2021

Basic Probability Theory

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

Basic Probability Theory [1]

Jose-Luis Menaldi[2]

Current Version: 19 November 2022[3]
First Version: – — 2014 [4]

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

[3]Long Title. Basic Probability Theory: Independent Random Variables and Sample Spaces.

[4]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 third one. The first part deals with measure and integration theory, while part two concerns basic function spaces, particularly the theory of distributions. In part four, stochastic integrals are studied in some details, and in part five, stochastic ordinary differential equations are discussed, 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.

The reader should be familiar with measure and integration to fully enjoy this **part three**, even if only certain pieces are actually needed. Therefore (at this point), most of the pieces are in place to begin a deep study on the basis of probability. First an update or translation (with simple examples) of the wording in measure theory is necessary to discuss probability as presented in Chapter 1. Next, Chapter 2 develops the continuation of measure theory necessary to study probability, in particular, conditional probability and infinite product of probabilities; while Chapter 3 is a beginning of more advanced topics, including suitable sample spaces used for stochastic processes. Then, Chapter 4 revisits probability is a more abstract way, ending with some examples of Markov processes. Finally Chapter 5 goes even deeper, including martingales and ending with brief comments on Hunt and standard processes. 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. Appendix A has a number of solutions for most of the exercises, while Appendix B is a detailed account (in a narrative way) of semi-group theory, which is very useful when dealing with Markov processes.

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 sometimes shorten some proofs, but with appropriate references. Thus, this book could be used as a second-semester course in Real Analysis, with a clear orientation to Probability Measures

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) to 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.

In the appendix, all exercises are re-listed by section, but now, most of them have a (possible) solution. **Certainly, this appendix is not for the first reading**, i.e., this part is meant to be read after having struggled (a little) with the exercises. Sometimes, there are many ways of solving problems, and depending of what was developed "in the theory", solving the exercises could have alternative ways. The instructor will find that some exercises are trivial while other are not simple. It is clear that what we may call "Exercises" in one textbook could be called "Propositions" in others. This part three (as part two) does not have a large number of exercises as in part one does, but the instructor may find a lot of exercises in some of the references quoted in the text.

The combination of parts I, II, and III is neither a comprehensive course in measure and integration (but a certain number of generalizations suitable for probability are included), nor a basic course in probability (but most of language used in probability is discussed), nor a functional analysis course (but function spaces and the three essential principles are addressed), nor a course in theory of distribution (but most of the key component are there). One of the objectives of these first three books is to show the reader a large open door to probability, without committing oneself to probability and without ignoring hard parts in measure and integration theory.

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

Actually, this book-project is unfinished. The reader will see that a revision is completely necessary, but still, in my opinion, there are some interesting material there!

Michigan (USA), *Jose-Luis Menaldi*, November 2022

Introduction

This is a continuation of the previous books, even if only certain material is essential to the understanding of what follow, e.g., it may be convenient for the reader to review Chapters 1, 2, 4 and 6 in our first part-book [95]. Note that measure theory is not really necessary for discrete probability. Moreover, there are a couple of sections that are repeated in four part-book [97], with the intention to cover a possible gap between 'elementary' and 'more advanced' concepts. As it should be clear now, that after part one, the reader may look at parts two, three and four in any order, but part five should follows part four.

In the second part, we begin by reinforcing some points regarding the theory of integrals, and the reader is recommended to review in in some details the concept of uniform integrability (which is very important various aspect of probability theory) and the integral theory as a Daniell functional. Moreover, at a certain point, Schwartz' theory of distributions and discuss elements of the Fourier analysis become useful, and eventually, the reader is referred to our second part-book [96].

Therefore, in this *third part*, we assume that the reader is familiar with the rudiments of metric, Banach and Hilbert spaces, and we declare our interest in probability. Thus, appealing to all these tools, probabilities are introduced as a particular case of measures. We are interested in the language used in Probability, starting from discrete random variables, passing throughout the concepts of tightness and weak convergence of probabilities, discussing the law of the large numbers and the central limit theorem. Characteristic functions is then well understood as the Fourier transform and a proof of Bochner's Theorem on existence of probability measures is given. Next we discuss conditional expectation and regular conditional probability, which are the building block of advanced probability arguments. Hence, we are ready to present a short analysis on the sample spaces, particularly, the Polish (separable complete metrizable) space of cad-lag (continuous from the right with left-hand limits) functions. Finally, we conclude with an initial discussion on stochastic processes, to prepare the way for a more advanced course.

In the next section, a quick overview on stochastic dynamical system models is used as a motivation regarding most of the material discussed in this book. Certainly, even if this some a vicious circle, this is addressed to a reader with some knowledge on probability. In any case, this following section may be useful when it is read after a couple of chapters.

Motivations and Initials Questions

Besides the natural interest in elementary probability, our intention is partially instruct the reader with the necessary 'probability background' to understand (comfortable) stochastic dynamical systems, first in a discreta-time setting and then, perhaps, begin a more delicate study on continuous-time models.

Assuming a rudimentary idea on dynamical system, a short discussion on *discrete-time dynamics* is perhaps necessary. Say, at a given time, the state of a dynamical systems changes according to a fix rule (or dynamical rule or dynamic). The state represents all the information necessary to evolve in time following the dynamic, and in the case of a controlled dynamical system, the dynamic includes a control (the information that is chosen by the controller before advancing the time). Moreover, this time-state-control model may be internal or theoretical, and an observation procedure should be discussed. Depending on the possible values taken by the time t and the state x (and the control v) the model is called discrete (a countable set of values) or continuous (a non-countable set of values).

Our interest is on discrete time and continuous state with full observation, i.e., due to the arrow of time, we may use the integer numbers to represent time $t = 0, \pm 1, \pm 2, \dots$ or $t = 0, 1, 2, \dots$ when an initial time $t = 0$ is convenient. In finite dimension, the state and the control are elements of a Euclidean space, i.e., x in $\mathbb{R}^d$ and v in $\mathbb{R}^m$. Therefore, $x_0, x_1, x_2, \dots$ are the states of our discrete-time dynamical system, x_t is the state at time t, and the state x_{t+1} is determined (via the one-step dynamical rule) by the state x_t and the control v_{t+1} chosen at time $t + 1$. By adding more dimensions to the Euclidean space, the case where x_t is determined by the previous states $x_t, x_{t-1}, \dots, x_0$ and previous controls (or control policy) $v_{t+1}, v_t, v_{t-1}, \dots, v_1$ can be included in the former one-step dynamic model. However, if x_t is the state at time t then all necessary information is contained in x_t and previous states $x_0, \dots, x_{t-1}$ should not be necessary to decide the values of the current control v_t.

Because the values of the state are composed by measurements (typically taken at successive points in time spaced at uniform time intervals), deterministic dynamical systems are considered a first degree of model approximation, i.e., the average system or the system without any disturbances, noises, unknowns or errors. This approximation is improved by adding more assumptions to model (the typical error due to the measurements and the disturbances, noises or unknown elements), and depending on the focus of interest, several points of view could be developed. For instance, worse-case scenario (when the bounds on the possible errors is the focus, which eventually becomes a max-min problem, sometimes called robust analysis) or a more statistical model (when the states are treated as a sequence of random variables -time series-). In this last approach, various questions could be addressed, e.g., forecasting analysis, regression analysis, etc, all of them used in applied science and engineering (such as statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering). In a way, stochastic

control problems is a more detailed model that include suitable assumptions that largely simplify the real model under consideration.

Discrete-Time Stochastic Models

The discrete-time t, the state x (and the control v) have been identified in a dynamical system. The control represents the parameters that can be chosen (a decision taken as time pass), but the arrow of time imposes causality, i.e., the control at the current time t is based on observation of the system up to the present time t. Because we assume full observation (or complete information), a decision (or control) at time t should involves only the states up to the present $x(s)$, for $s \leq t$ and perhaps, the previous controls $v(s)$, for $s < t$. Since the state $x(t)$ summarize all necessary information up to the time t, a feedback control of the form $v = v(t, x)$ suffices, as a function $v : \mathbb{T} \times \mathbb{X} \longrightarrow \mathbb{V}$, where $\mathbb{T}$ is the set of time, $\mathbb{X}$ the set of possible state-values and $\mathbb{V}$ the set of possible control-values. Once a feedback control (policy) has been chosen, the state of the system evolves as a Markov process, which is usually refer to as a controlled Markov process.

There are several ways to describe a (controlled) Markov process, e.g., the transition function

$$P(x_t \in A \,|\, x_{t-1} = x,\, v_t = v\} = \pi_t(A, x, v), \quad \text{for all Borel set } \; A \subset \mathbb{X}$$

may be given, and thus, to study the dependency on the feedback control, or alternatively, an evolution equation could be given (of which the transition function is its fundamental solution). In the last case, the evolution equation is what is called the dynamic of the system. For instance, a general model is a s follows.

Let S_t, C_t and D_t represent the possible states, the possible controls, and the possible disturbances (or noises) of the dynamical system at time $t = 0, 1, \ldots$. Suppose that ρ_t is a function from $S_{t-1} \times C_t \times D_t$ into S_t, for each $t = 1, 2, \ldots$. If the current state x_{t-1} and the control v_t are known, then the state x_t at time t is given by the relation

$$x_t = \rho_t(x_{t-1}, v_t, w_t), \qquad \text{for} \quad t = 1, 2, \ldots, \tag{1}$$

where the disturbances w_t should be suitable modeled.

In stochastic model, the disturbances is a sequence of random variables, while in a worse-case scenario, the disturbances are as bad as possible (within a priori bounds). The statistics of the sequence of random variables $\{w_1, w_2, \ldots\}$ should be prescribed in a way to preserve the structure of the dynamical system. For instance, an independent (and identically distributed, iid) sequence $\{w_t\}$ of random variables is a typical situation, but a random disturbance w_t which is independent of the state x_{t-1} and the control v_t (or more general, defined via $\pi_t(\cdot|x_{t-1}, v_t)$, the conditional probabilities) suffices.

A cost is associated with feedback policy (i.e., a sequence $v(\cdot) = \{v_1(\cdot), \dots\}$ of functions $v_t(\cdot)$ from S_{t-1} into C_t) given by a functional

$$J_{x_0,n}\big(v(\cdot)\big) = \mathbb{E}\big\{f_n(x_n) + \sum_{t=1}^{n} f_t(x_{t-1}, v_t, w_t)\big\}, \quad n = 1, 2, \dots, \tag{2}$$

which represents the cost of policy $v(\cdot)$ on the horizon $[0, n]$ with initial state x_0. The problem is to minimize this cost functional and find an optimal control policy.

All these problems are not studied here, but the interest reader may consult for instance, the book by Bertsekas [12], among many others. We prefer to present a short discussion on the dynamic of a system given by an equation of the form

$$x_t = g_t(x_{t-1}, v_t) + \sigma_t(x_{t-1}, v_t)w_t, \qquad \text{for} \quad t = 1, 2, \dots, \tag{3}$$

where g_t and σ_t are measurable functions defined on $\mathbb{R}^d \times \mathbb{V}$ (for each t in $\mathbb{N}$) with values in $\mathbb{R}^d$ and $\mathbb{R}^d \times \mathbb{R}^n$, respectively, and $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{V}$ is a compact set of some Euclidean space. This equation is a particular case of (1) with ρ_t a measurable functions defined on $\mathbb{R}^d \times \mathbb{V} \times \mathbb{R}^n$, and it takes place on a probability space $(\Omega, \mathcal{F}, P)$, where an independent (and identically distributed, iid) sequence $\{w_t\}$ of random variables with values in $\mathbb{R}^n$ is given.

It is clear that once an initial state x_0 is given, and the feedback is chosen, then either equation (1) or (3) determines a sequence $\{x_1, x_2, \dots\}$ of random variables (representing the state of the system). In this context, a measurable function $v(\cdot)$ defined on $\mathbb{R}^d \times \mathbb{N}$ with values in $\mathbb{V}$ represent a feedback control (policy). In the last model equation (3), the emphasis is on an average dynamic given by the functions g_t and the disturbances represented by $\sigma_t w_t$, where now w_t has zero-mean, i.e., $\mathbb{E}\{x_t\} = \mathbb{E}\{g_t(x_{t-1}, v_t)\}$, for every t.

Moreover, if $\mathcal{F}_t$ is the σ-algebra generated by the random variables x_0 for $t = 0$ and by $x_0, w_1, \dots, w_t$ for $t = 1, 2, \dots$, then the random variable x_t is $\mathcal{F}_t$-measurable, while the control $v_t = v(x_{t-1}, t)$ is $\mathcal{F}_{t-1}$-measurable. Thus the sequence of random variables $\{v_1, v_2, \dots\}$ is adapted to $\mathcal{F}_{t-1}$ (i.e, predictable), while the sequence of random variables $\{x_1, x_2, \dots\}$ is adapted to the filtration $\{\mathcal{F}_t\}$. There are almost not difficulties in replacing the possible state space $\mathbb{R}^d$ and the possible control space $\mathbb{V} \subset \mathbb{R}^m$ by a locally compact separable metric space $\mathbb{X}$ and a compact metric space $\mathbb{V}$, e.g., see Runggaldier and Stettner [115].

Now, recall the fact that if X and Y are two independent random variable with values in some Polish space (separable complete metrizable space), e.g., an Euclidean space, and $\varphi(x, y)$ is a non-negative Borel measurable function then the conditional expectation of the numerical random variable $\varphi(X, Y)$ given Y can be calculated as follows

$$\mathbb{E}\big\{\varphi(X, Y) \,|\, Y\big\} = \Phi_Y(X), \quad \text{with} \quad \Phi_Y(x) = \mathbb{E}\big\{\varphi(x, Y)\big\}. \tag{4}$$

Indeed, if $\varphi = \mathbb{1}_A \mathbb{1}_B$ then (4) is the definition of independence of the two variables X and Y. By linearity, this equality remains true for any finite combination, and with an argument of monotone class the conclusion follows. Another

way of staying this result is by expressing the conditional probability of (X, Y) given Y (or marginal distribution) as

$$P\big\{(X,Y) \in A \,|\, Y\big\} = P_X(A), \quad \text{with} \quad P_x(A) = P\big\{(x,Y) \in A\big\}. \tag{5}$$

for any Borel set A in the Polish space where (X, Y) takes values.

Based on the previous result, the sequence of random variable generated by either equation (1) or (3) defines what is called a controlled Markov chain, i.e., a Markov chain as soon as a measurable feedback policy is chosen. The transition probability functions $P\{x_{t+1} \in A \,|\, x_t\} = \pi_t(A, x_t)$ are the key elements of a Markov chain $\{x_t : t \in \mathbb{N}\}$. In the above construction, with a feedback policy $v_t(\cdot)$, we deduce $\pi_t(A, x) = P\{\rho_t(x, v_t(x), w_{t+1}) \in A\}$, for any x in $\mathbb{R}^d$ and any Borel set A in $\mathbb{R}^d$. Certainly, there is a huge literature regarding a lot of interesting questions to solve for this controlled Markov chain, for instance, the reader may take a look at the books by Bellman [7], Bensoussan [8], Bertsekas and Shreve [13], Bremaud [21, 22], Hernández-Lerma and Lasserre [60], Revuz [110], Yin and Zhang [135], even suitable books at a graduate level, such as Söderström [124], among many others.

Chapter 1

Elementary Probability

Probability is used as a mathematical model of random events or "controlled" experiments. The idea begins with a few basic principles which dictate about how "the laws of chances" should behave. We have an abstract non-empty set Ω, which represents all possible outcomes of the experiment to be studied. Initially, an event is an element of 2^Ω and an elementary event is a singleton $\{\omega\}$, a set of a single outcome ω in Ω. Thus events are handled mathematically by the set theory, and then (initially!) a probability is an finite additive function defined on events, normalized to have a finite total probability. These assumptions are well adapted to practical cases, where Ω is a finite set. However, the mathematical analysis imposes a certain continuity, and thus, we call a *probability* a (1) σ-additive function P defined on a σ-algebra $\mathcal{F} \subset 2^\Omega$ satisfying (2) $P(\Omega) = 1$, i.e., a probability (measure) is a finite measure normalized with the condition (2). For instance, a justification and discussion of the use of measure theoretic foundation can be found in the book Pollard [106]. There are several books treating integral, measure and probability simultaneously that the reader may check, e.g., Athreya and Lahiri [4] Capinski and Kopp [23], Doob [40], Rosenthal [113], Taylor [132].

Practically, the abstract probability space $(\Omega, \mathcal{F}, P)$ is only known through observations via random variables or observable functions, namely, real-valued measurable functions defined Ω or more general, $X : \Omega \to E$, where $(E, \mathcal{E})$ is a measurable space. The structure of E determines the model used for the experiment and mathematically the problem, tools and language to implement. For instance, (1) measurable functions taking values in $\mathbb{R}$ or $\mathbb{R}^d$ are called random variables, while (2) measurable functions taking values in some product space $\mathbb{R}^I$, I some index, are called random (or stochastic) processes. Mainly, we discuss the case of random variables (i.e., $E = \mathbb{R}^d$) in an abstract probability space. Similarly, the condition *almost everywhere* is called *almost surely* when dealing with probabilities.

Hence, in general, we forget about $(\Omega, \mathcal{F}, P)$ in the sense that the variable ω is not explicitly written in probability, but from the context, we understand that most objects are random, i.e., they depends on ω. Typically, we have a random variable X and we make assumptions (or we analyze) the distribution

induced by X, i.e., the probability measure

$$P_X(B) = P\big(X^{-1}(B)\big) = P\{X^{-1}(B)\} = P\{\omega \,:\, X(\omega) \in B\},$$

defined for every $B \in \mathcal{E}$. Usually, we exchange the notation $P\{\cdot\}$ and $P(\cdot)$.

The integral with respect to a probability measure is called the *expectation* and denoted by $\mathbb{E}\{\cdot\}$. Hence for any nonnegative measurable function $f : E \to \mathbb{R}$, we have

$$\mathbb{E}\{f(X)\} = \int_\Omega (f \circ X)\, dP = \int_\Omega f\big(X(\omega)\big)\, P(d\omega) = \int_E f(x)\, P_X(dx),$$

i.e., the composition $f \circ X$ is P-integrable if and only if f is P_X-integrable. Whenever it makes sense, $\mathbb{E}\{X\} = \overline{X}$ is called the *mean* and $\mathbb{E}\big\{(X - \overline{X})^2\big\}$ the *variance* of the random variable X. Note that $\mathbb{E}\big\{(X - \overline{X})^2\big\} = \mathbb{E}\big\{X^2\big\} - \overline{X}^2$. Moreover, if $f(x) = x^n$, with n an positive integer, then $\mathbb{E}\{X^n\}$ is called the moment of order n. If E is a topological space and g is a continuous linear functional on E, then the particular case where $f(x) = e^{i\langle g,x\rangle}$ yields the complex-valued function

$$g \mapsto \Phi_X(g) = \mathbb{E}\big\{e^{\mathtt{i}\langle g,X\rangle}\big\}, \quad \text{with } \mathtt{i} = \sqrt{-1},$$

is called the *characteristic function* or *Fourier transform* of either the random variable X (properly, of the distribution of the random variable X) or the measure P_X. For instance, note that $\Phi_X(0) = 1$, and as discussed later, under certain conditions, the characteristic function Φ_X identify the measure P_X. Moreover, for the particular of a real-valued random variables X we have

$$t \mapsto \Phi_X(t) = \mathbb{E}\{e^{\mathtt{i}tX}\}, \quad \forall t \in \mathbb{R},$$

and based on the estimate $|e^{\mathtt{i}h} - 1| \le |h|$, for every real value h, we can use the dominate convergence Theorem to show that

$$\frac{d^n \Phi_X(t)}{dt}\Big|_{t=0} = \mathtt{i}^n \mathbb{E}\{X^n\}, \quad \forall n \ge 1,$$

whenever X^n is integrable.

There are several ways to introduce and define the meaning of independence (which is of great importance in probability). Let $(E_i, \mathcal{E}_i)$ be a family measurable spaces for i in some set of indexes. A family $\{X_i : i \in I\}$ of random variables, $X_i : \Omega \to E_i$, is *independent* if for any finite family of indexes $J \subset I$ we have $P_{\{X_i : i \in J\}} = \prod_{i \in J} P_{X_i}$, where $\{X_i : i \in J\}$ is the (finite) product random variable with values in $\prod_{i \in J} E_i$. In particular, two random variables $X_1 : \Omega \to E_1$ and $X_2 : \Omega \to E_2$ are called *independent* if $P_Z = P_X \times P_Y$, where $Z = (X, Y)$ is the product random variable $Z : \Omega \to E \times F$.

Let $\mathcal{K}_i$ be π-classes generating $\mathcal{E}_i$, for every $i \in I$. In view of uniqueness of the extension of a measure initially defined on a π-classes, we deduce that a family $\{X_i : i \in I\}$ is independent if and only if

$$P\Big\{\bigcap_{i \in J} X_i^{-1}(A_i)\Big\} = \prod_{i \in J} P\{X_i^{-1}(A_i)\} \quad \forall A_i \in \mathcal{K}_i,$$

for any finite family $J \subset I$ of subindexes. This means that if $\mathcal{X}_i \subset \mathcal{F}$ is the σ-algebra generated by X_i then the family $\{X_i : i \in I\}$ is independent if and only if

$$P\Big\{\bigcap_{i\in J} F_i\Big\} = \prod_{i\in J} P\{F_i\} \quad \forall F_i \in \mathcal{X}_i, \quad \forall \text{ finite } J \subset I.$$

Hence, the concept of independence can be applied to an arbitrary family of sub σ-algebras of $\mathcal{F}$. For the particular case where $X_i = \mathbb{1}_{A_i}$ we deduce that $\{A_i : i \in I\}$ is a family of independent sets if and only if

$$P\Big\{\bigcap_{i\in J} A_i\Big\} = \prod_{i\in J} P\{A_i\} \quad \forall \text{ finite } J \subset I,$$

which is the elementary definition of independence.

1.1 Preliminary Examples

As noted early, probabilities (or measures) on countable spaces makes a great deal of simplification and called discrete probabilities. On the other hand, we have diffuse probability and probability with density. Some examples are presented in what follows.

1.1.1 Discrete Probabilities

This is the ideal situation where the random variable $X : \Omega \to E$ takes only a countable possible values, X and its distribution P_X are called discrete. Therefore, the analysis of a discrete probability P_X (or equivalently a discrete random variable X) is the discussion of probabilities in a countable measurable space $(\mathtt{A}, 2^{\mathtt{A}})$, $\mathtt{A} \subset E$. In view of the σ-additivity an discrete probability Q on $(\mathtt{A}, 2^{\mathtt{A}})$ is uniquely determined by the sequence $\{Q(a) : a \in \mathtt{A}\}$ of nonnegative number which yields a series satisfying $\sum_{a\in\mathtt{A}} Q(a) = 1$. Most of the time is a matter of notation, i.e.,

$$P_X(B) = \sum_{b\in B} P\big(X^{-1}(b)\big) = \sum_{i=1}^{\infty} P\big(a_i(X) \in B\big),$$

where $\{a_i(X) : i \geq 1\}$ is a enumeration of $\big\{a \in E : P\big(X^{-1}(a)\big) \neq 0\big\}$, (each $a_i = a_i(X)$ is called an *atom*); and the sum on $\{b \in B\}$ actually means on $\big\{b \in B : P\big(X^{-1}(b)\big) \neq 0\big\}$, since the zero term are ignored. In this case, we can take $\bigcup_{i\geq 1}\{a_i\} = \mathtt{A} \subset E$. Hence, for every nonnegative measurable function $f : E \to \mathbb{R}$, we have

$$\mathbb{E}\{f(X)\} = \sum_{a\in\mathtt{A}} f(a)\, P\big(X^{-1}(a)\big) = \sum_{i=1}^{\infty} f(a_i)\, P\big(X^{-1}(a_i)\big),$$

and composition $f \circ X$ is integrable if the above series is absolutely convergent. Perhaps the simplest example is the *Uniform distribution*, where the number of atoms is a finite set $\{a_1, \dots, a_n\} = \mathtt{A}$ and $P_X(B)$ is equal to the number of elements of $B \in 2^{\mathtt{A}}$ divided by n, the number of elements in $\mathtt{A}$.

Hence, a random variable $X : \Omega \to \mathbb{R}$ is *discrete* if and only if there exists a sequence of atoms $\{a_i : i \geq 1\} = \mathtt{A} = X(\Omega) =\subset \mathbb{R}$ such that $P\big(X^{-1}(a_i)\big) = p_i > 0$ and $\sum_{i=1}^{\infty} p_i = 1$. In this case $2^{\mathtt{A}} \subset \mathcal{B} = \mathcal{B}(\mathbb{R})$, $P_X(B) = \sum_{i=1}^{\infty} p_i \mathbb{1}_B(a_i)$, for every $B \in \mathcal{B}$ and $P_X(N) = 0$, for every $N \in \mathcal{B}$ such that $N \cap \mathtt{A} = \emptyset$. Typical examples are the following, with $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$:

(1) *Binomial distribution* with parameters (n, p), $0 < p < 1$: The atoms are $\{0, 1, \dots, n\} = \mathtt{A}$ and $P_X(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$, for every k in $\mathtt{A}$. The mean is $\mathbb{E}\{X\} = np$ and the variance is $np(1-p)$.

(2) *Poisson distribution* with parameter $\lambda > 0$: The atoms are $\{0, 1, \dots\} = \mathtt{A}$ (nonnegative integers) and $P_X(\{k\}) = \mathrm{e}^{-\lambda}\lambda^k/k!$ (recall $k! = k(k-1)\dots 1$), for every k in $\mathtt{A}$. The mean $\mathbb{E}\{X\}$ and the variance are equal to λ.

(3) *Geometric distribution* with parameter c, $0 \leq c < 1$: The atoms are $\{0, 1, \dots\} = \mathtt{A}$ (nonnegative integers) and $P_X(\{k\}) = (1-c)c^k$ (with the convention that $0^0 = 1$), for every k in $\mathtt{A}$. The mean is $\mathbb{E}\{X\} = c(1-c)^{-1}$ and the variance is $c(1-c)^{-2}$. Note that $P_X(\{0\}) = (1-c)$ and that sometimes, $p = 1 - c$ is the parameter and/or the values are shifted, i.e., the atoms are $\{1, 2, \dots\}$.

For any random variable X, the characteristic function (or the Fourier transform) is defined by

$$\Phi_X(t) = \mathbb{E}\{\mathrm{e}^{\mathrm{i}tX}\} = \sum_n \mathrm{e}^{\mathrm{i}tn} P\{X = n\}, \quad \forall t \in \mathbb{R},$$

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 Poisson distribution with parameter λ we get $G_X(t) = \exp[\lambda(t-1)]$, and for the Geometric distribution with parameter α we obtain $G_X(t) = (1-\alpha)/(1-\alpha t)$.

Typically, a probabilistic model of repeatedly tossing a coin (called a trial) that has probability p of coming up tails (and so, probability $1-p$ of coming up heads), with $0 < p < 1$, can be described by a sequence $\{X_k : k = 0, 1, \dots\}$ of independent Bernoulli random variables (i.e., X_k takes only the values 0 or

1 with probability $q = 1 - p$ or p). The random variable Y representing the number tails (i.e., the Bernoulli random variable takes value 1) appeared within the first n trials (i.e., $Y = X_0 + \cdots + X_{n-1}$) has a Binomial distribution with parameter (n, p). Also, the random variable Z indicating that for the first time that a tail appears (i.e., $Z = \inf\{k : X_k = 1\}$) has a Geometric distribution with parameter $c = 1 - p$. In general, if Y_k, $k = 1, \ldots, m$, are independent random variables with a Geometric distribution with parameters c_k then $\min_{k \le m} Y_k$ has also a Geometric distribution with parameter $c_1 \ldots c_m$.

In term, this Binomial distribution can be approximated by a Poisson distribution with parameter λ when n is very large and p very small so that np approximates λ, in fact, based on the factorial growth rate, namely $n! \approx \sqrt{2\pi n}(n/\mathrm{e})^n$, Poisson Theorem expresses precisely that if $n \to \infty$ and $p \to 0$ such that $np \to \lambda$ then

$$\frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \to \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}.$$

Thus, under the above limiting conditions, Binomial distributions are replaced by Poisson distributions.

