(\Phi_\alpha(t)\bar h, x) \leq \mathrm{e}^{-(\alpha-\alpha_0)t}p_0(\bar h, x), \tag{5.64}$$

for any $t \geq 0$, and

$$\|R_\alpha \bar h\| \leq \frac{1}{\alpha-\alpha_0}\,\|\bar h\|, \qquad p_0(R_\alpha \bar h, x) \leq \frac{1}{\alpha-\alpha_0}\,p_0(\bar h, x), \tag{5.65}$$

for any x in X and where the norm $\|\cdot\|$ and the semi-norms $p_0(\cdot, x)$ are given by (5.55) and (5.60), respectively. Notice that $\alpha_0 = 0$ for $p = 0$, and it is clear that for any $\bar{h} \leq h$ (pointwise) we have $R_\alpha \bar{h} \leq R_\alpha h$, which is a weak form of the maximum principle.

Limiting the operator to the space $C_u(X)$ of bounded uniformly continuous functions, we find the so-called π-semigroups as proposed in Priola [148]. When the $\Phi_\alpha(t)$ is a strongly continuous Markov-Feller semigroup (typically an stochastic ODE) the weak version of the semigroup is of limited importance, since the domain of the infinitesimal generator is dense (in norm) in the space $C_p^0(\mathcal{O})$ of locally uniformly continuous functions with a growth bounded by the p-power of the norm. In general, we only have a weakly continuous Markov-Feller semigroup (typically stochastic PDE) and this weak version is very useful.

5.2.4 Invariant Distribution

Let E be a (locally compact) Polish space and $\{S(t) : t \geq 0\}$ be a stochastically continuous Markov semigroup on Banach space $B(E)$ of all bounded Borel real-valued functions on E, with Markov transition function $\{P(t, x, \cdot) : t \geq 0,\, x \in E\}$,

$$S(t)f(x) = \int_E f(y)P(t, x, \mathrm{d}y), \quad \forall t \geq 0,\ x \in E.$$

We begin with the following

Definition 5.13 (invariant)**.** A probability measure μ on the Borel σ-algebra $\mathcal{B}(E)$ is called an *invariant distribution* or *invariant probability measure* of the Markov semigroup $\{S(t) : t \geq 0\}$ if

$$\int_E S(t)f \mathrm{d}\mu = \int_E f \mathrm{d}\mu,$$

or in term of the kernels

$$\int_E \mu(\mathrm{d}x) \int_E f(y)P(t, x, \mathrm{d}y) = \int_E f(x)\mu(\mathrm{d}x),$$

for every f in $B(E)$ and every $t > 0$. □

Notice that if an invariant distribution μ exits then the Markov semigroup satisfies $S(t)\mathbb{1} = \mathbb{1}$ or equivalently $P(t, x, E) = 1$, for every $t \geq 0$ and x in E, i.e, the semigroup has to be Markov, not sub-Markov, see Definition 5.8.

If $\{S(t) : t \geq 0\}$ is also a Markov-Feller semigroup then, besides $S(t) : B(E) \to B(E)$, we have $S(t) : C \to C$ for some closed subspace C of $C_b(E)$, the space of bounded continuous real-valued functions on E, e.g., C could be continuous functions vanishing at infinity or uniformly continuous functions or the whole space $C_b(E)$. Then $\{S(t) : t \geq 0\}$ is strongly continuous in C and the

infinitesimal generator $(\mathcal{D}(A), A)$ is densely defined on C and

$$S(t)f - f = \int_0^t AS(s)f\mathrm{d}s = \int_0^t S(s)Af\mathrm{d}s,$$
$$\int_E [S(t)f - f]\mathrm{d}\mu = \int_E A\Big(\int_0^t S(s)f\mathrm{d}s\Big)\mathrm{d}\mu,$$

for any probability measure μ. This proves that μ is an invariant distribution if and only if

$$\int_E Af\mathrm{d}\mu = 0, \quad \forall f \in \mathcal{D}(A),$$

provided $\{S(t) : t \geq 0\}$ is a Markov-Feller semigroup.

The following result give a condition for the existence of an invariant probability measure, see Doob, Khasminskii, Krylov-Bogoliubov theorems in Da Prato and Zabczyk [29, Chapters 3 and 4],

Theorem 5.14 (existence)**.** *Let $\{P(t, x, \cdot) : t \geq 0,\, x \in E\}$ be a stochastically continuous Markov transition function on a Polish space E. If the family of time-average probabilities $\{R(t, x, \cdot) : t \geq 0,\, x \in E\}$,*

$$R(t, x, \cdot) = \frac{1}{t}\int_0^t P(s, x, \cdot)\mathrm{d}s,$$

is tight for t in $[t_0, \infty)$, for some $x = x_0$ and $t_0 > 0$, then there exists an invariant distribution μ. □

Recall that a family of probabilities $\{R(t, x_0, \cdot) : t \geq t_0\}$ is tight when for every $\varepsilon > 0$ there exits a compact subset $K = K_\varepsilon$ of E such that $R(t, x_0, K) \geq 1 - \varepsilon$, for every $t \geq t_0$. Thus, any weak limit as $t \to \infty$ of the time-average probabilities is an invariant probability measure.

As it was defined early, $\{S(t) : t \geq 0\}$ is strongly Feller if $S(t_0) : B(E) \to C$, for some $t_0 > 0$. Also, a Markov transition function $\{P(t, x, B) : t \geq 0,\, x \in E,\, B \in \mathcal{B}(E)\}$ is called *irreducible* if there exist $t_0 > 0$ such that for every non-empty open subset $\mathcal{O}$ of E and for any x in $\mathcal{O}$ we have $P(t_0, x, \mathcal{O}) > 0$. Furthermore, it is called *regular* if there exists $t_0 > 0$ such that all transition probabilities $\{P(t_0, x, \cdot) : x \in E\}$ are mutually equivalent.

Theorem 5.15 (uniqueness)**.** *Let μ be an invariant distribution of a stochastically continuous Markov transition function $\{P(t, x, \cdot) : t \geq 0,\, x \in E\}$ on a Polish space E. If it is strongly Feller and irreducible then it is also regular, the invariant distribution is unique and*

(1) *for any x in E and B in $\mathcal{B}(E)$ we have $P(t, x, B) \to \mu(B)$ as $t \to \infty$,*

(2) *there exists $t_0 > 0$ such that all probabilities measures $\{P(t, x, \cdot) : t \geq t_0,\, x \in E\}$ are equivalent to μ.* □

A set B in $\mathcal{B}(E)$ is called invariant with respect to a stochastically continuous Markov transition function $\{P(t,x,B) : t \geq t_0,\ x \in E,\ B \in \mathcal{B}(E)\}$ having an invariant probability measure μ if except in a set of μ-measure zero, $P(t,\cdot,B) = \mathbb{1}_B$, for every $t > 0$. Then an invariant probability measure μ is called *ergodic* if the only invariant sets have μ measure 0 or 1, i.e., if $P(t,\cdot,B) = \mathbb{1}_B$ μ-a.s. implies $\mu(B) = 0$ or $\mu(B) = 1$. It can be proved that an invariant distribution μ is ergodic if and only if the time-average commute with the space average, i.e.,

$$\frac{1}{T}\int_0^T \mathrm{d}t \int_0^T f(y)P(t,\cdot,\mathrm{d}y) \to \int_E f(y)\mu(\mathrm{d}y) \ \text{ in } \ L^2(E,\mu),$$

as $T \to \infty$, for every f in $L^2(E,\mu)$.

Sometimes a stronger convergence than (1) in Theorem 5.15 is necessary, e.g., exponential convergence. Based on Doob's ergodicity Theorem on a compact space E, the so-called Doeblin's condition, namely, there exist $t_0 > 0$ and $\delta > 0$ such that

$$P(t_0,x,B) - P(t_0,y,B) \leq 1 - \delta, \quad \forall x,y \in E,\ B \in \mathcal{B}(E), \tag{5.66}$$

imply the existence of a unique invariant probability measure μ and the exponential convergence

$$\Big| \int_E f(y)P(t,x,\mathrm{d}y) - \int_E f(y)\mu(\mathrm{d}y) \Big| \leq C\mathrm{e}^{-\omega t} \sup_{y\in E} |f(y)|,$$

for some positive constants C and ω, and for every x in E, as long as E is compact.

Typical conditions to ensure the tightness of the probability measures needed in Theorem 5.14 are given in term of the existence of Liapunov functions, see Khasminskii [96]. For instance, if there exists a function φ in $\mathcal{D}(A)$ such that $\varphi \geq 0$ and satisfying $\varphi(x) \to -\infty$ as $|x| \to \infty$ (which means that for every $m > 0$ there is a compact set $K = K_m$ of E such that $\varphi(x) < -m$ for every x in $E \setminus K$) the family $\{P(t,x_0,\cdot) : t \geq 0,\ x \in E\}$ is tight for every x_0 fixed. The existence of a Liapunov function satisfying $\varphi(x) \to +\infty$ as $|x| \to \infty$ and $A\varphi - \alpha\varphi \leq C$ for some positive constants α and C, yields the uniqueness of the invariant probability measure.

If μ is an invariant distribution then Jensen's inequality yields

$$\Big| \int_E f(y)P(t,x,\mathrm{d}y) \Big|^p \leq \int_E |f(y)|^p P(t,x,\mathrm{d}y) = \int_E |f(y)|^p \mu(\mathrm{d}y),$$

for every p in $[1,\infty)$. The stochastically continuous Markov semigroup $\{S(t) : t \geq 0\}$ can be extended to a strongly continuous semigroup of contractions in the Lebesgue spaces $L^p(E,\mu)$. Moreover, any other probability measure ν which is equivalent to μ (i.e., ν is absolutely continuous with respect to μ and conversely) can be used to extend the semigroup to $L^p(E,\nu)$.

In a finite-dimensional setting, let D be a domain in $\mathbb{R}^d$ (i.e., the closure of the interior of D is equal to its closure) and let $(A,\mathcal{D})$ be a linear operator

defined on a linear sub-space $\mathcal{D}$ of $C^\infty(\bar{D})$ containing $C_0^\infty(D)$. Assume that there exist a probability measure μ on D such that

$$Av \in L^1(D,\mu) \quad \text{and} \quad \int_D Av(x)\,\mu(\mathrm{d}x) = 0, \quad \forall v \in \mathcal{D}.$$

Then, we want to find a unique extension of $(A,\mathcal{D})$ which generates a strongly continuous Markov semigroup $\{T(t) : t \geq 0\}$ in $L^p(D,\mu)$, $1 \leq p < \infty$, having μ as an invariant measure. Several conditions are given in the literature to ensure this construction, e.g., see Stannat [165] and references there in.

5.3 Integro-differential Operators

We are interested in integro-differential operators associated with diffusion processes with jumps, see Gikhman and Skorokhod [62, p. 245] and Bensoussan and Lions [6, p. 178]. For a comprehensive treatment on (elliptic/parabolic) integro-differential operators, we refer to the books Garroni and Menaldi [58, 59]. This is very similar to the so-called Waldenfels operators as considered in the paper Bony et al. [17] and the recent book Taira [172, Chapters 8 and 10, pp. 361–410, 477–546] [1].

A Radon measure $M(x,\mathrm{d}z)$ on $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$, for any x in $\mathbb{R}^d$, determines this operator. Depending on the assumptions on the singularity at the origin of the Lévy kernel $M(x,\mathrm{d}z)$ we may classify these integro-differential operators. The expression

$$I_1\varphi = \int_{|z|<1} [\varphi(\cdot + z) - \varphi] M_1(\cdot,\mathrm{d}z), \tag{5.67}$$

with

$$\int_{|z|<1} |z| M_1(\cdot,\mathrm{d}z) < \infty$$

define an integro-differential operator of order 1, since, in view of the mean value theorem, the expression (5.67) makes sense for bounded continuously differentiable functions. However, a form

$$I_2\varphi = \int_{|z|<1} [\varphi(\cdot + z) - \varphi - z\cdot\nabla\varphi] M_2(\cdot,\mathrm{d}z), \tag{5.68}$$

with

$$\int_{|z|<1} |z|^2 M_2(\cdot,\mathrm{d}z) < \infty$$

[1] note that Chapters 8 and 10 are part of the Second Edition!

gives an integro-differential operator of order 2, since, by Taylor's formula, the expression (5.68) makes sense for bounded twice-continuously differentiable functions. On the other hand,

$$I_0\varphi = \int_{\mathbb{R}^d_*} [\varphi(\cdot + z) - \varphi] M_0(\cdot, \mathrm{d}z), \quad \text{with } \int_{\mathbb{R}^d_*} M_0(\cdot, \mathrm{d}z) < \infty \tag{5.69}$$

provides a bounded (or order 0) integral (or non-local) operator, since the expression (5.69) makes sense for bounded functions.

Note that in the definitions (5.67) and (5.68) of the operators I_1 and I_2, we may replace the region of integration $\{|z| < 1\}$ by one of the form $\{|z| < r\}$, for any $r > 0$. The interesting part is the singularity at the origin, i.e., *small jumps*. On the other hand, in the definition (5.69) of the operator I_0 we may use $\{|z| \geq r\}$, for any $r > 0$, as the region of integration, instead of the whole space $\mathbb{R}^d_*$. Here the interest is on the integrability at infinity, i.e., *large jumps*.

Let us present some typical examples. First, an operator of order 0,

$$I_0\varphi = \lambda[\varphi(\cdot + \zeta) - \varphi],$$

for some constants $\lambda > 0$, $\zeta \in \mathbb{R}^d_*$. Here the Lévy kernel is $M_0(x, \cdot) = \lambda\delta_\zeta$, where δ_ζ denotes the Dirac measure at ζ. Second, two examples of order 1,

$$I_1\varphi = \int_{|z|<1} [\varphi(\cdot + z) - \varphi]|z|^{-d}\mathrm{d}z,$$

where the Lévy kernel $M_1(x, \mathrm{d}z) = \mathbb{1}_{|z|<1}|z|^{-d}\mathrm{d}z$, and

$$I_1\varphi = \sum_{n=1}^{\infty} \lambda_n[\varphi(\cdot + \zeta_n) - \varphi],$$

with

$$\lambda_n \geq 0, \quad \sum_{n=1}^{\infty} \lambda_n = \infty, \quad \sum_{n=1}^{\infty} \lambda_n|\zeta_n| < \infty.$$

Here $\zeta_n \to 0$ as $n \to \infty$ and the Lévy kernel $M_1(x, \cdot) = \sum_{n=1}^{\infty} \lambda_n\delta(\cdot - \zeta_n)$. Next, two examples of order 2,

$$I_0\varphi + I_2\varphi = \int_{R^d_*} [\varphi(\cdot + z) - \varphi - z \cdot \nabla\varphi \mathbb{1}_{|z|<1}]|z|^{-d-1}\mathrm{d}z,$$

where the Lévy kernels are $M_0(x, \mathrm{d}z) = \mathbb{1}_{|z|>1}|z|^{-d-1}\mathrm{d}z$ and $M_2(x, \mathrm{d}z) = \mathbb{1}_{|z|<1}|z|^{-d-1}\mathrm{d}z$, and

$$I_2\varphi = \sum_{n=1}^{\infty} \lambda_n[\varphi(\cdot + \zeta_n) - \varphi - \zeta_n \cdot \nabla\varphi], \quad \text{with}$$

$$\lambda_n \geq 0, \quad \sum_{n=1}^{\infty} \lambda_n(1 + |\zeta_n|) = \infty, \quad \sum_{n=1}^{\infty} \lambda_n|\zeta_n|^2 < \infty.$$

where the Lévy kernel $M_2(x,\cdot)=\sum_{n=1}^{\infty}\lambda_n\delta(\cdot-\zeta_n)$. Note that in all examples, the Lévy kernels $M(x,\mathrm{d}z)$ are independent of x.

Working with operators of the type (5.67) or (5.68), we see that the Lévy kernel $M_1(\cdot,\mathrm{d}z)$ and $M_2(\cdot,\mathrm{d}z)$ can be approximated by bounded kernels of the form

$$M_{i,\varepsilon}(\cdot,\mathrm{d}z)=\mathbb{1}_{|z|\geq\varepsilon}M_i(\cdot,\mathrm{d}z),\quad i=1,2.$$

We see that as ε goes to 0, the integro-differential operators (5.67) or (5.68) are limits of bounded non-local operators of the type (5.69).

Definition 5.16 (order γ)**.** We say that an integro-differential operator $I=I_\gamma$ is (a) of order $\gamma=0$ (or bounded) if

$$I\varphi=\int_{\mathbb{R}^d_*}[\varphi(\cdot+z)-\varphi]M(\cdot,\mathrm{d}z),\quad\text{with}\quad\int_{\mathbb{R}^d_*}M(\cdot,\mathrm{d}z)<\infty,$$

(b) of order γ in $(0,1]$ if

$$I\varphi=\int_{\mathbb{R}^d_*}[\varphi(\cdot+z)-\varphi]M(\cdot,\mathrm{d}z),\qquad\text{with}\quad\int_{|z|<1}|z|^\gamma M(\cdot,\mathrm{d}z)+\int_{|z|\geq1}M(\cdot,\mathrm{d}z)<\infty,$$

(c) of order γ in $(1,2]$ if

$$I\varphi=\int_{\mathbb{R}^d_*}[\varphi(\cdot+z)-\varphi-z\cdot\nabla\varphi\mathbb{1}_{|z|<1}]M(\cdot,\mathrm{d}z),\qquad\text{with}\quad\int_{|z|<1}|z|^\gamma M(\cdot,\mathrm{d}z)+\int_{|z|\geq1}M(\cdot,\mathrm{d}z)<\infty.$$

In all cases, γ is also referred to as the order of the Lévy kernel $M(\cdot,\mathrm{d}z)=M_\gamma(\cdot,\mathrm{d}z)$. □

Note that the order γ of an integro-differential operator does not (completely) characterize the behaviour of the singularity of the Lévy kernel $M(\cdot,\mathrm{d}z)$. Actually, the most significant values are $\gamma=0$ (where the operator is bounded), $\gamma=1$ (where the expression used to define operator changes), and in general $\gamma=2$. The use of "order" of the operator may be questionable, since an operator of order γ is also an operator of order γ', for any $\gamma\leq\gamma'\leq2$. For the sake of simplicity we use the expression "of order γ" instead of "of order at most γ".

It is hard to track the dependency on the variable x of the Lévy kernel $M(x,\mathrm{d}z)$ to ensure that the integro-differential operator I acts on Lebesgue (Sobolev) and Hölder spaces. We will make precise how the variable x intervenes on the Lévy kernel $M(x,\mathrm{d}z)$, allowing enough flexibility to include modulation of the amplitude (or intensity) of jumps (well adapted for stochastic differential equations, see Gikhman and Skorokhod [62, p. 215]) and the density (or size) of

jumps (better adapted for the martingale problem theory, see Bensoussan and Lions [6, p. 251]).

A priori the integro-differential operator is defined for functions $\varphi(x)$, with x in the whole space $\mathbb{R}^d$. However, we want to consider equations on a domain $\overline{\Omega}$ of $\mathbb{R}^d$, with either Dirichlet or Neumann boundary conditions, and even with oblique boundary conditions. We then need to localize the operator into $\overline{\Omega}$, e.g., by extending the data φ onto $\mathbb{R}^d \setminus \overline{\Omega}$. Thus $I\varphi$ becomes $I\tilde{\varphi}$, where $\tilde{\varphi}$ is a suitable extension of φ (defined only on $\overline{\Omega}$) to the whole space $\mathbb{R}^d$. The extension depends on the boundary value problem under consideration, which has a probabilistic interpretation. For instance, it is natural to use the zero-extension to study homogeneous Dirichlet boundary conditions. This corresponds to stopping the diffusion process with jumps (in the whole space $\mathbb{R}^d$) at the first exit time of the domain $\overline{\Omega}$. It is clear that the zero-extension will present some extra difficulties, e.g., if φ belongs to $W_0^{1,p}(\Omega) \cap W^{2,p}(\Omega)$ then the zero-extension $\tilde{\varphi}$ belongs to $W_0^{1,p}(\mathbb{R}^d)$ but in general, it is not an element of $W^{2,p}(\mathbb{R}^d)$.

As seen later, to treat the homogeneous Neumann (or oblique) boundary conditions, we will use a condition on the jumps (namely, no jumps outside of $\overline{\Omega}$) that will make the extension unnecessary, i.e., any extension $\tilde{\varphi}$ of φ will produce the same value for $I\tilde{\varphi}$.

5.3.1 The Epsilon-estimates

We need to describe the dependency of the variable x in the Lévy kernel $M(x, \mathrm{d}z)$. Suppose that there exist a σ-finite measure space $(F, \mathcal{F}, \pi)$, two Borel measurable functions $j(x, \zeta)$ and $m(x, \zeta)$ from $\mathbb{R}^d \times F$ into $\mathbb{R}^d_*$ and $[0, \infty)$, respectively, such that

$$M(x, A) = \int_{\{\zeta : \mathtt{j}(x,\zeta) \in A\}} m(x, \zeta)\pi(\mathrm{d}\zeta), \tag{5.70}$$

for any Borel measurable subset A of $\mathbb{R}^d_*$. The functions $\mathtt{j}(x, \zeta)$ and $m(x, \zeta)$ are called the *jump size (or amplitude)* and the *jump density (or intensity)*, respectively. The conditions (5.67), (5.68) or (5.69) on the singularity at the origin of the Lévy kernel $M(x, \mathrm{d}z)$ will be assumed to hold uniformly in x, so that for some measurable function $\bar{\jmath}(\zeta)$ from F into $(0, \infty)$ and some constant $C_0 > 0$ we have

$$\begin{aligned} &|\mathtt{j}(x, \zeta)| \le \bar{\jmath}(\zeta), \qquad 0 \le \mathtt{m}(x, \zeta) \le 1, \\ &\int_{\{\bar{\jmath}<1\}} [\bar{\jmath}(\zeta)]^\gamma \pi(\mathrm{d}\zeta) + \int_{\{\bar{\jmath}\ge 1\}} \pi(\mathrm{d}\zeta) \le C_0, \end{aligned} \tag{5.71}$$

where $0 \le \gamma \le 2$ is the order of the Lévy kernel. Actually, we may allow $0 \le m(x, \zeta) \le C$ if we re-define the measure $\pi(\mathrm{d}\zeta)$.

Thus, for any smooth function φ the integro-differential operator has the

form

$$I\varphi = \int_F [\varphi(\cdot + \mathtt{j}(\cdot,\zeta)) - \varphi]\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta) = \\ = \int_0^1 \mathrm{d}\theta \int_F \mathtt{j}(\cdot,\zeta)\cdot\nabla\varphi(\cdot + \theta\mathtt{j}(\cdot,\zeta))\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta), \quad (5.72)$$

for $0 \le \gamma \le 1$ and

$$I\varphi = \int_{\{\bar\jmath<1\}} [\varphi(\cdot + \mathtt{j}(\cdot,\zeta)) - \varphi - \mathtt{j}(\cdot,\zeta)\cdot\nabla\varphi]\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta) + \\ + \int_{\{\bar\jmath\ge1\}} [\varphi(\cdot + \mathtt{j}(\cdot,\zeta)) - \varphi]\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta), \quad (5.73)$$

for $1 < \gamma \le 2$, where the first term can be rewritten as

$$\int_0^1 (1-\theta)\mathrm{d}\theta \int_{\{\bar\jmath<1\}} \mathtt{j}(\cdot,\zeta)\cdot\nabla^2\varphi(\cdot + \theta\mathtt{j}(\cdot,\zeta))\mathtt{j}(\cdot,\zeta)\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta).$$

In order to study this integro-differential operator as acting on Lebesgue (Sobolev) spaces, we will need to perform a change of variables. Assume that the jump amplitude function $\mathtt{j}(x,\zeta)$ is continuously differentiable in x for any fixed ζ, and that there exist a constant $c_0 > 0$ such that for any x, x' and $0 \le \theta \le 1$ we have

$$c_0|x - x'| \le |(x - x') + \theta[\mathtt{j}(x,\zeta) - \mathtt{j}(x',\zeta)]| \le c_0^{-1}|x - x'|. \quad (5.74)$$

This implies that the change of variables $X = x + \theta\mathtt{j}(x,\zeta)$ is a diffeomorphism of class C^1 in R^d, for any θ in $[0,1]$ and ζ in F. Moreover, the Jacobian of the transformation satisfies

$$c_1^{-1} \le \det[I_d + \theta\nabla\mathtt{j}(x,\zeta)] \le C_1, \quad (5.75)$$

for any x, ζ, θ and some constants $C_1, c_1 \ge 1$. Here I_d is the identity matrix in $\mathbb{R}^d$, $\nabla\mathtt{j}(x,\zeta)$ is the matrix of the first partial derivatives in x, and $\det[\cdot]$ denotes the determinant of a matrix.

In order to study the integro-differential operator in the Hölder space C^α, we also need Hölder continuity of the amplitude and density of jumps. For some exponent $0 < \alpha < 1$ we assume that there exist a measurable function (again denoted by) $\bar\jmath(\cdot)$ from F into $(0,\infty)$ and some constant $M_0 > 0$ such that for any x, x' and ζ we have

$$\begin{aligned} &|\mathtt{j}(x,\zeta) - \mathtt{j}(x',\zeta)| \le \bar\jmath(\zeta)|x - x'|^\alpha, \\ &|\mathtt{m}(x,\zeta) - \mathtt{m}(x',\zeta)| \le M_0|x - x'|^\alpha, \\ &\int_{\{\bar\jmath<1\}} [\bar\jmath(\zeta)]^\gamma\pi(\mathrm{d}\zeta) + \int_{\{\bar\jmath\ge1\}} \pi(\mathrm{d}\zeta) \le M_0. \end{aligned} \quad (5.76)$$

Let $\mathcal{O}$ be a bounded subset of $\mathbb{R}^d$ and set $\mathcal{O}_\varepsilon = \{x \in \mathbb{R}^d : x = y + z,\ y \in \mathcal{O},\ |z| < \varepsilon\}$. Due to the non-local character of the integro-differential operator I we need a function φ to be defined in a neighborhood of the closure $\overline{\mathcal{O}}$ to consider $I\varphi$ in $\mathcal{O}$. Thus, we define the support of I as the closed subset $\overline{\mathcal{O}}_I$ of $\mathbb{R}^d$, where

$$\overline{\mathcal{O}}_I = \overline{\bigcup\{x + \operatorname{supp} M(x,\cdot) \,:\, x \in \overline{\mathcal{O}}\}} \tag{5.77}$$

and $\operatorname{supp} M(x,\cdot)$ means the support of the Lévy kernel (or measure)

$$M(x,B) = \int_{\mathtt{j}(x,\zeta)\in B} \mathtt{m}(x,\zeta)\pi(\mathrm{d}\zeta)\ , \quad B \subset \mathbb{R}^d_* \text{ measurable Borel}\ .$$

Proposition 5.17 (ε-estimates)**.** *If the integro-differential operator I has the form (5.72) or (5.73), and conditions (5.71) and (5.74) are satisfied then for every $\varepsilon > 0$ there exists constants C and $C(\varepsilon)$ depending only on ε, the dimension d, the bounds C_0 and c_1 of conditions (5.71) and (5.75) such that*

$$\|I\varphi\|_{L^p(\mathcal{O})} \le C\|\varphi\|_{L^p(\overline{\mathcal{O}}_I)}, \quad \text{if } \gamma = 0,$$

$$\|I\varphi\|_{L^p(\mathcal{O})} \le \varepsilon\|\nabla\varphi\|_{L^p(\mathcal{O}_\varepsilon)} + C(\varepsilon)\|\varphi\|_{L^p(\overline{\mathcal{O}}_I)}, \quad \text{if } 0 < \gamma \le 1$$

and, if $1 < \gamma \le 2$, then

$$\|I\varphi\|_{L^p(\mathcal{O})} \le \varepsilon\|\nabla^2\varphi\|_{L^p(\mathcal{O}_\varepsilon)} + C(\varepsilon)\Big[\|\varphi\|_{L^p(\overline{\mathcal{O}}_I)} + \|\nabla\varphi\|_{L^p(\mathcal{O})}\Big],$$

for $1 \le p \le \infty$. Moreover, if we also assume the Hölder condition (5.76) on the coefficients, then the above estimates are valid with the C^α-norm instead of the L^p-norm, and in this case the constants C and $C(\varepsilon)$ depends also on the bounds M_0 of assumption (5.76). □

At this point, it should be clear that the integro-differential operator I is naturally non-local, i.e., we need to use functions defined on the whole space $\mathbb{R}^d$. So, a direct approach to consider I as acting on functions φ defined only on a (proper) domain Ω of $\mathbb{R}^d$, is to extend first φ to the whole space. Thus, denoting by $\tilde{\varphi}$ a suitable extension of φ, we have $I\varphi = I\tilde{\varphi}$, by definition. However, if we assume that

$$\mathtt{m}(x,\zeta) \neq 0 \quad \text{implies} \quad x + \theta\mathtt{j}(x,\zeta) \in \overline{\Omega},\ \forall\theta \in [0,1], \tag{5.78}$$

valid for any (x,ζ) in $\overline{\Omega} \times F$, then we see that the value $I\varphi = I\tilde{\varphi}$ is independent of the extension $\varphi \mapsto \tilde{\varphi}$ used. Indeed, notice that $I\tilde{\varphi}$ is always defined as the limit $I_\varepsilon\tilde{\varphi}$, where the Lévy kernel of I_ε is $M_\varepsilon(\cdot,\mathrm{d}z) = \mathbb{1}_{(|z|>\varepsilon)}M(\cdot,\mathrm{d}z)$. Condition (5.78) means that all jumps from $\overline{\Omega}$ are within $\overline{\Omega}$. Hence, under this condition (5.78), we may consider $I\varphi$ without any reference to the extension used for its proper definition (included estimates on its norm).

From the stochastic process viewpoint, some action should be taken when the jumps are outside of the region under consideration, e.g., we may stop or reflect the jumps, so that condition (5.78) will be eventually satisfied for the actual (or modified) jumps. So that in general, this will take care of Dirichlet, Neumann and oblique boundary conditions. However, for homogeneous Dirichlet boundary conditions problems, we have a natural *zero-extension* which corresponds to stopping the stochastic process at the first exit time of $\overline{\Omega}$. The problem with zero-extension is that a function φ in $W^{2,p}(\Omega)\cap W^{1,p}_0(\Omega)$ gives a zero-extension φ^o in $W^{1,p}(\mathbb{R}^d)$, but the first order derivative may be discontinuous across the boundary $\partial\Omega$. To overcome this difficulty, we need to impose some integrability conditions on the functions

$$\mathtt{m}_\Omega(x,\zeta)=\mathbb{1}_{(x+\mathtt{j}(x,\zeta)\notin\Omega)}\mathtt{m}(x,\zeta)\quad\text{and}\quad\mathtt{m}^1_\Omega(x,\zeta)=\mathtt{j}(x,\zeta)\mathtt{m}_\Omega(x,\zeta),$$

as seen below. Indeed, let denote by I_Ω the integro-differential operator corresponding to the density m_Ω, i.e. for a smooth function v in the whole space $\mathbb{R}^d$ and for $1<\gamma\le 2$ we have

$$I_\Omega v(x)=\lim_{\varepsilon\to0}\int_{\bar\jmath\ge\varepsilon}[v(x+j(x,\zeta))-v(x)]\mathtt{m}_\Omega(x,\zeta)\pi(\mathrm{d}\zeta)$$
$$-\lim_{\varepsilon\to0}\int_{\varepsilon\le\bar\jmath<1}\nabla v(x+\mathtt{j}(x,\zeta))\cdot\mathtt{m}^1_\Omega(x,\zeta)\pi(\mathrm{d}\zeta).$$

Thus, we can write $Iv=I_\Omega v+(I-I_\Omega)v$ and, for any x in Ω, the first term (i.e., $I_\Omega v$) reduces to only one integral and the second term vanishes if the function v vanishes in Ω. Hence, if φ^o and $\tilde\varphi$ are two extensions to the whole space of a given function in Ω (e.g., φ^o the zero-extension and $\tilde\varphi$ a smooth extension) then $v=\varphi^o-\tilde\varphi$ vanishes in Ω. Assuming v smooth (which may not be the case!) we may use the previous argument to see that $I\varphi^o=I\tilde\varphi+I_\Omega(\varphi^o-\tilde\varphi)$. Thus, we have the following *localization* of the operator I, by imposing the above *vanishing* property for non-smooth functions.

Definition 5.18 (localization)**.** Let Ω be a bounded domain in $\mathbb{R}^d$ with smooth boundary, e.g., $C^{2+\alpha}$, and let I be the integro-differential operator given by (5.72) or (5.73) of order γ in $[0,2]$. For a smooth function φ defined on $\overline{\Omega}$ we denote by φ^o the zero-extension to whole space $\mathbb{R}^d$ and by $\tilde\varphi$ a smooth (say $C^{2+\alpha}$) extension. Under the conditions (5.71) we define the *localization* of I (to the domain Ω) as $I\varphi^o=I\tilde\varphi+I_\Omega(\varphi^o-\tilde\varphi)$, where the operator I_Ω is given by

$$I_\Omega(\varphi^o-\tilde\varphi)=-\int_{\{\zeta\in F\,:\,x+\mathtt{j}(x,\zeta)\notin\Omega\}}\tilde\varphi(\cdot+\mathtt{j}(\cdot,\zeta))\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta),$$

with the above notation. □

If I has the form (5.73) of order γ in $(1,2]$, then we assume that for some γ_1 in $[1,\gamma]$ there exist a measurable function $\lambda_1(\zeta)$ and a constant $K_1>0$ such that for every x in Ω and ζ with $x+\mathtt{j}(x,\zeta)$ belonging to $\mathbb{R}^d\setminus\Omega$,

$$|\mathtt{j}(x,\zeta)|\mathtt{m}(x,\zeta)\le\mathrm{d}^{1-\gamma_1}(x,\partial\Omega)\lambda_1(\zeta),\qquad\int_{\bar\jmath<1}\lambda_1(\zeta)\pi(\mathrm{d}\zeta)\le K_1,\tag{5.79}$$

where $\mathrm{d}(x,\partial\Omega)$ denotes the distance from x to the boundary $\partial\Omega$, and $\bar\jmath(\zeta)$ is the function in (5.71). Notice that if x is in Ω but $x+\mathtt{j}(x,\zeta)$ is not in Ω then

$$|\mathtt{j}(x,\zeta)|\mathtt{m}(x,\zeta) = |\mathtt{j}(x,\zeta)|^\gamma|\mathtt{j}(x,\zeta)|^{1-\gamma}\mathtt{m}(x,\zeta) \le \\ \le |\mathtt{j}(x,\zeta)|^\gamma \mathrm{d}^{1-\gamma}(x,\partial\Omega)\mathtt{m}(x,\zeta).$$

The function $\lambda(\zeta)=\sup_x|\mathtt{j}(x,\zeta)|^\gamma m(x,\zeta)$ is bounded by $[\bar\jmath(\zeta)]^\gamma$, which is integrable in view of assumption (5.71). This show that condition (5.79) is always satisfied with $\gamma_1=\gamma$. This γ_1 in $[1,\gamma]$ is called the *boundary order* of I (and of the Lévy Kernel) with respect to the bounded domain Ω.

In Hölder spaces, we need to assume that the function $m_\Omega(x,\zeta)=m(x,\zeta)$ only if $x+\mathtt{j}(x,\zeta)\notin\Omega$ and zero otherwise, satisfies the following inequalities for any x, x' and ζ

$$\begin{aligned}
&\int_F \big(\bar\jmath_\Omega(\zeta)\wedge 1\big)^{1-\alpha}\mathtt{m}_\Omega(x,\zeta)\pi(\mathrm{d}\zeta)\le M_1,\\
&\int_F \big(\bar\jmath_\Omega(\zeta)\wedge 1\big)|\mathtt{m}_\Omega(x,\zeta)-\mathtt{m}_\Omega(x',\zeta)|\pi(\mathrm{d}\zeta)\le M_1|x-x'|^\alpha,\\
&\Big|\int_{\bar\jmath<1}\mathtt{j}(x,\zeta)\mathtt{m}_\Omega(x,\zeta)\pi(\mathrm{d}\zeta)-\int_{\bar\jmath<1}\mathtt{j}(x',\zeta)\mathtt{m}_\Omega(x',\zeta)\pi(\mathrm{d}\zeta)\Big|\le M_1|x-x'|^\alpha,
\end{aligned}\tag{5.80}$$

where the function $\bar\jmath(\zeta)$ is as in assumption (5.71), $\bar\jmath_\Omega(\zeta)=\sup\{|\mathtt{j}(x,\zeta)|: x\in\Omega,\ x+\mathtt{j}(x,\zeta)\notin\Omega\}$, the constant M_1 is positive and the exponent α is the same as in condition (5.76).

We modified Proposition 5.17 as follows.

Proposition 5.19 (ε-loc-estimates)**.** *If the integro-differential operator I has the form (5.72) or (5.73), and conditions (5.71), (5.74) and (5.79) are satisfied then for any smooth function φ which vanishes on the boundary $\partial\Omega$ we have the following estimates:*
(1) *if $\gamma=0$ and $1\le p\le\infty$ then*

$$\|I\varphi\|_{L^p(\Omega)}\le C\|\varphi\|_{L^p(\Omega)},$$

(2) *if $0<\gamma\le 1$ and $1\le p\le\infty$ then*

$$\|I\varphi\|_{L^p(\Omega)}\le \varepsilon\|\nabla\varphi\|_{L^p(\Omega)}+C(\varepsilon)\|\varphi\|_{L^p(\Omega)},$$

(3) *if $1<\gamma_1\le\gamma\le 2$ and $1\le p<d/(\gamma_1-1)$ or if $\gamma_1=1$ and $1\le p\le\infty$ then*

$$\|I\varphi\|_{L^p(\Omega)}\le \varepsilon\|\nabla^2\varphi\|_{L^p(\Omega)}+C(\varepsilon)\Big[\|\nabla\varphi\|_{L^p(\Omega)}+\|\varphi\|_{L^p(\Omega)}\Big],$$

where $\varepsilon>0$ is arbitrary and the constant C and the function $C(\varepsilon)$ depend only on d, γ_1, Ω and the bounds in conditions (5.71) and (5.74). Moreover, if $\gamma>1-\alpha$, (5.76) and (5.80) are satisfied, then we have the following estimates:
(1) *if $\gamma=0$ then*

$$\|I\varphi\|_{C^\alpha(\overline\Omega)}\le C\|\varphi\|_{C^\alpha(\overline\Omega)},$$

(2) *if* $0 < \gamma \leq 1$ *then*

$$\|I\varphi\|_{C^\alpha(\overline{\Omega})} \leq \varepsilon \|\nabla\varphi\|_{C^\alpha(\overline{\Omega})} + C(\varepsilon)\|\varphi\|_{C^\alpha(\overline{\Omega})},$$

(3) *if* $1 < \gamma \leq 2$ *then*

$$\|I\varphi\|_{C^\alpha(\overline{\Omega})} \leq \varepsilon \|\nabla^2\varphi\|_{C^\alpha(\overline{\Omega})} + C(\varepsilon)\Big[\|\nabla\varphi\|_{C^\alpha(\overline{\Omega})} + \|\varphi\|_{C^\alpha(\overline{\Omega})}\Big],$$

where $\varepsilon > 0$ *is arbitrary and the constant* C *and the function* $C(\varepsilon)$ *depend only on* d, Ω *and the bounds in conditions* (5.71), (5.76), (5.79) *and* (5.80). □

5.3.2 A Priori Estimates

The starting point is a second order (uniformly) elliptic differential operator L of the form

$$L(x,\partial_x)\varphi(x) = -\sum_{i,j=1}^{d} a_{ij}(x)\partial_{ij}\varphi(x) + \\ + \sum_{i=1}^{d} a_i(x)\partial_i\varphi(x) + a_0(x)\varphi(x)\,, \quad (5.81)$$

and a boundary first order operator B of the form

$$B(x,\partial_x)\varphi(x) = -\sum_{i=1}^{d} b_i(x)\partial_i\varphi(x) + b_0(x)\varphi(x)\,, \quad (5.82)$$

where Ω is a domain with C^2 boundary and the coefficients satisfy

$$\begin{aligned} &\sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geq \mu|\xi|^2, \quad \forall\,\xi\in\mathbb{R}^d\,, \quad x\in\Omega\,, \\ &a_{ij}\in C^0(\Omega), \qquad a_i,\, a_0 \in L^\infty(\Omega), \\ &\sum_{i,j=1}^{d} a_{ij}n_in_j = \sum_{i=1}^{d} b_in_i \ \text{ in } \ \partial\Omega, \qquad b_i,\, b_0\in C^1(\Omega), \end{aligned} \quad (5.83)$$

When $\Omega = \mathbb{R}^d$ the second-order coefficients are uniformly continuous (and bounded) and certainly, for Dirichlet boundary conditions, the boundary operator B is not involved.

Consider L as an unbounded operator in $L^p(\Omega)$, with either Dirichlet boundary conditions or oblique B-boundary conditions. First, a priori elliptic (Agmon-Douglis-Nirenberg) estimates are obtained for (uniformly) elliptic differential operator of the following type: for any $1 < p < \infty$ there is a positive constant

$C = C_p$ depending only on p, μ, the bounds of the coefficients a_{ij}, a_i, a_0, the modulus of continuity of a_{ij} and the domain Ω such that

$$\begin{aligned} &\|u\|_{2,p} \le C\Big[\|Lu\|_{0,p} + \|u\|_p\Big], \quad \forall u \in W^{2,p}(\Omega), \\ &\text{satisfying either} \quad u = 0 \text{ on } \partial\Omega \quad \text{or} \quad Bu = 0 \text{ on } \partial\Omega, \end{aligned} \tag{5.84}$$

where $\|\cdot\|_p$ is the norm in $L^p(\Omega)$, $W^{2,p}(\Omega)$ is the Banach (Sobolev) space of functions whose derivatives up to the 2 belong to L^p, with the natural norm $\|\cdot\|_{2,p}$. When $\Omega = \mathbb{R}^d$, the same a priori bounds hold for u in $W^{2,p}(\mathbb{R}^d)$.

Next, based on the above a priori estimate applied to the (complex) elliptic operator $L(x,\partial_x) + \mathrm{e}^{\mathrm{i}\theta}\partial_t^2$ in $n+1$ variables (x,t), with θ in $(-\pi/2, \pi/2)$, we can deduce that for some constants $C_p > 0$ and $\omega_p > 0$ the following estimate holds

$$\|\partial_x^2 u\|_p + |\lambda|^{1/2}\|\partial_x u\|_p + |\lambda|\|u\|_p \le C_p\|\lambda u + Lu\|_p, \tag{5.85}$$

for every u belonging to $W^{2,p}(\Omega)$ satisfying either $u = 0$ on $\partial\Omega$ or $Bu = 0$ on $\partial\Omega$, and for any λ with $\Re(\lambda) \ge \omega_p$. Hence, the operator $A = -L$ with domain $\mathcal{D}(A)$ defined as the Sobolev space $W^{2,p}(\Omega)$ with one of the boundary conditions either $u = 0$ on $\partial\Omega$ or $Bu = 0$ on $\partial\Omega$, generates an analytic semigroup in $L^p(\Omega)$.

Once the a priori estimates have been obtained, the above argument applies to Hölder space $C^\alpha(\overline{\Omega})$, $0 < \alpha < 1$ and to some extend to $C^1(\overline{\Omega})$, $C^0(\overline{\Omega})$, $L^1(\Omega)$ and $L^\infty(\Omega)$, e.g., Lunardi [112, Chapter 3, pp. 69–119].

Now, consider $L - I$ as an unbounded operator in $L^p(\Omega)$, with either Dirichlet boundary conditions or B-oblique boundary conditions. Mixed boundary conditions can be used as long as the boundary $\partial\Omega$ is composed by two smooth (closed and disjointed) portions Γ and $\partial\Omega \setminus \Gamma$ on which Dirichlet and B-oblique boundary conditions are imposed. Unless Ω is the whole space $\mathbb{R}^d$, the integro-differential operator I need to be *localized* and assumptions should be such that the ε-estimates hold. For instance, besides hypotheses (5.71) and (5.74), if (homogeneous) Dirichlet boundary conditions are used then we need to impose also (5.79) with boundary order γ_1 such that $(\gamma_1 - 1)p < d$. However, for (homogeneous) B-oblique boundary conditions we need to impose (5.78), i.e., the localization is trivial since no jumps outside $\overline{\Omega}$ are allowed.

Set $A = I - L$, based on the ε-estimates of the Proposition 5.19 and the Agmon-Douglis-Nirenberg estimates (5.84) for (uniformly) elliptic differential operator L, we deduce that for any $1 < p < \infty$ there is a positive constant $C = C_p$ depending only on p, μ, the bounds of the coefficients a_{ij}, a_i, a_0, the modulus of continuity of a_{ij}, the domain Ω and the bounds in the assumptions on I, such that

$$\begin{aligned} &\|u\|_{2,p} \le C\Big[\|Au\|_{0,p} + \|u\|_p\Big], \quad \forall u \in W^{2,p}(\Omega), \\ &\text{satisfying either} \quad u = 0 \text{ on } \partial\Omega \quad \text{or} \quad Bu = 0 \text{ on } \partial\Omega, \end{aligned} \tag{5.86}$$

where $\|\cdot\|_p$ is the norm in $L^p(\Omega)$, $W^{2,p}(\Omega)$ is the Banach (Sobolev) space of functions whose derivatives up to the 2 belong to L^p, with the natural norm $\|\cdot\|_{2,p}$. When $\Omega = \mathbb{R}^d$, the same a priori bounds hold for u in $W^{2,p}(\mathbb{R}^d)$.