1.1.2 Other Probabilities

The probability distribution P_X of a real-valued random variable $X : \Omega \to \mathbb{R}$ is completely described by the distribution function $t \mapsto F(t) = F_X(t) = P\{X \le t\}$, which has the following properties: (1) F is an increasing function continuous from the right, (2) $F(t) \to 0$ as $t \to -\infty$, and (3) $F(t) \to 1$ as $t \to +\infty$. In the case of a discrete probability, the function F is piecewise constant.

A probability measure is called *diffuse* if there is not atoms, i.e., $P\{X = x\} = 0$, for every x. In term of the distribution function, this is equivalently to require that the function $t \mapsto F_X$ is continuous, i.e., $P\{X = x\} = F_X(x) - F_X(x-)$, where $F_X(x-)$ is the left-hand limit. In general, if $X : \Omega \to \mathbb{R}^d$ with $d > 1$ then we prefer to work directly with the image measures P_X, instead of using the distribution functions F_X.

For a $\mathbb{R}^d$-valued random variable, we say that P_X or X has a density (with respect to the Lebesgue measure) if there exists an integrable function $f = f_X$ such that

$$P_X(B) = P\{X \in B\} = \int_B f(x)\,\mathrm{d}x, \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

i.e., P_X is absolutely continuous with respect to the Lebesgue measure (denoted by m or simply $\mathrm{d}x$) and the Radon-Nikodym Theorem provides the density f. Note that a distribution may be diffuse and yet, without a density.

Perhaps the simplest example is when the density is constant on a set $K \subset \mathbb{R}^d$ of finite Lebesgue measure and vanishes on the complement $\mathbb{R}^d \setminus K$. This is called the *Uniform* distribution, and $P_X(B) = m(B \cap K)/m(K)$. For instance, if $d = 1$ and $K = [a, b]$ with $a < b$, then $F_X(t) = 0$ for $t \le a$, $F_X(t) = 1$ for $t \ge b$ and $F_X(t) = (t - a)/(b - a)$ for t in $[a, b]$.

Some typical examples for a one-dimensional random variable X, with $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$:

(1) *Gaussian* (o normal) with parameters m and $r > 0$, which is also denoted by $N(m, r)$: $X(\Omega) = \mathbb{R}$ and

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

The mean is $\mathbb{E}\{X\} = m$ and the variance is r^2. A simple change of variable shows that $(X - m)/r$ is indeed a normal random variable with mean 0 and variance 1. In a d-dimensional version, the mean m is a d-dimensional vector and the co-variance is a symmetric strictly positive d-dimensional matrix.

(2) *Exponential* with parameter $\alpha > 0$: $X(\Omega) = [0, +\infty) = \mathbb{R}_0^+$ and

$$P\{X \le t\} = \int_{0}^{t} \alpha \exp(-\alpha x) \, \mathrm{d}x, \quad \forall t \in \mathbb{R}.$$

The mean is $\mathbb{E}\{X\} = \alpha^{-1}$ and the variance is α^{-2}.

(3) *Cauchy* with parameters m and $c > 0$: $X(\Omega) = \mathbb{R}$ and

$$P\{X \le t\} = \pi^{-1} c \int_{-\infty}^{t} \big[(x-m)^2 + c^2 \big]^{-1} \, \mathrm{d}x, \quad \forall t \ge 0.$$

The particularity is that $\mathbb{E}\{(X-m)\, \mathbb{1}_{X>m}\} = \mathbb{E}\{(m-X)\, \mathbb{1}_{X<m}\} = +\infty$, i.e., this random variable is not integrable, and therefore the mean value is not defined.

(4) *Gamma* with parameters $c, \alpha > 0$: $X(\Omega) = [0, +\infty) = \mathbb{R}_0^+$ and

$$P\{X \le t\} = \frac{\alpha^c}{\Gamma(c)} \int_{0}^{t} x^{c-1} \mathrm{e}^{-\alpha x} \, \mathrm{d}x, \quad \forall t \ge 0,$$

where

$$\Gamma(c) = \int_{0}^{\infty} x^{c-1} \mathrm{e}^{-x} \, \mathrm{d}x, \quad \forall c > 0$$

is the Gamma function. Note that $c = 1$ reproduces the exponential distribution and also that $c = n/2$ and $\alpha = 1/2$ is referred to as the χ^2-distribution with n degrees of freedom. If c is an integer then the Gamma distribution is known as the *Erlang* distribution in queueing theory. The mean is $\mathbb{E}\{X\} = c\alpha^{-1}$ and the variance is $c(c+1)\alpha^{-2}$.

(5) *Beta* with parameters $a, b > 0$: $X(\Omega) = [0, 1]$ and

$$P\{X \le t\} = \frac{1}{B(a,b)} \int_{0}^{1} x^{a-1} (1-x)^{b-1} \, \mathrm{d}x, \quad \forall t \in [0, 1],$$

where

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x, \quad \forall a,b>0$$

is the Beta function, which satisfies $B(a,b)\Gamma(a+b) = \Gamma(a)\Gamma(b)$. The mean is $\mathbb{E}\{X\} = a/(a+b)$ and the variance is $ab(a+b)^{-2}(a+b+1)^{-1}$. The obvious case $a=b=1$ is called *Uniform* distribution on $[0,1]$, i.e., $P\{X\le t\} = t$ for any t in $[0,1]$.

Besides the characteristic function

$$\Phi_X(t) = \widehat{P}_X(t) = \mathbb{E}\{\mathrm{e}^{\mathtt{i}tX}\}, \quad \forall t\in\mathbb{R},$$

we may define the Laplace transform

$$\widetilde{P}_X(t) = \mathbb{E}\{\mathrm{e}^{-tX}\}, \quad \forall t\ge 0,$$

when $X(\Omega)\subset\mathbb{R}_0^+$. Certainly, assuming analytic extension, it is clear that $\widetilde{P}_X(-\mathtt{i}t) = \Phi_X(t)$ and $\Phi_X(-\mathtt{i}t) = \widetilde{P}_X(t)$. For instance, we compute the Fourier and Laplace transforms for the above distributions: (1) $\widehat{P}_X = \exp\big(-rt^2/2 + \mathtt{i}mt\big)$ for the normal; (2) $\widehat{P}_X = \alpha/(\alpha-\mathtt{i}t)$ and $\widetilde{P}_X = \alpha/(\alpha+t)$ for the exponential; (3) $\widehat{P}_X = \exp\big(-c|t|/2+\mathtt{i}mt\big)$ for the Cauchy; (4) $\widehat{P}_X = \big(1-\mathtt{i}\alpha^{-1}t\big)^{-c}$ and $\widetilde{P}_X = \big(1+\alpha^{-1}t\big)^{-c}$ for the Gamma. (5) The characteristic function of the Beta function involves the confluent hypergeometric function (a type of exponential function with rising factorials), namely

$$F_1(a,b,z) = \sum_{k=0}^{\infty}\Big(\prod_{i=0}^{k}\frac{a-i}{b-i}\Big)\frac{z^k}{k!},$$

which is an entire function in a,b,z with poles at $b=0,-1,-2,\ldots$ Thus $\widehat{P}_X = F_1(a,b,\mathtt{i}t)$ and $\widetilde{P}_X = F_1(a,b,-t)$.

Typically, if $X_0, X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m}$ is a finite sequence of independent normal random variables with parameter $m=0$ and $r=1$ then the random variable $Y = X_1^2+\cdots+X_n^2$ has a χ^2-distribution with n degrees of freedom. Also, the random variable $X_0/\sqrt{Y/n}$ has a Student's t-distribution with n degrees of freedom, which has the density

$$f_t(x) = \frac{\Gamma\big((n+1)/2\big)}{\sqrt{n\pi}\,\Gamma(n/2)}\Big(1+x^2/n\Big)^{-(n+1)/2}, \quad \forall x\in\mathbb{R}.$$

With the sam token, the random variable

$$\frac{(X_{n+1}^2+\cdots+X_{n+m}^2)/m}{(X_1^2+\cdots+X_n^2)/n}$$

has a F-distribution with (m, n) degrees of freedom, which has the density

$$f_F(x) = \frac{m}{nB(m/2, n/2)} \Big(\frac{m}{m}x\Big)^{m/2-1} \Big(1 + \frac{m}{n}x\Big)^{-(m+n)/2}, \quad \forall x \geq 0.$$

Certainly, the expression of the densities f_t and f_F of the t-distribution and the F-distribution can be extended to $n = a > 0$ and $m = b > 0$ non necessarily integers. By checking the densities, it becomes clear that if X is a random variable having a Beta-distribution with parameters a, b then $\sqrt{nX/(1-X)}$ with $a = 1/2$ and $b = n/2$ has a Student's t-distribution with n degrees of freedom, while $\big(nX\big)/\big((1-X)m\big)$ with $a = m/2$ and $b = n/2$ has a F-distribution with (m, n) degrees of freedom.

The exponential distribution is an way the continuous analogue of the geometric distribution, i.e, if X has an exponential distribution with parameter α then the floor $\lfloor X \rfloor$ has a geometric distribution with parameter $c = \mathrm{e}^{-\alpha}$, i.e., $P\{\lfloor X \rfloor = 0\} = P\{X < 1\} = 1 - \mathrm{e}^{-\alpha}$. A key property of the exponential distribution is the so-called memory-less property, namely, $P\{X > s + t,\ X > s\} = P\{X > t\}$. Similarly, X_i, $k = 1, 2, \ldots$ are independent distributed random variable exponentially with parameter α_i then (1) the random variable $\min_{i \leq k} X_i$ is also exponentially distributed with parameter $\alpha = \alpha_1 + \cdots + \alpha_k$; however, the random variable $\max_{i \leq k} X_i$ has not an exponential distribution. Now, assuming that $\alpha_i = \alpha$ for every $i = 1, 2, \ldots$, (2) the random variable $Y_k = X_1 + \cdots + X_k$ has a Gamma (or Erlang) distribution with parameters $c = k$ and α, and (3) the random variable $N_t = \sup\{k : Y_k \leq t\}$ with $t \geq 0$ has a Poisson distribution with parameter $\lambda = \alpha t$.

1.1.3 Independent Random Variables

First, remark that given a probability space $(\Omega, \mathcal{F}, P)$, it may not be possible to ensure the existence of a sequence of independent variables (having a prescribed distribution). However, the typical (universal) probability space where these constructions are possible is the Lebesgue space on the interval $[0, 1)$. A well known example is to write any ω in $\Omega = [0, 1)$ in binary form $\omega = \sum_k 2^{-k\omega_k}$, a observe that the sequence of variables $\{\pi_n(\omega) = \omega_n : n \geq 1\}$ provide independent coin-tossing variables each taking values 0 or 1 with probability $1/2$. Therefore, for any given injective mapping $(i, j) \mapsto k(i, j)$ from $1, 2, \ldots \times 1, 2, \ldots$ into $1, 2, \ldots$, the expressions $\xi_i = \sum_j 2^{k(i,j)}\omega_{k(i,j)}$, for $i \geq 1$, define an independent sequence of random variables uniformly distributed, i.e., $P(a < \xi_i < b) = b - a$, for every $0 \leq a < b \leq 1$.

Exercise 1.1. Let $\{F_i : i \geq 1\}$ be a sequence of distributions in $\mathbb{R}$, i.e., each F_i is a cad-lag non-decreasing function such that $\lim_{r \to -\infty} F_i(r) = 0$ and $\lim_{r \to \infty} F_i(r) = 1$. Show that there exits a sequence $\{X_i : i \geq 1\}$ of independent real-valued random variables defined on the universal probability space $(\Omega, \mathcal{F}, P)$, $\Omega = [0, 1)$, $\mathcal{F}$ the Borel σ-algebra and P the Lebesgue measure, such that $P(X_i \leq r) = F_i(r)$, i.e., each X_i has distribution F_i. Hint: First, complete the above arguments so that it is clear the construction of a sequence $\{\xi_i : i \geq 1\}$

of independent random variable uniformly distributed. Next, define the inverse of each F_i as $F_i^{-1}(s) = \inf\{r \in \mathbb{R} : s \le F_i(r)\}$, for every s in $[0,1)$ and verify that the sequence $\{X_i = F_i^{-1}(\xi_i) : i \ge 1\}$ has the required properties. □

If X and Y are $\mathbb{R}^d$-valued random variables then the (probability) distribution of the new random variable $Z = X + Y$ can be obtained form the joint distribution $P_{X,Y}$, i.e., the distribution of the $\mathbb{R}^{2d}$-valued random variable $\omega \mapsto (X, Y)$, i.e.,

$$P\{X + Y \in B\} = \int_{\mathbb{R}^{2d}} \mathbb{1}_{\{x+y\in B\}}\, P_{X,Y}(\mathrm{d}x, \mathrm{d}y), \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

but not necessarily in terms the single distributions P_X and P_Y, unless X and Y are independent. Hence, we define the convolution between two probabilities μ and ν on $\mathbb{R}^d$ as

$$(\mu \star \nu)(B) = \int_{\mathbb{R}^d} \mu(\mathrm{d}x) \int_{\mathbb{R}^d} \mathbb{1}_{\{x+y\in B\}}\, \nu(\mathrm{d}y), \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

to deduce that $P_{X+Y} = P_X \star P_Y$, if X and Y are independent, which can be generalized to a finite sum of variables.

We can also define the mean of a X as the d-dimensional value $\mathbb{E}\{X\} = \overline{X}$ (i.e., the mean by coordinates) and the variance (or co-variance matrix) as a $d \times d$ positive definite matrix $Q = (q_{ij}) = \mathbb{E}\{(X - \overline{X})^*(X - \overline{X})\}$ with $q_{ij} = \mathbb{E}\{(X_i - \overline{X}_i)(X_j - \overline{X}_j)\}$. It is easy to check that if X_i are independent then Q is a diagonal matrix. It is cleat that the converse is not valid in general.

For instance, if X_i, $i = 1, \dots, n$ are independent random variables normally distributed with parameters r_i^2 and m_i, then $X = \sum_{i=1}^n a_i X_i$ is normally distributed with parameters $m = \sum_{i=1}^n m_i$ and $r^2 = \sum_{i=1}^n r_i^2$. In general, a $\mathbb{R}^d$-valued random variable is normally distributed with mean m (in $\mathbb{R}^d$) and variance RR^* (an strictly positive matrix, but R may be an $d \times n$ matrix) if

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

Thus $\mathbb{E}\{X_i X_j\} = \sum_{k=1}^n r_{ik} r_{jk}$ and in this case, if the co-variance matrix is diagonal then X_i are independent.

Theorem 1.1 (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 $P(A) = 1$. □

Usually this theorem is called Borel-Cantelli's Lemma. 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 $A \subset \bigcup_{i=n}^{\infty} A_i$ we have $P(A) = 0$ o $P(A) = 1$.

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

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

where $\mathcal{F}_{\infty}$ is called the *tail* σ-algebra. It is clear that $\mathcal{F}_{\infty} \subset \mathcal{F}^{\infty} = \sigma\big(\cup_n \mathcal{F}^n\big)$.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 $A \subset \bigcup_{i=n}^{\infty} A_i$ belongs to the tail σ-algebra $\mathcal{F}_{\infty}$.

Theorem 1.2 (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, since the set $\{\omega : \lim_n X_n(\omega) \text{ exists}\}$ belongs to $\mathcal{F}_{\infty}$, for any sequence $\{X_n\}$ of independent random variables, we have (1) 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.

1.2 Laws of Large Numbers

Perhaps a typical example related to independent random variables is the so-called *random walk*, which in a simple way, is described as a particle moving on a one dimensional grid, i.e., along a line by steps, following the rule of taking each step in a unit of time and arbitrary to the right of the left with a prescribed probability. Moreover, each step is taken independently of all previous steps. Depending on where is the emphasis, we may represent the n-step taken with ξ_n, which assume only the values ± 1 with probability p or $q = 1-p$, and ξ_n are independent random variables. We can get a 2-dimensional graph of the sum $\xi_1 + \cdots + \xi_n$ and study its behavior, e.g., see Chung [25, Ch 8, pp. 240–304], or continue along a similar view that begins with the simple case of Bernoulli variables, i.e., random variables X with $P\{X = 1\} = p$ and $P\{X = 0\} = 1 - p$ for some $0 < p < 1$.

Theorem 1.3 (LLN, Binomial)**.** *Let $\{X_i\}$ be a sequence of independent random variable Bernoulli distributed with parameter $p \in (0, 1)$, i.e., the partial sum $S_n = X_1 + X_2 + \cdots + X_n$ is a Binomial distributed with parameter n, p. Then*

$$\lim_n P\{|S_n/n - p| > \varepsilon\} = 0,$$

for every $\varepsilon > 0$. Moreover, we have

$$P\{|S_n/n - p| > \varepsilon\} \le 2\exp[-nh(p, \varepsilon)], \quad \forall n = 1, 2, \ldots,$$

for some positive constant $h(p, \varepsilon)$.

Proof. First note that for any random variable X with finite variance, from $c^2 \mathbb{1}_{\{|X-\overline{X}|\ge c\}} \le |X - \overline{X}|^2$ we deduce

$$c\, P\{|X - \overline{X}| \ge c\} \le \mathbb{E}\{|X - \overline{X}|^2\} = \mathrm{Var}(X - \overline{X}), \quad \forall c > 0,$$

the so-called Chebishev inequality.

Because S_n is a binomial variable, the mean and the variance can be computed, namely, $\mathbb{E}\{S_n/n\} = p$ and $\mathrm{Var}(S_n/n) = p(1-p)/n$. Hence, Chebishev inequality yields

$$P\{|S_n/n - p| > \varepsilon\} \le \frac{\mathrm{Var}(S_n/n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2},$$

and the first part follows.

Now, set $q(s) = \mathbb{E}\{\exp(sX_1)\} = 1-p+pe^s$, and because the random variables X_i are independent we have $\mathbb{E}\{\exp(sS_n)\} = q(s)^n$. Hence, choice a in $(p, 1)$ so that

$$P\{S_n/n > a\} = P\{\exp[s(S_n/n - a)] > 1\} \leq \\ \leq \mathbb{E}\{\exp[s(S_n/n - a)]\} = q(s/n)^n e^{-as},$$

for every $s > 0$. In the last term, take $t = s/n$ and then the infimum in $t > 0$ to obtain

$$P\{S_n/n > a\} \leq \exp\big[- n \sup_{t>0} \big(at - \ln q(t)\big)\big].$$

Since the function $t \mapsto at - \ln(1 - p + pe^t)$ is concave, non positive as $t \to \infty$, and has an strictly positive derivative in 0, we deduce that it has a finite strictly positive maximum value in t belonging to $(0, \infty)$. Therefore, given $\varepsilon > 0$ with $p + \varepsilon < 1$, and denoting $h_1(p, \varepsilon)$ the maximum value of the above function with $a = p + \varepsilon$, we obtain

$$P\{S_n/n > p + \varepsilon\} \leq \exp\big[- nh_1(p, \varepsilon)\big].$$

Similarly, we deduce

$$P\{S_n/n < p + \varepsilon\} \leq \exp\big[- nh_2(p, -\varepsilon)\big].$$

Setting $h(p, \varepsilon) = \min\{h_1(p, \varepsilon), h_2(p, -\varepsilon)\}$ we conclude. □

We can check that by means of Borel-Cantelli Lemma 1.1 with $A_n = \{|S_n/n - p| > \varepsilon\}$ and the second part of the previous theorem, we can show that $S_n/n \to p$ almost surely. Also, we obtain the following

Corollary 1.4 (Bernstein polynomials)**.** *Let $X_{x,n}$ be a Binomial random variable with parameters n and x, $0 \leq x \leq 1$. Next, define $B_n(x) = \mathbb{E}\{f(X_{x,n}/n)\}$ for a real valued continuous function f on the interval $[0, 1]$. Then $B_n(x)$ is a polynomial od degree n, Bernstein polynomial, and the sequence $\{B_n\}_{n\geq 1}$ converges uniformly to the function f.* □

This corollary yields a probabilistic argument to show Weierstrass approximation theorem on any bounded interval.

• *Remark* 1.5. We can show that for any finite sequence $X_1, \ldots, X_n$ of independent random variables with finite variance, we have

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n),$$

i.e., we add the variances. □

Theorem 1.6 (Strong LLN)**.** *Let $\{X_i\}$ be a sequence of independent identically distributed random variables in some probability space with mean $\mu = \mathbb{E}\{X_i\}$ and finite variance $\sigma^2 = \mathrm{Var}(X_i)$. Then the partial sums $S_n = X_1 + X_2 + \cdots + X_n$ satisfies $\lim_n S_n/n = \mu$, almost surely and in the L^2 norm.*

Proof. First, by considering $Y_i = X_i - \mu$ in lieu of X_i, we may assume $\mu = 0$ without any lost of generality. Hence $\mathbb{E}\{S_n\} = 0$ and because the random variables X_i are independent, we have $\mathrm{Var}(S_n) = \mathbb{E}\{(S_n)^2\} = n\sigma^2$, which yields $\lim_n \mathbb{E}\{(S_n/n)^2\} = \lim_n \sigma^2/n = 0$.

Take the sub-sequence $Z_n = (X_1 + \cdots + X_{n^2})/n^2$ to have

$$\sum_{n=1}^{\infty} \mathbb{E}\{Z_n^2\} = \sum_{n=1}^{\infty} \sigma^2 n^{-2} < \infty,$$

which implies that $\sum_{n=1}^{\infty} Z_n < \infty$ almost surely. In particular, if $Y_n = Z_{p(n)}$, with $p(n)$ the integer such that $p(n)^2 \leq n < (p(n)+1)^2$, then $Y_n \to 0$ almost surely. However

$$\frac{S_n}{n} - \frac{p(n)^2}{n} Y_n = \frac{1}{n} \sum_{k=p(n)^2+1}^{n} X_k$$

and then

$$\mathbb{E}\Big\{ \Big(\frac{S_n}{n} - \frac{p(n)^2}{n} Y_n \Big)^2 \Big\} = \frac{n - p(n)^2}{n^2} \sigma^2 < \frac{(p(n)+1)^2 - p(n)^2}{n^2} \sigma^2 =$$

$$= \frac{1 + 2p(n)}{n^2} \sigma^2 \leq 3\sigma^2 n^{-3/2},$$

because $p(n) \leq \sqrt{n}$. This implies that

$$\lim_n \Big(\frac{S_n}{n} - \frac{p(n)^2}{n} Y_n \Big) = 0, \quad \text{a.s.},$$

i.e., the whole sequence $S_n/n \to 0$ almost surely. □

In the above prove, we use the fact that the independent random variables has a finite variance. However, this is not necessary, a more complicated proof can be produced without assuming finite variance, e.g., see Pollard [106, Chapter 4, pp. 77–98].

1.3 Convergence of Probabilities

As mentioned early, the observable elements in a model are the distribution of a random variable, i.e., the probability measure P_X on a measurable space $(E, \mathcal{E})$, with $X : \Omega \to E$, and $(\Omega, \mathcal{F}, P)$ and abstract probability space. Simple experiments use $E = \mathbb{R}^d$ and more sophisticate models require a (complete metrizable and separable) Polish space E, and $\mathcal{E} = \mathcal{B}(E)$ is the Borel σ-algebra. Thus, a sequence of random variables $\{X_n : n \geq 1\}$ can be viewed as a sequence of probability measures $\{P_n : n \geq 1\}$ on $(E, \mathcal{E})$.

Definition 1.7. A sequence $\{P_n : n \geq 1\}$ of probability measures on a metrizable space $(E, \mathcal{B}(E))$ converges *weakly* to a probability measure P, denoted by $P_n \rightharpoonup P$, if for every continuous and bounded function $f : E \to \mathbb{R}$ we have

$$\int_E f \, \mathrm{d}P_n \to \int_E f \, \mathrm{d}P \quad \text{as} \quad n \to \infty.$$

Thus, a sequence of random variables $\{X_n : n \geq 1\}$, with $X_n \colon \Omega_n \to E$ and $(\Omega_n, \mathcal{F}_n, P_n)$ probability spaces, converges *in distribution* or *in law* to a random variable, defined on some probability space $(\Omega, \mathcal{F}, P)$ and valued in E, if the distribution P_{X_n} converges to the distribution P_X weakly, i.e., if $P_{X_n} \rightharpoonup P_X$. □

The reader may consult Jacod and Protter [68, Chapter 18, pp. 151–166] for an accessible study on weak convergence in $E = \mathbb{R}^d$.

It is clear that weak convergence makes sense also for finite (signed) Borel measures, non necessarily probability measures. Denote by $\mathbb{M}^+(E)$ (or $\mathbb{M}$) and $\mathbb{M}^1(E)$ the sets of finite (or signed) Borel measures and probability measures, respectively, endowed with the weak convergence. Note that $\mathbb{M}$ is a topological vector space (i.e., the addition and the scalar multiplication are continuous operations), and $\mathbb{M}^1$ is a closed convex set of $\mathbb{M}$. In particular, if P_n, Q_n, P and Q are probabilities such that $P_n \rightharpoonup P$ and $Q_n \rightharpoonup Q$ then $P_n + Q_n \rightharpoonup P + Q$, even if $P_n + Q_n$ and $P + Q$ are not probability measures. However, if the sequences $\{X_n : n \geq 1\}$ and $\{Y_n : n \geq 1\}$ of random variables weakly converges to X and Y then we do not necessarily have $X_n + Y_n$ weakly convergent to $X + Y$, i.e., the space of random variables endowed with the convergence in law is not a topological vector space.

Let $C_b(E)$ be the Banach spaces of all real-valued continuous and bounded functions defined on E. Thus we may use the notation $P_n \rightharpoonup P$ iff $\mathbb{E}_n\{f\} \to \mathbb{E}\{f\}$ and $P_{X_n} \rightharpoonup P_X$ iff $\mathbb{E}_n\{f(X)\} \to \mathbb{E}\{f(X)\}$, for every f in $C_b(E)$.

1.3.1 Tightness

On a topological space E, an outer measure μ^* (i.e., a monotone and sub σ-additive set function satisfying $\mu^*(\emptyset) = 0$) is called a *Borel outer measure* if all Borel sets are μ^*-measurable and a *regular Borel outer measure* if for every $A \subset \Omega$ there exists $B \in \mathcal{B}(E)$ such that $A \subset B$ and $\mu^*(A) = \mu(B)$ (since E is a Borel set with $\mu^*(E) \geq \mu^*(A)$, this condition regards only the case where $\mu^*(A) < \infty$). Remark that if $\{A_n\}$ is a sequence of μ^*-measurable sets with finite measure $\mu(A_n) < \infty$, and $B_n \supset A_n$ are Borel sets satisfying $\mu(B_n) = \mu(A_n)$, then for $A = \bigcup_n A_n$ and $B = \bigcup_n B_n$ we have $B \smallsetminus A \subset \bigcup_n (B_n - A_n)$, which implies $\mu(B \smallsetminus A) = 0$. To make the name regular Borel outer measure more manageable, in many statement we omit the terms *regular* and/or *outer*, but unless explicitly stated, we really mean regular Borel outer measure. Moreover, in an equivalent way, a *Borel measure* μ is defined on the Borel σ-algebra $\mathcal{B}(E)$ of a topological space E and its corresponding outer measure

$$\mu^*(A) = \sup\{\mu(B) \,:\, A \subset B \in \mathcal{B}(E)\}, \qquad \forall A \subset 2^E,$$

a regular Borel outer measure on Ω. Furthermore, any Borel measure on a Polish space (separable, complete, and metrizable space) is inner regular, i.e., the representation

$$\mu(B) = \sup\{\mu(K) \,:\, K \subset B,\, K \text{ compact}\}, \qquad \forall B \in \mathcal{B}(E).$$

holds true.

Hence, the values $\mathbb{E}_n\{f\}$ and $\mathbb{E}\{f\}$ for any f in $C_b(E)$ are actually determined by the values for any f in a smaller space, namely, continuous functions with compact support. However, the sequence of probability measures $P_n(A) = \mathbb{1}_A(n)$ satisfies $\mathbb{E}_n\{f\} = f(n) \to 0$, for every f in $C_b(\mathbb{R})$, with a compact support, i.e., the limit is not a probability measure.