Therefore, we deduce that for some constants $C_p > 0$ and $\omega_p > 0$ the following estimate holds

$$\|\partial_x^2 u\|_p + |\lambda|^{1/2}\|\partial_x u\|_p + |\lambda|\|u\|_p \leq C_p\|Au - \lambda u\|_p, \tag{5.87}$$

for every u in $W^{2,p}(\Omega)$ satisfying either $u = 0$ on $\partial\Omega$ or $Bu = 0$ on $\partial\Omega$, and for any λ with $\Re(\lambda) \geq \omega_p$. Hence, the (elliptic) integro-differential operator A with domain $\mathcal{D}(A)$ defined as the Sobolev space $W^{2,p}(\Omega)$ with one of the boundary conditions either $u = 0$ on $\partial\Omega$ or $Bu = 0$ on $\partial\Omega$, generates an analytic semigroup in $L^p(\Omega)$.

Once a priori estimates have been obtained, the above argument applies to Hölder space $C^\alpha(\overline{\Omega})$, $0 < \alpha < 1$ and to some extend to $C^1(\overline{\Omega})$, $C^0(\overline{\Omega})$, $L^1(\Omega)$ and $L^\infty(\Omega)$.

5.3.3 Maximum Principles

In order to apply the theory of Markov-Feller semigroups we need to establish the *maximum principle* for (elliptic) integro-differential operators. There are several versions (depending on regularity imposed on the solution) of the maximum principle valid for elliptic second-order differential operators of the form (5.81). Moreover, the type of maximum principle we need to obtain a Markov-Feller semigroup is of a global character and related to an equation of the form

$$Iu - Lu = f \text{ in } \Omega, \qquad u = 0 \text{ on } \mathbb{R}^d \setminus \Omega, \tag{5.88}$$

and

$$Iu - Lu = f \text{ in } \Omega, \qquad Bu = 0 \text{ on } \partial\Omega, \tag{5.89}$$

or even mixed boundary conditions. The maximum principle is formally stated as follows: Let u be a function satisfying (5.88) or (5.89) with $f \geq 0$ then $u \geq 0$. Certainly, the function space where u belongs and the assumptions on the coefficients of the operators L and I determine the meaning of the above equations.

The interested reader should consult the books Garroni and Menaldi [58, 59] for a comprehensive study on second-order integro-differential problems, and Portenko [147] and Skubachevskii [162], among others, for more general boundary conditions.

For unbounded domains Ω an extra conditions of the type $a_0(x) \geq c > 0$ for some positive constant c is necessary to prevent ergodic situations. Really, to generate a Markov-Feller semigroup $S(t)$ satisfying $S(t)\mathbb{1} = \mathbb{1}$ for every $t \geq 0$ we need $a_0(x) = 0$, for any x, otherwise, we have a sub-Makovian Feller semigroup.

To conclude, let us mention that the analytic semigroup generated under the conditions of the previous section, is also a Feller-Markov semigroup in $C^0(\overline{\Omega})$.

5.4 Green and Poisson Functions

This is a short comment on (elliptic/parabolic) integro-differential operators with oblique boundary conditions as discussed in the books by Garroni and

Menaldi [58, 59]. For instance, for a comprehensive analysis on the oblique boundary conditions for parabolic second-order differential equations we can see Tsuchiya [174, 175], and for Wentzell boundary conditions we can check Menaldi and Tubaro [126], and even more general type of boundary conditions can be found in the books Portenko [147] and Skubachevskii [162], among others.

The modern (analytic) semigroup theory is a powerful method to treat many problems. Perhaps a classic application is to study parabolic equations from elliptic equations, where starting from a priori (elliptic) estimates the whole theory of parabolic equations can be developed. For elliptic and parabolic equations there are (classic) direct arguments based on so called Green and Poisson functions. Essentially, the inverse of an integro-differential problem is a Fredholm operator of an integral type, and in the case of parabolic equations this is related with a Markov process. Actually, the density probability transition function of a Markov process is the Green functions and the so-called *local time* is related with the Poisson function.

Let $L-I$ be an elliptic integro-differential operator as in the previous section. Given three functions $f(x,t)$, $\psi(x,t)$ and $\varphi(x)$ defined for $x\in\Omega$ and $t\in[0,T]$, we consider the second order integro-differential parabolic equation

$$\begin{aligned} &\partial_t u + Lu = Iu + f \quad \text{in } \Omega\times(0,T]\,,\\ &u=\varphi \quad \text{on } \Omega\times\{0\}\,,\\ &Bu=0 \quad \text{on } \partial\Omega\times[0,T]\,, \end{aligned} \tag{5.90}$$

with *homogeneous* oblique boundary conditions, and

$$\begin{aligned} &\partial_t v + Lv = Iv \quad \text{in } \Omega\times(0,T]\,,\\ &v=0 \quad \text{on } \Omega\times\{0\}\,,\\ &Bv=\psi \quad \text{on } \partial\Omega\times[0,T]\,, \end{aligned} \tag{5.91}$$

with *non homogeneous* oblique boundary conditions.

Definition 5.20 (Green/Poisson function)**.** A measurable function $G(x,t,\xi)$ defined in $\overline{\Omega}\times(0,T]\times\Omega$ and locally integrable in (t,ξ) is called a *Green function* for the parabolic second order integro-differential operator $L-I$ in Ω, with oblique boundary conditions given by first order differential operator B on $\partial\Omega$ if for any smooth functions $f(\xi,\tau)$ and $\varphi(\xi)$ with compact supports in $\Omega\times(0,T]$ and Ω, respectively, the potential function

$$u(x,t)=\int_0^t \mathrm{d}\tau\int_\Omega G(x,t-\tau,\xi)f(\xi,\tau)\mathrm{d}\xi+\int_\Omega G(x,t,\xi)\varphi(\xi)\mathrm{d}\xi$$

is either a *classic solution*, i.e., in the space $C^{2,1}(\Omega\times(0,T])\cap C^{1,0}(\partial\Omega\times(0,T])$ or a *strong solution*, i.e., in the space $W_p^{2,1}(\Omega\times(0,T))$ of the problem (5.90) with homogeneous oblique boundary conditions. Similarly, the *Poisson function* is a measurable function $P(x,t,\xi)$ defined in $\overline{\Omega}\times(0,T]\times\partial\Omega$ and locally integrable in (t,ξ) such that the potential function

$$v(x,t)=\int_0^t \mathrm{d}\tau\int_{\partial\Omega} P(x,t-\tau,\xi)\psi(\xi,\tau)\mathrm{d}\xi$$

is either a *classic solution*, i.e., in the space $C^{2,1}(\Omega\times(0,T])\cap C^{1,0}(\partial\Omega\times[0,T])$ or a *strong solution*, i.e., in the space $W_p^{2,1}(\Omega\times(0,T))$ of the problem (5.91) with non homogeneous oblique boundary conditions, for any smooth function $\psi(\xi,\tau)$ with a compact support in $\partial\Omega\times(0,T]$. □

The differential part of the Green function G_L is the piece of the Green function due to the differential operator L, i.e., the solution u of the equation

$$\begin{aligned}&\partial_t u + Lu = f \quad \text{in } \Omega\times(0,T]\,,\\ &u=\varphi \quad \text{on } \Omega\times\{0\}\,,\\ &Bu=0 \quad \text{on } \partial\Omega\times[0,T]\,,\end{aligned}$$

with *homogeneous* oblique boundary conditions, is given by the expression

$$u(x,t)=\int_0^t \mathrm{d}\tau\int_\Omega G_L(x,t-\tau,\xi)f(\xi,\tau)\mathrm{d}\xi+\int_\Omega G_L(x,t,\xi)\varphi(\xi)\mathrm{d}\xi\,,$$

for any smooth functions $f(\xi,\tau)$ and $\varphi(\xi)$ with compact supports in $\Omega\times(0,T]$ and Ω. Actually, in view of the estimates on G_L, the above representation formula remains valid for a more general class of functions, either in the Hölder space $C^{2+\alpha,1+\alpha/2}(\overline{\Omega}\times(0,T])$, $0<\alpha<1$ or in the Sobolev space $W_p^{2,1}(\Omega\times(0,T))$, $1<p<\infty$.

The following results are found in Solonnikov [163, 164] and Ivasišen [79]

Theorem 5.21. *Let Ω be a bounded domain in $\mathbb{R}^d$ with its boundary $\partial\Omega$ of class $C^{1,\alpha}$, with $0<\alpha<1$, and L and B be the operators as above, satisfying (5.83). Then the strong Green function for the parabolic second order differential operator ∂_t+L in $\Omega\times(0,T]$, with oblique boundary conditions given by first order differential operator B on $\partial\Omega\times[0,T]$ exists and satisfies the following estimate*

$$|\nabla^\ell G_L(x,t,\xi)|\le Ct^{-(d+\ell)/2}\exp(-c|x-\xi|^2/t)\,,$$

for every (x,t,ξ) in $\overline{\Omega}\times(0,T]\times\Omega$, for any $\ell=0,1,2$ and some positive constants C and c. Moreover, if the boundary $\partial\Omega$ is of class $C^{2,\alpha}$ and the lower order coefficients a_i are in $C^\alpha(\overline{\Omega})$ and the boundary coefficients b_i are in $C^{1+\alpha}(\partial\Omega)$, then G_L is the classic Green function and enjoys the estimates

$$\begin{aligned}|\nabla^\ell G_L(x,t,\xi)-\nabla^\ell G_L(y,t,\xi)|&\le M|x-y|^\alpha t^{-(d+\ell+\alpha)/2}\times\\ &\times[\exp(-m|x-\xi|^2/t)+\exp(-m|y-\xi|^2/t)]\,,\end{aligned}$$

$$\begin{aligned}|\nabla^\ell G_L(x,t,\xi)-\nabla^\ell G_L(x,s,\xi)|&\le M|t-s|^{\alpha/2}\times\\ &\times[t^{-(d+\ell+\alpha)/2}\exp(-m|x-\xi|^2/t)+s^{-(d+\ell)+\alpha/2}\exp(-m|x-\xi|^2/s)]\,,\end{aligned}$$

for every x,y in $\overline{\Omega}$, s,t in $(0,T]$ and ξ in Ω, and

$$\begin{aligned}|\nabla^\ell G_L(x,t,\xi)-\nabla^\ell G_L(x,t,\eta)|&\le M|\xi-\eta|^\alpha t^{-(d+\ell+\alpha)/2}\times\\ &\times[\exp(-m|x-\xi|^2/t)+\exp(-m|x-\eta|^2/t)]|\,,\end{aligned}$$

for every x in $\overline{\Omega}$, t in $(0,T]$ and ξ,η in Ω, for any $\ell = 0,1,2$ and some positive constants C, c, M and m. In all estimates, the constants C, c, M and m depend only on the bounds imposed on the coefficients (of the differential operators L and B) throughout the various assumptions and, on the domain $\Omega \times (0,T]$. Estimates similar to the above hold for the Poisson function. □

Let G_L be the Green function associated with the differential operator L. To construct the Green function G associated with the integro-differential operator $\partial_t + L - I$, we solve a Volterra equation

$$\begin{array}{lll} \text{either} & \text{find } Q_I \text{ such that} & Q_I = Q_L + Q_L \star Q_I\,, \\ \text{or} & \text{find } G \text{ such that} & G = G_L + G_L \star IG\,, \end{array} \tag{5.92}$$

with the relations $Q_L = IG_L$ and $G = G_L + G_L \star Q_I$. Recall that the bullet $\star$ means the kernel-convolution, i.e., for any $\varphi(x,t,y,s)$ and $\psi(x,t,y,s)$

$$(\varphi \star \psi)(x,t,y,s) = \int_0^T \mathrm{d}\tau \int_{\mathbb{R}^d} \varphi(x,t,z,\tau)\psi(z,\tau,y,s)\mathrm{d}z,$$

and, in particular for any $\varphi(x,t,y)$ and $\psi(x,t,y)$,

$$(\varphi \star \psi)(x,t,y) = \int_0^t \mathrm{d}\tau \int_{\mathbb{R}^d} \varphi(x,\tau,z)\psi(z,t-\tau,y)\mathrm{d}z,$$

for every $t > 0$, x and y in $\mathbb{R}^d$. Actually, we express Q_I as the following series

$$Q_I = \sum_{n=1}^{\infty} Q_n\,, \qquad Q_0 = Q_L, \quad Q_n = Q_L \star Q_{n-1}\,,\ n \geq 1, \tag{5.93}$$

where the convergence is in the sense of following Green spaces.

To estimate the Green function of the integro-differential operator $\partial_t + L - I$ we consider a number of semi-norms used to define the Green function spaces in the domain $\Omega \times (0,T]$. For any kernel $\varphi(x,t,\xi)$, with $x,\xi \in \Omega$, $t \in (0,T]$, $k \geq 0$ and $0 < \alpha < 1$, we define

$$C(\varphi,k) = \inf\{C \geq 0\ :\ |\varphi(x,t,\xi)| \leq Ct^{-1+(k-d)/2},\ \forall x,t,\xi\}, \tag{5.94}$$

$$K(\varphi,k) = K_1(\varphi,k) + K_2(\varphi,k), \tag{5.95}$$

$$K_1(\varphi,k) = \inf\{K_1 \geq 0 : \int_\Omega |\varphi(x,t,\xi)|\mathrm{d}\xi \leq K_1 t^{-1+k/2},\ \forall x,t\}, \tag{5.96}$$

$$K_2(\varphi,k) = \inf\{K_2 \geq 0 : \int_\Omega |\varphi(x,t,\xi)|\mathrm{d}x \leq K_2 t^{-1+k/2},\ \forall t,\xi\}, \tag{5.97}$$

$$M(\varphi,k,\alpha) = M_1(\varphi,k,\alpha) + M_2(\varphi,k,\alpha) + M_3(\varphi,k,\alpha), \tag{5.98}$$

$$M_1(\varphi,k,\alpha)=\inf\{M_1\geq 0:|\varphi(x,t,\xi)-\varphi(x',t,\xi)|\leq \\ \leq M_1|x-x'|^{\alpha}t^{-1+(k-d-\alpha)/2},\ \forall x,x',t\},\quad (5.99)$$

$$M_2(\varphi,k,\alpha)=\inf\{M_2\geq 0:|\varphi(x,t,\xi)-\varphi(x,t',\xi)|\leq \\ \leq M_2|t-t'|^{\alpha/2}[t^{-1+(k-d-\alpha)/2}\vee t'^{-1+(k-d-\alpha)/2}],\ \forall x,t,t',\xi\},\quad (5.100)$$

$$M_3(\varphi,k,\alpha)=\inf\{M_3\geq 0:|\varphi(x,t,\xi)-\varphi(x,t,\xi')|\leq \\ \leq M_3|\xi-\xi'|^{\alpha}t^{-1+(k-d-\alpha)/2},\ \forall x,t,\xi,\xi'\},\quad (5.101)$$

$$N(\varphi,k,\alpha)=N_1(\varphi,k,\alpha)+N_2(\varphi,k,\alpha)+N_3(\varphi,k,\alpha)+N_4(\varphi,k,\alpha),\quad (5.102)$$

$$N_1(\varphi,k,\alpha)=\inf\{N_1\geq 0:\int_{\Omega}|\varphi(x,t,\xi)-\varphi(x',t,\xi)|\mathrm{d}\xi\leq \\ \leq N_1|x-x'|^{\alpha}t^{-1+(k-\alpha)/2},\ \forall x,x',t,s\},\quad (5.103)$$

$$N_2(\varphi,k,\alpha)=\inf\{N_2\geq 0:\int_{\Omega}|\varphi(x,t,\xi)-\varphi(x,t',\xi)|\mathrm{d}\xi\leq \\ \leq N_2|t-t'|^{\alpha/2}[t^{-1+(k-\alpha)/2}\vee t'^{-1+(k-\alpha)/2}],\ \forall x,t,t'\},\quad (5.104)$$

$$N_3(\varphi,k,\alpha)=\inf\{N_3\geq 0:\int_{\Omega}|\varphi(x,t,\xi)-\varphi(x,t',\xi)|\mathrm{d}x\leq \\ \leq N_3|t-t'|^{\alpha/2}[t^{-1+(k-\alpha)/2}\vee t'^{-1+(k-\alpha)/2}],\ \forall t,t',\xi\},\quad (5.105)$$

$$N_4(\varphi,k,\alpha)=\inf\{N_4\geq 0:\int_{\Omega}|\varphi(x,t,\xi)-\varphi(x,t,\xi')|\mathrm{d}x\leq \\ \leq N_4|\xi-\xi'|^{\alpha}t^{-1+(k-\alpha)/2},\ \forall t,\xi,\xi'\},\quad (5.106)$$

$$R(\varphi,k,\alpha)=R_1(\varphi,k,\alpha)+R_2(\varphi,k,\alpha),\quad (5.107)$$

$$R_1(\varphi,k,\alpha)=\inf\{R_1\geq 0:\int_{\Omega}|\varphi(Z,t,\xi)-\varphi(Z',t,\xi)|J_{\eta}(Z,Z')\mathrm{d}z\leq \\ \leq R_1\eta^{\alpha}t^{-1+(k-\alpha)/2},\ \forall Z,Z',t,\xi\ \text{ and }\ \eta>0\},\quad (5.108)$$

$$R_2(\varphi,k,\alpha)=\inf\{R_2\geq 0:\int_{\Omega}|\varphi(x,t,Z)-\varphi(x,t,Z')|J_{\eta}(Z,Z')\mathrm{d}z\leq \\ \leq R_2\eta^{\alpha}t^{-1+(k-\alpha)/2},\ \forall x,t,Z,Z'\ \text{ and }\ \eta>0\},\quad (5.109)$$

where the change of variables $Z(z)$ and $Z'(z)$ are diffeomorphisms of class C^1 in $\mathbb{R}^d$, and the Jacobian

$$J_\eta(Z, Z') = |\det(\nabla Z)| \wedge |\det(\nabla Z')| \tag{5.110}$$

if $|Z - Z'| \le \eta$ and Z, Z' belong to $\overline{\Omega}$, and vanishing otherwise, here $\det(\cdot)$ means the determinant of a $d \times d$ matrix, ∇Z, $\nabla Z'$ stand for the matrices of the first partial derivatives of $Z(z), Z'(z)$ with respect to the variable z, and $\wedge, \vee$ denote the minimum, maximum (resp.) between two real numbers.

Definition 5.22 (Green function spaces)**.** Let us denote by $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}$ (or $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}(\overline{\Omega} \times (0,T], \mathbb{R}^n)$ when necessary), $k \ge 0$, $n \in \mathbb{N}$ and $0 < \alpha < 1$, the space of all continuous functions (or kernels) $\varphi(x,t,\xi)$ defined for x, ξ in $\Omega \subset \mathbb{R}^d$ and $0 < t \le T$, with values in $\mathbb{R}^n$ (usually $n = 1$ and $k \ge 0$) and such that the above infima (semi-norms) (5.94),..., (5.109) (of order k) are finite. Thus the maximum of the quantities (5.94),..., (5.109), denoted by $[\![\cdot]\!]_{k,\alpha} = [\![\cdot]\!]_{\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}}$, is the norm of the Banach space $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}$. When $\alpha = 0$, we denote by $\mathcal{G}_k^0$ (or $\mathcal{G}_k^0(\overline{\Omega} \times (0,T], \mathbb{R}^n)$ when necessary), $k \ge 0$, and $n \in \mathbb{N}$, the space of all measurable functions (or kernels) $\varphi(x,t,\xi)$ defined for x, ξ in $\Omega \subset \mathbb{R}^d$ and $0 < t \le T$, with values in $\mathbb{R}^n$ (usually $n = 1$ and $k \ge 0$) and such that the two infima (5.94) and (5.95) (of order k) are finite, with the norm $[\![\cdot]\!]_{k,0} = [\![\cdot]\!]_{\mathcal{G}_k^0}$. □

The Volterra equations (5.92) is solved in a Green function space $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}$. We have

Theorem 5.23 (Green function)**.** *Under suitable conditions on the coefficients as discussed above, and in particular if the boundary coefficients b_i belongs to $C^{1+\alpha}(\partial\Omega)$, for any $i = 1, \dots, d$, then there exists the (strong) Green function $G(x,t,\xi)$ for the parabolic second order integro-differential operator $\partial_t + L - I$ in $\Omega \times (0,T]$, with oblique boundary conditions given by first order differential operator B on $\partial\Omega \times [0,T]$. Moreover $G = G_L + G_L \star Q$, where Q is the solution of the Volterra equation (5.92) in the Green function space $\mathcal{G}_{2-\gamma}^0$, given by (5.93) with $Q_0 = IG_L$ and the semi-norms $C(\nabla^\ell G_L \star Q, 4-\ell-\gamma)$, $K(\nabla^\ell G_L \star Q, 4-\ell-\gamma)$, $M_i(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$, $i = 1,2$, $N_i(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$, $i = 1,2,3$ and $R_1(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$ are finite, for $\ell = 0,1$. Furthermore, if we assume Hölder continuous coefficients then $G(x,t,\xi)$ is also the classic Green function and solution of the Volterra equation (5.92) in the Green function space $\mathcal{G}_{2-\gamma}^{\alpha,\frac{\alpha}{2}}$. In this case, the semi-norms $M_3(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$, $N_4(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha$, $R_2(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$, for $\ell = 0,1$, and the semi-norms $C(\nabla^2 G_L \star Q, 2-\gamma)$, $K(\nabla^2 G_L \star Q, 2-\gamma)$, $M_2(\nabla G_L \star Q, 2-\gamma, 2\alpha)$, $M(\nabla^2 G_L \star Q, 2-\gamma, \alpha)$, $N(\nabla^2 G_L \star Q, 2-\gamma, \alpha)$, $N_i(\nabla G_L \star Q, 2-\gamma, 2\alpha)$, $i = 2,3$ and $R(\nabla^2 G_L \star Q, 2-\gamma, \alpha)$ are finite.* □

If $G(x,t,\xi)$ and $P(x,t,\xi)$ are the Green function and the Poisson kernel, respectively, then any smooth solution of the following (parabolic, differential)

boundary value problem

$$\begin{aligned}&\partial_t u + Lu = Iu + f \quad \text{in } \Omega \times (0,T]\,,\\ &u = \varphi \quad \text{on } \Omega \times \{0\}\,,\\ &Bu = \psi \quad \text{on } \partial\Omega \times [0,T]\,,\end{aligned}$$

is given by the expression

$$u(x,t) = \int_0^t \mathrm{d}\tau \int_\Omega G(x,t-\tau,\xi) f(\xi,\tau)\mathrm{d}\xi + \\ + \int_\Omega G(x,t,\xi)\varphi(\xi)\mathrm{d}\xi + \int_0^t \mathrm{d}\tau \int_{\partial\Omega} P(x,t-\tau,\xi)\psi(\xi,\tau)\mathrm{d}\xi\,,$$

and the *Chapman-Kolmogorov identity*

$$G(x,t+s,\xi) = \int_\Omega G(x,t,y)G(y,s,\xi)\mathrm{d}y\,,$$

for every x, ξ in Ω and t, s in $(0,T]$ is satisfied. In particular for $f = a_0$, $\varphi = 1$ and $\psi = b_0$ we obtain

$$1 - \int_\Omega G(x,t,\xi)\mathrm{d}\xi = \int_0^t \mathrm{d}\tau \int_\Omega G(x,t-\tau,\xi)a_0(\xi)\mathrm{d}\xi + \\ + \int_0^t \mathrm{d}\tau \int_{\partial\Omega} P(x,t-\tau,\xi)b_0(\xi)\mathrm{d}\xi\,.$$

In particular, if $a_0 = 0$ and $b_0 = 0$ then

$$\int_\Omega G(x,t,\xi)\mathrm{d}\xi = 1 \qquad \forall\,(x,t) \in \Omega \times (0,T]\,,$$

which is one of the key property of a *transition density* function, used to describe *Markov processes*. The *weak maximum principle* implies that $G \geq 0$ and sometime the *strong maximum principle* yields the strictly positivity of the Green (and Poisson) functions.

All the above estimates are valid on $\Omega \times [0,T]$ for any $T > 0$. In an unbounded time interval we have the following

Theorem 5.24 (time-unbounded)**.** *Let $G(x,t,\xi)$ be the Green function for the parabolic second order integro-differential operator $\partial_t + L - I$ in $\Omega \times (0,\infty]$, with oblique boundary conditions given by first order differential operator B on $\partial\Omega \times [0,\infty]$ as given by* Theorem 5.23*. Then we have the following estimates: for every $\delta > 0$ there exist positive constants C_0, M_0 such that for any $t, t' \geq \delta$*

$$|\nabla^\ell G(x,t,\xi)| \leq C_0, \quad \ell = 0,1,2, \tag{5.111}$$

$$|\nabla^\ell G(x,t,\xi) - \nabla^\ell G(x',t',\xi')| \leq M_0\big(|x-x'|^\alpha + \\ + |t-t'|^{\alpha/2} + |\xi-\xi'|^\alpha\big), \tag{5.112}$$

for any $\ell = 0, 1$, x, ξ *and* x', ξ' *in* $\overline{\Omega}$. *Moreover, if we assume Hölder continuous coefficients then for any* $\delta > 0$ *there exists a positive constant* $c = c(\delta) > 0$ *such that*

$$G(x,t,\xi) \geq c, \qquad \forall (x,t,\xi) \in \overline{\Omega} \times (\delta, \infty] \times \Omega\,, \tag{5.113}$$

we also have the estimate

$$|\nabla G(x,t,\xi) - \nabla G(x,t',\xi)| \leq M_0\, |t - t'|^{\alpha}, \tag{5.114}$$

and we may let $\ell = 2$ *in estimate of* (5.112). □

• *Remark* 5.25. Notice that from the technique used in proving of the above Theorem 5.24 we can estimate the constants C_0 and M_0 appearing in (5.111), (5.112) and (5.114) as t, t' become large, i.e., if we define

$$q(t) = \sup_x \int_\Omega G(x,t,y)\mathrm{d}y, \quad t \geq 1 \tag{5.115}$$

then we have for any $t' \geq t \geq T$ the estimates

$$C_0 \leq C_G\, q(T), \quad \text{and} \quad M_0 \leq M_G\, q(T), \quad T \geq 1 \tag{5.116}$$

where the constants C_G and M_G depend on the semi-norms $K_2(\nabla^\ell G, 2-\ell)$, $K_2(G,2)$, $M_1(\nabla^\ell G, 2-\ell, \alpha)$, $M_3(\nabla^\ell G, 2-\ell, \alpha)$ and $N_3(\nabla^\ell G, 2-\ell, \alpha)$, but all on the time interval $[0,1]$. This means that estimates for the Green function $G(x,t,\xi)$ on the Green spaces in the time interval $[0,\infty)$ are obtained from estimates on any bounded time interval $[0,\delta]$, with $\delta > 0$ plus a bound on the expression (5.115) of $q(t)$ as t becomes large. □

In all theses estimates applied to integro-differential (or Lévy-type) operators with a dominant second order differential part. Other situation can be found in Jacob [80, Vol II, Section 2.7, pp. 138–151], Kolokoltsov [99], Komatsu [100, 101], Mikulevicius and Pragarauskas [131, 132] and Mikulevicius and Rozovskii [133], among others.

5.5 Examples of Transition Functions

Green and transition functions are essentially the same objects, one is seen as the inverse of a functional operator (e.g., an integro-differential operator) and the other is the essence of a Markov-Feller processes.

Let us start with a couple of simple one-dimensional prototypes first in the whole real line and with boundary conditions in the real semi-line. First recall that given a locally compact separable complete metric space E, we define $C_0(E)$ as the Banach space of all continuous real functions on E vanishing at infinity, i.e., $f : E \to \mathbb{R}$, continuous and for any $\varepsilon > 0$ there exists a compact subset K of E such that $|f(x)| < \varepsilon$ for any x in $E \setminus K$. Note that we are using indistinctly $p(t,x,B)$ or $p(x,t,B)$ for the transition functions.

5.5.1 One-Dimensional

Example 5.1 (Wiener process)**.** On the state space $\mathbb{R}$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,B) = \frac{1}{\sqrt{2\pi t}}\Big\{\int_B \exp\big[-\frac{(y-x)^2}{2t}\big]\mathrm{d}y\Big\},$$

for any $t>0$, x in $\mathbb{R}$ and B in $\mathcal{B}$. This is the typical one-dimensional Brownian motion or Wiener process. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) = \int_{\mathbb{R}} f(y)\, p(t,x,\mathrm{d}y) = \\ = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x+\sqrt{t}z)\,\exp\big(-\frac{z^2}{2}\big)\mathrm{d}z,$$

for every $t>0$ and x in $\mathbb{R}$. Its infinitesimal generator A is the differential operator A

$$\mathcal{D}(A) = \{f\in C_0(\mathbb{R})\cap C^2(\mathbb{R}) : f''\in C_0(\mathbb{R})\},\quad Af = \frac{1}{2}f''.$$

The associated resolvent operator in $C_0(\mathbb{R})$ is given by

$$R(\lambda)f(x) = \int_{\mathbb{R}} f(y)\, r(\lambda,x,\mathrm{d}y) = \\ = \frac{1}{\sqrt{2\lambda}}\int_{-\infty}^{+\infty} f(x+\frac{z}{\sqrt{2\lambda}})\,\exp\big(-|z|\big)\mathrm{d}z,$$

where the resolvent kernel is

$$r(\lambda,x,B) = \frac{1}{\sqrt{2\lambda}}\int_B \exp\big(-\sqrt{2\lambda}|x-y|\big)\mathrm{d}y.$$

for every $\lambda>0$, x in $\mathbb{R}$, and B in $\mathcal{B}$. A constant drift b can be added so that $Af = \frac{1}{2}f''+bf'$ and a realization with continuous paths can be constructed. □

Example 5.2 (Poisson process)**.** On the state space $\mathbb{R}$ with its Borel σ-algebra $\mathcal{B}$ and for a given positive constant c, we consider

$$p(t,x,B) = \mathrm{e}^{-ct}\sum_{k=0}^{\infty}\frac{(ct)^k}{k!}\mathbb{1}_B(x+k),$$

for any $t>0$, x in $\mathbb{R}$ and B in $\mathcal{B}$. This is the typical one-dimensional Poisson process. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) = \int_{\mathbb{R}} f(y)\, p(t,x,\mathrm{d}y) = \mathrm{e}^{-ct}\sum_{k=0}^{\infty}\frac{(ct)^k}{k!}\, f(x+k),$$

for every $t > 0$ and x in $\mathbb{R}$. Its infinitesimal generator is

$$\mathcal{D}(A) = C_0(\mathbb{R}), \qquad Af(x) = c[f(x+1) - f(x)], \quad \forall x \in \mathbb{R},$$

Note that A is a nonlocal operator and that only a cad-lag realization of the above Poisson process can be constructed. We can generalize this example to a compound Poisson process $(P_t : t \geq 0)$, with parameters (c, μ), where $c > 0$ and μ is a probability distribution on $\mathbb{R}$. The probability transition function is

$$p(t, x, B) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!}\, \mu^k(B), \qquad \mu^0 = \delta_0, \quad \text{and}$$
$$\mu^k(B) = (\mu^{k-1} \star \mu)(B) = \int_{\mathbb{R}\times\mathbb{R}} \mathbb{1}_B(y+z)\, \mu^{k-1}(\mathrm{d}y)\, \mu(\mathrm{d}z),$$

for $k = 1, 2, \ldots$, for any $t > 0$, x in $\mathbb{R}$ and B in $\mathcal{B}$, where δ_0 is the Dirac measure at the origin. Since μ^k are all probability measures, the above series converges. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) = \int_{\mathbb{R}} f(y)\, p(t, x, \mathrm{d}y) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} \int_{\mathbb{R}} f(x+y)\, \mu^k(\mathrm{d}y),$$

for every $t > 0$ and x in $\mathbb{R}$. Its infinitesimal generator is the bounded (integral) linear operator on $C_0(\mathbb{R})$, defined by

$$Af(x) = c \int_{\mathbb{R}} [f(x+y) - f(x)]\, \mu(\mathrm{d}y), \quad \forall x \in \mathbb{R}.$$

Again, only a cad-lag realization of the above Poisson process can be constructed. □

Example 5.3 (Cauchy process)**.** On the state space $\mathbb{R}$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, B) = \frac{1}{\pi} \int_B \frac{t}{t^2 + (y-x)^2} \mathrm{d}y,$$

for any $t > 0$, x in $\mathbb{R}$ and B in $\mathcal{B}$. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) = \int_{\mathbb{R}} f(y)\, p(t, x, \mathrm{d}y) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(x+tz)\, \frac{1}{1+z^2}\, \mathrm{d}z,$$

for every $t > 0$ and x in $\mathbb{R}$. Its infinitesimal generator is of the form

$$Af(x) = \frac{1}{\pi} \int_0^{\infty} \frac{f(x+y) + f(x-y) - 2f(x)}{y^2} \mathrm{d}y, \quad \forall x \in \mathbb{R},$$

and the domain $\mathcal{D}(A)$ contains all twice-differentiable functions with compact support in $\mathbb{R}$. Note that A is a nonlocal operator and that only a cad-lag realization of the above Cauchy process can be constructed. □

Example 5.4 (Subordinator)**.** An increasing (and so, of bounded variation) Lévy process is called a subordinator (process). Its infinitesimal generator the bounded (integral) linear operator on $C_0(\mathbb{R}) \cap C^1(\mathbb{R})$, defined by

$$Af(x) = bf'(x) + \int_{\mathbb{R}} [f(x+y) - f(x)]\, \mu(\mathrm{d}y), \quad \forall x \in \mathbb{R}.$$

where b is a non-negative constant and μ is a measure on $\mathbb{R}$, with support in $[0, \infty[$, such that $\mu(\{0\}) = 0$ and

$$\int_{]0,\infty[} (y \wedge 1)\, \mu(\mathrm{d}y) < \infty,$$

the constants $b \geq 0$ and the measure μ are referred to as the drift (term) and the Lévy measure associated with the subordinator.

For a given subordinator $\{X_t : t \geq 0\}$ with characteristic (b, μ), if μ is bounded then the probability transition function $p(t, x, B)$, and the associated semigroup $S(t)$ in $C_0(\mathbb{R})$ can be expressed in a way similar to a compound Poisson process. In general, only the (infinitely divisible) distribution of X_1 or the *Laplace exponent* ψ of the subordinator is calculated (or given), i.e.,

$$\mathbb{E}\{\mathrm{e}^{-\xi X_t}\} = \mathrm{e}^{-t\psi(\xi)}, \quad \forall \xi \geq 0,$$

where

$$\psi(\xi) = b\xi + \int_{]0,\infty[} (1 - \mathrm{e}^{-y\xi}) \mu(\mathrm{d}y), \quad \forall \xi \geq 0.$$

For instance, an α-stable subordinator $\{X_t : t \geq 0\}$ has

$$\psi(\xi) = \xi^\alpha \quad \text{and} \quad \mu(\mathrm{d}y) = \frac{\alpha\, \mathrm{d}y}{\Gamma(1-\alpha) x^{1+\alpha}},$$

and in particular, for $\alpha = 1/2$, this is called the Lévy subordinator, which is the first hitting time for one-dimensional standard Brownian motion $\{B_t : t \geq 0\}$, i.e., $X_t = \inf\{s > 0 : B_s = t/\sqrt{2}\}$. Similarly, the inverse Brownian subordinator has

$$\psi(\xi) = b\xi + \frac{1}{\sigma^2}\big(\sqrt{2\xi\sigma^2 + \beta^2} - \beta^2\big) \quad \text{and}$$
$$\mu(\mathrm{d}y) = \frac{1}{\sqrt{2\pi\sigma^2 y^3}} \exp\Big(-\frac{y\beta^2}{2\sigma^2}\Big)\, \mathrm{d}y,$$

as its Laplace exponent and its Lévy measure. Moreover, if $b = 0$ then this can be interpreted as the hitting time for a Brownian motion with drift β and variance σ^2, and its probability transition function is given by

$$p(t, x, \mathrm{d}y) = \frac{t}{\sqrt{2\pi\sigma^2 (y-x)}} \exp\Big[-\frac{[\beta(y-x) - t]^2}{2(y-x)\sigma^2}\Big] \mathbb{1}_{\{y > x\}}\, \mathrm{d}y.$$

Analogously, the Gamma subordinator (process), with parameters $\alpha, \beta > 0$ and $b \geq 0$, has

$$\psi(\xi) = b\xi + \alpha \ln\Big(1 + \frac{\xi}{\beta}\Big) \quad \text{and} \quad \mu(\mathrm{d}y) = \frac{\alpha}{y} \exp\big(-y\beta\big)\,\mathrm{d}y,$$

and if $b = 0$ then its probability transition function is given by

$$p(t, x, \mathrm{d}y) = \frac{\beta^{\alpha t}}{\Gamma(at)} \exp\big[-(y-x)\beta\big](y-x)^{\alpha t - 1} \mathbb{1}_{\{y>x\}}\,\mathrm{d}y.$$

The interested reader may check, e.g., the books by Applebaum [1, Chapter 1, pp. 1-81], Bertoin [7, Chapter III, pp. 71-102], Itô [77, Section 1.11] and Sato [157, Chapter 6, pp. 197-236]. □

Example 5.5 (Wiener-Poisson)**.** On the state space $\mathbb{R}$ with its Borel σ-algebra $\mathcal{B}$ and for a given positive constant c, we consider

$$p(t, x, B) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} \frac{1}{\sqrt{2\pi t}} \Big\{ \int_B \exp\big[-\frac{(x+k-y)^2}{2t}\big]\mathrm{d}y \Big\},$$

for any $t > 0$, x in $\mathbb{R}$ and B in $\mathcal{B}$. This is the sum of independent Wiener and Poisson processes. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) = \int_{\mathbb{R}} f(y)\, p(t, x, \mathrm{d}y) = \\ = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{\sqrt{2\pi}\, k!} \int_{-\infty}^{\infty} f(x + k + \sqrt{t}z) \exp\big(-\frac{z^2}{2}\big)\mathrm{d}z,$$

for every $t > 0$ and x in $\mathbb{R}$. Its infinitesimal generator A is the closure of the (closable) operator $\mathring{A}$

$$\mathcal{D}(\mathring{A}) = C_0^0(\mathbb{R}) \cap C^2(\mathbb{R}),$$
$$\mathring{A}f(x) = \frac{1}{2}f''(x) + c[f(x+1) - f(x)], \quad \forall x \in \mathbb{R},$$

Only a cad-lag realization can be constructed as $(x + P_t + W_t : t \geq 0)$, where $(W_t : t \geq 0)$ is a standard Wiener process independent of the Poisson process $(P_t : t \geq 0)$. We can generalize this example to a $(W_t : t \geq 0)$ Wiener process (with drift b and covariance σ^2) and a $(P_t : t \geq 0)$ compound Poisson processes (with parameters (c, μ)), independent of each other. Thus b is a real constant, $\sigma, c > 0$, and μ is a probability distribution on $\mathbb{R}$. Again, a cad-lag realization is given by $(X_t = x + W_t + P_t : t \geq 0)$ and the probability transition function is

$$p(t, x, B) = \int_B \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{2\pi t\, k!}\, p_k(t, x-y)\mathrm{d}y,$$
$$p_0(t, x) = \Big\{ \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} \exp\big[-\frac{(x+bt-y)^2}{2t}\big]\, \mu(\mathrm{d}y) \Big\},$$
$$p_k(t, x) = \int_{\mathbb{R}} p_k(t, x-y)\, \mu(\mathrm{d}y), \quad k = 1, 2, \ldots,$$

for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$. Again, since p_0 is a probability density and μ^k is a probability measure the above series converges. Notice that if μ^k denotes the k convolution as defined in Example 5.2, then p_k can be expressed as $p_w \star \mu^k$, where p_w is the probability density transition function of a Wiener process. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) = \int_{\mathbb{R}} f(y)\, p(t,x,\mathrm{d}y) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{\sqrt{2\pi}\, k!} \times \\ \times \int_{\mathbb{R}} \mu^k(\mathrm{d}y) \int_{-\infty}^{\infty} f(x+y+\sqrt{t}z)\, \exp\big(-\frac{z^2}{2}\big)\mathrm{d}z,$$

for every $t > 0$ and x in $\mathbb{R}$. The infinitesimal generator is the closure of the (closable) integro-differential operator $\mathring{A}$

$$\mathcal{D}(\mathring{A}) = C_0^0(\mathbb{R}) \cap C^2(\mathbb{R}),$$
$$\mathring{A}f(x) = \frac{1}{2}\sigma f''(x) + bf'(x) + c\int_0^{\infty} [f(x+y) - f(x)]\, \mu(\mathrm{d}y),$$

for every x in $\mathbb{R}$. Again, notice the nonlocal character of this unbounded operator. □

Example 5.6 (reflecting barrier)**.** On the state space $\mathbb{R}_0^+ = [0,\infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,B) = \frac{1}{\sqrt{2\pi t}}\bigg(\int_B \Big\{\exp\big[-\frac{(y-x)^2}{2t}\big] + \exp\big[-\frac{(y+x)^2}{2t}\big]\Big\}\mathrm{d}y\bigg),$$

for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$. This represents Brownian motion with reflecting barrier at $x = 0$ and the process itself can be constructed as $(X_t = |x + W_t| : t \geq 0)$, where $(W_t : t \geq 0)$ is a standard Wiener process in $\mathbb{R}$. Its associated semigroup in $C_0(\mathbb{R}_0^+)$ is given by

$$S(t)f(x) = \int_{\mathbb{R}_0^+} f(y)\, p(t,x,\mathrm{d}y) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \check{f}(y)\, \exp\big[-\frac{(y-x)^2}{2t}\big]\mathrm{d}y,$$

where $\check{f}(y) = f(y)$ if $y \geq 0$ and $\check{f}(y) = f(-y)$ if $y \leq 0$, for every $t > 0$ and $x \geq 0$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) = \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = 0\},$$
$$Af = \frac{1}{2}f''.$$

Certainly, a constant drift b can be added so that $Af = \frac{1}{2}f'' + bf'$ and $X_t = |x + bt + W_t|$. □

The reflected Brownian motion of above Example 5.6 can also be constructed by means of local time as follows. First, for an given $x \geq 0$ we define τ_x the first

exit time of the open region $(0,\infty)$, i.e., $\tau_x = \inf\{t \geq 0 : x + W_t \geq 0\}$. Next, we look at its running maximum, i.e., $M_t = \max\{x + W_s : \tau_x \leq s \leq t\}$, which except for a factor $1/2$ is called the *local time* of $(x + W_t : t \geq 0)$ at the origin. It can be proved that the process $(|x + W_t| : t \geq 0)$ has the same law as the process $(x + M_t - W_t : t \geq 0)$, which gives another realization of the reflected Brownian motion.

Example 5.7 (absorbing barrier)**.** On the state space $\mathbb{R}^+ = (0,\infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,B) = \frac{1}{\sqrt{2\pi t}}\Big(\int_B \Big\{ \exp\big[-\frac{(y-x)^2}{2t}\big] - \exp\big[-\frac{(y+x)^2}{2t}\big] \Big\} \mathrm{d}y \Big),$$

for any $t > 0$, $x > 0$ and B in $\mathcal{B}$. This represents Brownian motion with absorbing barrier at 0, i.e., the Brownian motion particle *dies* at the first time when it hits the boundary $\{0\}$. The process itself can be constructed by stopping (or killing) the process $x + W_t$ at the first instant τ_x when it hits the boundary $\{0\}$, where W_t is a standard Wiener process in $\mathbb{R}$, i.e.,

$$\tau_x = \inf\{t > 0 : x + W_t = 0\}, \qquad X_t = x + W_t \quad t < \tau_x,\ x > 0,$$

thus $(X_t : t \geq 0)$ is the Brownian motion with initial value x at time $t = 0$ and absorbed (or otherwise said killed) at the origin. τ_x is the lifetime of the process X_t. Often we introduce an extra point (indicated by $\triangle$, or ∞ or ∂) to the state space $\mathbb{R}^+ = (0,\infty)$, called the *coffin* state, and defines X_t for all times by $X_t = \triangle$ for $t \geq \tau_x$. Its associated semigroup in $C_0(\mathbb{R}^+)$ is given by

$$S(t)f(x) = \int_{\mathbb{R}^+} f(y)\, p(t,x,\mathrm{d}y) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \hat{f}(y)\, \exp\big[-\frac{(y-x)^2}{2t}\big] \mathrm{d}y,$$

where $\hat{f}(y) = f(y)$ if $y \geq 0$ and $\hat{f}(y) = -f(-y)$ if $y \leq 0$, for every $t > 0$ and $x \geq 0$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) = \{f \in C_0(\mathbb{R}^+) \cap C^2(\mathbb{R}^+) : f'' \in C_0(\mathbb{R}^+)\},$$
$$Af = \frac{1}{2} f''.$$

A constant drift b can be added so that $Af = \frac{1}{2}f'' + bf'$. Notice that $p(t,x,\mathbb{R}^+) < 1$ so that the process X_t dies, i.e., it does hit the boundary $x = 0$ in a finite time. The semigroup $S(t)$ may be extended to the space of continuous and bounded function in $[0,\infty)$, where $S(t)\mathbb{1} = 0$ for all $t > 0$. Moreover, the Banach space $C_0(\mathbb{R}^+)$ includes a vanishing boundary condition at infinity and also at $x = 0$, so that actually the condition $f(0) = 0$ in the definition of $\mathcal{D}(A)$ is redundant. Generally, we look at this as a process in the closure $[0,\infty)$ and we use the Banach space $C_0([0,\infty[)$ instead of $C_0(]0,\infty[)$. □

Sometimes we may use the complementary error function

$$\mathrm{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-v^2} \mathrm{d}v, \tag{5.117}$$

which satisfies $\mathrm{Erfc}(0) = 1$ and

$$\frac{2}{x+\sqrt{x^2+2}} \le \sqrt{\pi}\, e^{x^2}\, \mathrm{Erfc}(x) \le \frac{2}{x+\sqrt{x^2+1}}. \tag{5.118}$$

Indeed, by considering the functions

$$f(x) = \frac{1}{x+\sqrt{x^2+1}} - \mathrm{e}^{x^2}\int_x^\infty \mathrm{e}^{-y^2}\mathrm{d}y,$$
$$g(x) = \mathrm{e}^{x^2}\int_x^\infty \mathrm{e}^{-y^2}\mathrm{d}y - \frac{1}{x+\sqrt{x^2+2}},$$

which satisfy $f(0) = 1 - \frac{\pi}{2} > 0$ and $g(0) = \frac{\pi}{2} - \frac{1}{2} > 0$, we can estimate

$$\mathrm{e}^{x^2}\int_x^\infty \mathrm{e}^{-y^2}\mathrm{d}y \le \frac{2}{x}\,\mathrm{e}^{x^2}\int_x^\infty 2y\,\mathrm{e}^{-y^2}\mathrm{d}y = \frac{1}{x},$$

so that $|f(x)| \le \frac{2}{x}$ and $|g(x)| \le \frac{2}{x}$. Calculations show that $f'(x) - 2xf(x) \le 0$ and $g'(x) - 2xg(x) \le 0$, for any $x \ge 0$, and the desired estimate (5.118) follows.