Since any probability (Borel) measure P on a Polish space E is inner regular, in particular, for $B = E$, we deduce (Ulam's Theorem) that for every $\varepsilon > 0$ there exists a compact set $K = K_\varepsilon$ such that $P\{E \setminus K\} < \varepsilon$ or equivalently $P(K) \geq 1 - \varepsilon$, which motivates the following

Definition 1.8 (Tightness)**.** A family of probability Borel measures $\{P_i : i \in I\}$ on a metrizable space E is *tight* if for every $\varepsilon > 0$ there exists a compact set $K = K_\varepsilon$ such that $P_i\{E \setminus K\} < \varepsilon$, for every i in I. □

When we say a probability (measure) on a metrizable (or Polish) space, we mean a probability Borel measure. Actually, weak convergence and tightness can be used on a more general topological space satisfying some conditions, e.g., locally compact spaces with a countable basis (which are indeed Polish spaces) and locally compact spaces countable at infinity, among others.

The key relation between tightness and the weak convergence is given by

Theorem 1.9 (Prohorov)**.** *Let $\{P_n : n \geq 1\}$ be a tight sequence of probability measures on a metrizable space E. Then there exists a subsequence weakly convergent, i.e., there exist $\{P_{n_k} : k \geq 1\}$ and a probability P on E such that $\mathbb{E}_n\{f\} \to \mathbb{E}\{f\}$, for every f in $C_b(E)$.* □

We have several ways to establish this result, depending on which tools are used. Perhaps a direct source is the book Billingsley [15, Section I.6, pp. 35–41] or Shiryayev [121, Section III.2, pp. 314–318].

Usually, the result is proved for $E = \mathbb{R}$ and extended later to $\mathbb{R}^d$. Next, it is extended again to $\mathbb{R}^\infty$, a σ-compact space and finally to a metric space, each time, by reducing to the precedent case. An alternative, is to use the fact that $\mathbb{M}(E)$, the space of finite signed Borel measures on E is identified with the dual space $C^*(E)$ of $C(E) = C_b(E)$ for a compact space E. For a direct proof in $\mathbb{R}^d$ see Dudley [41, Theorem 9.3.3, pp. 293–294].

• *Remark* 1.10. Prohorov Theorem 1.9 is the so-called *direct* statement, which is true for even more general topological spaces. The converse of Theorem 1.9 holds if E is a Polish space, i.e., relatively weakly compact family $\mathcal{P} = \{P_i : i \in I\}$ of probabilities measures is tight, where $\mathcal{P}$ is called *relatively weakly compact* or *weakly pre-compact* if any sequence in $\mathcal{P}$ has a weakly convergent subsequence. □

Note that in view of the previous results, a characterization compact sets of a Polish space E is a key factor in proving relatively compactness of a family of probabilities on E.

1.3.2 Approximation

A function $f\colon E \to \mathbb{R}$, where (E, d) is a metric space, is called Lipschitz continuous if there exists a constant $L = L_f$ (called a Lipschitz constant) such that $|f(x) - f(y)| \le L\,\mathrm{d}(x, y)$ for every x, y in E. Thus, let $C_b^{0,1}(E)$ be the space of all Lipschitz continuous real functions defined on E. We have

Lemma 1.11. *Let E be a metric space. If f is a nonnegative function in $C^b(E)$ then there exists an increasing sequence $\{f_n\}$ of nonnegative functions on $C_b^{0,1}(E)$ such that $f_n(x) \to f(x)$ for every x in E.*

Proof. Indeed, if $A \in \mathcal{B}(E)$ and $x \in E$ then the distance from x to A is defined as $\mathrm{d}(x, A) = \inf\{\mathrm{d}(x, y) : y \in A\}$. The triangular inequality for the distance d on E implies

$$|\mathrm{d}(x, A) - \mathrm{d}(y, A)| \le \mathrm{d}(x, y), \quad \forall x, y \in E,$$

Hence, the function $x \mapsto k(x, A, r, m) = \min\{r, m\mathrm{d}(x, A)\}$, with $r, m \ge 0$, is bounded by r and is Lipschitz continuous with Lipschitz constant $L = m$, i.e., $k(\cdot, A, r, m)$ belongs to $C_b^{0,1}(E)$. If $A = A_{f,r} = \{y \in E : f(y) \le r\}$ then $0 \le k(x, A_{f,r}, r, m) \le f(x)$, for every x in E. Let $\{(r_i, m_i) : i \ge 1|\}$ be an enumeration of $\mathbb{Q}^+ \times \mathbb{N}$, where $\mathbb{Q}^+$ are the positive rational numbers and $\mathbb{N}$ are the positive integer numbers. Define

$$f_n(x) = \sup\{k(x, A_{f,r_i}, r_i, m_i) : 1 \le i \le n\}, \quad \forall x \in E$$

to have an increasing sequence $\{f_n : n \ge 1\}$ of nonnegative functions in $C_b^{0,1}(E)$ with $0 \le f_n(x) \le f(x)$. To check that $f_n(x) \to f(x)$, with $f(x) > 0$, let $\varepsilon > 0$ and $r \in \mathbb{Q}^+$ such that $f(x) - \varepsilon < r < f(x)$. Since f is continuous, $f(y) > r$ for any y in some neighborhood of x, i.e., $\mathrm{d}(x, A_{f,r}) > 0$. Hence, $k(x, A_{f,r}, r, m) = r$ for m sufficiently large, i.e., $f_n(x) > f(x) - \varepsilon$ for n such that $(r, m) \in \{(r_i, m_i) : 1 \le i \le n\}$. □

Now, we can prove

Proposition 1.12. *Let $\{P_n : n \ge 1\}$ be a sequence of probability measures on a metric space (E, d). Then $P_n \rightharpoonup P$ weakly if and only if $\mathbb{E}_n\{f\} \to \mathbb{E}\{f\}$, for every f in $C_b^{0,1}(E)$.*

Proof. We have to show that $\mathbb{E}_n\{f\} \to \mathbb{E}\{f\}$, for every f in $C_b(E)$. Recalling that $\|f\| = \sup\{|f(x)| : x \in E\}$, let $\{g_k : k \ge 1\}$ be a sequence approximating the function $g = \|f\| - f$ as in Lemma 1.11. By assumption we obtain

$$\liminf_n \mathbb{E}_n\{g\} \ge \liminf_n \mathbb{E}_n\{g_k\} = \mathbb{E}\{g_k\}, \quad \forall k,$$

and by monotone convergence we deduce $\liminf_n \mathbb{E}_n\{g\} \le \mathbb{E}\{g\}$, i.e.,

$$\limsup_n \mathbb{E}_n\{f\} \le \mathbb{E}\{f\}, \quad \forall f \in C_b(E),$$

after canceling the term in $\|f\|$. Finally, by symmetry, applying the same argument to $-f$ we complete the proof. □

1.3.3 Various Types of Convergence

Besides the convergence in law, there are other types of convergence useful with random variables, for example:

(1) *Almost surely*: $X_n \to X$ almost surely if there exists a set N in $\mathcal{F}$ with $P(N) = 0$ such that $X_n(\omega) \to X(\omega)$ for every ω in $\Omega \setminus N$;

(2) *In probability* or *stochastic*: $X_n \to X$ in probability if for any $\varepsilon > 0$ there exists an index $N = N(\varepsilon)$ such that $P\{|X_n - X| \ge \varepsilon\} < \varepsilon$ for any $n \ge N$;

(3) *In p-mean* or *in* L^p: $X_n \to X$ in p-mean if for every $\varepsilon > 0$ there exists an index $N = N(\varepsilon)$ such that $\mathbb{E}\{|X_n - X|^p\} < \varepsilon$ for every $n \ge N$.

We have the following relations for a sequence of random variables

(a) Almost surely convergence or convergence in L^p implies convergence in probability, and the converse in false;

(b) Convergence in probability implies convergence in law, and the converse in false;

(c) Convergence in L^p implies convergence in L^q, for every $1 \le q < p$, and the converse in false;

(d) Convergence in probability implies the existence of an almost surely convergence subsequence.

(e) All three types of convergence (1), (2) and (3) are compatible with arithmetic operations, e.g., if $X_n \to X$ and $Y_n \to Y$ then $X_n + Y_n \to X + Y$. However, for the convergence in law, if $X_n \to X$ in law then it does not necessarily follow that $X_n - X \to 0$ in law.

Regarding the converse of (b) and related to (e), if X_n converge to $X = c$ (constant) in law then $P\{X_n \le c + \varepsilon\} \to 1$ and $P\{X_n \le c - \varepsilon\} \to 0$, for every $\varepsilon > 0$, which yields $X_n \to c$ in probability. Actually, the expression of convergence in law of a sequence of random variables is misleading, what we really mean is the weak convergence of the corresponding probability measures. Now, the only remaining point to verify is (b). Thus, if $X_n \to X$ in probability then, for every Lipschitz function $f : E \to \mathbb{R}$ with Lipschitz constant $L = L_f$ we have

$$\big\{\omega \in \Omega : |f(X_n(\omega)) - f(X(\omega))| \ge \varepsilon\big\} \subset \big\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \ge \frac{\varepsilon}{L_f}\big\}.$$

Since $P\{|X_n - X| \geq \delta\} \to 0$ for every $\delta > 0$, we deduce $P\{|f(X_n) - f(X)| \geq \varepsilon\} \to 0$ for every $\varepsilon > 0$, i.e., $f(X_n) \to f(X)$ in probability. Hence, the dominate convergence yields $\mathbb{E}\{|f(X_n) - f(X)|\} \to 0$, for every f in $C_b^{0,1}(E)$. In view of Proposition 1.12, we deduce $P_{X_n} \rightharpoonup P_X$, namely, X_n converges in law to X. Actually, we can show directly that, every continuous function g we have

(f) if $\{X_n : n \geq 1\}$ converges in probability to X then $\{g(X_n) : n \geq 1\}$ converges in probability to $g(X)$;

(g) if $\{X_n : n \geq 1\}$ converges in law to X then $\{g(X_n) : n \geq 1\}$ converges in law to $g(X)$.

First, (f) is proved by contradiction. Indeed, if $\{g(X_n) : n \geq 1\}$ does not converge in probability to $g(X)$ then there exists a $\varepsilon > 0$ and a subsequence $\{g(X_{n_k}) : k \geq 1\}$ such that $P(|g(X_{n_k}) - g(X)| \geq \varepsilon\} \geq \varepsilon$. Since $X_{n_k} \to X$ in probability there exists a subsequence which converges almost surely, and so, for this subsequence, the continuity of g yields a contradiction. Next, for the convergence in law, i.e., (g), we use the fact that composition of continuous functions is continuous.

• *Remark* 1.13. A typical application of the Borel-Cantelli Lemma 1.1 show that following assertion regarding the converse of (a). If a sequence $\{X_n : n \geq 1\}$ of random variables satisfies $\sum_n P\{|X_n - X_{n-1}| \geq \varepsilon_n\} < \infty$ for some convergence series of positive real values $\sum_n \varepsilon_n < \infty$ then $\{X_n : n \geq 1\}$ converges almost surely. □.

The relation between convergence in law (or weak convergence) and convergence in probability is clarify in the next

Theorem 1.14 (Skorohod)**.** *Let $\{Q_n : n \geq 1\}$ be an sequence of probabilities on a Polish space $(E, \mathcal{B}(E))$ which weakly converges to Q. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and random variables $X, X_1, X_2 \ldots$, with values in E such that $X_n \to X$ almost surely, and $P_X = Q$ and $P_{X_n} = Q_n$, for every $n \geq 1$.*

For instance, a proof can be found in Ash [3, Section 7.7, pp. 332–336], for $E = \mathbb{R}$, or in Da Prato and Zabczyk [30, Theorem 2.4, pp. 33–35], for a separable Banach space E or in Skorohod [122, Section 1.6, pp. 9–14] for a Polish space. Note that the space used is $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}(\Omega)$ and P is the Lebesgue measure, which is sometime called universal probability space. In particular, if $E = \mathbb{R}$ then we may work with the distributions F_n and F, instead of the probability measure, to define (initially) $X(\omega) = \min\{x : F(x) \geq \omega\}$ and similarly X_n, and to change (later) $X_n(\omega) = X(\omega)$ for every ω in the countable set of discontinuities of X.

Typical applications of Theorem 1.14 are the following:

(1) Let $g, g_1, g_2, \ldots$ be measurable functions (from a Polish space into another Polish space) such that for every convergent sequence $x_n \to x$ we have $g_n(x_n) \to g(x)$. If X_n converges in law to X then $g_n(X_n)$ converges in law to $g(X)$. In particular, if $a_n \to a$ and $b_n \to b$ in $\mathbb{R}$ then $a_n X_n + b_n$ converges in law to $aX + b$;

(2) If X_n converges in law to X then $\mathbb{E}\{|X|\} \le \liminf_n \mathbb{E}\{|X_n|\}$;

(3) If X_n converges in law to X and $\{X_n : n \ge 1\}$ is uniformly integrable then X is integrable and $\mathbb{E}\{X_n\} \to \mathbb{E}\{X\}$.

Indeed, by means of Skorohod representation Theorem 1.14, there exists a sequences of random variables $\{Y_n : n \ge 1\}$ (defined in the universal probability space) such that $Y_n \to Y$ almost surely, and $\mathbb{E}\{f(X_n)\} = \mathbb{E}\{f(Y_n)\}$ and $\mathbb{E}\{f(X)\} = \mathbb{E}\{f(Y)\}$, for every bounded measurable function f. Thus, for every continuous and bounded function f we have $f(g_n(Y_n)) \to f(g(Y))$ almost surely, which implies

$$\mathbb{E}\{f(g_n(X_n))\} = \mathbb{E}\{f(g_n(Y_n))\} \to \mathbb{E}\{f(g(Y))\} = \mathbb{E}\{f(g(X))\},$$

i.e., $g_n(X_n)$ converges in law to $g(X)$. To establish (2) and (3), we use the monotone sequence of bounded continuous functions $x \mapsto |x| \wedge k$, $k = 1, 2, \dots$ to check that

$$\mathbb{E}\{|X|\} = \lim_k \mathbb{E}\{|X| \wedge k\} = \lim_k \mathbb{E}\{|Y| \wedge k\}, \quad \text{and}$$

$$\mathbb{E}\{|Y| \wedge k\} \le \liminf_n \mathbb{E}\{|Y_n| \wedge k\} \le \liminf_n \mathbb{E}\{|Y_n|\},$$

this is (2). Moreover, if $\{X_n : n \ge 1\}$ is uniformly integrable, (i.e., for every $\varepsilon > 0$ there exists $a = a(\varepsilon)$ such that $\mathbb{E}\{X\, \mathbb{1}_{\{|X| \ge a\}}\} \le \varepsilon$ for every n) then by taking $f(x) = |x| \mathbb{1}_{\{a \le |x| \le a+k\}}$, $a > 0$ and $k \to \infty$ we deduce that $\{Y_n : n \ge 1\}$ is also uniformly integrable. Hence, by Vitali Theorem we deduce

$$\mathbb{E}\{X_n\} = \mathbb{E}\{Y_n\} \to \mathbb{E}\{Y\} = \mathbb{E}\{X\},$$

namely (3). Note that X_n converges in law to X does not means $X_n - X$ converges in law to 0, so in (3) we do not necessarily has $\mathbb{E}\{|X_n - X|\} \to 0$.

Exercise 1.2. Consider the Hilbert cube $\mathcal{H} = [0,1]^\infty$, i.e., h belongs to $\mathcal{H}$ if and only if $h\colon \{1, 2, \dots\} \to [0,1]$ endowed with the product norm $\mathrm{d}_{\mathcal{H}}(g,h) = \sum_i 2^{-i} |g(i) - h(i)|$. Verify that $\mathrm{d}_{\mathcal{H}}(h_n, h) \to 0$ if and only if $h_n(i) \to h(i)$ for every i. Let (X, d) be a metric space with a countable dense subset $\{e_i : i \ge 1\}$ and define the map $\Phi\colon X \to \mathcal{H}$ by the formula $h = \Phi(x)$, $h(i) = \min\{\mathrm{d}(x, e_i), 1\}$. Prove $\mathrm{d}(x_n, x) \to 0$ if and only if $\mathrm{d}_{\mathcal{H}}(\Phi(x_n), \Phi(x)) \to 0$. Deduce Urysohn's Theorem, namely, any separable metric space is homeomorphic to a subset of $\mathcal{H}$, i.e., Φ is injective continuous and open. The same map Φ can be used to convert the Borel measures on X to the Borel measure on $\mathcal{H}$. □

The reader interested in functional analysis oriented to probability, may check the book by Bobrowski [18]. Also take a look at the book by Pollard [106] for a guided tour to measure theoretic probability and to Gut [56] for a more statistical point of view.

Chapter 2

Basic Probability

To introduce the following concepts we do not need to pass first on distribution theory, but the Fourier transform is nicely treated on the space of tempered distributions. Instead of separating Probability from measure theory, the objective is to integrate both concepts into analysis, at least until conditional expectation is considered.

2.1 Characteristic Functions

As mentioned early, the characteristic function of a random variable X or, properly expressed of a probability measure μ on $\mathbb{R}^d$ (actually, any finite Radon measure μ suffices), is given by

$$\hat{\mu}(\xi) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathtt{i}x\cdot\xi}\, \mu(\mathrm{d}x), \quad \forall \xi \in \mathbb{R}^d.$$

We remark that based on the fact any finite Radon measure can be uniquely considered as a tempered distribution, we deduce that characteristic function uniquely determined the initial finite Radon measure. Actually this fact can be also proved directly, essentially based on the computation of the characteristic function of the measure with density $\mathrm{e}^{-\lambda|x|^2}$.

Exercise 2.1. Beside the computation of the characteristic function of the normal distribution in the real line, we should be able to verify the following calculations:

(1) if x is a Normal random variable with mean a and variance b, i.e., with (Lebesgue) density $(2\pi)^{-1/2}\exp\big(-(x-a)^2/(2b^2)\big)$ then $\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}tx}\big\} = \exp\big(\mathtt{i}at - b^2t^2/2\big)$,

(2) if x is a Poisson random variable, i.e., $P\{x = k\} = \mathrm{e}^{-\lambda}\lambda^k/k!$, for any $k = 0, 1, \ldots$, then $\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}tx}\big\} = \exp\big(\lambda(\mathrm{e}^{\mathtt{i}t} - 1)\big)$,

(3) if x is a random variable with a uniform distribution on (a, b), i.e., with (Lebesgue) density $\mathbb{1}_{(a,b)}/(b-a)$ then $\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}tx}\big\} = \big(\mathrm{e}^{\mathtt{i}bt} - \mathrm{e}^{\mathtt{i}at}\big)/\big(\mathtt{i}t(b-a)\big)$,

(4) if x is a random variable with a triangular distribution on $(-1/a, 1/a)$, i.e., with (Lebesgue) density $\mathbb{1}_{(-1/a,1/a)}(a-a^2|x|)$ then $\mathbb{E}\{e^{itx}\} = 2(1-\cos at)/(a^2t^2)$,

(5) if x is a random variable with an exponential distribution, i.e., with (Lebesgue) density $\mathbb{1}_{(0,\infty)}e^{-\lambda x}\lambda$ then $\mathbb{E}\{e^{itx}\} = \lambda/(\lambda - it)$,

(6) if x is a random variable with a bilateral distribution, i.e., with (Lebesgue) density $e^{-\lambda|x|}\lambda/2$ then $\mathbb{E}\{e^{itx}\} = \lambda/(\lambda^2 + t^2)$.

(7) if x is a Polya random variable, i.e., with (Lebesgue) density $(1-\cos x)(\pi x^2)$ then $\mathbb{E}\{e^{itx}\} = (1-|t|)^+$.

(8) if x is a Cauchy random variable, i.e., with (Lebesgue) density $1/\big(\pi(1+x^2)\big)$ then $\mathbb{E}\{e^{itx}\} = e^{-|x|}$.

Essentially, we should calculate (3) to deduce the following expressions by using linearity and convolution, e.g., see Durrett [42, Section 2.3, pp. 91-98]. □

On the other hand, it is not hard to check that (a) $\hat{\mu}$ is continuous and (b) $\hat{\mu}$ is positive definite, i.e., for every natural number k, any ζ_i in $\mathbb{R}^d$ and any complex number z_i, $i = 1, \dots, k$ we have

$$\sum_{i,j=1}^{k} \hat{\mu}(\zeta_i - \zeta_j) z_i \bar{z}_j \geq 0,$$

where $\bar{z}$ is the conjugate of a complex number. Moreover, Bochner's Theorem (proved in the context of the Fourier transform) states exactly the converse, i.e., if a complex-valued function Φ defined on $\mathbb{R}^d$ satisfies (a) and (b) then there exists a finite Radon measure μ on $\mathbb{R}^d$ such that $\hat{\mu} = \Phi$. Clearly, this is a complete description of the Fourier transform of finite Radon measures. Instead, we show a simpler result sufficient for our purpose.

Theorem 2.1 (Lévy Continuity Theorem)**.** *Let $\{\nu_n\}$ be a sequence of probability measures on $\mathbb{R}^d$ with characteristic functions $\{\Phi_n(\xi)\}$, i.e.,*

$$\Phi_n(\xi) = \int_{\mathbb{R}^d} e^{i x\cdot\xi} \nu_n(dx).$$

(1) If ν_n converges weakly to a probability measure ν then $\Phi_n(\xi) \to \Phi(\xi)$ for every ξ in $\mathbb{R}^d$, where $\Phi(\xi)$ is the characteristic function of ν;

(2) Conversely, if $\Phi_n(\xi) \to \Phi(\xi)$ for every ξ in $\mathbb{R}^d$, where Φ is a continuous function at 0, then Φ is the characteristic function of a probability measure ν, and moreover, ν_n converges weakly to ν.

Proof. Indeed, if ν_n converges weakly to a probability measure ν then

$$\lim_n \int_{\mathbb{R}^d} f(x)\, \nu_n(dx) = \int_{\mathbb{R}^d} f(x)\, \nu(dx), \quad \forall f \in C_b(\mathbb{R}^d).$$

Since $x \mapsto e^{i x\cdot\xi}$ is a continuous and bounded function, we deduce that $\Phi_n(\xi) \to \Phi(\xi)$, for every ξ in $\mathbb{R}^d$.

The converse assertion is harder. First, we claim that if μ is a probability measure on $\mathbb{R}$ then

$$a\,\mu\big(\{x\in\mathbb{R}: a|x|\geq 2\}\big)\leq \int_{-a}^{a}\big(1-\hat{\mu}(\xi)\big)\,\mathrm{d}\xi,\quad \forall a>0. \tag{2.1}$$

Indeed, exchanging the order of the integrals we have

$$\begin{aligned}\int_{-a}^{a}\big(1-\hat{\mu}(\xi)\big)\,\mathrm{d}\xi &= \int_{\mathbb{R}}\mu(\mathrm{d}x)\int_{-a}^{a}\big(1-\cos(x\xi)\big)\,\mathrm{d}\xi = \\ &= 2a\int_{\mathbb{R}}\big(1-\frac{\sin(ax)}{ax}\big)\,\mu(\mathrm{d}x).\end{aligned}$$

Since $(t-\sin t)\geq 0$ for any t, $2(t-\sin t)/t\geq 1$ for every $t\geq 2$ or $t\leq -2$, we deduce

$$\big(1-\frac{\sin(ax)}{ax}\big)\geq \mathbb{1}_{\{a|x|\geq 2\}},$$

i.e., the desired claim.

Since the function Ψ is continuous, for every $\varepsilon>0$ there exists $a>0$ such that $2|1-\Psi(\xi)|<\varepsilon/d$ if $|\xi|<a$. Thus, if $\Psi^{(i)}(\xi_i)=\Psi(0,\dots,0,\xi_i,0,\dots,0)$ then

$$\frac{1}{a}\int_{|\xi_i|\leq a}\big(1-\Psi^{(i)}(\xi_i)\big)\mathrm{d}\xi_i\leq 2\sup_{|\xi_i|\leq a}|1-\Psi^{(i)}(\xi_i)|<\frac{\varepsilon}{d}.$$

Now, we show that the sequence $\{\nu_n\}$ of probability measures on $\mathbb{R}^d$ is tight, see Definition 1.8. To this purpose, considering ν_n as a probability measure on only one variable, the i-coordinate, with characteristic function $\hat{\mu}_n^{(i)}(\xi_i)=\Psi_n(0,\dots,0,\xi_i,0,\dots,0)$, we can use the previous claim (2.1) to obtain

$$\nu_n(\{a|x_i|\geq 2\})\leq\frac{1}{a}\int_{|\xi_i|\leq a}\big(1-\hat{\mu}_n(\xi_i)\big)\mathrm{d}\xi_i.$$

In view of the dominate convergence,

$$\lim_n\int_{|\xi_i|\leq a}\big(1-\hat{\mu}_n(\xi_i)\big)\mathrm{d}\xi_i=\int_{|\xi_i|\leq a}\big(1-\Psi^{(i)}(\xi_i)\big)\mathrm{d}\xi_i.$$

Hence, there exists $N=N(\varepsilon)$ such that $\nu_n(\{a|x_i|\geq 2\})<\varepsilon/d$, for every $n>N$. On the other hand, the σ-additivity of the measure ν_n implies that there exists $M_n>0$ such that $\nu_n(\{|x_i|\geq M_n\})<\varepsilon/d$, for every $n=1,\dots,N$. This means that for $r=\max\{M_1,\dots,M_N,2/a\}$ we have $\nu_n(\{|x_i|\geq r\})<\varepsilon/d$, for every n, which yields

$$\nu_n(\{|x|\geq r\})<\varepsilon,\quad \forall n=1,2,\dots,$$

i.e., $\{\nu_n\}$ is tight.

Next, applying Prohorov's Theorem 1.9, there exits a subsequence $\{\nu_{n_k}\}$ and a probability measure ν such that ν_{n_k} weakly converges to ν. Therefore, the characteristic function of ν satisfies $\hat{\nu} = \Psi$. Because two different probabilities cannot have the same characteristic function, any weak limit of the initial sequence $\{\nu_n\}$ must be ν, which implies that the whole sequence weakly converges to ν. □

For instance, the reader may consult Lukacs [90], a book completely dedicated to characteristic functions.

2.2 Central Limit Theorem

This is a very deep and extensive subject, we develop only a simple example, say, a mathematical glimpse.

Theorem 2.2 (Central Limit)**.** *Let $\{X_{i,n} : i = 1, \dots, n,\ n \geq 1\}$ be a countable family of identically distributed random variables in some probability space with mean $\mu = \mathbb{E}\{X_{i,n}\}$ and finite variance $\sigma^2 = \mathrm{Var}(X_{i,n})$, and such that each $\{X_{i,n} : i = 1, \dots, n\}$ is a sub-family of independent random variables. If*

$$S_n = X_{1,n} + X_{2,n} + \cdots + X_{n,n} \quad \text{and} \quad Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

then

$$P\Big\{a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\Big\} \to \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-x^2/2} \mathrm{d}x, \quad \forall a < b,$$

i.e., the distribution of Y_n converges to the normal distribution $N(0, 1)$. □

Proof. Let $\phi_{i,n}$ and Φ_n be the characteristic function of $X_{i,n} - \mu$ and Y_n, respectively. Since $X_{i,n}$ are identically distributed we have $\phi_{i,n} = \phi$, independent of (i, n); and because they are independent we obtain

$$\Phi_n(\xi) = \Big[\phi\Big(\frac{\xi}{\sigma\sqrt{n}}\Big)\Big]^n, \quad \forall n, \xi.$$

Since $\mathbb{E}\{X_{i,n} - \mu\} = 0$ and $\mathbb{E}\{(X_{i,n} - \mu)^2\} = \sigma^2$ is finite, the function $\phi(\xi)$ has two continuous derivatives, i.e.,

$$\begin{aligned}
&\phi'(\xi) = \mathtt{i}\mathbb{E}\big\{(X_{i,n} - \mu)\mathrm{e}^{\mathtt{i}\xi(X_{i,n}-\mu)}\big\} \quad \text{with} \quad \phi'(0) = 0, \quad \text{and} \\
&\phi''(\xi) = -\mathbb{E}\big\{(X_{i,n} - \mu)^2\mathrm{e}^{\mathtt{i}\xi(X_{i,n}-\mu)}\big\} \quad \text{with} \quad \phi''(0) = \sigma^2.
\end{aligned}$$

Thus, we can write

$$\phi(\xi) = 1 + 0 - \frac{\sigma^2\xi^2}{2} + \xi^2 h(\xi),$$

where the function h satisfies $h(\xi) \to 0$ as $\xi \to 0$. Hence, using the principal part of the complex-valued logarithm, we obtain

$$\Phi_n(\xi) = \exp\left\{n \ln\left(\phi(\frac{\xi}{\sigma\sqrt{n}})\right)\right\} = \exp\left\{n \ln\left(1 - \frac{\xi^2}{2n} + \frac{\xi^2}{\sigma^2 n} h(\frac{\xi}{\sigma\sqrt{n}})\right)\right\}.$$

Next, as $n \to \infty$ we deduce

$$\lim_n \Phi_n(\xi) = \mathrm{e}^{-\xi^2/2},$$

and Theorem 2.1 implies the desired result. □

Exercise 2.2. **(1)** Consider the dyadic numbers $R_n = \{i2^{-n} : i = 1, \dots, 4^n\}$, $R = \bigcup_n R_n$ and prove that $\sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \le r} = r2^n$, for every r in R.

(2) Let $\{X_{i,n} : i = 1, \dots, 4^n, n \ge 1\}$ be a countable family of identically distributed random variables with $\mathbb{E}\{X_{i,n}\} = 0$ and $\mathbb{E}\{|X_{i,n}|^2\} = 1$, and such that for every index $n \ge 1$ fixed, $\{X_{i,n} : i = 1, \dots, 4^n\}$ is a set of independent random variables. Define

$$W_{n,r} = 2^{-n/2} \sum_{i=1}^{4^n} X_{i,n} \mathbb{1}_{i2^{-n} \le r}, \quad \forall n \ge 1,\ r \in R,$$

and revise the arguments in Theorem 2.2 to show that the distribution of the sequence $\{W_{n,r} : n \ge 1\}$ converges to the normal distribution $N(0, r)$.

(3) If, besides the condition on (2), we assume that $\{X_{i,n} : i = 1, \dots, 4^n, n \ge 1\}$ is a set of independent random variables then, use the technique of Theorem 2.2 to show that the distribution of series

$$W_r = \sum_n 2^{-n} \sum_{i=1}^{4^n} X_{i,n} \mathbb{1}_{i2^{-n} \le r}, \quad \forall r \in R,$$

converges to the normal distribution $N(0, r)$. □

There are much more to say about the Central Limit Theorem, many variations and a lot of applications, the reader may check almost any book in probability to enlarge this point, e.g., Billingsley [16, Section 5.27, pp. 366-382] or Durrett [42, Chapter 2, pp. 79-172].

2.3 Conditional Expectation

The concept of independent is fundamental for probability theory and in fact distinguishes it from classical measure theory. Recall

Definition 2.3 (independence)**.** A family $\mathcal{A}$ of measurable sets is (mutually) independent (relative to the probability P) if their elements are mutually independent, i.e., if for any finite number of sets $A_1, \dots, A_n$ in $\mathcal{A}$ we have

$$P(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i). \tag{2.2}$$

Now, a family of σ-algebras is (mutually) independent if any finite number of σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$ in the family and any sets A_i in $\mathcal{F}_i$ we have (2.2). Similarly, a family of random variables is (mutually) independent if the family of their generated σ-algebras is (mutually) independent. □

Remark that if $\mathcal{A}_i \subset \mathcal{F}$ is a family of σ-algebras on a probability space $(\Omega, \mathcal{F}, P)$ indexed by $i \in I$, we are defining $\{\mathcal{A}_i : i \in I\}$ as independent (sometimes called mutually independent) if any finite number of index $J \subset I$ and for any sets A_i in $\mathcal{A}_i$, $i \in J$, we have (2.2).

Exercise 2.3. If $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{A}$ is a sub σ-algebra of $\mathcal{F}$ then denote by $L^2_0(\mathcal{A})$ the closed subspace of $L^2(\Omega, \mathcal{F}, P)$ containing all $\mathcal{A}$-measurable functions with zero mean, i.e.,

$$L^2_0(\mathcal{A}) = \big\{ f \in L^2(\Omega, \mathcal{F}, P) : f \text{ is } \mathcal{A}\text{-measurable and } \mathbb{E}\{f\} = 0 \big\}.$$

Show that two sub σ-algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ of $\mathcal{F}$ are independent if and only if $L^2_0(\mathcal{A}_1)$ is orthogonal to $L^2_0(\mathcal{A}_2)$, i.e.,

$$\mathbb{E}\{fg\} = 0, \quad \forall f \in L^2_0(\mathcal{A}_1),\ g \in L^2_0(\mathcal{A}_2).$$

Prove or disprove an analogue result for a family of σ-algebras $\{\mathcal{A}_i : i \in I\}$. □

It is clear that if $\mathcal{H}$ and $\mathcal{G}$ are two sub σ-algebras of $\mathcal{F}$, which are generated by the π-systems $\mathcal{H}_0$ and $\mathcal{G}_0$ (i.e., $\sigma(\mathcal{H}_0) = \mathcal{H}$ and $\sigma(\mathcal{G}_0) = \mathcal{G}$, recall that a π-system means a collection of subsets closed or stable under finite intersections) then $\mathcal{H}$ and $\mathcal{G}$ are independent if and only if $\mathcal{H}_0$ and $\mathcal{G}_0$ are independent, i.e., if and only if $P(H \cap G) = P(H)P(G)$ for any H in $\mathcal{H}_0$ and G in $\mathcal{G}_0$, actually

Exercise 2.4. Let $(\Omega, \mathcal{F})$ be a measurable space. Recall that a π-systems $\mathcal{F}_0$ is a subset of $\mathcal{F}$ which is stable under finite intersections, i.e., if A and B belongs to $\mathcal{F}_0$ then $A \cap B$ also belongs to $\mathcal{F}_0$. Also, we denote by $\sigma(\mathcal{F}_0)$ the minimal sub σ-algebra of $\mathcal{F}$ containing all the elements of $\mathcal{F}_0$, i.e. the σ-algThe concept of independent is fundamental for probability theory and in fact distinguishes it from classical measure theory.ebra generated by $\mathcal{F}_0$. Prove that if $\mathcal{H}$ and $\mathcal{G}$ are two sub σ-algebras which are generated by the π-systems $\mathcal{H}_0$ and $\mathcal{G}_0$, then $\mathcal{H}$ and $\mathcal{G}$ are independent if and only if $\mathcal{H}_0$ and $\mathcal{G}_0$ are independent, i.e., if and only if $P(H \cap G) = P(H)P(G)$ for any H in $\mathcal{H}_0$ and G in $\mathcal{G}_0$ (e.g., see the book by Bauer [6, Section 5.1, pp. 149–154]). □

It should be clear that given a probability space $(\Omega, \mathcal{F}, P)$, it is not possible to ensure the existence of (independent) random variables (or stochastic processes) 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 $\{1, 2, \ldots\} \times \{1, 2, \ldots\}$ into $\{1, 2, \ldots\}$, the expression $X_i = \sum_j 2^{-k(i,j)}\omega_{k(i,j)}$ for $i = 1, 2, \ldots$ 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 $[0, 1]$. 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 $i = 1, 2, \ldots$ then there exists a sequence $\{\xi_1, \xi_2, \ldots\}$ 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 $\mathcal{B}_i(S_i)$, $i = 1, 2, \ldots$, i.e., the distribution of ξ_i is exactly P_i, e.g., see Kallenberg [71, Theorem 3.19, pp. 55–57].

Definition 2.4 (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 $\mathbb{E}\{X \,|\, Z\} = \mathbb{E}\{X \,|\, \sigma(Z)\}$, 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 $\mathbb{E}\{X \mid Z = z\} = \mathbb{E}\{X \mid Z^{-1}(z)\}$ for any value z 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 2.4 should be complemented with the following existence result:

Theorem 2.5. *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 $\mathbb{E}\{X \mid \mathcal{G}\}$ given $\mathcal{G}$ is (uniquely determined up to null sets) a $\mathcal{G}$-measurable random variable satisfying

$$\int_A \mathbb{E}\{X \mid \mathcal{G}\}(\omega) P(\mathrm{d}\omega) = \int_A X(\omega) P(\mathrm{d}\omega), \quad \forall A \in \mathcal{G}.$$

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. □

Actually, it is very instructive to discuss the details on following points.

Exercise 2.5. Establish the existence for the conditional expectation on a given probability space $(\Omega, \mathcal{F}, P)$ for an integrable random variable X with respect to a sub σ-algebra $\mathcal{G}$ by two ways. Firstly (a) by means of the Radon-Nikodym theorem, i.e., on the measurable space $(\Omega, \mathcal{G})$ consider the probability measures $\nu(G) = \mathbb{E}\{X\mathbb{1}_G\}$ and $\mu(G) = \mathbb{E}\{\mathbb{1}_G\}$ satisfying $\nu \ll \mu$. Secondly (b) by means of the orthogonal projection π from the Lebesgue space $L^2(\Omega, \mathcal{F}, P)$ into the closed subspace $L^2(\Omega, \mathcal{G}, P)$, i.e., π satisfies $(X - \pi(X), Y) = 0$, for any Y in $L^2(\Omega, \mathcal{G}, P)$, where $(\cdot, \cdot)$ denotes the scalar product. □

Exercise 2.6. Let $G_1, \ldots, G_n$ be a measurable disjoint sets in probability space $(\Omega, \mathcal{F}, P)$ with $P(G_i) > 0$. If $\mathcal{G}$ is the σ-algebra generated by $\{A_1, \ldots, A_n\}$ then show that $\mathbb{E}\{X \,|\, \mathcal{G}\} = \sum_{i=1}^n p_i(X)\mathbb{1}_{G_i}$, where $p_i(X) = \mathbb{E}\{X\mathbb{1}_{A_i}\}/P(A_i)$. Finally, discuss the validity of the expression $\mathbb{E}\{X \,|\, \mathcal{G}\} = \sum_{i=1}^\infty (X, g_i) g_i$, where $(\cdot, \cdot)$ denotes the scalar product in $L^2(\Omega, \mathcal{F}, P)$, and now $\mathcal{G}$ is the σ-algebra generated by a sequence of random variables $\{g_i,\, i \geq 1\}$, which is assumed to be an orthonormal system. Perhaps, one should consider first the case when g_i assumes only a finite number of values or even $f_i = \mathbb{1}_{A_i}/\sqrt{P(A_i)}$ or when $\{g_i\}$ have disjoint supports. □

Exercise 2.7. Let X, Y be real random variables on a complete probability space. If Z is a random variable with values in some Polish space E then prove that the relation $X = \mathbb{E}\{Y \,|\, Z\}$ is characterized by the condition $\mathbb{E}\{Y\varphi(Z)\} = \mathbb{E}\{X\varphi(Z)\}$, for all $\varphi : E \longrightarrow \mathbb{R}$ which is bounded and continuous. Moreover, if E is locally compact, then the class of continuous function with compact support is sufficient to characterized the conditional expectation. Furthermore, any class of Borel functions that approximate any continuous and bounded function in

the pointwise and bounded topology is sufficient. In particular simple functions, i.e., $\mathbb{E}\{Y\mathbb{1}_{a<Z\leq b}\} = \mathbb{E}\{X\mathbb{1}_{a<Z\leq b}\}$, for every $b > a$ in $\mathbb{R}$. □

There are a couple of properties that are inherited from the integral:

(a) $X \leq Y$ a.s. implies $\mathbb{E}\{X \mid \mathcal{G}\} \leq \mathbb{E}\{Y \mid \mathcal{G}\}$ 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 $\phi\big(\mathbb{E}\{X \mid \mathcal{G}\}\big) \leq \mathbb{E}\{\phi(X) \mid \mathcal{G}\}$ 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 $-|X| \leq X \leq |X|$ yields $|Y| \leq \mathbb{E}\{|X| : \mathcal{G}\}$ with $Y = \mathbb{E}\{X \mid \mathcal{G}\}$, which can be used to deduce (i). Indeed, the definition of conditional expectation implies that $\mathbb{E}\{|Y|\mathbb{1}_{|Y|>k}\} \leq \mathbb{E}\{|X|\mathbb{1}_{|Y|>k}\}$ and $kP\{|Y| > k\} \leq \mathbb{E}\{|Y|\} \leq \mathbb{E}\{|X|\}$, 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 yields

Theorem 2.6 (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* $\liminf_{n\to\infty} \mathbb{E}\{X_n \mid \mathcal{G}\} \leq \mathbb{E}\{\liminf_{n\to\infty} X_n \mid \mathcal{G}\}$*, a.s. Moreover, if the sequence $\{X_n\}$ is uniformly integrable then* $\limsup_{n\to\infty} \mathbb{E}\{X_n \mid \mathcal{G}\} \geq \mathbb{E}\{\limsup_{n\to\infty} X_n \mid \mathcal{G}\}$*, a.s.* □

This means that, essentially, the conditional expectation behaves like 'an integral'. It is clear that in particular the dominated convergence is valid, i.e., if the sequence $\{X_n\}$ converges almost surely to limit X and $\mathbb{E}\{\sup_n |X_n|\} < \infty$ then $\lim_{n\to\infty} \mathbb{E}\{X_n \mid \mathcal{G}\} = \mathbb{E}\{X \mid \mathcal{G}\}$, almost surely. However, a more subtle key point is the following:

Theorem 2.7. *If $\{\dots,\mathcal{G}_{-2},\mathcal{G}_{-1},\mathcal{G}_0,\mathcal{G}_1,\mathcal{G}_2,\dots\}$ is a monotone increasing sequence (i.e., $\mathcal{G}_k\subset\mathcal{G}_{k+1}$, for every n) of sub σ-algebras on the probability space $(\Omega,\mathcal{F},P)$ and $\mathcal{G}_{\pm\infty}=\lim_{k\to\pm\infty}\mathcal{G}_k$, (i.e., $\mathcal{G}_{+\infty}$ is the* sup *σ-algebra, namely, generated by $\{\mathcal{G}_n,\mathcal{G}_{n+1},\mathcal{G}_{n+2},\dots\}$ for some fixed n, and $\mathcal{G}_\infty$ is the* inf *σ-algebra, namely, the intersection $\bigcap_k\mathcal{G}_k$) then $\mathbb{E}\{X\mid\mathcal{G}_k\}\to\mathbb{E}\{X\mid\mathcal{G}\}$ almost surely and in $L^1(\Omega,\mathcal{F},P)$ as $k\to\infty$, for every X in $L^1(\Omega,\mathcal{F},P)$.*

Proof. First, remark that if X belong to L^2 then the sequence $\{X_k\}$, with $X_k=\mathbb{E}\{X\mid\mathcal{G}_k\}$ satisfies $\|X_k\|_{L^2}\le\|X_{k+1}\|_{L^2}\le\|X\|_{L^2}$ and a L^2-convergence could be studied. Nevertheless, in view of Vitali's Theorem and the fact that the sequence $\{X_k\}$ is uniformly integrable, we must prove only the almost surely pointwise convergence of the sequence $\{X_k\}$ to deduce that it also converges in the L^1-norm. To this purpose, consider the case $k\ge1$, i.e., $k\to+\infty$, and choose positive integers $i<j<m<n$ and real numbers $a<b$ and define $\kappa(m;b)=\inf\{k\ge m:X_k\ge b\}$, with $\kappa=\infty$ if $X_k<b$ for every k, and

$$\begin{cases}[i,j;m;r]=\big\{\min\limits_{i\le k\le j}X_k\le a,\ \kappa(m;b)=r\big\},\\ [i,j;m,n]=\big\{\min\limits_{i\le k\le j}X_k\le a,\ \max\limits_{m\le k\le n}X_k\ge b\big\},\\ [a;b]=\big\{\liminf\limits_{n\to+\infty}X_k\le a,\ \limsup\limits_{n\to+\infty}X_k\ge b\big\},\end{cases}\tag{2.3}$$

where the set $[i,j,m;r]$ is $\mathcal{G}_r$-measurable and $[i,j;m,n]=\sum_{r=1}^n[i,j;m;r]$. Since $X_r\ge b$ on each $[i,j,m;r]$, we can write

$$\begin{cases}\displaystyle\int_{[i,j;m,n]}X\mathrm{d}P=\sum_{r=1}^n\int_{[i,j;m;r]}X\mathrm{d}P\ge\\ \qquad\displaystyle\ge\sum_{r=1}^n\int_{[i,j;m;r]}X\mathrm{d}P b\mathrm{d}P=bP\big([i,j;m,n]\big).\end{cases}\tag{2.4}$$

In the above relations, the inequality "$\le a$" or "$\ge b$" used in the definition (2.3) could be changed into a strict inequality, e.g., $\kappa'(m;b)=\inf\{k\ge m:X_k>b\}$, and

$$\begin{aligned}&[i,j;m;r)=\big\{\min_{i\le k\le j}X_k\le a,\ \kappa'(m;b)=r\big\},\\ &(i,j;m,n]=\big\{\min_{i\le k\le j}X_k<a,\ \max_{m\le k\le n}X_k\ge b\big\},\\ &[a;b)=\big\{\liminf_{n\to+\infty}X_k\le a,\ \limsup_{n\to+\infty}X_k>b\big\}.\end{aligned}$$

Hence, as $j\to\infty$, $i\to\infty$, $n\to\infty$, and $m\to\infty$ into the estimate (2.4), we truly deduce

$$\int_{[a;b]}X\mathrm{d}P\ge bP\big([a;b]\big),$$

and by symmetry (i.e., replacing x, a, b with $-x, -b, -a$)

$$\int_{[a;b]} X \mathrm{d}P \le aP\big([a;b]\big),$$

which means that $P\big([a;b]\big) = 0$, for every $a < b$. Since the set where the limit $\lim_{k\to+\infty} X_k$ does not exist as a finite value is actually equal to the countable union of all subset $[a;b]$ with $a < b$ and rational, we deduce that the sequence $\{X_k\}$ converges (as $k \to +\infty$) almost surely to some random variable $\lim_{k\to+\infty} X_k$, which certainly can be taken to be $\mathcal{G}(+\infty)$ measurable.

Proceed similarly to study the case $k \to -\infty$, choose negative integers $-n < -m$ and real numbers $a < b$ to define

$$[-n,-m] = \big\{ \liminf_{n\to-\infty} X_k \le a,\ \max_{-n\le k\le -m} X_k \ge b\big\},$$
$$[a;b] = \big\{ \liminf_{n\to-\infty} X_k \le a,\ \limsup_{n\to-\infty} X_k \ge b\big\},$$

and eventually to deduce that the sequence $\{X_k\}$ converges (as $k \to -\infty$) almost surely to some random variable $\lim_{k\to-\infty} X_k$, which certainly can be taken to be $\mathcal{G}(-\infty)$ measurable.

Finally, if A is a $\mathcal{G}_n$-measurable set then

$$\mathbb{E}\{X\mathbb{1}_A\} = \mathbb{E}\big\{\mathbb{E}\{X\mathbb{1}_A \,|\, \mathcal{G}_k\}\big\} = \mathbb{E}\big\{\mathbb{1}_A \mathbb{E}\{X \,|\, \mathcal{G}_k\}\big\},$$

for every $k \ge n$, i.e., $\mathbb{E}\{X\mathbb{1}_A\} = \mathbb{E}\{\mathbb{1}_A(\lim_{k\to+\infty} X_k)\}$. This equality remains true for any bounded $\mathcal{G}_{+\infty}$-measurable set A, by means of a monotone class argument, which proves that the random variable $\lim_{k\to+\infty} X_k$ is indeed a conditional expectation of X given $\mathcal{G}(+\infty)$.

Similarly, if A is a $\mathcal{G}_{-\infty}$-measurable set then A is also $\mathcal{G}_k$-measurable for every k, and therefore,

$$\mathbb{E}\{X\mathbb{1}_A\} = \mathbb{E}\big\{\mathbb{E}\{X\mathbb{1}_A \,|\, \mathcal{G}_k\}\big\} = \mathbb{E}\big\{\mathbb{1}_A \mathbb{E}\{X \,|\, \mathcal{G}_k\}\big\}.$$

This yields $\mathbb{E}\{X\mathbb{1}_A\} = \mathbb{E}\{\mathbb{1}_A(\lim_{k\to-\infty} X_k)\}$, which proves that the random variable $\lim_{k\to-\infty} X_k$ is indeed a conditional expectation of X given $\mathcal{G}(-\infty)$.

For instance, the reader may find interesting checking Doob [39, Part 2, Sections 1.5–7, pp. 393–397] and Schilling [117, Theorem 23.15, pp. 266–268] for more comments. □

Exercise 2.8. On a probability space $(\Omega, \mathcal{F}, P)$, let X be a real-valued random variable independent of a sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$, and f be a bounded Borel measurable function in $\mathbb{R}^2$. Define $f_1(y) = \mathbb{E}\{f(X,y)\}$. Prove that f_1 is Borel measurable and $f_1(Y) = \mathbb{E}\{f(X,Y) \,|\, \mathcal{G}\}$ almost surely. □

Sometimes, if X is a integrable random variable and $\{X_i : i \in I\}$ is a family of random variables, then $\mathbb{E}\{X \,|\, X_i,\ i \in I\}$ denotes the conditional expectation with respect to the σ-algebra $\mathcal{G} = \sigma\{X_i : i \in I\}$, generated by the family $\{X_i : i \in I\}$.

It is rather relevant to insist that the *conditional expectation* is an operator defined and valued on classes of equivalence of random variables, i.e., an operator on Lebesgue spaces, from $L^p(\Omega, \mathcal{F}, P)$ into $L^p(\Omega, \mathcal{G}, P)$, for any $1 \leq p \leq \infty$. It can be extended to *functions* such that the positive (or negative) part belongs to the above Lebesgue spaces.

Remark the notation for random variables: X or X_n or $X(n)$ or $X(\omega)$ or $X(n, \omega)$, and sometimes with lowercase letters, such as x, or x_n or $x(n)$ or $x(\omega)$ or $x(n, \omega)$, which is usually understood from the context. Now, we can discuss the concept of conditional independence (for two events or σ-algebras or random variables) given another σ-algebra or random variable).

Definition 2.8 (conditional independence)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{C}$ be sub σ-algebras of $\mathcal{F}$. We say that two measurable sets A and B are (conditional) independent given $\mathcal{C}$ if

$$\mathbb{E}\{\mathbb{1}_A \mathbb{1}_B \,|\, \mathcal{C}\} = \mathbb{E}\{\mathbb{1}_A \,|\, \mathcal{C}\}\, \mathbb{E}\{\mathbb{1}_B \,|\, \mathcal{C}\}, \quad \text{a.s.} \tag{2.5}$$

holds. Moreover, two sub σ-algebras $\mathcal{H}$ and $\mathcal{G}$ are (conditional) independent given $\mathcal{C}$ (relative to the probability P) if (2.5) 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., $\mathcal{H} = \sigma(X(i) : t \in I)$, $\mathcal{G} = \sigma(Y(j) : j \in J)$ and $\mathcal{C} = \sigma(Z(k) : k \in K)$, then (2.5) 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.

Recall that $\mathbb{E}\{\prod_i h_i(X(i)) \,|\, \mathcal{C}\}$ and $\mathbb{E}\{\prod_j g_j(Y(j)) \,|\, \mathcal{C}\}$ are defined (almost surely) as $\mathcal{C}$ measurable integrable (also, bounded because h_i and g_j are so) functions satisfying

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

and

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

for any functions h_i, g_i and c_k as above.

The definition of conditional independence applies to two random variables and a σ-algebra, i.e., a random variable X is (conditional) independent of another random variable Y given a sub σ-algebra $\mathcal{C}$ in the probability space $(\Omega, \mathcal{F}, P)$ if for any bounded and measurable functions f and g we have

$$\mathbb{E}\{f(X)\, g(Y) \,|\, \mathcal{C}\} = \mathbb{E}\{f(X) \,|\, \mathcal{C}\}\, \mathbb{E}\{g(Y) \,|\, \mathcal{C}\}, \quad \text{a.s.}$$

This means that the concept of two measurable sets A and B being (conditional) independent given (another measurable set) C (relative to the probability P) is properly defined by means of $\mathbb{1}_A$ and $\mathbb{1}_B$ as random variables and $\mathcal{C} = \{\Omega, C, \Omega \setminus C, \emptyset\}$, the σ-algebra generated by C (or equivalently $\mathbb{1}_C$). Indeed, if we use the equality $P(A \cap B \cap C)\, P(C) = P(A \cap C)\, P(B \cap C)$ then we need to add the complement equality $P(A \cap B \setminus C)\, (1 - P(C)) = P(A \setminus C)\, P(B \setminus C)$, to deduce the validity of condition (2.5). In particular, when $C = \Omega$ the conditional independence coincides with the independence concept of Definition 2.3 and each of the previous equalities is trivially satisfied.

In analogy with elementary conditional probability, where $P(A \,|\, C) = P(A \cap C)/P(C)$, we define the *conditional expectation* of a random variable X relative to a set C (with positive probability), instead of a σ-algebra $\mathcal{C}$, by means of

$$\mathbb{E}\{X \,|\, C\} = \frac{\mathbb{E}\{X\, \mathbb{1}_C\}}{P(C)},$$

i.e., expectation with respect to the conditional probability $P(\cdot \,|\, C)$. Thus assuming this notation, two measurable sets A and B are conditional independent given another set C if $\mathbb{E}\{\mathbb{1}_A\, \mathbb{1}_B \,|\, C\} = \mathbb{E}\{\mathbb{1}_A \,|\, C\}\, \mathbb{E}\{\mathbb{1}_B \,|\, C\}$, i.e., $P(A \cap B \,|\, C) = P(A \,|\, C)\, P(B \,|\, C)$ or equivalently $P(A \cap B \cap C)\, P(C) = P(A \cap C)\, P(B \cap C)$. Similarly, two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ (or two random variables X and Y, where $\mathcal{A}$ and $\mathcal{B}$ are the generated by X and y) are conditional independent given a set C if the previous condition holds for any A in $\mathcal{A}$ and B in $\mathcal{B}$. However, we cannot use the condition $\mathbb{E}\{\mathbb{1}_A\, \mathbb{1}_B \,|\, C\} = \mathbb{E}\{\mathbb{1}_A \,|\, C\}\, \mathbb{E}\{\mathbb{1}_B \,|\, C\}$, for any C in $\mathcal{C}$, as definition of conditional independent given a σ-algebra $\mathcal{C}$, since this would include $C = \Omega$ and then A and B would be independent, not just conditional independent. Thus, we need to recall that conditioning with respect to a set C yields a number, an evaluation operator. While, conditioning with respect to a σ-algebra (or a random variable) is an operator (with values into the sets of random variables) defined almost surely.

As mentioned early, conditional expectation can be derived from the orthogonal projection, i.e., if $\mathcal{G}$ is a sub σ-algebra in $(\Omega, \mathcal{F}, P)$, x is an element in the Lebesgue space $L^2(\Omega, \mathcal{F}, P)$, and $L^2(\mathcal{G})$ denotes the subspace of $L^2(\Omega, \mathcal{F}, P)$ composed by all $\mathcal{G}$-measurable functions (actually, equivalent classes) then $Y = \mathbb{E}\{X \,|\, \mathcal{G}\}$ if and only if $X = Y + Z$ where Y belongs to $L^2(\mathcal{G})$ and Z is orthogonal to $L^2(\mathcal{G})$, namely

$$Y \in L^2(\mathcal{G}) \qquad \text{and} \qquad \mathbb{E}\{(Y - X)g\} = 0, \quad \forall g \in L^2(\mathcal{G}).$$

Clearly, if X belongs only to $L^1(\mathcal{G})$ then, by density, the above condition be-

comes

$$Y \in L^1(\mathcal{G}) \qquad \text{and} \qquad \mathbb{E}\{(Y - X)g\} = 0, \quad \forall g \in L^\infty(\mathcal{G}),$$

where $L^\infty(\mathcal{G})$ is the space of bounded $\mathcal{G}$-measurable functions (actually, equivalent classes). Note that a simple argument of monotone class shows that if $\mathcal{G}$ and $\mathcal{H}$ are two sub σ-algebras and X is an element in $L^1(\Omega, \mathcal{F}, P)$, then $Y = \mathbb{E}\{X \,|\, \mathcal{G} \vee \mathcal{H}\}$ if and only if

$$Y \in L^1(\mathcal{G} \vee \mathcal{H}) \qquad \text{and} \qquad \mathbb{E}\{(Y - X)\mathbb{1}_G \mathbb{1}_H\} = 0, \quad \forall G \in \mathcal{G},\ H \in \mathcal{H},$$

where $\mathcal{G} \vee \mathcal{H}$ denotes the σ-algebra generated by $\mathcal{G}$ and $\mathcal{H}$.