Example 5.8 (sticking barrier)**.** On the state space $\mathbb{R}_0^+ = [0,\infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,B) = \frac{1}{\sqrt{2\pi t}}\Big(\int_B \Big\{\exp\big[-\frac{(y-x)^2}{2t}\big] - \exp\big[-\frac{(y+x)^2}{2t}\big]\Big\}\mathrm{d}y\Big) + \\ + \ \mathrm{Erfc}\Big(\frac{x}{\sqrt{2\pi t}}\Big)\ \mathbb{1}_B(0),$$

for any $t > 0$, $x \ge 0$ and B in $\mathcal{B}$. This represents Brownian motion with sticking barrier at $x = 0$, i.e., when the Brownian motion particle reaches $x = 0$ for the first time, it sticks there forever. The infinitesimal generator of its associated semigroup in $C_0(\mathbb{R}_0^+)$ is the differential operator

$$\mathcal{D}(A) = \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f''(0) = 0\},$$
$$Af = \frac{1}{2}f''.$$

A constant drift b can be added so that $Af = \frac{1}{2}f'' + bf'$. Notice that comparing with the previous Example 5.7, now the state space $\mathbb{R}_0^+$ includes the barrier $x = 0$ and $p(t,x,\mathbb{R}_0^+) = 1$ for any $t > 0$ and x in $\mathbb{R}_0^+$. The semigroup takes the form

$$S(t)f(x) = \int_{\mathbb{R}_0^+} f(y)\,p(t,x,\mathrm{d}y) = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^\infty \hat{f}(y)\,\exp\big[-\frac{(y-x)^2}{2t}\big]\mathrm{d}y,$$

for every x in $\mathbb{R}_0^+$ and $t > 0$. Notice that the function $y \mapsto \hat{f}(y)$, where $\hat{f}(y) = f(y)$ if $y \ge 0$ and $\hat{f}(y) = 2f(0) - f(-y)$ if $y < 0$, is continuously differentiable whenever f is so. Thus, the function $x \mapsto S(t)f(x)$ can be defined as a smooth

function, for every x in $\mathbb{R}$. The process itself can be constructed by stopping the process $x + W_t$ at the first instant τ_x when it hits the boundary $\{0\}$, where W_t is a standard Wiener process in $\mathbb{R}$, i.e.,

$$\tau = \inf\{t > 0 : x + W_t = 0\}, \qquad X_t = x + W_{t\wedge\tau_x} \quad t \geq 0,\ x > 0,$$

thus $(X_t : t \geq 0)$ is the Brownian motion with initial value x at time $t = 0$ and stopped at the origin. □

We may combine the reflecting barrier Example 5.6 and this sticking barrier to get a process where the domain of the infinitesimal generator $\mathcal{D}(A)$ has a boundary condition of the form $f'(0) - cf''(0) = 0$ with a positive constant c instead of just $f''(0) = 0$. This is called *sticky* barrier. Similarly, we may combine the reflecting barrier Example 5.6 with the absorbing barrier to get a process where the domain of the infinitesimal generator $\mathcal{D}(A)$ has a boundary condition of the form $f'(0) - cf(0) = 0$ with a positive constant c. This is called *elastic* barrier. The construction of a sticky Brownian motion or an elastic Brownian motion is more delicate, it starts with the reflecting Brownian motion and its local time at the barrier, the reader is referred to the classic book by Itô and McKean [78] for a complete analysis.

Example 5.9 (sticky Wiener)**.** On the state space $\mathbb{R}_0^+ = [0, +\infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}}\left(\exp\Big[-\frac{(y-x)^2}{2t}\Big] - \exp\Big[-\frac{(y+x)^2}{2t}\Big]\right) + \\ + \delta(y)\, e^{\frac{t+2cx}{2c^2}}\operatorname{Erfc}\Big(\frac{t+cx}{c\sqrt{2t}}\Big) + \frac{1}{c}\, e^{\frac{t+2c(x+y)}{2c^2}}\operatorname{Erfc}\Big(\frac{t+c(x+y)}{c\sqrt{2t}}\Big),$$

for any $t > 0$, $x \geq 0$, $y \geq 0$. This represents a *slowly* reflecting Brownian motion on $[0, +\infty)$, i.e., when the Brownian motion particle reaches $x = 0$, it sticks there for some time. The infinitesimal generator of its associated semigroup in $C_0(\mathbb{R}_0^+)$ is the differential operator

$$\mathcal{D}(A) = \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = cf''(0)\},$$
$$Af = \frac{1}{2}f''.$$

We can visualize this process as a Brownian motion with a suitable time change, more specifically, starting with a standard Wiener process $x + W_t$ we have $X_t = x + W_{\ell_x(t)}$, where

$$\ell_x(t) = \inf\{s : A_s \leq t\}, \qquad A_t = \int_0^t \mathbb{1}_{\{x+W_s>0\}}\, \mathrm{d}s + c\, M_t,$$

with $M_t = \max\{x + W_s : \tau_x \leq s \leq t\}$ and $\tau_x = \inf\{t \geq 0 : x + W_t \geq 0\}$. □

Example 5.10 (elastic Wiener)**.** On the state space $\mathbb{R}_0^+ = [0, +\infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}}\Big(\exp\Big[-\frac{(y-x)^2}{2t}\Big] + \exp\Big[-\frac{(y+x)^2}{2t}\Big]\Big) - \\ - c\ \exp\Big(c(x+y) + \frac{c^2 t}{2}\Big)\ \mathrm{Erfc}\Big(\frac{x+y+ct}{\sqrt{2t}}\Big),$$

for any $t > 0$, $x \geq 0$, $y \geq 0$. This represents reflecting Brownian motion on $[0, +\infty)$ killed elastically at $x = 0$. The infinitesimal generator of its associated semigroup in $C_0(\mathbb{R}_0^+)$ is the differential operator

$$\mathcal{D}(A) = \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = cf(0)\},$$
$$Af = \frac{1}{2} f''.$$

We can visualize this process as a reflected Brownian motion killed at a random time r, where $P(\{r > t\} \cap B) = \mathbb{E}\{\mathbb{1}_B \exp(-c\tau(t))\}$, B is any Borel set of $C(\mathbb{R}_0^+)$ and $\tau(t)$ is the local time of the Wiener process. □

Example 5.11 (doubly reflected Wiener)**.** We consider now a Brownian motion with state space $[0, a]$. where a is a positive real number. On the state space $[0, a]$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,B) = \sum_{k\in\mathbb{Z}} \frac{1}{\sqrt{2\pi t}}\Big(\int_B \Big\{ \exp\big[-\frac{(2ka + y - x)^2}{2t}\big] + \\ + \exp\big[-\frac{(2ka + y + x)^2}{2t}\big]\Big\} \mathrm{d}y\Big),$$

for any $t > 0$, $0 \leq x \leq a$ and B in $\mathcal{B}$, where $\mathbb{Z}$ is the set of all integer numbers. This represents Brownian motion with reflecting barrier at $x = 0$ and at $x = a$. The process itself can be constructed as $(X_t = \varphi_a(x + W_t) : t \geq 0)$, where $(W_t : t \geq 0)$ is a standard Wiener process in $\mathbb{R}$ and φ_a is the function $x \mapsto \min\{(2a - x), x\}$ for x in $[0, 2a]$ and extended to $\mathbb{R}$ as a continuous periodic function with $2a$-period, so that φ_a maps $\mathbb{R}$ onto $[0, a]$. The associated semigroup in $C([0, a])$ is given by

$$S(t)f(x) = \int_0^a f(y)\, p(t,x,\mathrm{d}y) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f[\varphi_a(y)]\ \exp\big[-\frac{(y-x)^2}{2t}\big]\mathrm{d}y,$$

for every $t > 0$ and $0 \leq x \leq a$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) = \{f \in C^2([0, a]) : f'(0) = 0,\ f'(a) = 0\},$$
$$Af = \frac{1}{2} f''.$$

A constant drift b can be added so that $Af = \frac{1}{2} f'' + bf'$ and $X_t = \varphi_a(x + bt + W_t)$. Notice that $p(t, x, [0, a]) = 1$ for any $t > 0$ and x in $[0, a]$. □

Example 5.12 (doubly absorbed Wiener)**.** We consider now a Brownian motion with state space $(0, a)$, where a is a positive real number. On the state space $(0, a)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, B) = \sum_{k\in\mathbb{Z}} \frac{1}{\sqrt{2\pi t}} \Big(\int_B \Big\{ \exp\big[-\frac{(2ka + y - x)^2}{2t}\big] - \\ - \exp\big[-\frac{(2ka + y + x)^2}{2t}\big] \Big\} \mathrm{d}y \Big),$$

for any $t > 0$, $0 \le x \le a$ and B in $\mathcal{B}$, where $\mathbb{Z}$ is the set of all integer numbers. This represents Brownian motion with absorbing barrier at $x = 0$ and at $x = a$, i.e., the Brownian motion particle *dies* at the first time when it hits the boundary $x = 0$ or $x = a$. The process itself can be constructed as $(X_t = x + W_{t\wedge\tau} : t \ge 0)$, where $(W_t : t \ge 0)$ is a standard Wiener process in $\mathbb{R}$ and τ is the first exit time from the open set $(0, a)$ for the process $x + W_t$, i.e.,

$$\tau = \inf\{t > 0 : x + W_t \le 0 \ \text{ or } \ x + W_t \ge a\}, \quad t \ge 0, \ x > 0,$$

The associated semigroup in $C_0(]0, a[)$ is given by

$$S(t)f(x) = \int_{\mathbb{R}_0^+} f(y)\, p(t, x, \mathrm{d}y) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \hat{f}_a(y) \, \exp\big[-\frac{(y - x)^2}{2t}\big] \mathrm{d}y,$$

for every $t > 0$ and $0 \le x \le a$, where now $\hat{f}_a(y) = f[\varphi_a(y)]$ if $y \ge 0$ and $\hat{f}_a(y) = -f[\varphi_a(y)]$ if $y \le 0$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) = \{f \in C_0(]0, a[) \cap C^2(]0, a[) : f'' \in C_0(]0, a[), \ f(0) = 0, \ f(a) = 0\},$$
$$Af = \frac{1}{2} f''.$$

Technically, the points 0 and a does not belong to the domain of definition of functions f in the Banach space $C_0(]0, a[)$, but we identify $C_0(]0, a[)$ with the subspace of $C([0, a]) = C_0([0, a])$ satisfying $f(0) = f(a) = 0$. Again notice that $p(t, x,]0, a[) < 1$ and that the semigroup $S(t)$ may be considered as defined on the Banach $C([0, a])$ where $S(t)\mathbb{1} = 0$ for all $t > 0$, so that the state of the process could be regarded as $[0, a]$. A constant drift b can be added so that $Af = \frac{1}{2} f'' + bf'$. and $X_t = x + bt \wedge \tau + W_{t\wedge\tau})$, where τ is now the first exit time from the open set $(0, a)$ for the process $(bt + W_t : t \ge 0)$. □

Some details on Brownian motion on a finite interval relative to Examples 5.11 and 5.12 can be found in Karatzas and Shreve [91, Section 2.8.c, pp. 97–100].

Example 5.13 (periodic Wiener)**.** We consider now a Brownian motion with state space $[0, a]$. where a is a positive real number. On the state space $[0, a]$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, B) = \sum_{k\in\mathbb{Z}} \frac{1}{\sqrt{2\pi t}} \Big(\int_B \Big\{ \exp\big[-\frac{(ka + y - x)^2}{2t}\big] \Big\} \mathrm{d}y \Big),$$

for any $t > 0$, $0 \le x \le a$ and B in $\mathcal{B}$, where $\mathbb{Z}$ is the set of all integer numbers. This represents Brownian motion on a circle (the interval $[0, a]$, with 0 and a identified). The process itself can be constructed as $(X_t = \psi_a(x + W_t) : t \ge 0)$, where $(W_t : t \ge 0)$ is a standard Wiener process in $\mathbb{R}$ and $\psi_a(x) = x - a[x/a]$ (where $[x]$ denotes the integral part of x), which maps $\mathbb{R}$ onto $[0, a]$. The associated semigroup in $C_\#(0, a) = \{f \in C([0, a]) : f(0) = f(a)\}$ is given by

$$S(t)f(x) = \int_0^a f(y)\, p(t, x, \mathrm{d}y) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f[\psi_a(y)] \exp\big[- \frac{(y-x)^2}{2t}\big] \mathrm{d}y,$$

for every $t > 0$ and $0 \le x \le a$. The infinitesimal generator is a closed extension of the differential operator $Af = \frac{1}{2} f''$ with domain

$$\mathcal{D}(A) = \{f \in C_\#(0, a) \cap C^2([0, a]) :\ f''(0) = f''(a)\}.$$

Note that even if ψ_a is not continuous, the composition $x \mapsto f[\psi_a(x)]$ is continuously differentiable for any continuous function f in the domain $\mathcal{D}(A)$. Also, a constant drift b can be added so that $Af = \frac{1}{2} f'' + bf'$ and $X_t = \psi_a(x + bt + W_t)$. □

More generally, we may consider a Sturm-Liouville problem in $[0, a]$ of the form

$$\frac{1}{2} u'' + \lambda u = 0, \text{ in } (0, a), \tag{5.119}$$

$$\alpha_0 u(0) - \beta_0 u'(0) = \alpha_a u(a) + \beta_a u'(a) = 0, \tag{5.120}$$

where $\alpha_0, \beta_0, \alpha_a, \beta_a \ge 0$, $\alpha_0 + \beta_0 > 0$ and $\alpha_a + \beta_a > 0$. A periodic condition of the form $u(0) - u(a) = u'(0) - u'(a) = 0$ can also be used. There is a sequence of (positive) eigenvalues $(0 < \lambda_0 < \lambda_1 < \cdots)$, $\lambda_n \to \infty$ as $n \to \infty$, with its corresponding eigenfunctions $(u_0, u_1, \ldots)$, satisfying the boundary value problem (5.119) and form an orthonormal basis in $L^2(0, a)$. Certainly, u_n is a linear combination of $\sin(x\sqrt{2\lambda_n})$ and $\cos(x\sqrt{2\lambda_n})$ and if $a = \pi$ then when $\alpha_0\alpha_a = \beta_0\beta_a = 0$ and $\alpha_0\beta_a \ne \alpha_a\beta_0$ we have $2\sqrt{2\lambda_n} = 2n + 1$ and when $\alpha_0\beta_a = \alpha_a\beta_0$ and $\alpha_0\alpha_a \ne 0$ or $\beta_0\beta_a \ne 0$ we have $\sqrt{2\lambda_n} = n$. In the case of periodic boundary conditions, if $a = 2\pi$ then $\sqrt{2\lambda_n} = n$. Define

$$p(t, x, B) = \sum_{n=0}^{\infty} \mathrm{e}^{-t\lambda_n} u_n(x) \int_B u_n(y)\mathrm{d}y, \tag{5.121}$$

for every x in $[0, a]$, $t > 0$ and B in $\mathcal{B}$. The maximum principle ensures that p in nonnegative, i.e, $0 \le p(t, x, B) \le 1$, for every t, x, B. Parseval equality yields

$$\int_0^a |p(t, x, B)|^2 \mathrm{d}x = \sum_{n=0}^{\infty} \mathrm{e}^{-t\lambda_n} |B|, \quad \forall x \in [0, a],\ t > 0,\ B \in \mathcal{B},$$

where $|B|$ denotes the Lebesgue measure of the Borel set B. Some more details are needed to discuss the convergence of the series (5.121), which is the eigenvalues and eigenfunctions expansion of the Green function or Green operator relative to the boundary value problem (5.119). From here, the associate semigroup and the stochastic process can be constructed. The interested reader may consult the pioneer paper Feller [47] related to parabolic differential equations and the associated semigroups.

Example 5.14 (reflecting Wiener-Poisson)**.** On the state space $\mathbb{R}_0^+ = [0, \infty)$ with its Borel σ-algebra $\mathcal{B}$ and for a given positive constant c, we consider

$$p(t,x,B) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{2\pi t\, k!} \Big(\int_B \Big\{ \exp\big[-\frac{(x+k-y)^2}{2t} \big] + \\ + \exp\big[-\frac{(x+k+y)^2}{2t} \big] \Big\} \mathrm{d}y \Big),$$

for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$. This is a Wiener-Poisson process with reflecting barrier at $x = 0$ and the process itself can be constructed as $(X_t = |x + W_t + P_t| : t \geq 0)$, where $(W_t : t \geq 0)$ is a Wiener process independent of the Poisson process $(P_t : t \geq 0)$, both in in $\mathbb{R}$. Its associated semigroup in $C_0(\mathbb{R}_0^+)$ is given by

$$S(t)f(x) = \int_{\mathbb{R}} f(y)\, p(t,x,\mathrm{d}y) = \\ = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{\sqrt{2\pi}\, k!} \int_{-\infty}^{\infty} \check{f}(x+k+\sqrt{t}z)\, \exp\big(-\frac{z^2}{2}\big) \mathrm{d}z,$$

where $\check{f}(y) = f(y)$ if $y \geq 0$ and $\check{f}(y) = f(-y)$ if $y \leq 0$, for every $t > 0$ and $x \geq 0$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) = \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = 0\},$$
$$Af(x) = \frac{1}{2} f''(x) + c[f(x+1) - f(x)], \quad \forall x \in \mathbb{R}_0^+.$$

Only a cad-lag realization can be constructed. We can generalize this example to a $(W_t : t \geq 0)$ Wiener process (with drift b and covariance σ) and a $(P_t : t \geq 0)$ compound Poisson processes (with parameters (c, μ)), independent of each other and with reflecting barrier at $x = 0$. The compound process is indeed a *subordinator*, i.e., increasing in t so that all jumps of the sum process $(x + W_t + P_t : t \geq 0)$ are inside the real semi-line $[0, \infty)$. Thus $c > 0$ and μ is a probability distribution on $(0, \infty)$. Again, a cad-lag realization is given by $(X_t = |x + W_t + P_t| : t \geq 0)$ and the probability transition function is

$$p(t,x,B) = \int_B \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{2\pi t\, k!}\, p_k(t,x,y) \mathrm{d}y,$$

$$p_k(t,x,y) = \frac{1}{\sqrt{2\pi t}}\Big(\int_0^\infty \Big\{ \exp\big[-\frac{(x-y-z)^2}{2t}\big] + \\ + \exp\big[-\frac{(x+y-z)^2}{2t}\big] \Big\} \mu^k(\mathrm{d}z) \Big),$$
$$\mu^0 = \mu, \quad \mu^k(B) = \int_{\mathbb{R}\times\mathbb{R}} \mathbb{1}_B(y+z)\, \mu^{k-1}(\mathrm{d}y)\, \mu(\mathrm{d}z),$$

for every $k = 1, 2, \dots$, for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$. Here some work is necessary to ensure the proper convergence of the above series. Again notice that $\mu^k = \mu^{k-1} \star \mu$ is the k convolution of μ. Its associated semigroup in $C_0(\mathbb{R}_0^+)$ is given by

$$S(t)f(x) = \int_{\mathbb{R}} f(y)\, p(t,x,\mathrm{d}y) = \mathrm{e}^{-ct} \sum_{k=0}^\infty \frac{(ct)^k}{\sqrt{2\pi}\, k!} \times \\ \times \int_0^\infty \mu^k(\mathrm{d}y) \int_{-\infty}^\infty \check{f}(x+y+\sqrt{t}z)\, \exp\big(-\frac{z^2}{2}\big)\mathrm{d}z,$$

for every $t > 0$ and $x \geq 0$. The infinitesimal generator is the integro-differential operator

$$\mathcal{D}(A) = \{ f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = 0 \},$$
$$Af(x) = \frac{1}{2}\sigma f''(x) + bf'(x) + c\int_0^\infty [f(x+y) - f(x)]\, \mu(\mathrm{d}y),$$

for every x in $\mathbb{R}_0^+$. If the compound Poisson process P_t has the parameter-distribution μ in the whole space $\mathbb{R}$ then the sum process $x + W_t + P_t$ may have a jumps outside of the semi-line $[0,\infty)$. In this case, we may keep the expression $(X_t = |x + W_t + P_t| : t \geq 0)$ and make appropriated modifications. For instance the semigroup takes the form

$$S(t)f(x) = \mathrm{e}^{-ct} \sum_{k=0}^\infty \frac{(ct)^k}{\sqrt{2\pi}\, k!} \times \\ \times \int_{\mathbb{R}} \mu^k(\mathrm{d}y) \int_{-\infty}^\infty \check{f}(x+y+\sqrt{t}z)\, \exp\big(-\frac{z^2}{2}\big)\mathrm{d}z,$$

for every $t > 0$ and x in $\mathbb{R}$ and $\check{f}$ as above, however, the boundary condition for the domain of the infinitesimal generator needs more work. It is clear that absorbing and sticking barriers can be considered for Wiener-Poisson processes by means of the expression with $\hat{f}$ and the stopping argument. □

As in Examples 5.7 and 5.8, we can discuss absorbing and sticking barriers for Wiener-Poisson processes by means of arguments similar to Example 5.14. This is on the space either $\mathbb{R}^+ = (0,\infty)$ or $\mathbb{R}_0^+ = [0,\infty)$ with its Borel σ-algebra $\mathcal{B}$ and for a given positive constant c, we consider

$$p(t,x,B) = \mathrm{e}^{-ct} \sum_{k=0}^\infty \frac{(ct)^k}{2\pi t\, k!}\, p_i(t, x+k, B), \tag{5.122}$$

for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$, where $p_i(t, x, B)$ is the transition function for absorbing barrier with $i = 1$ or for sticking barrier with $i = 2$, as in previous examples. Notice that

$$p_2(t, x, B) = p_1(t, x, B) + [1 - p_1(t, x, \mathbb{R})]\mathbb{1}_B(0).$$

In the case of the Wiener-Poisson process, the boundary condition for the absorbing barrier is clearly $f(0) = 0$. However, for the sticking barrier boundary condition is

$$\frac{1}{2}f''(0) + c[f(0+1) - f(0)] = 0,$$

i.e., the equation is satisfied up to the boundary.