Exercise 2.9. Prove that two σ-algebras $\mathcal{G}$ and $\mathcal{H}$ are independent in a probability space $(\Omega, \mathcal{F}, P)$ if and only if the subspace $L^2(\mathcal{G})$ and $L^2(\mathcal{H})$ are orthogonal on the constant functions, i.e., X in $L^2(\mathcal{G})$, Y in $L^2(\mathcal{H})$, and $\mathbb{E}\{X\} = \mathbb{E}\{Y\} = 0$ imply $\mathbb{E}\{XY\} = 0$, this is a rewording of Exercise 2.3. Next, deduce that $\mathcal{G}$ and $\mathcal{H}$ are independent if and only if $\mathbb{E}\{XY\} = \mathbb{E}\{X\}\mathbb{E}\{Y\}$ for every X in $L^2(\mathcal{G})$ and Y in $L^2(\mathcal{H})$. □

Based on the previous Exercise, we may introduce the concept of *independence given* a set, a function or an σ-algebra, e.g., we may say that two σ-algebras $\mathcal{G}$ and $\mathcal{H}$ are independent *given another σ-algebra* $\mathcal{C}$ if X in $L^2(\mathcal{G})$, Y in $L^2(\mathcal{H})$, and $\mathbb{E}\{XZ\} = \mathbb{E}\{YZ\} = 0$ for every Z in $L^2(\mathcal{C})$ imply $\mathbb{E}\{XY\} = 0$. This should agree with the concept of conditional independence.

Exercise 2.10. Show that a family of σ-algebras $\{\mathcal{G}_i : i \in I\}$ is independent (sometimes called mutually independent) if and only if for any finite subset J of indexes I, and for any random variables X_i in $L^\infty(\mathcal{G}_i)$ we have $\mathbb{E}\{\prod_{i\in J} X_i\} = \prod_{i\in J} \mathbb{E}\{X_i\}$, e.g., see Malliavin [92, Section IV.3, pp. 190–198]. □

Sometimes, we need to extend the notion of conditional expectation to random variables X which are only σ-integrable with respect to a given sub σ-algebra $\mathcal{G}$, i.e., X is a measurable functions on Ω such X is integrable on G_n, for any n, where $\{G_n\}$ is some increasing sequence of $\mathcal{G}$-measurable set satisfying $\Omega = \bigcup_n G_n$. In this case, and assuming $X \geq 0$, we define $Y = \mathbb{E}\{X \,|\, \mathcal{G}\}$ as the monotone limit of $\mathbb{E}\{X\mathbb{1}_{G_n} \,|\, \mathcal{G}\}$. Certainly, the random variable Y is the unique $\mathcal{G}$-measurable function satisfying $\mathbb{E}\{X\mathbb{1}_G\} = \mathbb{E}\{Y\mathbb{1}_G\}$, for any G in $\mathcal{G}$ with $\mathbb{E}\{|X|\mathbb{1}_G\} < \infty$, e.g., see He et al. [59, Section I.4, pp. 10–13].

On the other hand, we may consider random variables X defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in some Banach space B with norm $\|\cdot\|_B$ (not just $\mathbb{R}^d$) with its Borel σ-algebra $\mathcal{B}$ and if $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$ then $\mathcal{G}$ measurable random variable Y with values in $(B, \mathcal{B})$ is called a *conditional expectation* of X given $\mathcal{G}$ if for every element f in the dual space B' of B we have $\langle f, X\rangle = \mathbb{E}\{\langle f, Y\rangle \,|\, \mathcal{G}\}$. The uniqueness (almost surely) of follows from the definition, however, the existence of the conditional expectation $Y = \mathbb{E}\{X \,|\, \mathcal{G}\}$ needs some discussion. Indeed, if X is an integrable function (i.e., the real-valued random variable $\langle f, Y\rangle$ is integrable for every f in B') and B is separable and

reflexive Banach (i.e., the double dual space $B'' = B$) then given dense subspace B_0' of B', we can construct a maps $G : \Omega \times B_0' \to \mathbb{R}$ such that (a) for every ω in Ω the function $f \mapsto G(\omega, f)$ is linear, (b) for every f in B_0' the function $\omega \mapsto G(\omega, f)$ is $\mathcal{G}$-measurable, (c) for any ω outside of a negligible set and for every f in B_0' we have $G(\cdot, f) = \mathbb{E}\{\langle f, Y\rangle \,|\, \mathcal{G}\}$ and $|G(\cdot, f)| \le \mathbb{E}\{\|Y\|_B \,|\, \mathcal{G}\}\, \|G\|_{B'}$. Then, $G(\omega, \cdot)$ can be uniquely extended to an element in the double dual space B'' for each omega, and because B is reflexive, $G = Y$ with the desired properties.

The reader may benefice from a look at the viewpoint in Schilling [117, Chapters 22–24, pp. 248–312].

2.4 Regular Conditional Probability

As before, let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$ and consider the conditional expectation of $\mathbb{E}\{f \,|\, \mathcal{G}\}$ for the special case where the random variable f is the indicator function $\mathbb{1}_A(\cdot)$ of a set A in $\mathcal{F}$. We will refer to the conditional expectations as the *conditional probability* and denote it by $P\{A \,|\, \mathcal{G}\}$. On the other hand, we may begin with the conditional probability, i.e., $A \mapsto \mathbb{E}\{\mathbb{1}_A \,|\, \mathcal{G}\} = P\{A \,|\, \mathcal{G}\}$ a linear operator with values in $[0, 1]$ defined almost surely such that

$$P\{A \cap B\} = \mathbb{E}\big\{P\{A \,|\, \mathcal{G}\}\mathbb{1}_B\big\}, \quad \forall A \in \mathcal{F},\, B \in \mathcal{G},$$

then we define $\mathbb{E}\{f \,|\, \mathcal{G}\}$ for simple functions f and we pass to the limit for any integrable f, by using the fact that the operator $f \mapsto \mathbb{E}\{f \,|\, \mathcal{G}\}$ is a contraction in $L^1(\Omega, \mathcal{F}, P)$.

In any way, the conditional probability has some elementary properties inherited from the properties of the conditional expectation. For instance, if A and B are two disjoint sets in $\mathcal{F}$ then

$$P\{A \cup B \,|\, \mathcal{G}\} = P\{A \,|\, \mathcal{G}\} + P\{B \,|\, \mathcal{G}\} \quad \text{a.s.}$$

However, $P\{A \,|\, \mathcal{G}\}$ can be altered on a set of measure zero for each A in $\mathcal{F}$, we cannot conclude that $P\{A \,|\, \mathcal{G}\}$ (which is a random variable for each fixed A) is a countably (or finitely) additive probability measure on A in $\mathcal{F}$ for each ω outside of a null set. Technically, we have a function $P\{A \,|\, \mathcal{G}\}(\omega)$ of two variables A and ω, which is defined $A \mapsto P\{A \,|\, \mathcal{G}\}(\cdot)$ as a function of A taking values in a "class-of-equivalence" space in ω and now we want to consider this function as $\omega \mapsto P\{\cdot \,|\, \mathcal{G}\}(\omega)$ taking values in the space of probability measures, for each ω or even for almost every ω. For this to work, we need first to define the function $P\{A \,|\, \mathcal{G}\}(\omega)$ in a "dense" countable set of (A, ω) and then to extend its definition in a suitable way. A countably generated sub σ-algebra $\mathcal{G}$ is a suitable choice to handle the variable A, but some topology is required in the base space Ω to deal with ω. In short, this means that we look for a member from the above equivalence class of functions in such a way that *additivity* property (in particular order preserving and positivity) is preserved, e.g., see Taylor [132, pp. 210–226].

Let G be a set in $\mathcal{F}$ such that both G and $\Omega \setminus G$ have positive probability. In elementary probability, we define the conditional probability of a set A (in $\mathcal{F}$) *given* G by the formula $P(A \,|\, G) := P\{A \cap G\}/P\{G\}$. On the other hand, if $\mathbb{1}_A(\cdot)$ denotes the indicator (or characteristic) function of the set A, and $\sigma(G) = \{\Omega, \emptyset, G, \Omega \setminus G\}$ the σ-algebra generated by the set G, then $P\{A \,|\, \sigma(G)\} = \mathbb{E}\{\mathbb{1}_A \,|\, \sigma(G)\}$ and

$$P\{A \,|\, \sigma(G)\}(\omega) = \begin{cases} P\{A \cap G\}/P\{G\} & \text{if } \omega \in G, \\ P\{A \setminus G\}/P\{\Omega \setminus G\} & \text{if } \omega \in \Omega \setminus G, \end{cases}$$

so that both concepts are reconcilable. However, we should recall that the conditional probability given a set C is an evaluation, while given a σ-algebra is an operator (with values into the set of probability measures) defined almost surely. Simple considerations on the random variable $g(\omega) = P(A \,|\, G)\mathbb{1}_G(\omega) + P(A \,|\, \Omega \setminus G)\mathbb{1}_{\Omega \setminus G}(\omega)$ establishes that g is $\sigma[G]$-measurable and uniquely determined (almost surely) by the condition

$$\int_A g(\omega) P(\mathrm{d}\omega) = P(A \cap G), \quad \forall A \in \sigma[G].$$

It is remarkable to note that the above expression makes perfectly sense when G is negligible and gives the precise generalization quoted in the previous section. Moreover, this is better seen if the σ-algebra $\mathcal{G}$ is finitely-generated, i.e., $\mathcal{G} = \sigma\{G_1, \ldots, G_n\}$, where $P\{A \,|\, \mathcal{G})\}$ can be explicitly defined and the σ-additive condition is easily checked.

Given any event A, when $P(B) > 0$ and $B \in \mathcal{G}$, we have

$$P\{A \,|\, B\} = \frac{1}{P(B)} \int_B P\{A \,|\, \mathcal{G}\}(\omega) P(\mathrm{d}\omega).$$

Moreover we recall that two events A and B are said conditionally independent with respect to (or given) the *sub σ-algebra* $\mathcal{L}$ if

$$P(A \cap B \,|\, \mathcal{L}) = P(A \,|\, \mathcal{L})\, P(B \,|\, \mathcal{L}). \tag{2.6}$$

Analogously, $\mathcal{H}$ and $\mathcal{G}$ are called independent given (or with respect to) $\mathcal{L}$, a sub σ-algebra of $\mathcal{F}$, (relative to the probability P) if (2.6) is true for any sets $A \in \mathcal{H}$, $B \in \mathcal{G}$, see Definition 2.8.

It is interesting to note that given two random variables X and Y with a joint probability density function $f_{X,Y}(x, y)$, the functions

$$f_X(y) = \int f_{X,Y}(x, y)\mathrm{d}y, \qquad f_Y(y) = \int f_{X,Y}(x, y)\mathrm{d}x,$$

are the probability density for X and Y, and the *elementary conditional* probability density function $f_{X|Y}$ of X given Y is defined by

$$f_{X|Y}(x, y) = \begin{cases} \dfrac{f_{X,Y}(x, y)}{f_Y(y)} & \text{if } \ f_Y(y) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then for any Borel measurable function h such that

$$\mathbb{E}\{|h(X)|\} = \int |h(x)|\, f_X(x)\mathrm{d}x < \infty,$$

we can define the function

$$g(y) = \int h(x) f_{X|Y}(x, y)\mathrm{d}x$$

which provides a version of the conditional expectation of $h(X)$ given $\sigma(Y)$, i.e., $g(Y) = \mathbb{E}\{h(X) \,|\, Y\}$. Moreover, the function

$$\omega \mapsto \int_A f_{X|Y}(x, Y(\omega))\mathrm{d}x$$

is a regular version of the conditional probability of X given Y or given $\sigma(Y)$, usually denoted by $P\{X \in A \,|\, Y\}$.

Another way of looking at the same problem is to discuss *conditional distributions* of a given random variable X and a sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$.

Exercise 2.11. Let $\mathcal{G}$ be a finitely-generated σ-algebra, i.e., $\mathcal{G} = \sigma(F_1, \ldots, F_n)$. First, show that $\mathcal{G}$ can be expressed as $\sigma(G_1, \ldots, G_m)$, where the sets $G_1, \ldots, G_m$ are disjoint and minimal in the sense that any proper subset of $\{G_1, \ldots, G_m\}$ does not generate $\mathcal{G}$. Actually, $\{G_1, \ldots, G_m\}$ is a partition and the set G_i are called *atoms* of $\mathcal{G}$, which has exactly 2^m elements. Second, gives an explicit expression of $P\{A \,|\, \mathcal{G})\}(\omega)$ in term of the family of sets $G_1, \ldots, G_m$. Third, if X is a simple random variable (i.e., having a finite number of values, say $x_1, \ldots, x_m$ with $P\{X = x_i\} > 0$ and $\sum_i P\{X = x_i\} = 1$) then show that $\sigma(X)$ (i.e., the minimal σ-algebra for which X is measurable) is finitely-generated, calculate $P\{A \,|\, X = x_i\}$, for $i = 1, \ldots, m$ and consider the function $x \mapsto P(x, A)$ defined as $P(x, A) = P\{A \,|\, X = x_i\}$ if $x = x_i$ for some $i = 1, \ldots, m$, and $P(x, A) = P(A)$ otherwise. Fourth, show that the expression $P(X, A)$ is a regular conditional probability of A given X, i.e., for any A measurable set we have $P\{A \,|\, X\} = P(X, A)$ almost surely. □

Definition 2.9 (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 $\mathbb{E}\{\mathbb{1}_A \,|\, Z\} = \mathbb{E}\{\mathbb{1}_A \,|\, \sigma(Z)\}$. 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 2.10 (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:

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

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

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

4.- 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

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

and this 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 $P\{A_i \smallsetminus K_i\,|\,\mathcal{G}\}(\omega)| < \delta 3^{-i}$. 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 $\Omega \smallsetminus N$, 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. □

The reader is referred to Morimoto [100, Sec 2.3, pp. 61–64] for some more detailed and self-contained arguments on probing the previous theorem about regular conditional probability in Polish spaces. Also, note that the condition $P\{B\,|\,\mathcal{G}\}(\omega) = \mathbb{1}_B(\omega)$, 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 $P\{\Omega\,|\,\mathcal{G}\}(\omega) = 1$ and $P\{\emptyset\,|\,\mathcal{G}\}(\omega) = 0$ on the condition 2 of the construction given on the above proof to obtain a regular conditional probability.

• *Remark* 2.11 (conditional distribution). 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 $P_X\{A\,|\,\mathcal{G}\}$ i.e., (a) for each A in $\mathbb{E}$ the function $\omega \mapsto P\{X^{-1}(A)\,|\,\mathcal{G}\}(\omega)$ is $\mathcal{G}$-measurable, (b) for every $A \in \mathbb{E}$ and $B \in \mathcal{G}$ we have

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

and (c) for each ω in Ω the function $A \mapsto P\{X^{-1}(A)\,|\,\mathcal{G}\}(\omega)$ 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 $\mathcal{G}$. □

• *Remark* 2.12. It is clear that the concept of *conditional expectation* or *regular conditional probability* can be applied to σ-finite measures, instead of just probability measures. □

• *Remark* 2.13 (regular conditional distribution). Let $(\Omega, \mathcal{F}, P)$ be a probability space as in Theorem 2.10, and let X be a given a random variable with values in some Polish space $(E, \mathcal{E})$. Then the regular conditional probability (in this case, also called regular conditional distribution given X) exists for $\mathcal{G} = X^{-1}(\mathcal{F})$, the

σ-algebra generated by X. Thus, for any A in $\mathcal{F}$, the function $\omega \mapsto P\{A \,|\, \mathcal{G}\} = P\{A \,|\, X\}$ is a real-valued measurable with respect to $X^{-1}(\mathcal{F})$ and therefore there exists a (real valued) Borel measurable function $x \mapsto P(x, A)$ on $(E, \mathcal{E})$ (which depends on X) such that $P\{A \,|\, X\}(\omega) = P(X(\omega), A)$ almost surely. This is called the transition probability function $P(x, A)$ given X, and usually denoted by $P\{A \,|\, X = x\}$. Note the two defining properties: (a) for each A in $\mathcal{F}$, the function $x \mapsto P(x, A)$ is measurable from $(E, \mathcal{E})$ into $[0, 1]$, and (b) for any x in E, the function $A \mapsto P(x, A)$ is a probability measure on $(\Omega, \mathcal{F})$. Clearly, if P_X denotes the probability distribution of X then for any A in $\mathcal{F}$ the function $x \mapsto P\{A \,|\, X = x\}$ is uniquely determinate outside of a P_X-negligible set. Moreover, condition (b) in Theorem 2.10 can be rewritten as

$$P\big\{A \cap \{X \in C\}\big\} = \int_{\{X \in C\}} P\{A \,|\, X\} \, \mathrm{d}P, \quad \forall A \in \mathcal{F},\, C \in \mathcal{E},$$

and in particular, for $C = \{x\}$, we have

$$P\{A \,|\, X = x\} = \frac{P\big\{A \cap \{X = x\}\big\}}{P\{X = x\}},$$

for any x in E with $P\{X = x\} > 0$ and any A in $\mathcal{F}$. For instance, the reader may take a look at Taira [131, Chapter 2], among other books, to read a more detailed account on this point. □

• *Remark* 2.14. Related to the previous discussion on regular conditional distribution, of particular interest is the case of two E-valued random variables X and Y, for which we consider the joint distribution P_{XY} on the product Polish space E^2. Independently of the initial probability space, the image $(E^2, \mathcal{E}^2, P_{XY})$ is a probability space satisfying the assumptions of Theorem 2.10. Thus, we may consider the conditional probability distribution given the sub σ-algebra $\mathcal{E}_x$ generated by the projection $(x, y) \mapsto x$ of for the first coordinate, i.e., $P_{XY}\{A \,|\, \mathcal{E}_x\} = P\{(X, Y) \in A \,|\, X\}$, with A in $\mathcal{E}^2$. Thus, we obtain the conditional probability distribution of Y given X, namely $P_{Y|X}(B) = P\{Y \in B \,|\, X\}$, after choosing $A = E \times B$. Hence, there exists a transition function $P(x, B)$ of Y given X, which depends on X and Y and is denoted by $P\{Y \in B \,|\, X = x\}$, i.e., $P\{Y \in B \,|\, X\}(\omega) = P(X(\omega), B)$, for any B in $\mathcal{E}$, see Exercise 2.11. □

It is now clear that 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 any 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 re-stated 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. Any Polish space possess these properties.

Chapter 3

Canonical Sample Spaces

Let $\mathtt{S}_\infty$ be the set of non-decreasing divergent sequences with values in $[0,\infty]$, i.e., $\{s_k : k \geq 1\}$, $0 \leq s_k \leq s_{k+1}$, for every $k \geq 1$, $s_k \to \infty$, and the value $s_k = \infty$ is allowed. This is a locally compact Polish (complete, separable and metrizable) space, with the product topology, and the sum is defined (and continuous) but the subtraction is not always possible, similar to the interval $[0,\infty]$. For a given locally compact Polish space E not containing the symbol ∂, denote by $\overline{E}_{\mathtt{S}} = ([0,\infty] \times E) \cup \{\partial\}$ the one-point compactification of the locally compact Polish space $[0,\infty[\times E$, and let $S_\partial(E) \subset \overline{E}_{\mathtt{S}}^{\infty}$ be the space of functions ω from $\mathbb{N} = \{1,2,\ldots\}$ into $\overline{E}_{\mathtt{S}}$ with the property that $s \mapsto \omega_k(s,e)$ belongs to $\mathtt{S}_\infty$ and $\omega_k(\infty,e) = \partial$, for every e in E. Similarly, if $\mathtt{S}_\infty$ is replaced by the space $\widetilde{S}_\infty \subset [0,\infty]^\infty$ of all divergent sequences with values in $[0,\infty]$ then the Polish $\widetilde{S}_\partial(E)$ is also defined. Each element ω in $\widetilde{S}_\partial(E)$ can be regarded as either the jumps or the value at discontinuity points of a function without discontinuities of the second class from $[0,\infty[$ into E, since they can have only a countable number of discontinuities. Moreover, the product spaces $C([0,\infty[;E) \times \widetilde{S}_\partial(E)$, or $C([0,\infty[;E) \times \widetilde{S}_\partial(E^2)$, or $C([0,\infty[;E) \times \widetilde{S}_\partial(E^2)$, or $D([0,\infty[;E) \times \widetilde{S}_\partial(E)$, can be used to represent function without discontinuities of the second class, for instant, $C([0,\infty];E) \times \widetilde{S}_\partial(E) \sim D([0,\infty];E)$, i.e., a function ω' in $D([0,\infty];E)$ can be regarded

========== ** TO BE CHECKED ** ==========

First, on a measurable space $(\Omega,\mathcal{F})$, we recall that a E-valued ($E \subset \mathbb{R}^d$, some $d \geq 1$) random variable x is a measurable on $(\Omega,\mathcal{F})$. Thus a E-valued "general" random (or stochastic) process is a family $X = \{x_t : t \in T\}$ of random variables. The sample paths of a random process are the functions $\omega \mapsto \{x_t(\omega) : t \in T\}$, i.e., the graphs of a function $t \mapsto x(t,\omega)$ for ω in Ω. Thus the paths are elements in the Cartesian product space E^T. As we may expect, the case of a continuous parameter set T (e.g., an interval) presents measurability problems. Without giving full details, we have to work with the product Borel σ-algebra $\mathcal{B}^T(E)$, i.e., the smallest σ-algebra on E^T such that the projections form E^T into E are Borel measurable. Actually, we can show that a set A belongs to $\mathcal{B}^T(E)$ if

and only if there exists a countable subset of indices, $T_0 \subset T$, and Borel sets $B_t \in \mathcal{B}(E)$, for every $t \in T_0$ such that $A = \{X \in E^T : x_t \in B_t,\ \forall t \in T_0\}$, i.e., only countable many constraints are allowed. By contrast, we say that $\mathcal{B}^T(E)$ is too small for the big space E^T, e.g., a simpleton (a set with only one point) is a closed set in the product topology E^T, but it is not $\mathcal{B}^T(E)$-measurable.

Thus $(E, \mathcal{B})$-valued general random (or stochastic) process X is a $(E^T, \mathcal{B}^T)$-valued random variable. When a probability measure P is given on $(\Omega, \mathcal{F})$, we say that $P_{X,I}(B) = P\big(X^{-1}(B \times E^{T \setminus I})\big)$ is the family of finite-distribution of X, for any B in $B^I(E)$ and for I a finite subset of indices of T, i.e., $P_{X,I}$ is the X image of P on E^I. The family of finite-distribution carries all the "practical" statistic properties of X, but all properties of X are identifies by the X image of P on E^T, i.e., the measure PX^{-1} on $\mathcal{B}^T(E)$.

To make this manageable, we desire to replace the product space E^T with some better space, which are referred to as canonical sample spaces, and a E-valued random (or stochastic) process with sample path in a topological space $S \subset E^T$ is regarded as a random variable with values in the canonical space S. A natural question (which is reserved for a more advance course) is realizations of random processes, i.e., the construction of a probability space $(\Omega, \mathcal{F}, P)$ and a process X when the finite-dimensional distributions are given. In what follows, we discuss some common candidates for sample spaces with $E = \mathbb{R}$ and $T = [0, \infty)$, other cases ($E \subset \mathbb{R}^d$ and T an interval of $\mathbb{R}$) are treated analogously.

The main purpose of this chapter is to go behind the classic setting of real valued continuous functions on $T = [0, \infty[$ with the locally uniformly convergence, and give a quick discussion on the space of right-continuous functions having left-hand limits (cad-lag) with the so-called Skorokhod Topology. The interested reader may take a look at more advance books, e.g., Billingsley [15, Chapters 2 and 3], Jacod and Shiryaev [69, Chapter VI], Liptser and Shiryayev [88, Chapter VI], and Pollard [105, Chapter VI], among others.

3.1 Continuous and cad-lag Functions

A large category of sample spaces are the so-called Polish spaces, i.e., complete separable metrizable spaces $S \subset \mathbb{R}^{[0,\infty)}$. The fact that 'metrizable' is mostly used instead of 'metric' is related to the marginal interest in the metric itself (which could be very complicated to treat), since a good characterization of closed sets is given by the understanding of the convergence of sequences (the space is assumed to be separable), and also a clear understanding of totally bounded sets is necessary to deal with compact subsets, where the metric is not really involved, mainly the topology on the space is to be understood.

A convenient and requested property is that the product σ-algebra $\mathcal{B}^{[0,\infty)}(\mathbb{R})$ (i.e, the σ-algebra generated by cylindrical sets) coincides with the Borel σ-algebra of S. Therefore, the discussion of probability measures on S is treatable and, in particular, a good characterizations of compact sets is necessary.

The most typical canonical space is $C([0, \infty[)$, the space of real-valued continuous functions on $[0, \infty)$, which becomes a Polish space with the locally uniform

convergence. The reader can verify that a subset K of $C([0,\infty[)$ is pre-compact if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|\omega(t)| \le 1/\varepsilon$ and $|\omega(t)-\omega(s)| \le \varepsilon$, for every s,t in $[0,1/\varepsilon]$ and $|t-s| < \delta$, for every ω in K. Recall that $C([0,\infty[)$ is also a separable Fréchet space. Certainly, random variable with values in $C([0,\infty[)$ are called *random processes with continuous path.*

On the other side of the spectrum, a large sample space is the the Schwartz space of tempered distributions $\mathcal{S}'([0,\infty[)$, which is a Polish space,[1] this is a separable complete locally convex topological vector space, where bounded sets are pre-compacts. Usually, random variable with values in $\mathcal{S}'([0,\infty[)$ are call *generalized random processes.* A negative aspect is a partial (instead of a complete) order in the time (index) variable, i.e., instead of looking at a random process as a family X_t of random variables indexed by the time t in $[0,\infty[$, a generalized random process is a family X_φ of random variables indexed by the time φ in $\mathcal{S}([0,\infty[)$. This means that the random variable $t \mapsto X(t)$ is now regarded as an element in $\mathcal{S}'([0,\infty[)$, defined by

$$X_\varphi = \int_0^\infty X_t \varphi(t) \mathrm{d}t, \quad \forall \varphi \in \mathcal{S}([0,\infty[),$$

which assumes local integrability in t with a slow growth at ∞.

Another sample space is $L^2(]0,\infty[, \mathrm{d}\alpha)$, where α is an increasing function. This space is useful when studying stochastic integrals. There is also the space $L^2(\mathbb{R}^n)$ to study processes with finite second moment, and all the theory of spectrum via the Fourier transform.

However, we want to consider the sample space $D([0,\infty[)$ of all cad-lag functions from $[0,\infty[$ into $\mathbb{R}$, i.e., continuous functions from the right having limits from the left, for every ω in $D(]0,\infty[)$ we have

$$\omega(t) = \omega(t+) = \lim_{s\to t,\, s>t} \omega(s) \quad \forall t \ge 0, \quad \text{and} \quad \omega(t-) = \lim_{s\to t,\, s<t} \omega(s)$$

exits and is finite for every $t > 0$, and sometimes, we complete the definition by adding $\omega(0-) = 0$. We may endow $D([0,\infty[)$ with the locally uniform convergence to obtain a Fréchet space, which is not separable and contains $C([0,\infty[)$ as a closed subspace. Indeed, the uncountable family $\{\omega_r : r > 0\}$ with $\omega_r(t) = \mathbb{1}_{t<r}$ are at mutual distance $\|\omega_r - \omega_s\|_\infty = 1$ for $r \ne s$.

In analysis, functions having one-sided limits (but not necessarily cad-lag) have been called regulated functions (e.g., see Dieudonne [37, Section VII.6, pp. 145-146]).

Lemma 3.1. *Let (E, d_E) be a complete metric space. A function $\varphi : [a,b] \to E$ has one-sided limits (i.e., $\varphi(t+)$ exists and is finite for every t in $[a,b[$, and $\varphi(t-)$ exists and is finite for every t in $]a,b]$) if and only if φ is a uniform limit of a sequence $\{\varphi_n : n \ge 1\}$ of step functions (i.e., for each φ_n there exists a partition $a = x_0 < x_1 < \cdots < x_k = b$ such that φ_n is constant on the open interval $]x_{i-1}, x_i[$, for $i = 1, \ldots, k$).*

[1] Note that the full space of distributions $\mathcal{D}'([0,\infty[)$ is not a Polish space, because it not metrizable.

Proof. To show the necessity, the existence of lateral limits for φ implies that for every $\varepsilon > 0$ and t in $[a, b]$ there exists an open interval $U(t, \varepsilon) =]x(t, \varepsilon), y(t, , \varepsilon)[$ containing t such that $\mathrm{d}_E\big(\varphi(s), \varphi(r)\big) < \varepsilon$ if both times s, r belong to either $]x(t, \varepsilon), t[\cap[a, b]$ or $]t, y(t, \varepsilon)[\cap[a, b]$. Because $[a, b]$ is compact, there exist a finite number of open intervals $U(t_i, \varepsilon)$ covering $[a, b]$ and so we can relabel the times $a, b, t_i, x(t_i, \varepsilon), y(t_i, \varepsilon)$ as a strictly increasing finite sequence $a = x_0 < x_1 < \cdots < x_k = b$ such that for every s, r in $]x_{i-1}, x_i[$ we have $\mathrm{d}_E\big(\varphi(s), \varphi(r)\big) < \varepsilon$. Thus, choose x_i^* in $]x_{i-1}, x_i[$ and define the step function $\varphi_\varepsilon(t) = \varphi(x_i^*)$ for every t in $]x_{i-1}, x_i[$ and $\varphi_\varepsilon(x_i) = \varphi(x_i)$, to deduce that $\mathrm{d}_E\big(\varphi_\varepsilon(t), \varphi(t)\big) < \varepsilon$, for every t in $[a, b]$.

To check the converse, if φ is the uniform limit of a sequence φ_n of step function, then for every $\varepsilon > 0$ there exists n such that $\mathrm{d}_E\big(\varphi_n(t), \varphi(t)\big) < \varepsilon$, for every t in $[a, b]$. Now, for every t there is an open interval $]c, d[$ containing t and such that $\mathrm{d}_E\big(\varphi_n(s), \varphi_n(r)\big) < \varepsilon$, whenever both s, r belong to either $]c, t[\cap[a, b]$ or $]t, d[\cap[a, b]$. This implies for the same s, r that $\mathrm{d}_E\big(\varphi(s), \varphi(r)\big) < 2\varepsilon$ and because E is complete, the lateral limit exists at any t. □

• *Remark* 3.2. The proof in previous Lemma 3.1 allows us to affirm that if $\varphi_n \to \varphi$ uniformly then the cad-lag version $\varphi_n(t+)$ converges to the cad-lag version $\varphi(t+)$ uniformly. □

We may insists in working on a non-separable complete metrizable space, but measurability problems appear, the Borel σ-algebra does not agree with the σ-algebra $\mathcal{B}$ generated by all cylindrical sets

$$\big\{\omega \in D([0, \infty[) : \omega(t_i) \in B_i,\ i = 1, \ldots, n\big\},$$

for any n, and Borel subsets B_i of $\mathbb{R}$, e.g., see Pollard [105, Chapter V, pp. 89–121]. As seen in a more advanced course of Probability, we prefer a topology that makes $D([0, \infty[)$ a Polish space, i.e., a complete separable metrizable space. This is known as the Skorokhod topology.

If your interest is Markov processes with continuous paths (e.g., diffusion processes) then you may skip the sequel, what was developed early is sufficient to handle most situations. Processes are realized in the space of the continuous functions $C([0, \infty[)$, which is a separable complete and metrizable space and also, a vector topological space. Completeness and separability are two very important properties when dealing with probability measures defined on the Borel σ-algebra. Indeed, suitable characterizations of compact sets depend on completing the space and the cylindrical and Borel σ-algebras are actually the same one, under the separability condition.

To study Markov process with possible discontinuities of the first kind (e.g., diffusion processes with jumps), we are forced to consider the space $D([0, \infty[)$, which is also separable complete and metrizable space, but not a topological vector space, i.e., the addition is not necessarily a continuous operation. This is not a serious handicap from the probabilistic viewpoint.

In most of the cases, the question of having a convergent sequence of processes is transformed into a convergent of probability measures on some Polish

space. Thus, characterizations of compact sets in Polish space is a key point to establish the compactness of a sequence of probability measures. Several very well tools are available (and well known) when dealing with the Polish space of the continuous functions $C([0,\infty[)$. However, we need to develop a little more the tools for the Polish space cad-lag functions $D([0,\infty[)$.

3.2 Modulus of Continuity

In what follows in this Chapter, we discuss briefly (only with partial proofs) the canonical space $D([0,\infty[))$. For a complete study with full proofs the reader is refer to, e.g., Billingsley [15, Chapter 3, pp. 109–153], for a comprehensive study on the Skorokhod space $D([0,1])$, and to Jacod and Shiryaev [69, Sections VI.1 and VI.2, pp. 288–310] or Ethier and Kurtz [47, Section 3.5, pp. 116–154]. for the case $D([0,\infty[;\mathbb{R}^d)$. Also the reader may check the book Bass [5, Chapters 34, pp. 259–268], Gikhman and Skorokhod [53, Chapter 4, pp. 144–173], among others.

All arguments below apply to cad-lag functions with values in some Polish space E, i.e., in $D([0,\infty[;E)$ instead of just $D([0,\infty[)$. Moreover, the time $t=0$ plays an special role (any ω is virtually continuous at 0), and for $t=\infty$, it is convenient to consider $D([0,\infty[)=\bigcup_{n>0} D([0,n])$, as long as no special role is played by the times $t=n$. Also, there are small differences when discussing the spaces $D([0,\infty[)$, $D(]-\infty,+\infty[)$ and $D([0,T])$, just a matter of a good role for the finite end-points.

As mentioned early, it is interesting to remark that $D([0,\infty),\mathbb{R}^d)$ is not a topological vector space, i.e., in the Skorokhod topology, the convergences $\alpha_n\to\alpha$ and $\beta_n\to\beta$ does not necessarily imply that $\alpha_n+\beta_n$ converges to $\alpha+\beta$, unless α (or β) belongs to $C([0,\infty),\mathbb{R}^d)$. Moreover, if $\{\alpha_n\}$ are continuous functions then the limit (in the Skorokhod topology) function α, is also continue. Moreover, the topology in $D([0,\infty),\mathbb{R}^d)$ is strictly stronger that the product topology in $D([0,\infty),\mathbb{R}^{d_1})\times D([0,\infty),\mathbb{R}^{d_2})$, $d=d_1+d_2$.

Recall that a function ω is said to have a discontinuity of the first kind at t if $\omega(t-)$ and $\omega(t+)$ exists but differ and $\omega(t)$ is between them. Any element in $D([0,\infty[)$ is continuous or at most has a discontinuity of the first kind. Any cad-lag function ω defined on $[0,T]$ can be (canonically) extended to a cad-lag function in $[0,+\infty[$ by means of the expression $\omega(t)=\omega(t\wedge T)$, i.e., $\omega(t)=\omega(T)$ for $t>T$. Sometimes and mainly for a notation preference, we may add the condition either $\omega(0-)=0$ or $\omega(0-)=\omega(0)$.

For a continuous function, the modulus of continuity is defined by

$$w_c(\omega,r,T)=\sup_{0\le s<t\le T,\ |t-s|<r}|\omega(t)-\omega(s)|, \tag{3.1}$$

with the key property that ω belongs to $C([0,\infty[)$ if and only if $w_c(\omega,h,T)\to 0$ as $h\to 0$, for every fixed $T>0$. In general, the expression

$$\mathrm{osc}(\omega,I)=\sup_{s,t\in I}|\omega(t)-\omega(s)| \tag{3.2}$$

defines the oscillation of ω on the interval I of $\mathbb{R}$, and

$$w_c(\omega, h, T) = \sup_{0\le t\le T-h} \operatorname{osc}(\omega, [t, t+h])$$

is an alternative definition for the modulus of continuity.

Note that a piecewise continuous function (i.e., for every $T > 0$ there exists a finite partition $0 = t_1 < t_2 < \cdots < t_n = T$ such that ω can be extended to a continuous function on each closed subinterval $[t_{i-1}, t_i]$, $i = 1, \ldots, n$) may have at most discontinuities of the first kind, however, a function having at most discontinuities of the first kind is not necessarily a piecewise continuous function. Certainly, a cad-lag function may be only discontinuous at a countable set, but this set of discontinuity may be dense in $[0, \infty[$. For instance, let $\{r_k : k \ge 1\}$ be an enumeration of the positive rational numbers and define $\omega_n(t) = \sum_{k=1}^{n} a_k \mathbb{1}_{\{t\ge r_k\}}$ and $\omega(t) = \sum_{k=1}^{\infty} a_k \mathbb{1}_{\{t\ge r_k\}}$, where $\{a_k : k \ge 1\}$ is a numerical sequence such that $\sum_{k=1}^{\infty} |a_k| < \infty$. It is clear that the cad-lag function ω_n satisfies $\omega_n(r_k) - \omega_n(r_k-) = a_k$, for every $n \ge k$. Because $\sup_{t\ge 0} |\omega_n(t) - \omega(t)| \le \sum_{k>n} |a_k| \to 0$, the limit function ω is also a cad-lag (purely jumping bounded variation) function, which is discontinuous at every rational number and is continuous at each irrational number. However, we have

Lemma 3.3. *If ω belongs to $D([0, \infty[)$ then for every $\varepsilon > 0$ there exists a finite sequence of times $0 = t_0 < t_i < \cdots < t_{n-1} < 1/\varepsilon \le t_n$ such that $\operatorname{osc}(\omega, [t_{i-1}, t_i[) \le \varepsilon$, for every $i = 1, \ldots, n$. In other words, this means that any function in $D([0, \infty[)$ can be approximated by right-continuous step functions, namely, if $\omega_\varepsilon(t) = \omega(t_i)$ for every t in $[t_{i-1}, t_i)$ then $|\omega(t) - \omega_\varepsilon(t)| \le \varepsilon$ for every t in $[0, 1/\varepsilon]$.*

Proof. Indeed, first we take ω in $D([0, \infty[)$ and some $\varepsilon > 0$. For any $T > 0$, consider a ε-decomposition of the form $0 = t_0 < t_1 < \cdots < t_{n-1} < T \le t_n$ such that $\operatorname{osc}(\omega, [t_{i-1}, t_i[) \le \varepsilon$, for every $i = 1, \ldots, n$. Since $\omega(0) = \omega(0+)$ there exists $T > 0$ sufficiently small so that such a decomposition (with n=1) is possible. Now, define T^* the supremum of all those T, where a finite ε-decomposition $[0, T) = \bigcup_i [t_{i-1}, t_i)$ is possible. If T^* is finite, then because $\omega(T^*-)$ exists, we can decomposes $[0, T^*)$ and since $\omega(T^*) = \omega(T^*+)$, we would be able to decompose some interval $[0, T) \supset [0, T^*]$, which is a contradiction. Hence $T^* = \infty$, i.e, finite ε-decompositions are always possible. □

Another way of re-phasing the previous Lemma 3.3 is given by the following concept. Let $N_T^{a,b}(\omega)$ be the number of up-crossing (of ω) of the interval $[a, b]$ within the time interval $[0, T]$, i.e., $N_T^{a,b}(\omega)$ is the infimum of all $k \ge 0$ such that there exist $0 \le t_1 < t_2 < \ldots < t_{2k-1} < t_{2k}$ satisfying $x(t_{2i-1}) < a < b < x(t_{2i})$ for any $i = 1, 2, \ldots k$. We can verify that a function ω has has one-sided limits (or equivalent, has at most discontinuities of the first kind) at within the interval $[0, T]$ if and only if $N_T^{a,b}(\omega) < +\infty$, for each $a < b$.

Clearly, we can review the previous argument for functions ω having at most discontinuities of the first kind, by replacing $\operatorname{osc}(\omega, [t_{i-1}, t_i[) \le \varepsilon$ with

$\operatorname{osc}(\omega,]t_{i-1}, t_i[) \le \varepsilon$. In particular, any cad-lag function can have only a finite number of jumps greater than $\varepsilon > 0$ within a bounded interval $[0, T]$.

Therefore, we can define a cad-lag modulus of continuity as follows

$$w(\omega, r, T) = \inf_{\{t_i\}} \max_i \sup_{t,s \in [t_{i-1}, t_i[} |\omega(t) - \omega(s)|, \tag{3.3}$$

where $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \cdots < t_{n-1} < T \le t_n$, with $t_i - t_{i-1} \ge r > 0$, $i = 1, \dots, n$ and $T > r > 0$. This means that $w(\omega, r, T) = \inf_{\{t_i\}} \max_i \operatorname{osc}(\omega, [t_i, t_{i+1}[)$, and certainly, we should use $]t_i, t_{i+1}]$ instead of $[t_i, t_{i+1}[$ to work with continuous functions from the left having limits from the right. If the interest is on functions having finite lateral limits, but either continuous from the right or from the left, we should use the expression

$$w'(\omega, r, T) = \sup \big\{ |\omega(t') - \omega(s)| \wedge |\omega(s) - \omega(t)| : \\ : t, s, t' \in [0, T],\ t < s < t' < t + r \big\}, \tag{3.4}$$

where $\wedge$ means the minimum between two real numbers. Note that for a cad-lag jump function of the form $\mathbb{1}_{[0,a[}(t)$ or $\mathbb{1}_{[0,a]}(t)$ we have $w'(1_{[0,a[}, r, T) = w'(1_{[0,a]}, r, T) = 0$, $w(1_{[0,a[}, r, T) = 0$ and $w(1_{[0,a]}, r, T) = 1$, if $r < a$ and $a < T - r$. A relation among the three moduli of continuity $w_c(\omega, r, T)$ $w(\omega, r, T)$ and $w'(\omega, r, T)$ is given below.

Lemma 3.4. *For every function ω defined on $[0, \infty[$, and with the notation (3.1), (3.2), (3.3) and (3.4), we have $w'(\omega, r, T) \le w(\omega, r, T)$. Moreover, if ω is a cad-lag function then*

$$w_c(\omega, r, T) \le 2 \max \big\{ w'(\omega, r, T), \operatorname{osc}(\omega, [0, r[), \operatorname{osc}(\omega, [T - r, T[) \big\} + \\ + \sup_{r \le t \le T} |\omega(t) - \omega(t-)| \tag{3.5}$$

and

$$w_c(\omega, r, T) \le 2 w(\omega, r, T) + \sup_{r \le t \le T} |\omega(t) - \omega(t-)|, \tag{3.6}$$

for every $0 < r < T$.

Proof. First, if $\varepsilon > w(\omega, r, T)$ then there exist a partition $\{t_i\}$ on $[0, T]$ such that $\operatorname{osc}(\omega, [t_i, t_{i+1}[) < \varepsilon$, which implies that $w'(\omega, r, T) \le \varepsilon$. This shows that $w'(\omega, r, T) \le w(\omega, r, T)$, for every function ω defined on $[0, \infty[$. By the way, note that if only $\operatorname{osc}(\omega,]t_i, t_{i+1}[) < \varepsilon$ and the value $\omega(t_i)$ is between $\omega(t_i-)$ and $\omega(t_i+)$, then we also have $w'(\omega, r, T) \le \varepsilon$.

To check estimate (3.5) or

$$w_c(\omega, r, T) \le 2 \max \big\{ w'(\omega, r, T), \operatorname{osc}(\omega, [0, r[), \operatorname{osc}(\omega, (T - r, T]) \big\} + \\ + \sup_{r \le t \le T - r} |\omega(t) - \omega(t-)|,$$

for a given interval $[a,b[$ with $0 \le a < b \le T$ and $b-a \le r$ define $t^* = t^*([a,b[)$ as the supremum of all τ in $[a,b[$ such that

$$|\omega(a) - \omega(t)| \le \max\big\{w'(\omega, r, T), \mathrm{osc}(\omega, [0,r[), \mathrm{osc}(\omega, [T-r, T[)\big\},$$

for every t in $[a,\tau[$.

If $t^*([a,b[) = b$ for every interval $[a,b[$ then (3.5) holds true. Otherwise, for some interval $[a,b[$ with $r \le a < b \le T-r$ and $b-a \le r$ we have $t^* = t^*([a,b[)$ in $[a,b[$, which yields

$$|\omega(t^*-) - \omega(a)| \le \max\big\{w'(\omega, r, T), \mathrm{osc}(\omega, [0,r[), \mathrm{osc}(\omega, [T-r, T[)\big\}.$$

Now, since

$$|\omega(a) - \omega(t^*)| \wedge |\omega(t^*) - \omega(b)| \le w'(\omega, r, T)$$

and

$$|\omega(a) - \omega(t^*)| \ge w'(\omega, r, T) + \varepsilon$$

we must have

$$|\omega(t^*) - \omega(b)| \le w'(\omega, r, T).$$

Hence

$$\begin{aligned} |\omega(b) - \omega(a)| &\le |\omega(t^*) - \omega(b)| + |\omega(t^*) - \omega(t^*-)| + |\omega(t^*-) - \omega(a)| \le \\ &\le \max\big\{2w'(\omega, r, T), \mathrm{osc}(\omega, [0,r[), \mathrm{osc}(\omega, [T-r, T[)\big\} + \\ &\qquad + \sup_{r \le t \le T} |\omega(t) - \omega(t-)|, \end{aligned}$$

which proves (3.5) and (3.6). □

Along these same lines, consider the following argument:
For a cad-lag function ω there is only a finite number of jumps greater than $r > 0$ within the interval $[0,T]$. Hence, there exits a partition $0 = t_0 < t_1 < \cdots < t_n$ with $t_n \to \infty$ with the properties (a) $t_k - t_{k-1} \le r$ for very k and (b) if a time t has a jump $|\omega(t) - \omega(t-)| > r$ then $t = t_k$ for some k. Thus any jump which is not included in this partition must have size less or equal than r. Thus, based on estimate (3.6), define

$$\omega_r(t) = \sum_{k=0}^{\infty} \omega(t_k) \mathbb{1}_{\{t_k \le t < t_{k+1}\}}, \quad t \ge 0,$$

to deduce

$$|\omega(t) - \omega_r(t_k)| \le 2w(\omega, r, T) + r, \quad \forall t \in [t_k, t_{k+1}[, \ \forall k,$$

i.e., $\sup_{0\le t\le T}|\omega-\omega_r|\le 2w(\omega,r,T)+r$, for every $r,T>0$. Actually, we can improve this point by using the following remark.

For any interval $[a,a+r[$ there exists at most one single time τ with a jump larger than $2w'(\omega,2r,T)$, with $T\ge a+3r$. Indeed, if $|\omega(\tau)-\omega(\tau-)|\ge 2w'(\omega,2r,T)$ then for any t in $[a,a+r[$ we have either

$$|\omega(t)-\omega(\tau-)|=\min\big\{|\omega(t)-\omega)\tau-)|,|\omega(\tau)-\omega)\tau-)|\big\}\le w'(\omega,r,T),$$

for every t in $[a,\tau]$ or

$$|\omega(t)-\omega(\tau)|\le w'(\omega,r,T),\quad \forall t\in]\tau,a+r[.$$

This implies that for any s in $[a,a+r[$ we have either $|\omega(s)-\omega(s-)|\le|\omega(s)-\omega(\tau)|+|\omega(s-)-\omega(\tau)|$ or $|\omega(s)-\omega(s-)|\le|\omega(s)-\omega(\tau-)|+|\omega(s-)-\omega(\tau-)|$, so that in both cases we deduce $|\omega(s)-\omega(s-)|\le 2w'(\omega,r,T)\le 2w'(\omega,2r,T)$.

Hence, considering the partition $k2^{-n}$ with $k=0,1,\dots$, we obtain at most one τ_k in $[k2^{-n},(k+1)2^{-n}[$ such that $|\omega(\tau_k)-\omega(\tau_k-)|\ge 2w'(\omega,2r,T)$. Therefore, the piecewise linear function λ_n defined by the equations $\lambda_n(0)=0$ and $\lambda_n(k2^{-n})=\tau_{k-1}$, for $k\ge 1$ is a continuous strictly increasing function from $[0,\infty[$ onto itself with the property $t-2^{-n}\le\lambda_n(t)\le t$, i.e., $\sup_{t\ge0}|\lambda_n(t)-t|\le 2^{-n}$. Thus, in the precedent construction we can use this dyadic partition with $\omega_n(t)=\omega(\lambda_n(k2^{-n}))$ if t belongs to $[k2^{-n},(k+1)2^{-n}[$, for some $k=0,1,\dots$, to deduce that

$$\big|\omega\big(\lambda_n(t)\big)-\omega_n(k2^{-n})\big|\le 2w'(\omega,2^{-n},T)+2w'(\omega,2^{-n}2,T),$$

for every t in $[k2^{-n},(k+1)2^{-n}[$ and $T\ge(k+1)2^{-n}$, i.e.,

$$\sup_{0\le t\le T}\big|\omega\big(\lambda_n(t)\big)-\omega_n(t)\big|\le 4w'(\omega,2^{-n}2,T+2^{-n}2),\tag{3.7}$$

for any $n=1,2,\dots$ and for every $T>0$. This approximation is used to show the separability of the space $D([0,\infty[)$, with a suitable metric.

On the other hand, we have

Lemma 3.5. *The elements ω in $D([0,\infty[)$ are locally bounded functions, i.e., for every $T>0$ there exists $C=C(T)$ such that $|\omega(t)|\le C$, for every t in $[0,T]$. Moreover a function ω has finite lateral limits and is either continuous from the right or from the left if and only if $w'(\omega,r,T)\to 0$ as $r\to 0$, for every fixed $T>0$. Furthermore, ω belongs to $D([0,\infty[)$ if and only if $w(\omega,r,T)\to 0$ as $r\to 0$, for every fixed $T>0$.*

Proof. If ω belongs to $D([0,\infty[)$ then there exists a partition $\{t_i\}$ on $[0,T]$ such that ω is continuous on $[t_{i-1},t_i[$ and $\omega(t_i-)$ exits for every $i=1,\dots,n$. Thus $\max_i\sup_{[t_{i-1},t_i]}|\omega(t)|$ is finite. This is to say that the range of a cad-lag function is a pre-compact set in $\mathbb{R}$.

For any ω in $D([0,\infty[)$ and $\varepsilon>0$ define $\tau_0=0$ and $\tau_{i+1}=\inf\{t>\tau_i:|\omega(t)-\omega(\tau_i)|>\varepsilon/2\}$. The arguments in Lemma 3.3 shows that for every $T>0$ there exists n such that $\tau_{n-1}\le T<\tau_n$. By construction we have $\operatorname{osc}(\omega,[\tau_{i-1},\tau_i[)\le\varepsilon$

and thus $w(\omega, r, T) \le \varepsilon$ if $r \le \inf_{i \le n}(\tau_i - \tau_{i-1})$. This proves that $w(\omega, r, T) \to 0$ as $r \to 0$, for every fixed $T > 0$.

To prove the converse, take a function ω satisfying $w'(\omega, r, T) \to 0$ as $r \to 0$, for every fixed $T > 0$. If one of the lateral limits at $t > 0$ is infinite or does not exist then there exist a constant $a > 0$ and a sequence $t_n \to t$ with either $t_n > t_{n+1} > t$ (or $t_n < t_{n+1} < t$) for every n such that $|\omega(t_n) - \omega(t_m)| > a$. Thus $\mathrm{osc}(\omega,]t, t + r[) > a$ (or $\mathrm{osc}(\omega,]t - r, t[) > a$), i.e., $w'(\omega, r, T) > a$ for every $T > t + r$ (or $r < t < T$). Hence, both lateral limits must exits at every point. However, if the value $\omega(t)$ is neither $\omega(t-)$ nor $\omega(t+)$ then $w'(\omega, r, T) \le |\omega(t-) - \omega(t)| \wedge |\omega(t) - \omega(t+)| > 0$. Moreover, we complete the argument by observing that if $\omega(t) \ne \omega(t+)$ then $\mathrm{osc}(\omega, [t, t + r) \le |\omega(t) - \omega(t+)| > 0$. □

Below are some related arguments concerning the oscillation. Beginning with $\mathrm{osc}(\omega, I) \le \mathrm{osc}(\omega, J)$, for every sets $I \subset J$, we can rewrite (3.3) as

$$w(\omega, r, T) = \inf \Big\{ \max_i \sup_{t,s \in [t_{i-1}, t_i[} |\omega(t) - \omega(s)| : \\ : t_0 = 0,\ t_{n-1} < T \le t_n,\ r \le t_i - t_{i-1} \le 2r,\ \forall i \Big\}, \quad (3.8)$$

where $T > r > 0$. If ω were defined only on $[0, T]$ then either we may use $\tilde{\omega}(t) = \omega(T \wedge t)$ instead of ω or only impose $t_i - t_{i-1} \le 2r$ for $i = n$.

By means of the inequality

$$a \wedge b \le \sqrt{ab} \le \sqrt{a \wedge b}\,\sqrt{a \vee b}$$

we can show that the expression

$$w''(\omega, r, T) = \sup \Big\{ \sqrt{|\omega(t') - \omega(s)||\omega(s) - \omega(t)|} : \\ : t, s, t' \in [0, T],\ t < s < t' < t + r \Big\}, \quad (3.9)$$

is equivalently to (3.4) in the sense that

$$w''(\omega, h, T) \le 2\sqrt{w'(\omega, h, T)} \sup_{0 \le t \le T} \sqrt{|\omega(t)|}$$

and $w'(\omega, h, T) \le w''(\omega, h, T)$.

3.3 Skorokhod Topology

For any function having at most discontinuities of the first kind, denote by $\delta\omega(t) = \delta(\omega, t)$ the jump of ω at time t, i.e., $\delta(\omega, t) = \omega(t+) - \omega(t-)$ for any $t > 0$. A cad-lag functions ω may has jumps, say times $t > 0$ where $\delta(\omega, t) \ne 0$ with size $|\omega(t) - \omega(t-)| > 0$.