Also the case of a Wiener-Poisson process with periodic conditions can be easier studied, e.g., a Wiener-Poisson process in $\mathbb{R}$ is combined with the operation modulo $[0, a]$, ψ_a, as in Example 5.13, which maps $\mathbb{R}$ into $[0, a]$.

Trying to extend the doubly reflected Wiener in an interval, Example 5.11, to a Wiener-Poisson process, we encounter a new difficulty, we may *jump outside the interval*. This forces us to make a decision on the jumps, e.g., a natural extension or reflection. This is a more delicate issue. For instance, if we want the reflection on an interval $[0, a]$, first we make a periodic condition on $[-a, a]$ and then we take the absolute value. However, if we want a natural extension, first we make a constant and continuous extension outside of the given interval $[0, a]$ and then we use the process in the whole line.

For instance, the reader may consult the books Mandl [117] for a comprehensive treatment of one-dimensional Markov processes. On the other hand, several examples (without jumps) can be found in Borodin and Salminen [18, Appendix 1, pp. 102-119]

5.5.2 Multi-Dimensional

In the whole space $\mathbb{R}^d$, $d \geq 2$ we have more difficulties. A central role is played by the *Gauss kernel* Γ_0 defined by

$$\Gamma_0(x, t, a) = \frac{1}{(2\pi t)^{d/2}\sqrt{\det a}} \exp\big(-\frac{x \cdot a^{-1}x}{2t}\big), \tag{5.123}$$

for every $t > 0$ and x in $\mathbb{R}^d$, where a is an invertible symmetric nonnegative $d \times d$-matrix, if we write by components $a = (a_{ij})$ then its inverse $a^{-1} = (a^{ij})$ and $x \cdot a^{-1}x = \sum_{ij} a^{ij}x_i x_j$. When a is the identity matrix $\mathbb{1}$ we write $\Gamma_0(x, t) = \Gamma_0(x, t, \mathbb{1})$. In analysis the constant $1/2$ is replaced by $1/4$ and called *heat kernel*. This function is infinitely differentiable in all its arguments and in particular, for any derivative ∂^ℓ, with $\ell = (\ell_1, \ell_2, \dots, \ell_d, \ell_t, \ell_a)$ and $|\ell| = \ell_1 + \ell_2 + \cdots + \ell_d + 2\ell_t$ we have

$$|\partial^\ell \Gamma_0(x, t, a)| \leq Ct^{-(d+|\ell|)/2} \exp\big(-c\frac{|x|^2}{t}\big), \tag{5.124}$$

for every $t > 0$ and x in $\mathbb{R}^d$, for some positive constants $C = C(d, \ell, \delta)$ and $c = c(d, \ell, \delta)$, where the symmetric matrix a satisfies $\delta|\xi|^2 \le \xi \cdot a\xi \le |\xi|^2/\delta$ for any ξ in $\mathbb{R}^d$, for some $\delta > 0$. Also we have

$$\int_{\mathbb{R}^d} \partial^\ell \Gamma_0(x, t, a)\, \mathrm{d}x = \begin{cases} 1 & \text{if } \ell = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.125}$$

This $\Gamma_0(x, t, a)$ is the probability density transition function of a Wiener process in $\mathbb{R}^d$, with zero mean and co-variance a. The corresponding resolvent kernel is given by

$$R(\lambda, x) = \frac{1}{(2\pi t)^{-d/2}\sqrt{\det a}} \int_0^\infty \exp\big(-\frac{x \cdot a^{-1}x}{2t} - \lambda t\big)\mathrm{d}t = \\ = \frac{2}{(2\pi t)^{-d/2}} \Big(\frac{2\lambda}{x \cdot a^{-1}x}\Big)^{(d/4-1/2)} \mathrm{K}_{d/2-1}\big(\sqrt{2\lambda x \cdot a^{-1}x}\big),$$

for every $\lambda > 0$ and x in $\mathbb{R}^d$, where K_ν is the modified Bessel function of 2nd kind. In particular,

$$\mathrm{K}_{n-1/2}(z) = \sqrt{\frac{\pi}{2z}} z^n \Big(-\frac{1}{z}\frac{\mathrm{d}}{\mathrm{d}z}\Big)^n e^z, \quad n = 0, 1, \dots,$$

and so

$$R(\lambda, x) = \frac{1}{4\pi|x|} \exp(-\sqrt{2\lambda}|x|),$$

for $d = 3$ and $a = \mathbb{1}$, the identity matrix.

Example 5.15 (d-dimensional Wiener)**.** A Wiener process with vector mean b and co-variance matrix a has a transition probability function on the state space space $\mathbb{R}^d$ with its Borel σ-algebra $\mathcal{B}$ defined by

$$p(t, x, B) = \int_B \Gamma_0(x - bt - y, t, a)\mathrm{d}y,$$

for every x in $\mathbb{R}^d$, $t > 0$ and B in $\mathcal{B}$, where $\Gamma_0(x, t, a)$ is the Gauss kernel (5.123). Notice that if $(W_t : t \ge 0)$ is a standard Wiener process starting at the origin, i.e., with $W_0 = 0$, zero mean and co-variance matrix $\mathbb{1}$ or equivalently $p(t, x, \mathrm{d}y)\Gamma(x - \mathrm{d}y, t)$ as its transition probability function, then the process $X_t = x + bt + \sqrt{a}W_t$ is a realization of the above Wiener process starting at x. Also this can be constructed as the product of d independent one dimensional Brownian motions, i.e., the probability transition *density* function $\Gamma_0(x, t)$ is the product of d similar one dimensional expressions $\Gamma_0(x_i, t)$ as the one used in Example 5.1. The associated semigroup in $C_0(\mathbb{R}^d)$ is given by

$$S(t)f(x) = \int_{\mathbb{R}^d} f(y)\, p(t, x, \mathrm{d}y) = \int_{\mathbb{R}^d} f(x + bt + \sqrt{ta}\, z)\, \Gamma_0(z, 1)\, \mathrm{d}z,$$

for every $t > 0$ and x in $\mathbb{R}^d$. Its infinitesimal generator A is the closure of the (closable) differential operator $\mathring{A}$

$$\mathcal{D}(\mathring{A}) = C_0^0(\mathbb{R}^d) \cap C^2(\mathbb{R}^d),$$
$$\mathring{A}f = \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}\, \partial_{ij}^2 f + \sum_{i=1}^{d} b_i\, \partial_i f,$$

where a_{ij} and b_i are the entries of the matrix a and the vector b. □

Example 5.16 (Ornstein-Uhlenbeck)**.** This is a modification of a Wiener process where a linear drift is added. Two matrices a and b describe the process X, namely,

$$X_t(x) = \mathrm{e}^{bt}x + \int_0^t \sqrt{a}\, \mathrm{e}^{b(t-s)}\, \mathrm{d}W_t,$$

where $(W_t : t \geq 0)$ is a standard Wiener process. The process $X_t(x)$ has a Gaussian distribution with mean $\mathrm{e}^{bt}x$ and covariance

$$q_t = \int_0^t \mathrm{e}^{bs} a \mathrm{e}^{b^* s} \mathrm{d}s, \quad t > 0,$$

where b^* is the adjoint matrix. Thus, the transition probability function of the Ornstein-Uhlenbeck process $(X_t(x) : t \geq 0)$ on state space $\mathbb{R}^d$ with its Borel σ-algebra $\mathcal{B}$ is given by

$$p(t, x, B) = \int_B \Gamma_0(\mathrm{e}^{bt}x - y, 1, q_t^{-1})\, \mathrm{d}y,$$

for every x in $\mathbb{R}^d$, $t > 0$ and B in $\mathcal{B}$, where $\Gamma_0(x, t, a)$ is the Gauss kernel (5.123). The associated semigroup in $C_0(\mathbb{R}^d)$ is given by

$$S(t)f(x) = \int_{\mathbb{R}^d} f(y)\, p(t, x, \mathrm{d}y) = \int_{\mathbb{R}^d} f(\mathrm{e}^{bt}x + q_t\, z)\, \Gamma_0(z, 1)\, \mathrm{d}z,$$

for every $t > 0$ and x in $\mathbb{R}^d$. Its infinitesimal generator is the closure of the (closable) differential operator $\mathring{A}$

$$\mathcal{D}(\mathring{A}) = C_0^0(\mathbb{R}^d) \cap C^2(\mathbb{R}^d),$$
$$\mathring{A}f = \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}\, \partial_{ij}^2 f + \sum_{i,j=1}^{d} b_{ij} x_j\, \partial_i f,$$

where a_{ij} and b_{ij} are the entries of the matrices a and b. □

Example 5.17 (compound Poisson)**.** A compound poisson process with parameter $c > 0$ and μ, where μ is a distribution in $\mathbb{R}_*^d = \mathbb{R}^d \setminus \{0\}$ has a transition

probability function on the state space space $\mathbb{R}^d$ with its Borel σ-algebra $\mathcal{B}$ defined by

$$p(t,x,B) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} \int_{\mathbb{R}^d_*} \mathbb{1}_B(x+y)\, \mu^k(\mathrm{d}y),$$
$$\mu^0 = \mu, \quad \mu^k(B) = \int_{\mathbb{R}^d_* \times \mathbb{R}^d_*} \mathbb{1}_B(y+z)\, \mu^{k-1}(\mathrm{d}y)\, \mu(\mathrm{d}z),$$

for every $k = 1, 2, \dots$, for any $t > 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$. The probability measures $\mu^k = \mu^{k-1} \star \mu$ are called the k-convolution of μ. Based on two sequences of independent identically distributes random variables with exponential and μ distribution, a canonical realization of the compound Poisson process can be constructed. The associated semigroup in $C_0(\mathbb{R}^d)$ is given by

$$S(t)f(x) = \int_{\mathbb{R}^d} f(y)\, p(t,x,\mathrm{d}y) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} \int_{\mathbb{R}^d_*} f(x+y)\, \mu^k(\mathrm{d}y),$$

for every $t > 0$ and x in $\mathbb{R}^d$. Its infinitesimal generator is the integral operator

$$\mathcal{D}(A) = C_0(\mathbb{R}^d), \qquad Af(x) = c \int_{\mathbb{R}^d_*} [f(x+y) - f(x)]\, \mu(\mathrm{d}y),$$

which is clearly a nonlocal operator. If the distribution μ has support in an open semi-space $\mathbb{R}^d_+ = \mathbb{R}^{d-1} \times (0, \infty)$ then we may consider the compound Poisson process only in $\mathbb{R}^d_+$, which is called *subordinator* in the one dimensional case. □

It is clear that we may mix Examples 5.15 and 5.17 to produce a d-dimensional Wiener-Poisson process with probability density transition function defined by

$$\Gamma(x,t) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} (\Gamma_0 \star \mu^k)(x,t),$$
$$\mu^0 = \delta_0, \qquad \mu^k(B) = \int_{\mathbb{R}^d_* \times \mathbb{R}^d_*} \mathbb{1}_B(x+y)\, \mu(\mathrm{d}x)\, \mu^{k-1}(\mathrm{d}y), \tag{5.126}$$
$$(\Gamma_0 \star \mu^k)(x,t) = \int_{\mathbb{R}^d_*} \Gamma_0(x-y,t)\, \mu^k(\mathrm{d}y),$$

for every x in $\mathbb{R}^d$ and $t > 0$, where δ_0 is the Dirac measure at the origin. Since μ is a probability measure on $\mathbb{R}^d_*$, so is μ^k and the above series is clearly convergent. The infinitesimal generator takes the form

$$Af(x) = \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}\, \partial^2_{ij} f(x) + \sum_{i=1}^{d} b_i\, \partial_i f(x) + \\ + c \int_{\mathbb{R}^d_*} [f(x+y) - f(x)]\, \mu(\mathrm{d}y), \tag{5.127}$$

which is a second order integro-differential (non-local) operator.

Example 5.18 (Neumann)**.** This is a half-space normal reflecting barrier, i.e., on the semi-space $\bar{\mathbb{R}}^d_+ = \mathbb{R}^{d-1}\times[0,\infty)$, with the notation $x = (\tilde{x}, x_d)$, we consider the function

$$G^N_0(\tilde{x}, x_d, t, \xi_d) = \Gamma_0(\tilde{x}, x_d - \xi_d, t) + \Gamma_0(\tilde{x}, x_d + \xi_d, t),$$

for every $t > 0$, $x_d, \xi_d \geq 0$, and $\tilde{x}$ in $\mathbb{R}^{d-1}$. As in Example 5.6, we may define a transition probability function on the state space $\bar{\mathbb{R}}^d_+$ with its Borel σ-algebra $\mathcal{B}$

$$p(t, x, B) = \int_B G^N_0(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d)\,\mathrm{d}\xi,$$

for any $t > 0$, x in $\bar{\mathbb{R}}^d_+$ and B in $\mathcal{B}$. The arguments are the same, even the construction of the (standard) *normal reflected* Wiener process in a d-dimensional half-space is simple, since this is a $(d-1)$-dimensional Wiener process and an independent one-dimensional Wiener process with reflecting barrier at $x_d = 0$. Expressions for the associated semigroup and its infinitesimal generator can be obtained, e.g.,

$$\mathcal{D}(A) = \{f \in C_0(\bar{\mathbb{R}}^d_+) : \Delta f \in C_0(\bar{\mathbb{R}}^d_+),\ \partial_d f(\tilde{x}, 0) = 0\},$$
$$Af = \frac{1}{2}\Delta f,$$

where Δ is the usual Laplacian operator $\sum_{i=1}^d \partial_i^2$, here in the sense of Schwartz distribution derivative. Except for the $1/2$ factor, the local time correspond to the so-called *Poisson kernel* which is $P_0(t, x, \tilde{\xi}) = -2\,\Gamma_0(\tilde{x} - \tilde{\xi}, x_d, t)$, for any $t > 0$, $x = (\tilde{x}, x_d)$ in $\mathbb{R}^{d-1} \times (0,\infty)$ and $\tilde{\xi}$ in $\mathbb{R}^{d-1}$. □

Example 5.19 (Dirichlet)**.** This is a half-space normal reflecting barrier, i.e., on the semi-space $\mathbb{R}^d_+ = \mathbb{R}^{d-1} \times (0,\infty)$, with the notation $x = (\tilde{x}, x_d)$, we consider the function

$$G^D_0(\tilde{x}, x_d, t, \xi_d) = \Gamma_0(\tilde{x}, x_d - \xi_d, t) - \Gamma_0(\tilde{x}, x_d + \xi_d, t),$$

for every $t > 0, x_d, \xi_d > 0$, and $\tilde{x}$ in $\mathbb{R}^{d-1}$. As in Example 5.7, we may define a transition function on the state space $\mathbb{R}^d_+$ with its Borel σ-algebra $\mathcal{B}$

$$p(t, x, B) = \int_B G^D_0(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d)\,\mathrm{d}\xi,$$

for any $t > 0$, x in $\mathbb{R}^d_+$ and B in $\mathcal{B}$. The arguments are the same, even the construction of the (standard) *stopped* Wiener process in d-dimensional half-space is simple, since this is a $(d-1)$-dimensional Wiener process and an independent one dimensional Wiener process with absorbing barrier at $x_d = 0$. If the barrier $x_d = 0$ is of some interest, then we may proceed as in Example 5.8 and convert p into a probability transition function. To that effect, we note that

$$G^D_0(\tilde{x}, x_d, t, \xi_d) = \frac{1}{\sqrt{2\pi t}}\Big\{\exp\Big[-\frac{(x_d - \xi_d)^2}{2t}\Big] - \exp\Big[-\frac{(x_d + \xi_d)^2}{2t}\Big]\Big\}\Gamma_{d-1}(\tilde{x}, t),$$

where $\Gamma_{d-1}(\tilde{x},t)$ has the same expression (5.123) with the identity matrix $a=\mathbb{1}$ in dimension $(d-1)$. Then we define

$$p(t,x,B)=\int_B\Big\{\frac{1}{\sqrt{2\pi t}}\Big(\exp\big[-\frac{(x_d-\xi_d)^2}{2t}\big]-\exp\big[-\frac{(x_d+\xi_d)^2}{2t}\big]\Big)+ \\ +\Big[1-\frac{1}{\sqrt{2\pi t}}\int_{-x_d}^{x_d}\exp\big(-\frac{z^2}{2t}\big)\mathrm{d}z\Big]\mathbb{1}_B(\tilde{\xi},0)\Big\}\,\Gamma_{d-1}(\tilde{x}-\tilde{\xi},t)\mathrm{d}\xi,$$

for any $t>0$, x in $\bar{\mathbb{R}}^d_+$ and B in $\mathcal{B}$. This yields the (standard) *sticking* Wiener process in a d-dimensional half-space. Expressions for the associated semigroup and its infinitesimal generator are obtained immediately from the one-dimensional case. □

A reflected Wiener process with zero mean and co-variance matrix a in a d-dimensional half-space presents more difficulties. After a rotation of coordinates, we can reduce the general case of a Wiener process with zero mean and covariance matrix a to the case where a is the identity matrix, but the boundary condition is a *oblique* reflection, i.e., instead of the condition $\partial_d f(\tilde{x},0)=0$ on the domain of the infinitesimal generator $A=\Delta/2$ we have $b\cdot\nabla f(\tilde{x},0)=0$, where ∇ is the gradient operator in the first d-dimensional variable, i.e., x, and $b=(b_1,\dots,b_d)$ is a vector with $b_d>0$. For the boundary value problem

$$\begin{aligned}&\partial_t u-\tfrac{1}{2}\Delta u(x,t)=0,\quad \forall x\in\mathbb{R}^d_+,\ t>0,\\ &u(x,0)=0,\quad \forall x\in\mathbb{R}^d_+,\\ &b\cdot\nabla u(x,t)=\psi(x,t),\qquad\qquad \forall x\in\partial\mathbb{R}^d_+,\ t>0,\end{aligned}\tag{5.128}$$

where $\partial\mathbb{R}^d_+=\mathbb{R}^{d-1}\times\{0\}$, we can calculate the Poisson kernel P_0^b as

$$P_0^b(\tilde{x},x_d,t)=\varphi_b(\tilde{x},x_d,t)\,\Gamma_0(\tilde{x},x_d,t),\tag{5.129}$$

and

$$\varphi_b(\tilde{x},x_d,t)=-\frac{1}{|b|^2}\Big\{b_d+ \\ +\frac{|b|^2x_d-b_d(b\cdot x)}{|b|\sqrt{t/2}}\exp\Big[\frac{(b\cdot x)^2}{2t|b|^2}\Big]\int_{(b\cdot x)/|b|\sqrt{2t}}^{+\infty}\mathrm{e}^{-r^2}dr\Big\},$$

for any $t>0$, $x=(\tilde{x},x_d)$ in $\mathbb{R}^d_+$. Actually, we use $P_0^b(\tilde{x}-\tilde{\xi},x_d,t)$ with $\tilde{\xi}$ in $\partial\mathbb{R}^d_+$.

Example 5.20 (oblique)**.** This is a half-space oblique reflecting barrier in the direction of the vector $b=(b_1,\dots,b_d)$ with $b_d>0$. On the semi-space $\bar{\mathbb{R}}^d_+=\mathbb{R}^{d-1}\times[0,\infty)$, with the notation $x=(\tilde{x},x_d)$, we consider the function

$$G_0^b(\tilde{x},x_d,t,\xi_d)=\Gamma_0(\tilde{x},x_d-\xi_d,t)-\Gamma_0(\tilde{x},x_d+\xi_d,t)- \\ -2\,b_dP_0^b(x,x_d+\xi_d,t),$$

for every $t > 0$, $x_d, \xi_d \geq 0$, and $\tilde{x}$ in $\mathbb{R}^{d-1}$. This yields a transition probability function on the state space $\bar{\mathbb{R}}^d_+$ with its Borel σ-algebra $\mathcal{B}$

$$p(t, x, B) = \int_B G^b_0(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d) \mathrm{d}\xi,$$

for any $t > 0$, x in $\bar{\mathbb{R}}^d_+$ and B in $\mathcal{B}$. This is not a product of $(d-1)$ independent Brownian motions in $\mathbb{R}$ with an independent reflected Brownian motion in $[0, \infty)$, certainly, the function φ_b in (5.129) makes the coupling. Expressions for the associated semigroup and its infinitesimal generator can be obtained, e.g.,

$$\mathcal{D}(A) = \{f \in C_0(\bar{\mathbb{R}}^d_+) : \Delta f \in C_0(\bar{\mathbb{R}}^d_+),\ b \cdot \nabla f(\tilde{x}, 0) = 0\},$$
$$Af = \frac{1}{2}\Delta f,$$

but a realization of a d-dimensional (standard) Wiener process in $\bar{\mathbb{R}}^d_+$ with oblique reflection at the barrier $x_d = 0$ is obtained from general existence theorems based on the above transition probability function. □

The fact that an explicit expression can be found for the above transition function give specific estimates allowing the construct Green function for variable coefficients and integro-differential operators, the reader may consult the books Garroni and Menaldi [58, 59]. The case of a normal reflected Wiener-Poisson process can be treated as in the one dimensional case Example 5.14, however, the oblique reflection needs another method. Let us consider the case of an integro-differential operator of the form $A = \Delta/2 + I$, where

$$I\varphi(x) = c \int_{\mathbb{R}^d_+} [\varphi(x + y) - \varphi(x)]\, \mu(\mathrm{d}y), \quad \forall x \in \mathbb{R}^d_+, \tag{5.130}$$

where μ is now a probability measure in $\bar{\mathbb{R}}^d_+$ and $c > 0$. We define

$$G^b = G^b_0 + G^b_1 + \cdots + G^b_k + \cdots, \qquad G^b_k = G^b_0 \star IG^b_{k-1}, \tag{5.131}$$

where I is considered acting on the first d-dimensional variables, i.e., for any fixed $t > 0$, $\xi_d \geq 0$

$$IG^b_k(\tilde{x}, x, t, \xi_d) = c \int_{\mathbb{R}^d_+} [G^b_k(x + y, t, \xi_d) - G^b_k(x, t, \xi_d)]\, \mu(\mathrm{d}y),$$

for every $x = (\tilde{x}, x_d)$ in $\mathbb{R}^d_+$, for any $k \geq 0$, and the *kernel-convolution*

$$(\varphi \star \psi)(\tilde{x}, x_d, t, \xi_d) = \\ = \int_0^t ds \int_{\mathbb{R}^d_+} \varphi(\tilde{x} - \tilde{y}, x_d, t - s, y_d)\, \psi(\tilde{y}, y_d, s, \xi_d) \mathrm{d}y, \tag{5.132}$$

for $\tilde{x}$ in $\mathbb{R}^{d-1}$, $x_d, \xi_d \geq 0$ and $t > 0$.