Suppose that a sequence $\{\omega_n\}$ converges to ω in the locally uniform topology, i.e., $\mathrm{d}_c(\omega_n, \omega) \to 0$, where

$$\mathrm{d}_c(\omega, \omega') = \sum_{n=1}^{\infty} 2^{-n}(1 \wedge \|\omega - \omega'\|_{[0,n]}), \qquad \|\omega\|_I = \sup_{t \in I} |\omega(t)\|. \quad (3.10)$$

If ω is continuous at a time t_0 then ω_n may be discontinuous at t_0 but the size of the jumps of ω_n should vanish, i.e., $\delta(\omega_n, t_0) \to 0$. A typical example is the approximation of a continuous function by a sequence of step functions. However, if ω has a jump at time t_0 then all, except for a finite number of n, ω_n should have a jump at the same time t_0 and $\delta(\omega_n, t_0) \to \delta(\omega, t_0)$. For instance, if $\omega_n = \mathbb{1}_{[0,t_n[}$ for $0 < t_n \le t_0$ and $t_n \to t_0$ then for $\omega = \mathbb{1}_{[0,t_0[}$ we have $\omega_n(t) \to \omega(t)$ for every t, but $\mathrm{d}_c(\omega_n, \omega) = 1$, for every n. This suggests the following

Definition 3.6. A *change of time* is a continuous strictly increasing function $\lambda\colon [0,\infty[\to [0,\infty[$ satisfying $\lambda(0) = 0$ and $\lambda(t) \to \infty$ as $t \to \infty$, and Λ denotes the set of all change of time. A sequence $\{\omega_n\}$ of elements in $D([0,\infty[)$ converges to ω in the *Skorokhod topology* if there exists a sequence $\{\lambda_n\}$ in Λ such that $\|\lambda_n - I\|_{[0,\infty[} \to 0$ and $\|\omega_n \circ \lambda_n - \omega\|_{[0,T]} \to 0$, for every $T > 0$, where $I(t) = t$, for every $t \ge 0$, is the identity mapping, $\|\cdot\|_I$ is the sup-norm on the interval I and $\omega_n \circ \lambda_n$ is the composition of the functions ω_n and λ_n. □

We gain more insight on this convergence with

Proposition 3.7. *Let $\{\omega_n\}$ be a sequence in $D([0,\infty[)$. If ω_n converges to ω locally uniform, i.e., $\mathrm{d}_c(\omega_n, \omega) \to 0$, then $\omega_n \to \omega$ in the Skorokhod topology. Moreover, if ω is continuous and $\omega_n \to \omega$ in the Skorokhod topology then $\omega_n \to \omega$ in the locally uniform topology.*

Proof. It is clear that by taking $\lambda_n(t) = t$, i.e., $\lambda_n = 1$, we deduce the first assertion. Now, suppose that $\omega_n \to \omega$ in the Skorokhod topology and that ω is continuous. Thus, there exists a sequence $\{\lambda_n\}$ of change of times such that $\lambda_n \to 1$ and $\omega_n \circ \lambda_n \to \omega$. By means of the inequalities

$$|\omega_n(t) - \omega(t)| \le |\omega_n \circ \lambda_n \circ \lambda_n^{-1}(t) - \omega \circ \lambda_n^{-1}(t)| + |\omega \circ \lambda_n^{-1}(t) - \omega(t)|,$$

$$\|\omega_n \circ \lambda_n \circ \lambda_n^{-1} - \omega \circ \lambda_n^{-1}\|_{[0,T]} = \|\omega_n \circ \lambda_n - \omega\|_{[0,\lambda_n(T)]},$$

and, with w_c is the modulus of continuity (3.1),

$$\|\omega \circ \lambda_n^{-1} - \omega\|_{[0,T]} \le w_c\big(\omega, \|\lambda_n - I\|_{[0,T]}, T \vee \lambda_n(T)\big)$$

we prove the second statement. □

A couple of examples may help. For numerical sequences $\{a_n\}$, $\{b_n\}$, $\{x_n\}$ and $\{y_n\}$ with $a_n < b_n$, $x_n \to x \ne 0$, $y_n \to y \ne 0$, and $x \ne y$, define $\omega_n = x_n \mathbb{1}_{[a_n,\infty]} + y_n \mathbb{1}_{[b_n,\infty[}$. We can check that $\{\omega_n\}$ converges to some limit ω in $D([0,\infty[)$ if and only if either (1) $a_n \to \infty$ and $b_n \to \infty$, with $\omega = 0$, or (2) $a_n \to a$ and $b_n \to \infty$, with $\omega = x\mathbb{1}_{[a,\infty[}$, or (3) $a_n \to a$, $b_n \to b$ and $a < b$, with $\omega = x\mathbb{1}_{[a]} + y\mathbb{1}_{[b,\infty[}$. Note that if $a = b$ then $\omega_n(a_n) = x_n \to x$ and $\omega_n(b_n) = y_n \to y$ so that $\omega_n(a_n) - \omega_n(b_n) \to (x - y) \ne 0$ but $a_n - b_n \to 0$.

These examples show that we may have $\omega_n \to \omega$ and $\omega'_n \to \omega'$, but $\omega_n + \omega'_n$ does not converges to $\omega + \omega'$ in the Skorokhod topology. Thus $D([0,\infty[)$ is a vector space, but it is not a topological vector space with the Skorokhod topology.

Moreover, this same construction applies to functions with values in a Polish space, in particular, cad-lag functions with values in $\mathbb{R}^d$, i.e., $D([0,\infty[;\mathbb{R}^d)$. Again the previous examples show that the Skorokhod topology in $D([0,\infty[;\mathbb{R}^d)$ is stronger that the product Skorokhod topology $\big(D([0,\infty[)\big)^d = D([0,\infty[;\mathbb{R}^d)$. Nevertheless, based on Proposition 3.7, we can shows that if $\omega_n \to \omega$, $\omega'_n \to \omega'$ in the Skorokhod topology and ω is continuous then $\omega_n + \omega'_n \to \omega + \omega'$.

The function

$$\mathrm{d}_1(\omega,\omega') = \inf_{\lambda\in\Lambda}\Big\{\sup_{t\geq 0}\big(1\wedge|\lambda(t)-t|\big)+ \\ +\sum_{n=1}^{\infty}2^{-n}\sup_{t\geq 0}\big(1\wedge|\omega(n\wedge\lambda(t))-\omega'(n\wedge t)|\big)\Big\}$$

is a metric in $D([0,\infty[)$ yielding the Skorokhod topology, but it fails to be complete. The problem is with the λ, actually, the distance d_1 may be regarded as acting on the couple (λ,ω), namely,

$$\mathrm{d}_1\big((\lambda,\omega),(\lambda',\omega')\big) = \sup_{t\geq 0}\big(1\wedge|\lambda(t)-\lambda'(t)|\big)+ \\ +\sum_{n=1}^{\infty}2^{-n}\sup_{t\geq 0}\big(1\wedge|\omega(n\wedge\lambda(t))-\omega'(n\wedge\lambda'(t)|\big),$$

and then taking distance from (λ,ω) to (I,ω'), but there is no condition in this metric to force λ to remains in λ. A possibility could be

$$\tilde{\mathrm{d}}_1(\omega,\omega') = \inf_{\lambda\in\Lambda}\Big\{\sup_{t\geq 0}\big(1\wedge|\lambda(t)-t|\big)+\sup_{t\geq 0}\big(1\wedge|\lambda^{-1}(t)-t|\big)+ \\ +\sum_{n=1}^{\infty}2^{-n}\sup_{t\geq 0}\big(1\wedge|\omega(n\wedge\lambda(t))-\omega'(n\wedge t)|\big)\Big\},$$

or as normally done, we localize the sup-norm with the function $k_n(t) = 1$ for $0\leq t\leq n$, $k_n(t) = n+1-t$ for $n<t<n+1$ and $k_n(t)=0$ for $t\leq n+1$, to define

$$\begin{aligned}
&[\![\lambda]\!]_s = \sup_{t>s\geq 0}\big|\ln\big(\frac{\lambda(t)-\lambda(s)}{t-s}\big)\big|,\\
&\mathrm{d}_s(\omega,\omega',n,\lambda) = [\![\lambda]\!]_s + \sup_{t\geq 0}\big|k_n(\lambda(t))\omega(\lambda(t)) - k_n(t)\omega'(t)\big|,\\
&\mathrm{d}_s(\omega,\omega') = \sum_{n=1}^{\infty}2^{-n}\inf_{\lambda\in\Lambda}\big\{1\wedge\mathrm{d}_s(\omega,\omega',n,\lambda)\big\},
\end{aligned} \tag{3.11}$$

which is a complete metric yielding the Skorokhod topology, i.e., $D([0,\infty[)$ is a complete metric (metrizable, because the specific metric is seldom used) space. This metric (3.11) has been introduced by Prokhorov, and it takes some doing to show that d_s is indeed a metric, and even more to show that d_s is complete.

Actually, a simple but important point is the construction a sequence $\{\lambda_n\}$ in Λ to verify that $\mathrm{d}_s(\omega_n, \omega) \to 0$ implies $\omega_n \to \omega$ in the Skorokhod topology, see Definition 3.6.

Consider the coordinate mappings $X_t \colon D([0,\infty[) \to \mathbb{R}$ defined by $X_t(\omega) = \omega(t)$, and denote by $\mathcal{D}_t^0$ (respectively, by $\mathcal{D}$, the cylindrical σ-algebra) the σ-algebra in $D([0,\infty[)$ generated by the family of functions $\{X_s : 0 \le s \le t\}$ (respectively, $\{X_s : s \ge 0\}$). Next define $\mathcal{D}_t = \bigcap_{s>t} \mathcal{D}_s^0$ and $\mathcal{D}_{t-}$ the σ-algebra generated by $\{\mathcal{D}_s : s < t\}$. The family $\{\mathcal{D}_t : t \ge 0\}$ is called a filtration. Then, without proof, we can state

Theorem 3.8. *The Skorokhod topology, as given by* Definition 3.6, *makes $D([0,\infty[)$, the space of cad-lag functions, a complete separable metrizable space, i.e., a Polish space. Moreover, a subset K of $D([0,\infty[)$ is pre-compact if and only if it is bounded and equi-cad-lag, i.e., if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|\omega(t)| \le 1/\varepsilon$ and $w(\omega, r, [0.1/\varepsilon]) \le \varepsilon$, for every $0 < r < \delta$ and any ω in K, where w is the modulus* (3.3). *Furthermore, $\mathcal{D}_{t-}$ is generated by all real-valued continuous functions on $D([0,\infty)$ which are $\mathcal{D}_{t-}$-measurable, and the Borel σ-algebra of $D([0,\infty)$ coincides with $\mathcal{D}$, the cylindrical σ-algebra.* □

• *Remark* 3.9. Suppose that $\omega_n \to \omega$ in the Skorokhod topology. For a given $t \ge 0$, consider the sequence $\{t_n : n \ge 1\}$, where $t_n = \lambda_n(t)$ and λ_n are the change of time in Definition 3.6. This sequence has the following properties: (a) $t_n \to 0$, $\omega_n(t_n) \to \omega(t)$ and $\omega_n(t_n-) \to \omega(t-)$; (b) if $\omega(t) \neq \omega(t-)$ then any other sequence $\{t'_n : n \ge 1\}$ satisfying (a) coincides with $\{t_n : n \ge 1\}$ for all n large enough; (c) if $\omega(t) = \omega(t-)$ then for any sequence $\{t'_n : n \ge 1\}$ satisfying $t'_n \to t$ we have $\omega_n(t'_n) \to \omega(t)$ and $\omega_n(t'_n-) \to \omega(t-)$; (d) for any other sequence $\{s_n : n \ge 1\}$ such that $s_n \to t$ we deduce that

$$\begin{aligned}
&s_n < t_n,\ \forall n \quad \text{imply} \quad \omega_n(s_n) \to \omega(t-),\\
&s_n \le t_n,\ \forall n \quad \text{imply} \quad \omega_n(s_n-) \to \omega(t-),\\
&s_n \ge t_n,\ \forall n \quad \text{imply} \quad \omega_n(s_n) \to \omega(t),\\
&s_n > t_n,\ \forall n \quad \text{imply} \quad \omega_n(s_n-) \to \omega(t).
\end{aligned}$$

Moreover, if δ is the jump operator $\delta\omega(t) = \omega(t) - \omega(t-)$ the define $\omega'_n(s) = \omega_n(s) - \delta\omega_n(t_n)\mathbb{1}_{\{t_n \le s\}}$ and $\omega'(s) = \omega(s) - \delta\omega(t)\mathbb{1}_{\{t_n \le s\}}$ and use the inequalities

$$\begin{aligned}
&\big|\omega'_n\big(\lambda_n(s)\big) - \omega'(s)\big| \le \big|\omega_n\big(\lambda_n(s)\big) - \omega(s)\big| + \big|\delta\omega_n(t_n) - \delta\omega(s)\big|,\\
&\big|\omega_n\big(\lambda_n(s)\big) - \omega(s)\big| \le \big|\omega_n\big(\lambda_n(s)\big) - \omega\big(\lambda_n(s)\big)\big| + \big|\omega\big(\lambda_n(s)\big) - \omega(s)\big|
\end{aligned}$$

to deduce (1) $\omega'_n \to \omega'$ in the Skorokhod topology and (2) $\mathrm{osc}(\omega'_n, [t-r, t+r]) \to 0$ as $r \to 0$, where osc is given by (3.2). Furthermore, we have

$$\limsup_n \sup_{a \le t \le b} |\delta\omega_n(t)| \le \sup_{a \le t \le b} |\delta\omega(t)|, \quad \forall b \ge a \ge 0.$$

Indeed, if L denote the value of the limit in the left-hand side then there exist a subsequence $\{n_k : k \ge 1\}$ and sequences $\{t_k : k \ge 1\}$ and $\{t'_k : k \ge 1\}$ in

$[a,b]$ (if $b>a$ otherwise $t=a=b$) such that $t_k<t\le t'_k$, $t'_k-t_k\to 0$ and $\omega_{n_k}(t'_k)-\omega_{n_k}(t_k)\to L$. Thus define the sequences $\{\tau_k : k\ge 1\}$ and $\{\tau'_k : k\ge 1\}$ by $t_k=\lambda_{n_k}(\tau_k)$ and $t'_k=\lambda_{n_k}(\tau'_k)$ satisfying $\tau_k<\lambda_{n_k}(t)\le\tau'_k$, $\tau'_k\to t$ and $\tau_k\to t$ to obtain the inequality

$$\begin{aligned}\big|\omega_{n_k}(t'_k)-\omega_{n_k}(t_k)\big|\le\big|\omega_{n_k}\big(\lambda_{n_k}(\tau'_k)\big)-\omega(\tau'_k)\big|+\\ +\big|\omega_{n_k}\big(\lambda_{n_k}(\tau_k)\big)-\omega(\tau_k)\big|+\big|\omega(\tau'_k)-\omega(\tau_k)\big|,\end{aligned}$$

Since the first two terms in the right-hand side vanish as $k\to\infty$, the previous claim is proved. □

3.4 Skorokhod Topology for BV functions

This section is not standard, only a number of comments are presented and, certainly, it can be skipped, and for instance, the reader may take a look at the books Jacod and Shiryaev [69, Sections VI.1 and VI.2, pp. 288–310], Ethier and Kurtz [47, Section 3.5, pp. 116–154],

The dual of the (vector) locally convex topological space $C([0,\infty[)$ can be identified as the space of Radon measures on $[0,\infty[$, which has a one-to-one relation with the space of cad-lag functions having locally bounded variation $V([0,\infty[)$. Thus, we want to discuss a little how this relate to the Skorokhod Topology. There are other possible good topologies for $D([0,\infty[)$, e.g, Jakubowski [70].

Consider the semi-space $V^+([0,\infty[)$ of cad-lag monotone increasing functions vanishing at the origin, i.e.,

$$V^+([0,\infty[)=\big\{\alpha\in D([0,\infty[):\alpha(0)=0,\ \alpha(t)\le\alpha(s),\ \forall 0\le t\le s\big\}.$$

Recall that for any given countable subset Q of $[0,\infty[$ and a locally uniformly bounded sequence of monotone increasing functions we can extract a convergence subsequence, i.e., by means of Cantor diagonal procedure and the compactness of any bounded closed set of real number, if (1) $\{\alpha_n : n\ge 1\}$ is a sequence in $V^+([0,\infty[)$, and (2) for every $T>0$ there exists a constant $C_T>0$ such that $\alpha_n(T)\le C_T$, for every $n\ge 0$, then there exists a subsequence $\{\alpha_{n_k} : k\ge 1\}$ such that $\alpha_n(q)\to\tilde\alpha(q)$, for every q in Q (implying that $\tilde\alpha$ is monotone increasing as a function defined on Q). Hence, if Q is a (countable) dense set in $[0,\infty[$ then define $\alpha(t)=\inf_{q\in Q,\ q>t}\tilde\alpha(q)$ to deduce that α belongs to $V^+([0,\infty[)$ and $\alpha_n(t)\to\alpha(t)$, for every t in $[0,\infty[$ such that $\alpha(t)=\alpha(t-)$.

The above argument can be applied to difference of monotone increasing functions, i.e., to functions with locally bounded variation. Moreover, if β is a cad-lag function with locally bounded variation then we can define its discontinuous part and its continuous part as

$$\beta_d(t)=\sum_{s\le t}\big(\beta(t)-\beta(t-)\big),\qquad \beta_c(t)=\beta(t)-\beta_d(t),\quad \forall t>0,$$

where the series over $s \le t$ is absolutely convergence (certainly, under the convention that we discard only zero terms and if no term meets the condition $s \le t$ then the sum is empty and equal to 0, moreover, the order in which terms are added is of no importance). For a general cad-lag function, only a finite number of jumps larger than $\varepsilon > 0$ may occur within a bounded interval $[0, T]$, i.e., for every $\varepsilon > 0$ the set $\{t \in [0, 1/\varepsilon] : |\omega(t) - \omega(t-)| \ge \varepsilon\}$ is finite, but the above series cannot be defined for ω, since it is not necessarily absolutely convergence.

An increasing *counting function* is an element α in $V^+([0,\infty[)$ with all jumps of size 0 (i.e., no jump, meaning continuity) or 1, i.e., for every $t > 0$, if $\alpha(t) - \alpha(t-) > 0$ then $\alpha(t) - \alpha(t-) = 1$. On $V^+([0,\infty[)$ the Skorokhod convergence becomes

Theorem 3.10. *Let α and $\{\alpha_n : n \ge 1\}$ be in $V^+([0,\infty[))$. Then $\alpha_n \to \alpha$ in the Skorokhod topology if and only if for some dense subset Q of $[0,\infty[$ we have* (1) *$\alpha_n(q) \to \alpha(q)$, for every q in Q and* (2)

$$\sum_{0<s\le q} |\alpha_n(s) - \alpha_n(s-)|^2 \to \sum_{0<s\le q} |\alpha(s) - \alpha(s-)|^2, \quad \forall q \in Q,$$

and in this case, we may take $Q = \{t > 0 : \alpha(t) = \alpha(t-)\}$, i.e., all points of continuity of α. Moreover, the precedent condition (2) *can be replaced by either* (3) *there exists a strictly convex function $h\colon [0,\infty) \to [0,\infty)$ such that $h(0) = h'(0) = 0$ and*

$$\sum_{0<s\le q} \big|h\big(\alpha_n(s) - \alpha_n(s-)\big) - h\big(\alpha(s) - \alpha(s-)\big)\big| \to 0, \quad \forall q \in Q,$$

or (4) *for every $t \ge 0$ there exists a sequence $\{t_n : n \ge 1\}$ satisfying $t_n \to t$ with $t_n \le t$ if t belongs to Q and $\big(\alpha_n(t_n) - \alpha_n(t_n-)\big) \to \big(\alpha(t) - \alpha(t-)\big)$. Furthermore, only the previous condition* (1) *is required to obtain Skorokhod convergence in the following cases:* (a) *either α_n and α are counting functions or α is continuous;* (b) *there exists a sequence $\{\beta_n : n \ge 1\}$ in $V^+([0,\infty[))$ relatively compact in the Skorokhod topology such that $\beta_n - \alpha_n$ belongs to $V^+([0,\infty[))$, for every $n \ge 1$.*

Proof. Only some ideas are given. Since the Skorokhod topology is metrizable, a sequence ω_n converges to ω if and only if from any subsequence we can extract another subsequence convergent to ω. Thus, by means of Cantor diagonal procedure, we may assume (without any loss of generality) that condition (1) is satisfied also for q in some larger $Q_0 \supset Q$, provided $Q_0 \setminus Q$ is countable, i.e., we may assume that 0 and n belongs to Q, see the metric (3.10), and that Q contains all times of continuity of α (and α_n if necessary).

First we establish that the convergence on Q, i.e., condition (1) implies

$$\limsup_n \Big(\sup_{t\le T} \big|\alpha_n(t) - \alpha(t-)\big|\Big) \le 2\Big(\sup_{t\le T} \big|\alpha(t) - \alpha(t-)\big|\Big), \tag{3.12}$$

and

$$\limsup_n \Big[\sup_{q<t\le T} \big(\alpha_n(t) - \alpha_n(t-)\big)\Big] \le \Big[\sup_{q<t\le T} \big(\alpha(t) - \alpha(t-)\big)\Big], \tag{3.13}$$

for every q in Q and $T > q$. For instance, to establish (3.13) we proceed as follows. First, denote by a the limit in the left-hand term of (3.13) and find a subsequence $\{n_k : k \geq 1\}$ and a sequence $\{t_k : k \geq 1\}$ in $(q, T]$ such that $t_k \to t^*$ and $\delta\alpha_{n_k}(s_k) \to a$, where δ is the jump operator, i.e., $\delta\alpha(t) = \alpha(t) - \alpha(t-)$. Since α is cad-lag, the convergence (1) implies that for any $\varepsilon > 0$ there exits $q \leq q' < q''$ in Q satisfying $q' \leq t^* < q''$ if $t^* = q$ and $q' < t^* < q''$ if $t^* > q$, and such that $\alpha(q'') - \alpha(q') \leq \varepsilon + \mathbb{1}_{\{q<t^*\}}\delta\alpha(t^*)$. Because s_k belongs to $(q', q'']$ (for k sufficiently large) and α is increasing, the convergence on Q, i.e., condition (1), yields

$$a = \lim_k \delta\alpha_{n_k}(s_k) \leq \lim_k \big(\alpha_{n_k}(q'') - \alpha_{n_k}(q')\big) = \alpha(q'') - \alpha(q') \leq$$
$$\leq \varepsilon + \mathbb{1}_{\{q<t^*\}}\delta\alpha(t^*) \leq \varepsilon + \sup_{q<t\leq T}\big(\delta\alpha(t)\big),$$

which proves estimate (3.13).

Now, if α is continuous then condition (1) implies condition (2). Indeed, the previous estimate shows that $\alpha_n \to \alpha$ locally uniformly. Hence $\sup_{s\leq t}\big(\alpha_n(t) - \alpha_n(t-)\big) \to 0$, for every $t > 0$, and then

$$\sum_{s\leq t}\big|\alpha_n(s) - \alpha_n(s-)\big|^2 \leq \alpha_n(t)\sup_{s\leq t}\big(\alpha_n(s) - \alpha_n(s-)\big)$$

so does.

On the other hand, if α_n and α are counting functions then we have

$$\alpha(t) = \sum_{s\leq t}\big(\alpha_n(s) - \alpha_n(s-)\big) = \sum_{s\leq t}\big|\alpha_n(s) - \alpha_n(s-)\big|^2$$

and similarly for α_n. This shows that condition (1) implies condition (2).

The longest part is to show that if $\alpha_n(q) \to \alpha(q)$, for every q in Q then conditions (3) and (4) are equivalent. For instance, we separate the small jumps, namely, for a given $\varepsilon > 0$ define s_i by induction as $s_0 = 0$, and $s_i = \inf\{t > s_{i-1} : \alpha(t) - \alpha(t-) > \varepsilon\}$, e.g., if α is continuous then $s_1 = \infty$. For each s_i find a sequence $\{t_{i,n} : n \geq 1\}$ satisfying condition (4) for $t = s_i$ to define

$$\alpha'(t) = \alpha(t) - \sum_{s_i\leq t}\delta\alpha(s_i), \qquad \alpha'_n(t) = \alpha_n(t) - \sum_{s_i\leq t}\delta\alpha_n(s_{i,n}),$$

where δ is the jump operator. First, it is clear that $\alpha'_n(q) \to \alpha'(q)$ for every q in Q, and by means of estimate (3.13), we deduce that $\limsup_n \sup_{s\leq t}|\delta\alpha_n(s)| \leq \varepsilon$. Thus

$$\sum_{s\leq t}\big|h(\delta\alpha_n(s)) - h(\delta\alpha(s))\big| \leq \sum_{s_i\leq t}\big|h(\delta\alpha_n(s_{i,n})) - h(\delta\alpha(s_i))\big| +$$
$$+ \sum_{s\leq t}\big(|h(\delta\alpha'_n(s))| + |h(\delta\alpha'(s))|\big).$$

Taking limit as $n \to \infty$, the first sum is finite and so it vanishes, while from

$$|h(\delta\alpha_n'(s))| \le g\big(\sup_{s\le t} \delta\alpha_n'(s)|\big)\, |\delta\alpha_n'(s)|, \qquad \sum_{s\le t} |\delta\alpha_n'(s)| \le \alpha_n(t),$$

with $h(x) = |x|\, g(x)$, $g(x) \to 0$ as $x \to 0$ (recall that $h(0) = h'(0) = 0$) and similar inequalities with α' instead of α_n', we obtain

$$\limsup_n \sum_{s\le t} \big(|h(\delta\alpha_n'(s))| + |h(\delta\alpha'(s))|\big) \le g(\varepsilon) \limsup_n \big(|\alpha_n(t)| + |\alpha(t)|\big).$$

Hence, condition (3) is satisfied.

Another delicate point is to establish that conditions (1) and (4) imply that the sequence $\{\alpha_n : n \ge 1\}$ is relatively compact in the Skorokhod topology. All these properties and, because condition (1) yields α as the only possible limit, we conclude proving the "if" part of the main assertion.

Finally, suppose that $\alpha_n \to \alpha$ in the Skorokhod topology. By means of Remark 3.9 we obtain condition (4), which implies that condition (2) is satisfied. □

• *Remark* 3.11. Note that if

$$\sum_{0<s\le q} \big||\alpha_n(s) - \alpha_n(s-)| - |\alpha(s) - \alpha(s-)|\big| \to 0, \quad \forall q \in Q,$$

then $|\alpha_n(s) - \alpha_n(s-)| \le C$ for every n, any $s \le T$, and some constant $C = C_T$. Hence

$$\sum_{0<s\le q} \big||\alpha_n(s) - \alpha_n(s-)|^2 - |\alpha(s) - \alpha(s-)|^2\big| \le$$
$$\le 2C_T \sum_{0<s\le q} \big||\alpha_n(s) - \alpha_n(s-)| - |\alpha(s) - \alpha(s-)|\big|, \quad \forall q \le T,$$

yields condition (2). □

Now, to consider the space of functions with local bounded variation and cad-lag, note that for an element ω in $D([0,\infty[)$ we can define the variation as

$$\mathrm{var}(\omega,[0,t]) = \sup\Big\{\sum_{i=1}^n |\omega(t_i) - \omega(t_{i-1}| : 0 = t_0 < t_1 < \cdots < t_n = t\Big\},$$

and the positive and negative variation $\mathrm{var}^+(\omega,[0,t])$ and $\mathrm{var}^-(\omega,[0,t])$ by replacing the absolute value $|\cdot|$ with the positive and negative parts, $[\,\cdot\,]^+$ and $[\,\cdot\,]^-$, respectively. Moreover, because the function ω is cad-lag, we may only consider partitions of the form $t_i = t \wedge 2^{-n}i$, for $i = 0, 1, \ldots, 4^n$. If $\mathrm{var}(\omega,[0,t]) < \infty$ then $\mathrm{var}(\omega,[0,t]) = \mathrm{var}^+(\omega,[0,t]) + \mathrm{var}^-(\omega,[0,t])$ and $\omega(t) = \mathrm{var}^+(\omega,[0,t]) -$

$\mathrm{var}^-(\omega,[0,t])$. This means that a cad-lag function having locally bounded variation can be regarded as the increasing functions, $t \mapsto \mathrm{var}^+(\omega,[0,t])$ and $t \mapsto \mathrm{var}^-(\omega,[0,t])$. It is also clear that

$$\sum_{s\le t} \big|\omega(t)-\omega(t-)\big|^2 \le 2\Big(\sup_{s\le t}\big|\omega(t)\big|\Big)\,\mathrm{var}(\omega,[0,t]), \quad \forall t>0.$$

Therefore, if $V([0,\infty[)$ denote the subspace of $D([0,\infty[)$ of all cad-lag functions having locally bounded variation vanishing at 0, then $V([0,\infty[) = V^+([0,\infty[) \ominus V^+([0,\infty[)$ and we can apply Theorem 3.10 to deduce (for instance) that for any ω and $\{\omega_n : n \ge 1\}$ in $V([0,\infty[)$ we have: $\omega_n \to \omega$ in the Skorokhod topology if and only if the exists a dense set Q of $(0,\infty)$ such that $(1)'$ $\omega_n(q) \to \omega(q)$ and $\mathrm{var}(\omega_n,[0,q]) \to \mathrm{var}(\omega,[0,q])$, for every q in Q, and $(2)'$

$$\sum_{0<s\le q} \big||\omega_n(s)-\omega_n(s-)|^2 - |\omega(s)-\omega(s-)|^2\big| \to 0, \quad \forall q\in Q,$$

and in this case, we may take $Q = \{t>0 : \omega(t) = \omega(t-)\}$, i.e., all points of continuity of ω or equivalently, of $t \mapsto \mathrm{var}(\omega,[0,t])$. Note that by means of Remark 3.11, the conditions $(1)'$ and $(2)'$ are satisfied if $\mathrm{var}(\omega_n-\omega,[0,q]) \to 0$, for every q in Q (and so for every q not necessarily in Q).