The semigroup property or Chapman-Kolmogorov identity for the transition function G_0^b, namely

$$G_0^b(\tilde{x}, x_d, t+s, \xi_d) = \int_{\mathbb{R}^d_+} G_0^b(\tilde{x}-\tilde{y}, x_d, t, y_d)\, G_0^b(\tilde{y}, y_d, s, \xi_d)\mathrm{d}y,$$

for every $\tilde{x}$ in $\mathbb{R}^{d-1}$, $x_d, \xi_d \geq 0$ and $s, t > 0$, and the explicit form of the function G_0^b given in Example 5.20 yield the identity $G_k(x,t,\xi_d) = (t^k/k!)\, I^k G_0^b(x,t,\xi_d)$, where I^k is the k-iteration of the integral operator I.

Certainly, we use the technique of Section 5.4 to check in what sense the above series (5.131) converges. First we define the Green space $\mathcal{G}_k^0$ of continuous kernels $\varphi(x,t,\xi_d)$ for x in $\mathbb{R}^d_+$, $t > 0$ and $\xi_d \geq 0$ such that

$$|\varphi(x,t,\xi_d)| \leq C_0\, t^{-1+k-d/2}, \quad \forall x, t, \xi_d, \tag{5.133}$$

$$\int_{\mathbb{R}^d_+} \big[|\varphi(\tilde{y}-\tilde{\xi}, y_d, t, \xi_d)| + |\varphi(\tilde{x}-\tilde{y}, x_d, t, y_d)|\big]\mathrm{d}y \leq K_0\, t^{-1+k/2}, \quad \forall x, t, \xi,$$

for some constants C_0 and K_0, and the infimum of all such constants, denoted by $C(\varphi,k)$ and $K(\varphi,k)$, are semi-norms for $k > 0$.

It is easy to check that I maps the Green space $\mathcal{G}_k^0$ into itself,

$$C(I\varphi, k) \leq 2c\, C(\varphi,k) \quad \text{and} \quad K(I\varphi,k) \leq 2c\, K(\varphi,k), \tag{5.134}$$

for every φ, k, and that G_0^b belongs to $\mathcal{G}_2^0$ in view of (5.124), which is valid for G_0^b instead of Γ_0.

Therefore, G_k^b belongs to $\mathcal{G}_{2k+2}^0$ and

$$C(\partial^\ell G_k^b, 2k+2-|\ell|) \leq \frac{(2c)^k}{k!}\, C(\partial^\ell G_0^b, 2-|\ell|), \tag{5.135}$$

$$K(\partial^\ell G_k^b, 2k+2-|\ell|) \leq \frac{(2c)^k}{k!}\, K(\partial^\ell G_0^b, 2-|\ell|), \tag{5.136}$$

for every $k = 1, 2, \dots$ and ∂^ℓ, with $\ell = (\ell_1, \ell_2, \dots, \ell_d, \ell_t, \ell_a)$ and $|\ell| = \ell_1 + \ell_2 + \cdots + \ell_d + 2\ell_t$. Because of the identity $G_k(x,t,\xi_d) = (t^k/k!)\, I^k G_0^b(x,t,\xi_d)$, the expression $G_k^b = G_0^b \star I G_{k-1}^b$ is not really used, not integration in the time variable is needed in this explicit case and the semi-norms (5.133) are meaningful even for $k \leq 0$. Recall that $c > 0$ is the constant used in the definition of the integral operator I in (5.130).

At this point we have proved that the remainder of the series (5.131) defining $\partial^\ell G^b$, i.e., $\partial^\ell G_k^b + \partial^\ell G_{k+1}^b + \cdots$ converges in the Green space $\mathcal{G}_{2k+2-|\ell|}^0$, for $\ell \geq 0$, so that G^b is infinitely many time differential in all its arguments. Moreover, G^b satisfies the Volterra equation $G^b = G_0^b + G_0^b \star I G^b$, the Chapman-Kolmogorov identity, and

$$\int_{\mathbb{R}^d_+} G^b(\tilde{x}-\tilde{\xi}, x_d, t, \xi_d)\, d\xi = \int_{\mathbb{R}^d_+} G_0^b(\tilde{x}-\tilde{\xi}, x_d, t, \xi_d)\mathrm{d}\xi = 1,$$

since IG_k^b has means zero for any $k \geq 1$.

To complete this explicit calculation, we denote by Δ_0 the Dirac measure at the origin to have

$$I^k\varphi(x) = \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-1} \frac{(c\,t)^k}{k!} \int_{\mathbb{R}^d_+} \varphi(x+y)\, \bar{\mu}^k(\mathrm{d}y),$$

$$\mu^0 = \delta_0, \qquad \mu^k(B) = \int_{\mathbb{R}^d_+ \times \mathbb{R}^d_+} \mathbb{1}_B(x+y)\, \mu(\mathrm{d}x)\, \mu^{k-1}(\mathrm{d}y),$$

for every $k \geq 1$, which implies

$$G^b(x,t,\xi_d) = \sum_{k=0}^{\infty} \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-1} \frac{(c\,t)^k}{k!}\, \mu^k G_0^b(x,t,\xi_d),$$

$$\mu^k G_0^b(x,t,\xi_d) = \int_{\mathbb{R}^d_+} G_0^b(x+y,t,\xi_d)\, \mu^k(\mathrm{d}y),$$

and interchanging the order of the summation we obtain

$$G^b(x,t,\xi_d) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(c\,t)^k}{k!}\, \mu^k G_0^b(x,t,\xi_d),$$

$$\mu^0 = \delta_0, \quad \mu^k(B) = \int_{\mathbb{R}^d_+ \times \mathbb{R}^d_+} \mathbb{1}_B(x+y)\, \mu(\mathrm{d}x)\, \mu^{k-1}(\mathrm{d}y), \tag{5.137}$$

$$\mu^k G_0^b(x,t,\xi_d) = \int_{\mathbb{R}^d_+} G_0^b(x+y,t,\xi_d)\, \mu^k(\mathrm{d}y),$$

for every $k \geq 1$, for any x in $\bar{\mathbb{R}}^d_+$, $\xi_d \geq 0$ and $t > 0$. Since μ is a probability measure on $\bar{\mathbb{R}}^d_+$, so is μ^k and the above series is clearly convergent as the initial one given by (5.131). These arguments complement the one dimensional examples.

Example 5.21 (oblique Wiener-Poisson)**.** This is a half-space oblique reflecting barrier in the direction of the vector $b = (b_1, \ldots, b_d)$ with $b_d > 0$, for a standard Wiener process in $\mathbb{R}^d$ and a compound Poisson process with parameters $c > 0$ and μ, where μ is a distribution on the open semi-space $\mathbb{R}^d_+ = \mathbb{R}^{d-1} \times (0,\infty)$, with the notation $x = (\tilde{x}, x_d)$. In the state space $\bar{\mathbb{R}}^d_+ = \mathbb{R}^{d-1} \times [0,\infty)$, the closed semi-space, we consider the function G^b defined by (5.131) or (5.137). This yields a transition probability function on the state space $\bar{\mathbb{R}}^d_+$ with its Borel σ-algebra $\mathcal{B}$

$$p(t,x,B) = \int_B G^b(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d)\mathrm{d}\xi,$$

for any $t > 0$, x in $\bar{\mathbb{R}}^d_+$ and B in $\mathcal{B}$. Expressions for the associated semigroup and its infinitesimal generator can be obtained, e.g.,

$$\mathcal{D}(A) = \{f \in C_0(\bar{\mathbb{R}}^d_+) : \Delta f \in C_0(\bar{\mathbb{R}}^d_+),\ b \cdot \nabla f(\tilde{x}, 0) = 0\},$$

$$Af = \frac{1}{2}\Delta f + If,$$

where the integral operator I is given by (5.130). A realization of a d-dimensional (standard) Wiener-Poisson process in $\mathbb{R}^d_+$ with parameter $c > 0$ and μ, and oblique reflection at the barrier $x_d = 0$ is obtained from general existence theorems based on the above transition probability function. □

It is possible to use an integral operator I of the form

$$I\varphi(x) = \int_{\mathbb{R}^d_+} [\varphi(x+y) - \varphi(x) - y \cdot \nabla\varphi(x)]\, m(\mathrm{d}y),$$

$$\forall x \in \mathbb{R}^d_+, \quad \text{with} \quad \int_{\mathbb{R}^d_+} \frac{|y|^2}{1+|y|}\, m(\mathrm{d}y) < \infty. \quad (5.138)$$

The definition (5.131) of G^b still valid but not (5.137). Because of the *constant coefficients* we can make explicit calculations and $G^b_k = G^b_0 \star IG^b_{k-1} = (t^k/k!)I^k G^b_0$ but we need to work harder to show the convergence of the series. For instance, if we assume

$$\int_{\mathbb{R}^d_+} \frac{|y|^{2-\alpha}}{1+|y|}\, m(\mathrm{d}y) < \infty, \quad \alpha \in (0,2],$$

then the integral operator I maps the Green space $\mathcal{G}^2_k$ (kernel φ satisfying condition (5.133) for $\partial_\ell\varphi$ of order $k - |\ell|$, with $|\ell| \le 2$) into the Green space $\mathcal{G}^0_{k+\alpha}$, with appropriate estimates, see previous Section 5.4 and the books Garroni and Menaldi [58, 59] for details.

On the other hand, the spectral theory of compact operators can be used to give an eigenvalues and eigenfunction expansion of the Green function or Green operator as in the Sturm-Liouville case.

Notation

Some Common Uses:

$\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$: natural, rational, real and complex numbers.

$\mathtt{i}$, $\Re(\cdot)$, I: imaginary unit, the real part of complex number and the identity (or inclusion) mapping or operator.

P, $\mathbb{E}\{\cdot\}$: for a given measurable space $(\Omega, \mathcal{F})$, P denotes a probability measure and $\mathbb{E}\{\cdot\}$ the expectation (or integration) with respect to P. As customary in probability, the random variable ω in Ω is seldom used in a explicit notation, this is understood from the context.

$\mathcal{F}(t)$, $\mathcal{F}_t$, $\mathcal{B}(t)$, $\mathcal{B}_t$: usually denote a family increasing in t of σ-algebra (also called σ-fields) of a measurable space $(\Omega, \mathcal{F})$. If $\{x_t : t \in T\}$ is a family of random variables (i.e., measurable functions) then $\sigma(x_t : t \in T)$ usually denotes the σ-algebra generated by $\{x_t : t \in T\}$, i.e., the smallest sub σ-algebra of $\mathcal{F}$ such that each function $\omega \to x_t(\omega)$ is measurable. Usually $\mathbb{F}$ denotes the family of σ-algebras $\{\mathcal{F}(t) : t \in T\}$, which is referred to as a filtration.

$X(t)$, X_t, $x(t)$, x_t: usually denote the same process in some probability space $(\Omega, \mathcal{F}, P)$. One should understand from the context when we refer to the value of the process (i.e., a random variable) or to the generic function definition of the process itself.

$\mathbb{1}_A$: usually denotes the characteristic function of a set A, i.e., $\mathbb{1}_A(x) = 1$ if x belongs to A and $\mathbb{1}_A(x) = 0$ otherwise. Sometimes the set A is given as a condition on a function τ, e.g., $\tau < t$, in this case $\mathbb{1}_{\tau<t}(\omega) = 1$ if $\tau(\omega) < t$ and $\mathbb{1}_{\tau<t}(\omega) = 0$ otherwise.

δ: most of the times this is the δ function or Dirac measure. Sometimes one write $\delta_x(\mathrm{d}y)$ to indicate the integration variable y and the mass concentrated at x. On certain occasions, δ denotes the jumps operator, defined be $\delta X(0) = 0$ and $\delta X = X(t+) - X(t-)$, $t > 0$, any process X without discontinuity of the second kind.

$\mathrm{d}\mu$, $\mu(\mathrm{d}x)$, $\mathrm{d}\mu(x)$: together with the integration sign, usually these expressions denote integration with respect to the measure μ. Most of the times $\mathrm{d}x$

means integration respect to the Lebesgue measure in the variable x, as understood from the context.

E^T, $\mathcal{B}(E^T)$, $\mathcal{B}^T(E)$: for E a Hausdorff topological (usually a separable complete metric, i.e., Polish) space and T a set of indexes, usually this denotes the product topology, i.e., E^T is the space of all function from T into E and if T is countable then E^T is the space of all sequences of elements in E. As expected, $\mathcal{B}(E^T)$ is the σ-algebra of E^T generated by the product topology in E^T, but $\mathcal{B}^T(E)$ is the product σ-algebra of $\mathcal{B}(E)$ or generated by the so-called cylinder sets. In general $\mathcal{B}^T(E) \subset \mathcal{B}(E^T)$ and the inclusion may be strict.

$C([0,\infty),\mathbb{R}^d)$ **or** $D([0,\infty),\mathbb{R}^d)$ canonical sample spaces of continuous or cadlag (continuous from the right having left-hand limit) and functions, with the locally uniform or the Skorokhod topology, respectively. Sometimes the notation $\mathbb{C}_d$ or $C([0,\infty[,\mathbb{R}^d)$ or $\mathbb{D}_d$ or $D([0,\infty[,\mathbb{R}^d)$ could be used.

Most Commonly Used Function Spaces:

$C(X)$: for X a Hausdorff topological (usually a separable complete metric, i.e., Polish) space, this is the space of real-valued (or complex-valued) continuous functions on X. If X is a compact space then this space endowed with sup-norm is a separable Banach (complete normed vector) space. Sometimes this space may be denoted by $C^0(X)$, $C(X,\mathbb{R})$ or $C(X,\mathbb{C})$ depending on what is to be emphasized.

$C_b(X)$: for X a Hausdorff topological (usually a complete separable metric, i.e., Polish) space, this is the Banach space of real-valued (or complex-valued) continuous and bounded functions on X, with the sup-norm.

$C_0(X)$: for X a *locally compact* (but not compact) Hausdorff topological (usually a complete separable metric, i.e., Polish) space, this is the separable Banach space of real-valued (or complex-valued) continuous functions vanishing at infinity on X, i.e., a continuous function f belongs to $C_0(X)$ if for every $\varepsilon > 0$ there exists a compact subset $K = K_\varepsilon$ of X such that $|f(x)| \leq \varepsilon$ for every x in $X \setminus K$. This is a proper subspace of $C_b(X)$ with the sup-norm.

$C_0(X)$: for X a *compact* subset of a locally compact Hausdorff topological (usually a Polish) space, this is the separable Banach space of real-valued (or complex-valued) continuous functions vanishing on the boundary of X, with the sup-norm. In particular, if $X = X_0 \cup \{\infty\}$ is the one-point compactification of X_0 then the boundary of X is only $\{\infty\}$ and $C_0(X) = C_0(X_0)$ via the zero-extension identification.

$C_0(X)$, $C_0^0(X)$: for X a proper *open* subset of a locally compact Hausdorff topological (usually a Polish) space, this is the separable Fréchet (complete

locally convex vector) space of real-valued (or complex-valued) continuous functions with a compact support X, with the inductive topology of uniformly convergence on compact subset of X. When necessary, this Fréchet space may be denoted by $C_0^0(X)$ to stress the difference with the Banach space $C_0(X)$, when X is also regarded as a locally compact Hausdorff topological. Usually, the context determines whether the symbol represents the Fréchet or the Banach space.

$C_b^k(E)$, $C_0^k(E)$**:** for E a *domain* in the Euclidean space $\mathbb{R}^d$ (i.e, the closure of the interior of E is equal to the closure of E) and k a nonnegative integer, this is the subspace of either $C_b(E)$ or $C_0^0(E)$ of functions f such that all derivatives up to the order k belong to either $C_b(E)$ or $C_0^0(E)$, with the natural norm or semi-norms. For instance, if E is open then $C_0^k(E)$ is a separable Fréchet space with the inductive topology of uniformly convergence (of the function and all derivatives up to the order k included) on compact subset of E. If E is closed then $C_b^k(E)$ is the separable Banach space with the sup-norm for the function and all derivatives up to the order k included. Clearly, this is extended to the case $k = \infty$.

$B(X)$**:** for X a Hausdorff topological (mainly a Polish) space, this is the Banach space of real-valued (or complex-valued) Borel measurable and bounded functions on X, with the sup-norm. Note that $\mathcal{B}(X)$ denotes the σ-algebra of Borel subsets of X, i.e., the smaller σ-algebra containing all open sets in X, e.g., $B(\mathbb{R}^d)$, $\mathcal{B}(\mathbb{R}^d)$, or $B(E)$, $\mathcal{B}(E)$ for a Borel subset E of d-dimensional Euclidean space $\mathbb{R}^d$.

$L^p(X, m)$**:** for $(X, \mathcal{X}, m)$ a complete σ-finite measure space and $1 \leq p < \infty$, this is the separable Banach space of real-valued (or complex-valued) $\mathcal{X}$-measurable (class) functions f on X such that $|f|^p$ is m-integrable, with the natural p-norm. If $p = 2$ this is also a Hilbert space. Usually, X is also a locally compact Polish space and m is a Radon measure, i.e., finite on compact sets. Moreover $L^\infty(X, m)$ is the space of all (class of) m-essentially bounded (i.e., bounded except in a set of zero m-measure) with essential-sup norm.

$L^p(\mathcal{O})$, $H_0^m(\mathcal{O})$, $H^m(\mathcal{O})$**:** for $\mathcal{O}$ an open subset of $\mathbb{R}^d$, $1 \leq p \leq \infty$ and $m = 1, 2, \ldots$, these are the classic Lebesgue and Sobolev spaces. Sometimes we may use vector-valued functions, e.g., $L^p(\mathcal{O}, \mathbb{R}^n)$.

$\mathcal{D}(\mathcal{O})$, $\mathcal{S}(\mathbb{R}^d)$, $\mathcal{D}'(\mathcal{O})$, $\mathcal{S}'(\mathbb{R}^d)$**:** for $\mathcal{O}$ an open subset of $\mathbb{R}^d$, these are the classic test functions (C^∞ functions with either compact support in $\mathcal{O}$ or rapidly decreasing in $\mathbb{R}^d$) and their dual spaces of distributions. These are separable Fréchet spaces with the inductive topology. Moreover, $\mathcal{S}(\mathbb{R}^d) = \cap_m H^m(\mathbb{R}^d)$ is a countable Hilbertian nuclear space. Thus its dual space $\mathcal{S}'(\mathbb{R}^d) = \cup_m H^{-m}(\mathbb{R}^d)$, where $H^{-m}(\mathbb{R}^d)$ is the dual space of $H^m(\mathbb{R}^d)$. Sometimes we may use vector-valued functions, e.g., $S(\mathbb{R}^d, \mathbb{R}^n)$.

Bibliography

[1] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, Cambridge, 2004. 62, 66, 185, 210, 331, 353, 406

[2] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkäuser, Boston, 1997. x

[3] R.F. Bass. *Stochastic Processes*. Cambridge University Press, Cambridge, 2011. 57

[4] R. Bellman. *Dynamic programming*. Princeton University Press, Princeton, NJ, 2010. Reprint of the 1957 edition, With a new introduction by Stuart Dreyfus. 332

[5] A. Bensoussan. *Dynamic programming and inventory control*. IOS Press, Amsterdam, 2011. 55

[6] A. Bensoussan and J.L. Lions. *Impulse Control and Quasi-Variational Inequalities*. Gauthier-Villars, Paris, 1984. (pages are referred to the French edition). 200, 385, 388

[7] J. Bertoin. *Lévy processes*. Cambridge University Press, Cambridge, 1996. 66, 69, 406

[8] M. Bertoldi and L. Lorenzi. *Analytical Methods for Markov Semigroups*. Chapman & Hall / CRC Press, Boca Raton (FL), 2007. 331, 374

[9] D.P. Bertsekas. *Dynamic programming and Stochastic Control*. Prentice Hall Inc., Englewood Cliffs, NJ, 1987. Deterministic and stochastic models. 55, 332

[10] D.P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, 2nd, two-volume set edition, 2001. x

[11] K. Bichteler. *Stochastic Integration with jumps*. Cambridge University Press, Cambridge, 2002. 31, 70, 108, 123, 127, 135, 167, 185, 216

[12] K. Bichteler, J.B. Gravereaux, and J. Jacod. *Malliavin Calculus for Processes with Jumps*. Gordon and Breach Science Publishers, New York, 1987. 119

[13] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968. 29, 30, 35

[14] P. Billingsley. *Probability and Measure*. Wiley, New York, 2nd edition, 1986. 23

[15] R.M. Blumenthal and R.K. Getoor. *Markov Processes and Potential Theory*. Academic Press, New York, 1968. 331, 348, 351, 353, 355, 371

[16] A. Bobrowski. *Functional analysis for probability and stochastic processes*. Cambridge University Press, Cambridge, 2005. 57

[17] J-M. Bony, P. Courrège, and P. Priouret. Semi-groupes de Feller sur une variété à bord compacte et problèmes aux limites intégro-différentiels du second ordre donnant lieu au principe du maximum. *Ann. Inst. Fourier (Grenoble)*, 18(fasc. 2):369–521 (1969), 1968. 385

[18] A. N. Borodin and P. Salminen. *Handbook of Brownian Motion - Facts and Formulae*. Birkäuser, Boston, 1996. 57, 416

[19] P. Bremaud. *Point Processes and Queues*. Springer-Verlag, New York, 1981. 51, 69, 97, 134

[20] P. Bremaud. *Markov Chains*. Springer-Verlag, New York, 1998. 53, 55, 97

[21] T.M. Chao and C.S. Chou. Anticipative stochastic integrals equations driven by semimartingales. *Stochastic Anal. Appl.*, 20(4):755–782, 2002. 238

[22] T. S. Chihara. *An Introduction to Orthogonal Polynomials*. Gordon and Breach, New York, 1978. 117

[23] K.L. Chung. *Markov Chains with Stationary Transition Probabilities*. Springer-Verlag, New York, 2nd edition, 1980. 51, 53, 54

[24] K.L. Chung. *From Markov Processes to Brownian Motion*. Springer-Verlag, New York, 1982. 353

[25] K.L. Chung and R.J. Williams. *Introduction to Stochastic Integration*. Birkhäuser, Boston, 2nd edition, 1990. 70, 135, 168, 170, 176, 188, 190, 201

[26] E. Çınlar. *Probability and stochastics*. Springer, New York, 2011. 133, 353

[27] G. Da Prato. *Introduction to stochastic analysis and Malliavin calculus*. Edizioni della Normale, Pisa, 2007. 135

[28] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimension*. Cambridge University Press, Cambridge, 1992. 25, 44, 105

[29] G. Da Prato and J. Zabczyk. *Ergodicity for Infinite Dimensional Systems*. Cambridge University Press, Cambridge, 1996. 67, 373, 383

[30] M.H.A. Davis. *Markov Models and Optimization*. Chapman and Hall, London, 1993. 296, 341, 343, 355, 359, 361

[31] C. Dellacherie, B. Maisonneuve, and P.A. Meyer. *Probabilités et Potentiel (XVII-XXIV), Processus de Markov (fin), Compl'ements de calcul stochastique*. Hermann, Paris, 1992. 331

[32] C. Dellacherie and P.A. Meyer. *Probabilités et Potentiel*. Hermann, Paris. Chapter I–IV, 1976, Chapters V–VIII (Théories des martingales), 1980, Chapters IX–XI (Théories discréte du potentiel), 1983, Chapters XI–XVI (Théories des processus de Markov), 1987, Chapters XVII-XXIV (with B. Maisonneuve, Processus de Markov - fin), 1992. 51, 73, 77, 79, 115, 122, 125, 151, 165, 170, 176, 201, 229, 331, 353, 367, 371, 373, 374

[33] J.L. Doob. *Stochastic Processes*. Wiley, New York, 1953. 8, 23, 51, 52, 57, 138, 144, 149, 169, 336

[34] J.L. Doob. *Classical Potential Theory and its Probabilistic Counterpart*. Springer-Verlag, New York, 1984. 22, 34, 351, 355

[35] J.L. Doob. *Measure Theory*. Springer-Verlag, New York, 1984. 51

[36] R. M. Dudley and R. Norvaisa. *Concrete functional calculus*. Springer Monographs in Mathematics. Springer-Verlag New York, 1 edition, 2011. 145

[37] R.M. Dudley. *Real Analysis and Probability*. Cambridge University Press, Cambridge, 2002. 8, 29, 70, 115, 320

[38] N. Dunford and J.T. Schwartz. *Linear Operators, Three Volumes*. Wiley, New York, 1988. 372

[39] J. Duoandikoetxea. *Fourier Analysis*. American Mathematical Society, Providence, RI, 2001. 37

[40] R. Durrett. *Stochastic Calculus. A practical introduction*. CRC Press, Boca Raton, 1996. 27, 70, 178

[41] E.B. Dynkin. *Theory of Markov Processes*. Prentice-Hall, Englewood Cliffs (NJ), 1961. 57, 353

[42] E.B. Dynkin. *Markov Processes*, volume 1 and 2. Springer-Verlag, New York, 1965. 331, 353, 355, 378

[43] R.J. Elliott. *Stochastic Calculus and Applications*. Springer-Verlag, New York, 1982. 69, 70, 122

[44] M. Errami, F. Russo, and P. Vallois. Itô's formula for $C^{1,\lambda}$-functions of a càdlàg process and related calculus. *Probab. Theory Related Fields*, 122:191–221, 2002. 238

[45] S.N. Ethier and T.G. Kurtz. *Markov Processes*. Wiley, New York, 1986. 35, 302, 316, 331, 353, 361, 371

[46] H. Federer. *Geometric Measure Theory*. Springer-Verlag, New York, 1996. 115

[47] W. Feller. The parabolic differential equations and the associated semigroups of transformations. *Ann. of Math.*, 55:468–519, 1952. 414

[48] W. Feller. *An Introduction to Probability Theory and its Applications*, volume Vols I and II. Wiley, New York, 1960 and 1966. 8, 54, 333

[49] D.L. Fisk. The parabolic differential equations and the associated semigroups of transformations. *Trans. of Am. Math. Soc.*, 120:369–389, 1967. 238

[50] W.H. Fleming and R.W. Rishel. *Deterministic and Stochastic Control Theory*. Springer-Verlag, New York, 1975. x

[51] W.H. Fleming and H.M. Soner. *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag, New York, 1992. xi

[52] G.B. Folland. *Real Analysis*. Wiley, New York, 2nd edition, 1999. 8, 372

[53] H. Föllmer. *Calcul d'Itó sans probabilités. Séminaire de Proba. XV*, volume 850 of *Lectures Notes in Mathematics*, pages 144–150. Springer-Verlag, Berlin, 1981. 238

[54] D. Freedman. *Brownian Motion and Diffusion*. Springer-Verlag, New York, 1983. 57

[55] A. Friedman. *Stochastic Differential Equations and Applications*, volume I and II. Academic Press, New York, 1975 and 1976. 67

[56] P.K. Friz and N.B. Victoir. *Multidimensional Stochastic Processes as Rough Paths*. Cambridge University Press, Cambridge, 2010. Theory and applications. 57

[57] M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet Forms and Symmetric Markov Processes*. De Gruyter, Berlin, 1994. 331

[58] M.G. Garroni and J.L. Menaldi. *Green Functions for Second Order Integral-Differential Problems*. Pitman Research Notes in Mathematics Series No 275. Longman, Harlow, 1992. 385, 395, 396, 422, 425

[59] M.G. Garroni and J.L. Menaldi. *Second Order Elliptic Integro-Differential Problems*. Research Notes in Mathematics Series No 430. Chapman & Hall / CRC Press, Boca Raton (FL), 2002. 385, 395, 396, 422, 425

[60] I.I. Gihman and A.V. Skorohod. *The Theory of Stochastic Processes*, volume I, II and III. Springer-Verlag, New York, 1974, 1975 and 1979. 57

[61] I. Gikhman and A. Skorokhod. *Introduction to the Theory of Random Process*. Dover Publications, New York, 1996. 24, 25, 26, 138, 144

[62] I.I. Gikhman and A.V. Skorokhod. *Stochastic Differential Equations*. Springer-Verlag, Berlin, 1972. 200, 385, 387

[63] I.I. Gikhman and A.V. Skorokhod. *The theory of stochastic processes. II*. Springer-Verlag, Berlin, 2004. 341

[64] B.V. Gnedenko. *The Theory of Probability*. Chelsea, New York, 1989. 8, 333, 335

[65] L. Grafakos. *Classical Fourier analysis*. Springer, New York, 2nd edition, 2008. 37

[66] L. Grafakos. *Modern Fourier analysis*. Springer, New York, 2nd edition, 2009. 37

[67] P.R. Halmos. *Measure Theory*. Springer-Verlag, New York, 1974. 8, 18

[68] S.W. He, J.G. Wang, and J.A. Yan. *Semimartingale theory and stochastic calculus*. Kexue Chubanshe (Science Press), Beijing, 1992. 79, 127, 131, 229, 279

[69] O. Hernández-Lerma and J.B. Lasserre. *Discrete-time Markov control processes*. Springer-Verlag, New York, 1996. Basic optimality criteria. 55

[70] O. Hernández-Lerma and J.B. Lasserre. *Further topics on discrete-time Markov control processes*. Springer-Verlag, New York, 1999. 55

[71] O. Hernández-Lerma and J.B. Lasserre. *Markov chains and invariant probabilities*. Birkhäuser Verlag, Basel, 2003. 55

[72] T. Hida, H. Kuo, J. Potthoff, and L. Streit. *White Noise: An Infinite-Dimensional Calculus*. Kluwer Academic Publishers, Dordrecht, 1993. 106

[73] F.S. Hillier and G.J. Lieberman. *Introduction to operations research*. Holden-Day Inc., Oakland, Calif., 3rd edition, 1980. 332

[74] H. Holden, B. Oksendal, J. Uboe, and T. Zhang. *Stochastic Partial Differential Equations*. Birkhäuser, Boston, 1996. 38

[75] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North Holland, Amsterdam, 2nd edition, 1989. 158

[76] Y. Ishikawa. *Stochastic calculus of variations for jump processes*. De Gruyter, Berlin, second edition, 2016. 119, 185

[77] K. Itô. *Stochastic Processes*. Springer-Verlag, New York, 2004. 57, 66, 406

[78] K. Itô and H.P. McKean. *Diffusion Processes and their Sample Paths*. Springer-Verlag, New York, 1965. 410

[79] S.D. Ivasišen. Green's matrices of boundary value problems for Petrovskii parabolic systems of general form, I and II. *Math. USSR Sbornik*, 42:93–144, and 461–489, 1982. 397

[80] N. Jacob. *Pseudo-Differential Operators and Markov Processes*, volume 1: Fourier Analysis and Semigroups, 2: Generators and Potential Theory, 3: Markov Processes and Applications. Imperial College Press - World Scientific Publishing, Singapur, 2001, 2002 and 2005. 331, 375, 378, 402

[81] Martin Jacobsen. *Point process theory and applications*. Birkhäuser, Boston, Boston, MA, 2006. Marked point and piecewise deterministic processes. 97

[82] J. Jacod. Semi-groupes et mesures invariantes pour les processus semi-markoviens à espace d'état quelconque. *Ann. Inst. H. Poincaré Sect. B*, 9:77–112, 1973. 341

[83] J. Jacod and P. Protter. *Probability Essentials*. Springer-Verlag, New York, 2nd edition, 2004. 22

[84] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York, 1987. 32, 73, 78, 120, 122, 127, 129, 134, 158, 159, 160, 161, 170, 171, 176, 201, 204, 210, 211, 229, 278, 308, 309, 316, 328

[85] A. Jakubowski. A non-Skorohod topology on the Skorohod space. *Electronic J. Probab.*, 2:1–21, 1997. 321, 322, 326

[86] A. Jakubowski, J. Mémin, and G. Pagès. Convergence en loi des suites d'intégrales stochastiques sur l'espace ^{1} de skorokhod. *Probab. Theory Related Fields*, 81:111–137, 1989. 325

[87] O. Kallenberg. *Random Measures*. Akademie-Verlag, Berlin, 1983. 127

[88] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, New York, 2nd edition, 2002. 8, 21, 37, 77, 78, 80, 207, 214, 229

[89] G. Kallianpur and R.L. Karandikar. *White Noise Theory of Prediction, Filtering and Smothing*. Gordon and Breach, London, 1988. 106

[90] G. Kallianpur and J. Xiong. *Stochastic Differential Equations in Infinite Dimensional Spaces*, volume 26 of *IMS Lecture Notes-Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1995. 46, 108, 143

[91] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 2nd edition, 1991. 21, 25, 33, 73, 85, 179, 190, 412

[92] S. Karlin and H.M. Taylor. *A First Course in Stochastic Processes*. Academic Press, New York, 1975. 51, 57

[93] S. Karlin and H.M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, New York, 2nd edition, 1981. 51, 57

[94] A.F. Karr. *Probability*. Springer-Verlag, New York, 1993. 47

[95] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*. Springer-Verlag, New York, 1976. 53

[96] R.Z. Khasminskii. *Stochastic Stability of Differential Equations*. Sijthoff and Noordhoff, The Netherlands, 1980. 384

[97] D. Khoshnevisan. *Multiparameter processes*. Springer-Verlag, New York, 2002. An introduction to random fields. 35

[98] L. Kleinrock. *Queueing Systems*, volume 1 and 2. Wiley, New York, 1975. 53

[99] V.N. Kolokoltsov. *Semiclasical Analysis for Diffusions and Stochastic Processes*, volume 1724 of *Lecture Notes in Mathematics*. Springer-Varlag, Berlin, 2000. 402

[100] T. Komatsu. Markov processes associated with certain integro-differential operator. *Osaka J. Math.*, 10:271–303, 1973. 402

[101] T. Komatsu. Uniform estimates for fundamental solutions associated with non-local Dirichlet form. *Osaka J. Math.*, 32:833–860, 1995. 402

[102] N.V. Krylov. *Introduction to the Theory of Diffusion Processes*. American Mathematics Society, Providence, Rhode Island, 1995. 67, 138, 144

[103] N.V. Krylov. *Introduction to the Theory of Random Processes*. American Mathematics Society, Providence, Rhode Island, 1995. 64

[104] H. Kunita. *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge, 1990. 35, 275, 280, 294, 302

[105] H. Kunita. Stochastic differential equations based on Lévy processes and stochastic flows of diffeomorphisms. In M.M. Rao, editor, *Real and stochastic analysis*, Trends Math., pages 305–373. Birkhäuser Boston, Boston, MA, 2004. 280, 294, 302

[106] H.H. Kuo. *White Noise Distribution Theory*. CRC Press, Boca Raton, 1996. 106

[107] H.H. Kuo. *Introduction to Stochastic Integration*. Springer-Verlag, New York, 2006. 70, 135

[108] J. Lamperti. *Stochastic Processes. A survey of the mathematical theory*. Springer-Verlag, New York, 1977. 57

[109] W. Laukajtys and L. Slominski. Penalization methods for reflecting stochastic differential equations with jumps. *Stochastics and Stochastics Reports*, 75:275–293, 2003. 322

[110] T.M. Liggett. *Continuous time Markov processes*. American Mathematical Society, Providence, R.I., 2010. 331, 353

[111] R.S. Liptser and A.N. Shiryayev. *Theory of Martingales*. Kluwer Academic Publ., Dordrecht, 1986. 76, 77, 78, 79, 109, 170, 279

[112] A. Lunardi. *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Birkhäuser, Boston, 1995. 394

[113] T. Lyons and Z. Qian. *System control and rough paths*. Clarendon Press; Oxford University Press, 2002. 145

[114] T.J. Lyons, M. Caruana, and T. Lévy. *Differential Equations Driven by Rough Paths*. Lecture Notes in Mathematics 1908. Springer-Verlag Berlin Heidelberg, 1 edition, 2007. 145

[115] P. Malliavin. *Integration and Probability*. Springer-Verlag, New York, 1995. 37, 102, 372

[116] P. Malliavin. *Stochastic Analysis*. Springer-Verlag, New York, 1997. 21

[117] P. Mandl. *Analytical Treatment of One-Dimensional Markov Processes*. Springer-Verlag, New York, 1968. 416

[118] M.B. Marcus and J. Rosen. *Markov processes, Gaussian processes, and local times*. Cambridge University Press, Cambridge, 2006. 353

[119] H.P. McKean. *Stochastic Integral*. Academic Press, New York, 1971. 85, 145

[120] P. Medvegyev. *Stochastic integration theory*. Oxford University Press, Oxford, 2007. 70, 135

[121] J.L. Menaldi. Stochastic hybrid optimal control models. *Aportaciones Matemáticas*, 16:205–250, 2001. 378

[122] J.L. Menaldi. *Distributions and Function Spaces*. Don't Know Yet!, Someplace, Somewhere, 2014. ix

[123] J.L. Menaldi. *Measure and Integration*. Wayne State University, link: http://digitalcommons.wayne.edu/mathfrp/22, 2015. ix

[124] J.L. Menaldi and S.S. Sritharan. Stochastic 2-d Navier-Stokes equation. *Appl. Math. Optim.*, 46:31–53, 2002. 379

[125] J.L. Menaldi and S.S. Sritharan. Impulse control of stochastic Navier-Stokes equations. *Nonlinear Anal.*, 52:357–381, 2003. 379

[126] J.L. Menaldi and L. Tubaro. Green and Poisson functions with Wentzell boundary conditions. *J. Differential Equations*, 237:77–115, 2007. 396

[127] M. Métivier. *Semimartingales: A course on Stochastic Processes*. De Gruyter, Berlin, 1982. 57, 108

[128] M. Métivier and J. Pellaumail. *Stochastic integration*. Academic Press, New York, 1980. 122, 135

[129] P.A. Meyer. *Probability and potentials*. Blaisdell Publ. Co., Waltham, Mass., 1966. 23

[130] P.A. Meyer. *Un cours sus le intégrales stochastiques. Séminaire de Proba. X*, volume 511 of *Lectures Notes in Mathematics*, pages 246–400. Springer-Verlag, Berlin, 1976. 238

[131] R. Mikulevicius and H. Pragarauskus. On the Cauchy problem for certain integro-differential operators in Sobolev and Hölder spaces. *Liet. Matem. Rink*, 32:299–331, 1992. 402

[132] R. Mikulevicius and H. Pragarauskus. On the martingale problem associated with non-degenerate Lévy operators. *Lithunian Math. J.*, 32:297–311, 1992. 402

[133] R. Mikulevicius and B.L. Rozovskii. *Martingale problems for stochastic PDE's*, volume 64 of *Mathematical Surveys and Monographs, eds: R.A. Carmona and B.L. Rozovskii*, pages 243–325. Amer. Math. Soc., Providence, 1999. 402

[134] H. Morimoto. *Stochastic Control and Mathematical Modeling*. Cambridge Univ Press, New York, 2010. xi

[135] R. Nelson. *Probability, Stochastic Processes and Queueing Theory*. Springer-Verlag, New York, 1995. 51, 53

[136] J. Neveu. *Bases Mathématiques du Calcul des Probabilités*. Masson, Paris, 1970. 18, 23, 24

[137] J. Neveu. *Discrete-Parameters Martingales*. North Holland, Amsterdam, 1975. 51, 73

[138] D. Nualart. *Malliavin Calculus and Related Topics*. Springer-Verlag, New York, 1995. 119

[139] B. Oksendal and A. Sulem. *Applied Stochastic Control of Jump Diffusions*. Springer-Verlag, Berlin, 2005. xi

[140] R. Pallu de la Barrière. *Optimal Control Theory*. Dover, New York, 1967. 11

[141] K.R. Parthasarathy. *Probability Measures on Metric Spaces*. Academic Press, New York, 1967. 115

[142] A. Pazy. *Semigroups and Linear Operators and Applications to Partial Differential Equations*. Springer-Verlag, New York, 2nd edition, 1983. 378

[143] G. Peskir and A. Shiryaev. *Optimal Stopping and Free-Boundary Problems*. Birkhäuser Verlag, Basel, 2006. 55

[144] M.A. Pinsky. *Introduction to Fourier Analysis and Wavelets*. Brooks/Cole Series in Advanced Mathematics. Brooks/Cole, Pacific Grove, CA, 2002. 37

[145] D. Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, New York, 1984. 29, 57

[146] D. Pollard. *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge, 2002. 1

[147] N.I. Portenko. *Generalized Diffusion Processes*. Amer. Math. Soc., Providence, Rhode Island, 1990. 395, 396

[148] E. Priola. On a class of markov type semigroups in spaces of uniformly continuous and bounded functions. *Studia Math.*, 136(3):271–295, 1999. 382

[149] P. Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, New York, 1990. 69, 70, 78, 79, 122, 135, 165, 176, 185, 221, 229

[150] M.M. Rao. *Stochastic Processes: General Theory*. Kluwer Academic, Dordrecht, 1995. 57

[151] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer-Verlag, New York, 3rd edition, 1999. 21, 73, 77, 108, 178

[152] M. Robin. Contrôle impulsionnel des processus de Makov. Thèse d'état, 1978, 353 pp. Link at https://hal.archives-ouvertes.fr/tel-00735779/document. 341

[153] L.C.G. Rogers and D. Williams. *Diffusions. Markov Processes and Martingales*, volume 1: Foundations, 2: Itô calculus. Cambridge Univ. Press, Cambridge, 2nd edition, 2000. 22, 72, 75, 79, 122, 165, 166, 171, 176, 248, 331, 353, 355

[154] J.S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second edition, 2006. 51

[155] B. Rudiger. Stochastic integration with respect to compensated Poisson random measures on separable Banach spaces. *Stochastics and Stochastics Reports*, 76:213–242, 2004. 111

[156] M. Sanz-Sole. *Malliavin calculus*. CRC Press, 2005. with applications to stochastic partial differential. 119

[157] K.I. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Univ. Press, Cambridge, 1999. 43, 62, 66, 67, 68, 69, 367, 371, 406

[158] M. Sharpe. *General Theory of Markov Processes*. Academic Press, New York, 1988. 331

[159] P.S. Shields. *The Ergodic Theory of Discretes Sample Paths*. Am. Math. Soc., Providence, Rhode Island, 1996. 54

[160] A.N. Shiryayev. *Probability*. Springer-Verlag, New York, 1984. 18

[161] A.V. Skorokhod. *Studies in the Theory of Random Processes*. Dover, New York, 1982. 302, 310

[162] A.L. Skubachevskii. *Elliptic Functional Differential Equations and Applications*. Birkhäuser, Boston, 1997. 395, 396

[163] V.A. Solonnikov. A priori estimates for second order equations of parabolic type. *Trudy Math. Inst. Steklov*, 70:133–212, 1964. 397

[164] V.A. Solonnikov. *On Boundary Value Problems for Linear General Parabolic Systems of Differential Equations*. Trudy Math. Inst. Steklov. English transl., Amer. Math. Soc., Providence, Rhode Island, 1967. 397

[165] W. Stannat. (Nonsymmetric) dirichlet operators on l^1: existence, uniqueness and associated Markov processes. *Ann. Scuola Norm. Sup. Pisa Cl. Sci*, **(4)** 28:99–140, 1999. 385

[166] E.M. Stein and R. Shakarchi. *Fourier analysis; An introduction*. Princeton University Press, Princeton, NJ, 2003. 36

[167] K.R. Stromberg. *Probability for Analysts*. Chapman and Hall, New York, 1994. 8

[168] D.W. Stroock. *Probability Theory: An Analytic View*. Cambridge University Press, Cambridge, 1999. revised edition. 8, 64

[169] D.W. Stroock and S.R. Varadhan. *Multidimensional Diffusion Process*. Springer-Verlag, Berlin, 1979. 364

[170] G. Szegö. *Orthogonal Polynomials*. Amer. Math. Soc., Providence, Rhode Island, 4th edition, 1975. 117

[171] K. Taira. *Diffusion Processes and Partial Differential Equations*. Academic Press, Boston, 1988. 370, 374, 375

[172] K. Taira. *Semigroups, Boundary Value Problems and Markov Processes*. Springer-Verlag, Berlin, second edition, 2014. 331, 353, 385

[173] M.E. Taylor. *Partial Differential Equations*, volume Vols. 1, 2 and 3. Springer-Verlag, New York, 1997. 377

[174] M. Tsuchiya. On the oblique derivative problem for diffusion processes and diffusion equations with Hölder continuous coefficients. *Trans. Amer. Math. Soc.*, 346(1):257–281, 1994. 396

[175] M. Tsuchiya. Supplement to the paper: "On the oblique derivative problem for diffusion processes and diffusion equations with Hölder continuous coefficients". *Ann. Sci. Kanazawa Univ.*, 31:1–52, 1994. 396

[176] Jan A. van Casteren. *Markov processes, Feller semigroups and evolution equations*. World Scientific Publishing, 2011. 331

[177] A.D. Wentzell. *A Course in the Theory of Stochastic Processes*. McGraw-Hill, New York, 1981. 57

[178] D. Williams. *Probabiliy with Martingales*. Cambridge Univ. Press, Cambridge, 1991. 22, 51

[179] E. Wong. *Stochastic Processes in Information and Dynamical Systems*. McGraw-Hill, New York, 1971. 26

[180] E. Wong and B. Hajek. *Stochastic Processes in Engineering Systems*. Springer-Verlag, New York, 1985. 57

[181] J. Yeh. *Martingales and Stochastic Analysis*. World Scientific Publishing, Singapur, 1995. 207

[182] G.G. Yin and Q. Zhang. *Continuous-Time Markov Chains and Applications*. Springer-Verlag, New York, 1998. 53

[183] J. Yong and X.Y. Zhou. *Stochastic Controls*. Springer-Verlag, New York, 1999. xi

[184] J. Zabczyk. *Mathematical Control Theory: An Introduction*. Birkhäuser, Boston, 1992. xi

Index