On the other hand, we may define the r-quadratic variation as

$$\mathrm{var}_2(\omega,[0,t],r) = \sup\Big\{\sum_{i=1}^n |\omega(t_i)-\omega(t_{i-1}|^2 \; : \\ : t_0=0,\, 0<t_i-t_{i-1}\le r,\, t_n=t\Big\},$$

which may be bounded for functions which are not of bounded variation. However, with the cad-lag modulus (3.1), we have

$$\mathrm{var}_2(\omega,[0,t],r) \le w_c(\omega,r,t)\,\mathrm{var}(\omega,[0,t]), \quad \forall t,r.$$

Therefore, defining the decreasing limit

$$\mathrm{var}_2(\omega,[0,t]) = \lim_{r\to 0}\mathrm{var}_2(\omega,[0,t],r),$$

we have

$$\sum_{0<s\le t} |\omega(s)-\omega(s-)|^2 \le \mathrm{var}_2(\omega,[0,t]), \quad \forall t,$$

so that $\mathrm{var}_2(\omega,[0,t]) = 0$ if ω is continuous or has locally bounded variation.

If a cad-lag purely jumps function ω has the following characterizing property

$$\omega(t) = \omega(0) + \lim_{\varepsilon\to 0}\sum_{s\le t}\big(\delta\omega(s)\big)\mathbb{1}_{\{|\delta\omega(s)|\ge\varepsilon\}}, \quad \forall t>0,$$

where the series is convergent (not necessarily absolutely convergent). It is clear that for a (cad-lag) purely jumps function ω we have

$$\mathrm{var}_2(\omega,[0,t]) = \sum_{s\le t} \big|\delta\omega(s)\big|^2, \quad \forall t>0.$$

For instance, consider a decreasing sequence of positive numbers $\{a_i : i\ge 1\}$ such that $\sum_i a_i^2 < \infty$ and $\sum_i a_i^p = \infty$, for every $p<2$, and an increasing sequence of times $\{s_i : i\ge 1\}$ such that $s_i \to 1$. We may define a purely jumps function by

$$\omega(t) = \sum_{s_i\le t} (-1)^i a_i, \forall t\ge 0.$$

It is clear that the only jumps are $\delta\omega(t) = a_i$ for every $t = s_i$, with $i\ge 1$, while $\omega(t)$ and $t\mapsto \mathrm{var}_2(\omega,[0,t])$ are a finite sum of jumps for $t<1$, and

$$\omega(1-) = \sum_i (-1)^i a_i = \omega(t), \quad \forall t\ge 1,$$
$$\mathrm{var}_2(\omega,[0,t]) = \sum_{0<s\le t} |\omega(s)-\omega(s-)|^2 = \sum_i a_i^2, \quad \forall t\ge 1.$$

The increasing process $t\mapsto \mathrm{var}_2(\omega,[0,t])$ has jumps $\delta\mathrm{var}_2(\omega,[0,t]) = |\delta\omega(t)|^2$ for every $t\ge 0$, the function ω has infinite variation on $[0,1]$ and

$$\omega(t) = \sum_{s_{2k}\le t} a_{2k} - \sum_{s_{2k-1}\le t} a_{2k-1}, \quad \forall t<1,$$

where the sums have only a finite number of terms, but certainly, we would have $\infty-\infty$ for $t\ge 1$.

Let $\{\omega_n : n\ge 1\}$ be a sequence of cad-lag functions converging in the Skorokhod topology to ω. We separate the small jumps as in the proof of property (3) in Theorem 3.10, i.e., for a given $\varepsilon>0$ define s_i by induction as $s_0=0$, and $s_i = \inf\{t>s_{i-1} : |\omega(t)-\omega(t-)|>\varepsilon\}$, e.g., if ω is continuous then $s_1=\infty$. For each s_i find a sequence $\{t_{i,n} : n\ge 1\}$ satisfying condition (4) for $t=s_i$ to define

$$\omega'(t) = \omega(t) - \sum_{s_i\le t} \delta\omega(s_i), \qquad \omega_n'(t) = \omega_n(t) - \sum_{s_i\le t} \delta\omega_n(s_{i,n}),$$

with the properties (see Remark 3.9) that $\omega_n'\to\omega'$ in the Skorokhod topology and $\limsup_n \sup_{s\le t} |\delta\omega_n(s)| \le \varepsilon$. Remarking that $\delta\omega_n(s) = \delta\omega_n'(s)$ for $s\ne s_{i,n}$ and $\delta\omega(s) = \delta\omega'(s)$ for $s\ne s_i$ we can write

$$\sum_{s\le t} \big|h(\delta\omega_n(s)) - h(\delta\omega(s))\big| \le \sum_{s_i\le t} \big|h(\delta\omega_n(s_{i,n})) - h(\delta\omega(s_i))\big| + \\ + \sum_{s\le t} \big(|h(\delta\omega_n'(s))| + |h(\delta\omega'(s))|\big),$$

Taking limit as $n \to \infty$, the first sum is finite and so it vanishes as long as h is continuous, while for $h(x) = |x|^3$,

$$|\delta\omega'_n(s)|^3 \leq \big(\sup_{s\leq t} |\delta\omega'_n(s)|\big)\, |\delta\omega'_n(s)|^2, \qquad \sum_{s\leq t} |\delta\omega'_n(s)|^2 \leq \mathrm{var}_2(\omega_n, [0,t]),$$

implies

$$\limsup_n \sum_{s\leq t} \big(|\delta\omega'_n(s)|^3 + |\delta\omega'(s)|^3\big) \leq \\ \leq \varepsilon \sup_n \big(\mathrm{var}_2(\omega_n, [0,t]) + \mathrm{var}_2(\omega, [0,t]|\big),$$

which shows that

$$\lim_n \sum_{s\leq t} \big||\delta\omega_n(s)|^3 - |\delta\omega(s)|^3\big| = 0, \quad \forall t > 0,$$

and so $|\delta\omega_n(s)| \to |\delta\omega(s)|$, uniformly in s belonging to $[0,T]$, for every $T > 0$.

On the other hand, the continuous (but not absolutely continuous near 0) function $t \mapsto \sqrt{t}$ satisfies

$$\sum_{i=1}^n \big|\sqrt{t_i} - \sqrt{t_{i-1}}\big|^2 \leq \sum_{i=1}^n |t_i - t_{i-1}| = T,$$

for every partition $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$, which proves that $\mathrm{var}_2(\sqrt{\cdot}, [0,t]) \leq t$, for every $t > 0$.

Perhaps, we may expect to deduce that $\omega_n \to \omega$ in the Skorokhod topology if the exists a dense set Q of $(0,\infty)$ such that $(1)''$ $\omega_n(q) \to \omega(q)$ and $\mathrm{var}_2(\omega_n, [0,q]) \to \mathrm{var}_2(\omega, [0,q])$, and $(2)''$

$$\sum_{0<s\leq q} \big||\omega_n(s) - \omega_n(s-)|^2 - |\omega(s) - \omega(s-)|^2\big| \to 0,$$

for every q in Q.

3.5 Integer Measures

The term *integer measure* refers to a measure with integer values, or equivalently, a series of Dirac measures. However, the same name integer measure is used to refer to a random variable whose values are integer measures. Moreover, if a *time* variable is singled-out then this is loosely referred to (until some topological property is assigned to the paths) as a process whose values are integer measures. Furthermore, all these instruments are used to study the jumps of cad-lag processes, i.e., random variables with values in the canonical space $D([0,\infty[)$ or equivalently, real-valued process whose paths are cad-lag almost surely.

In our discussion, the time-variable is singled-out so that (deterministic) integer measures are Borel σ-finite measures defined on $\mathcal{B}(E\times[0,\infty[)$, the Borel σ-algebra of the abstract space $E\times[0,\infty[$. Actually, to simplify arguments without sacrificing too much generality, the base space E is assumed locally compact, moreover, $E=\mathbb{R}^m_*$, where $\mathbb{R}^m_*=\mathbb{R}^m\smallsetminus\{0\}$.

Therefore, a (deterministic) integer measure ν is a Radom measure on $\mathbb{R}^m_*\times[0,\infty[$ which takes only integer values and satisfies some extra properties due to the special time-variable, i.e., besides ν being a measure on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^m_*\times[0,\infty[)$ the following properties are assumed: (a) $\nu(\mathbb{R}^m_*\times\{0\})=0$; (b) $\nu(\mathbb{R}^m_*\times\{t\})$ is either 0 or 1, for any t in $]0,\infty[$; (c) $\nu(K\times]0,t])$ is a positive integer number, for any Borel set K of $\mathbb{R}^m_*$ separated from the origin (i.e., with a positive distance to $\{0\}$) and any t in $]0,\infty[$. All this means that ν is indeed a series of Dirac measures of the following form

$$\nu(B\times]a,b])=\sum_k \mathbb{1}_{z_k\in B}\mathbb{1}_{s_k\in]a,b]},\quad \forall B\in\mathcal{B}(\mathbb{R}^m_*),\ a,b\in[0,\infty[, \tag{3.14}$$

where $\{(z_k,s_k):k\}$ is a sequence (possible finite) of points in $\mathbb{R}^m_*\times]0,\infty[$ such that all times $\{s_k:k\}$ are distinct. What should be remarked that there is not a particular order in the (strictly positive) real-valued sequence $\{s_k\}$, i.e., the index k is anyone of the several enumeration, but condition (c) requires that (d) for every $\varepsilon>0$ there is only a finite number of $\{(z_k,s_k):k\geq 1\}$ satisfying $\varepsilon\leq|z_k|$ and $s_k\leq 1/\varepsilon$. Thus, if desired the sequence can be completed to an infinite sequence by adding either $s_k=s_{k_0}$ or $s_k=\infty$, and $z_k=0$ for any $k>k_0$. A sequence (possible finite) $\{(z_k,s_k):k\geq 1\}$ of points in $\mathbb{R}^m_*\times]0,\infty[$ satisfying the property (d) and such that $\{s_k:k\}$ are distinct when finite (allowing $s_k=\infty$ as a technically) is called a *point-sequence*. The quantity (z_k,s_k) represents the k-jump of size z_k at time s_k, and if $(z,s)\mapsto f(z,s)$ is a non-negative Borel measurable function on $R^m_*\times]0,\infty[$ then the integral

$$\int_{R^m_*\times]0,\infty[} f(z,s)\,\nu(\mathrm{d}z,\mathrm{d}s)=\sum_k f(z_k,s_k),$$

and this gives rise to a cad-lag *purely jump* $\mathbb{R}^m$-valued function

$$\alpha_\varepsilon:t\mapsto\sum_{s_k\leq t} z_k\mathbb{1}_{|z_k|\geq\varepsilon}=\int_{\{(z,s):|z|\geq\varepsilon,\,0<s\leq t\}} z\,\nu(\mathrm{d}z,\mathrm{d}s), \tag{3.15}$$

for any $\varepsilon>0$, of which in general, the limit as ε vanishes may not exits or perhaps exits as a singular integral.

At this point it may be useful to recall some properties of cad-lag functions, for instance, the space $D([0,\infty[;\mathbb{R}^m)$ of all R_m-valued cad-lag functions, which becomes a Polish (metrizable, separable and complete) with a suitable topology. A key property of a cad-lag function ω is that for every $\varepsilon>0$ there exists a partition $0=t_0<t_1<\cdots<t_n=1/\varepsilon$ such that the oscillation of $s\mapsto\omega(s)$ on each subinterval $[t_{i-1},t_i[$ is small, precisely, $\sup\{|\omega(s)-\omega(s')|:s,s'\in[t_{i-1},t_i[\}<\varepsilon$ for every $i=1,\ldots,n$. Thus, if ω_ε is the piecewise constant

jump function define by $\omega_\varepsilon(s) = \omega(t_i) - \omega(t_i-)$ for any $t_i < s < t_{i+1}$ then the oscillation of the function $\omega - \omega_\varepsilon$ is smaller than ε within any closed subinterval $[t_{i-1}, t_i]$. Therefore, the limiting function $\omega_0(s) = \lim_{\varepsilon\to 0}\omega_\varepsilon(s)$ exists, uniformly on any bounded time interval $[0,T]$, and $\omega - \omega_0 = \omega_c$ is a continuous function, each cad-lag function ω can be written as the sum of a continuous function ω_c and a cad-lag purely jump function ω_0. In a symmetric way, the cag-lad functions are discussed.

Back to integer measure (it may be convenient for the reader to check Kallenberg [72, Introduction, pp. 1–13] for a full description and implications), if our interest is on the jump a given cad-lag function α (or equivalently, a cad-lag purely jump function α is given, not necessarily piecewise constant, since the Cantor-type function is an extreme example) then an integer measure ν can be defined by counting the jumps, i.e., $\nu(B\times]a,b])$ is equal to the number of jumps in the time-interval $]a,b]$ with size in B, for any Borel set in $\mathbb{R}^m_*$ separated from the origin. Note that property (c) or equivalently (d) is satisfied due to the cad-lag assumption on α. Thus an integer measure is constructed and associated to a cad-lag function, and in this case, the integer measure ν corresponding to the jumps of a cad-lag function α satisfies

$$\nu\big(\{(z,s) : |z| \geq \varepsilon,\ s \leq t\}\big) < \infty, \quad \forall \varepsilon > 0.$$

Usually, as $\varepsilon \to 0$, the jumps of size $|z| > 1/\varepsilon$ are called *large jumps*, while the jumps of size $|z| < \varepsilon$ are referred to as *small jumps*. Any cad-lag function has a finite number of jumps of size larger than $\varepsilon > 0$ within any bounded time-interval $]a,b]$, which implies that there is only a countable number of jumps on $[0,\infty[$, but this does not means that the jumps-times are discrete (i.e., separated, one from each other), actually, there could be accumulation of small jumps as ε vanishes.

For any cad-lag function α with a corresponding integer measure ν_α the quantities $\alpha(t)$ and $\nu_\alpha(\mathbb{R}^m_*\times]0,t])$ are related, and if

$$\alpha(t) = \alpha(0) + \lim_{\varepsilon\to 0}\int_{\{(z,s):|z|\geq\varepsilon,\,0<s\leq t\}} z\,\nu_\alpha(\mathrm{d}z,\mathrm{d}s).$$

then α is called a cad-lag purely jump function. If a cad-lag function has bounded variation on any bounded time-interval then the series of jumps $\sum_k z_k$ is absolutely convergent and the limit of α_ε in (3.15) exists as ε vanishes. However, if the integer measure ν is initially given and the limit as $\varepsilon \to 0$ of the piecewise constant function α_ε in (3.15) exists as a singular integral, i.e.,

$$\alpha_0(t) = \lim_{\varepsilon\to 0}\sum_k z_k \mathbb{1}_{|z_k|\geq\varepsilon}\mathbb{1}_{s_k\leq t}, \quad \forall t \in [0,\infty[, \tag{3.16}$$

either uniformly in every bounded time interval $[0,T]$ or in the topology of the (canonical) Polish space $D([0,\infty[;\mathbb{R}^m)$, then the cad-lag purely jump function α_0 can be defined. Certainly, this limit may not exist, and it could be regarded as an extra property imposed on the integer measure.

Because the time plays a important role, recall that the Lebesgue-Stieltjes (LS) and Riemann-Stieltjes (RS) integrals. If $s \mapsto \alpha$ is a cad-lag non-decreasing function then $\mathrm{d}\alpha(s)$ represents a Borel measure on $[0,\infty[$, finite on every bounded time-interval. Any Borel function bounded on any time-interval is LS-integrable but not necessarily RS-integrable. However, any cag-lad function is both, RS- and LS-integrable and both integral values agree. Moreover, if f is a cad-lag function and $f_- : s \mapsto f(s-)$ denotes its cag-lad counterpart, then

$$\int_{]0,t]} f(s)\mathrm{d}\alpha(s) = \int_{]0,t]} f_-(s)\mathrm{d}\alpha(s) + \sum_{0<s\leq t} [f(s)-f(s-)][\alpha(s)-\alpha(s-)],$$

where the integral on the left is considered in the LS-sense, while the integral on the right can be interpreted in either sense. Each cad-lag non-decreasing function α can be decomposed into its continuous-part α_c and its jump-part α_0. The integer measure is associated with the jump-part α_0, while the Lebesgue-Stieltjes measure on $[0,\infty[$ corresponding to $\mathrm{d}\alpha_c$ can be regarded as a measure on $\{0\}\times]0,\infty[$ associated with the continuous-part α_c. By means of signed measures, the same can be done with cad-lag functions with a continuous-part having bounded variation on every time-interval. Note that the jump-part may have unbounded variation, the cad-lag regularity allows the construction of the associated integer measure.

It should be clear that if ν is an integer measure with its corresponding point-sequence $\{(z_k,s_k) : k \geq 1\}$ then a function f on $\mathbb{R}^m_*\times]0,t]$ is ν-integrable if and only if the series $\sum_k |f(z_k,s_k)|\mathbb{1}_{s_k\leq t} < \infty$ and in this case

$$\int_{\mathbb{R}^m_*\times]0,t]} f(z,s)\,\nu(\mathrm{d}z,\mathrm{d}s) = \sum_k f(z_k,s_k)\mathbb{1}_{s_k\leq t}, \quad \forall t>0.$$

Thus is $|f|^p$ is integrable for some $p>0$ then $|f|^q$ is integrable for every $q \geq p$.

Perhaps the simplest case of purely jumps functions are piecewise constant functions, which normalized to be cad-lag take the form $t \mapsto \sum_k z_k \mathbb{1}_{s_k\leq t}$, where the jumps satisfy $|z_k| \to \infty$ and $0 < s_k \to \infty$ as $k \to \infty$. In this case, the time-jumps can be ordered as $0 < s_1 < s_2 < \cdots < s_k < \cdots$, and if $\dot{s}_k = s_k - s_{k-1}$ represents the time between two consecutive jumps then $\sum_k \dot{s}_k = \infty$. The corresponding integer measure ν is finite on any set of the form $\mathbb{R}^m_*\times]0,t]$, for every t in $[0,\infty[$. Next order of difficulty are the purely jumps functions having bounded variation on any bounded time-interval, where the corresponding integer measures integrate the function $(z,s) \mapsto |z|\mathbb{1}_{s\leq t}$ for any $t>0$. More general, an interesting situation is the case when the cad-lag function α has finite (or bounded) p-variation (with $p \geq 1$, and particularly when $p=2$) on any bounded time-interval, i.e., $\sum_{i=1}^n |\alpha(t_i)-\alpha(t_{i-1})|^p \leq C(\alpha,t)$, for any partition $0 = t_0 < t_1 < \cdots < t_n = t$, where now, the corresponding integer measures integrate the function $(z,s) \mapsto |z|^p\mathbb{1}_{s\leq t}$ for any $t>0$.

The following example may help to clarify some difficulties, consider the point-function given by $z_k = (-1)^{k+1}/k$ and $s_k = 2 - 1/k$ or equivalently the

integer measure on $\mathbb{R}_*\times]0,\infty]$ define by

$$\nu\big(K\times]a,b]\big) = \sum_{k=1}^{\infty} \mathbb{1}_{(-1)^{k+1}/k\in K}\,\mathbb{1}_{a<2-1/k\le b},$$

so that $\nu(\mathbb{R}_*\times]0,1]) = 0$, $\nu(\mathbb{R}_*\times]2,b]) = 0$, for every $b \ge 2$. In this case,

$$\alpha(t) = \lim_{\varepsilon\to 0} \sum_{k\varepsilon\le 1} \frac{(-1)^{k+1}}{k}\,\mathbb{1}_{2-1/k\le t} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}\,\mathbb{1}_{2-1/k\le t}$$

is the associated cad-lag purely jump function. Certainly, $\alpha(t) = 0$ if $0 \le t < 1$, $\alpha(t) = 1$ for $1 \le t < 3/2$, $\alpha(t) = 1/2$ for $3/2 \le t < 5/3$, $\alpha(t) = 1 - 1/2 + 1/3$ for $2 - 1/3 \le t < 2 - 1/4$, and so on, and $\alpha(t) = \sum_k (-1)^{k+1}/k$ for $t \ge 2$. This cad-lag purely jump function α has infinite variation on $[a,b]$ if $a < 2 < b$, but it has p-variation bounded if $p > 1$. Note that

$$\int_{\mathbb{R}\times]0,t]} |z|\,\nu(\mathrm{d}z,\mathrm{d}t) = \sum_k \frac{1}{k} = \infty, \quad \forall t \ge 2,$$

$$\text{but} \quad \alpha(t) = \lim_{\varepsilon\to 0} \int_{\{|z|\ge\varepsilon\}\times]0,t]} z\,\nu(\mathrm{d}z,\mathrm{d}s), \quad \forall t \ge 0.$$

Hence, the expression $\mathrm{d}\alpha$ is not a measure on $[0,\infty[$, but if $\alpha_- : s \mapsto \alpha(s-)$ denote its cag-lad counterpart then

$$\int_{]0,t]} \alpha_-(s)\,\mathrm{d}\alpha(s) = \sum_{0<s\le t} \alpha(s-)[\alpha(s) - \alpha(s-)], \quad \forall t > 0$$

as a singular Riemann-Stieltjes integral, which suggests

$$\int_{]0,t]} [\alpha(s) - \alpha(s-)]\,\mathrm{d}\alpha(s) = \sum_{0<s\le t} [\alpha(s) - \alpha(s-)]^2, \quad \forall t > 0,$$

as a definition, since the series is absolutely convergent. In other words, if a cad-lag purely jump process α is given and its integer measure ν_α integrates the function $z \mapsto |z|^2$ then it make sense to define

$$\int_{]0,t]} [\alpha(s) - \alpha(s-)]\cdot \mathrm{d}\alpha(s) = \int_{\mathbb{R}_*^m} |z|^2 \nu_\alpha(\mathrm{d}z,\mathrm{d}s) = \sum_{0<s\le t} [\alpha(s) - \alpha(s-)]^2,$$

as an integral in a reasonable sense, which is the quadratic variation of the function α on the time-interval $]0,t]$. Certainly, more sophisticated examples can be produced with analogous arguments.

If h is a Borel measurable function from $\mathbb{R}_*^n\times]0,+\infty[$ into $\mathbb{R}_*^m$ then any point-function $\{(z_k,s_k) : k\}$ or integer measure ν in $\mathbb{R}_*^m\times]0,\infty[$ is transformed into another point-function $\{(h(z_k,s_k),s_k) : k\}$ or integer measure $\nu_h(B\times]a,b]) = \nu(\{(z,s) : h(z,s) \in B,\, a < s \le b\})$, where the key properties (a),...,(d) are

preserved. However there some difficulties to change the time scale, and an adaptation is necessary.

Until now, integer measures and point-sequences have a one-to-one relation and both are loosely interpreted as a series of Dirac measures. Due to the condition (b) imposed on integer measures (i.e., $\nu(\mathbb{R}^m_* \times \{s\})$ is either 0 or 1) a finite sum of integer measures is not necessarily another integer measure. On the other hand, for a point-function $\{(z_k, s_k) : k\}$, the condition that $\{s_k : k\}$ should be a sequence (possible finite) of positive *distinct* numbers sounds odd on a second analysis.

Definition 3.12. A *point-sequence* on $\mathbb{R}^m \times [0,\infty]$ is a sequence $\{(z_k, s_k) : k\}$ of points in $\mathbb{R}^m \times [0,\infty]$ such that for any $\varepsilon > 0$ there is only a finite number of indexes k satisfying $|z_k| \geq \varepsilon$ and $1/\varepsilon \leq s_k < \infty$. An *integer measure* ν on $\mathbb{R}^m_* \times [0,\infty[$ is a Borel measure such that $\nu(\{(z,s) : |z| \geq \varepsilon,\, s \leq t\})$ is a nonnegative integer, for any $\varepsilon > 0$. If $\nu(\mathbb{R}^m_* \times \{t\})$ is either 0 or 1, for every $t > 0$, then the integer measure is called *simple*, and if $\nu(\mathbb{R}^m_* \times]0,t]) < \infty$, for every $t > 0$, then the integer measure is called *bounded*. In general, ν is of bounded variation on any bounded time-interval if the function $z \mapsto 1 \wedge |z|$ is ν-integrable and ν is an integer measure of Lévy-type if the function $z \mapsto 1 \wedge |z|^2$ is ν-integrable. *Sometimes*, it may be necessary to add a normalizing condition at time $t = 0$, namely, either $s_k > 0$ or $\nu(\mathbb{R}^m_* \times \{0\}) = 0$, in this case, ν is considered a measure on $\mathbb{R}^m_* \times]0,\infty[$. Regarding a point-sequence, it is clear that the points in $\{0\} \times [0,\infty[$ or $\mathbb{R}^m \times \{\infty\}$ does not intervene and can be eliminated from the sequence, in this case, the remaining points form a sequence that may be finite or empty. Thus, allowing points in either $\mathbb{R}^m_* \times]0,\infty]$ or $\mathbb{R}^m \times]0,\infty[$, the indexes of the sequence $\{(z_k, s_k) : k\}$ can be assumed infinite. It may be convenient to regard this sequence as a *point-function*, i.e., a function from $]0,\infty[$ into $\mathbb{R}^m$ vanishing except in a countable (possible finite) set, i.e., mapping a time s into z_k if $s = s_k$ and into 0 otherwise. Using the space $\mathbb{R}^m_* \times]0,\infty]$ instead of $\mathbb{R}^m \times]0,\infty]$ for the integer measure helps to avoid the temptation of assigning a non-zero value to the origin. □

With this definition, there is a one-to-to correspondence between point-functions and integer measures, and a finite sum of integer measures is again an integer measure. However, only a simple integer measure is associated with the jumps of a cad-lag function; and a cad-lag function can be associated to a integer measure only if the cad-lag purely jumps function (3.16) can be defined. Note that because ν takes integer values, if the function $z \mapsto 1 \wedge |z|^p$ is ν-integrable then so is $z \mapsto 1 \wedge |z|^q$, for every $q \geq p$. Because there is only a finite number of indexes k such that $|z_k| \geq 1$ and $\tau_k \leq T$, the function $z \mapsto |z|^p \mathbb{1}_{|z|\geq 1}$ is ν-integrable for every $p \geq 0$. Note that point-sequences and point-functions are alternative viewpoints of countable sets, i.e., either as a sequence $\{(z_k, s_k) : k\}$ indexed by the natural numbers $\{k = 1, 2, \ldots\}$ or as a countable (possible finite) set of times $S \subset [0,\infty]$ (or just $S \subset]0,\infty[$) used as index in $\{z_s : s \in S\} \subset \mathbb{R}^m_*$.

Therefore, the series expression

$$\nu(B\times]a,b]) = \sum_{s\in S} \mathbb{1}_{z_s\in B}\,\mathbb{1}_{a<s\leq b}, \quad \forall B\in\mathcal{B}(\mathbb{R}^m_*),\ a,b\in[0,\infty[,$$

defines the integer measure ν. Clearly, for a given cad-lag function α, the set of time-jumps is $S=\{s\in]0,\infty[\colon \alpha(s)-\alpha(s-)\neq 0\}$ and the jump-set is $z_s=\alpha(s)-\alpha(s-)$.

Hence, to change the time use a Borel measurable function τ from $\mathbb{R}^m_*\times]0,\infty]$ into $]0,\infty]$ such that for every $\varepsilon>0$ there is only a finite number of indexes k satisfying $|z_k|\geq\varepsilon$ and $\tau(z_k,s_k)\leq 1/\varepsilon$. Therefore, an integer measure ν is transformed as follows:

$$\nu_\tau\big(B\times]a,b]\big) = \nu\big(\{(z,s)\colon z\in B,\ a<\tau(z,s)\leq b\}\big),$$

for every B in $\mathcal{B}(\mathbb{R}^m_*)$ and $0\leq a<b<\infty$, while a point-function $\{(z_k,s_k)\colon k\}$ becomes another point-function $\{(z_k,\tau(z_k,s_k))\colon k\}$. Note that the index k could be any positive integer, and that nonsense empty (or constantly infinite-valued) point-function or the null integer measure is allowed, as well as the transformation $\tau(z,s)=\infty$, which produces $\nu_\tau=0$. This means that the possibility of canceling a jumps (z_k,s_k) is accomplished by mapping it to (z_k,∞). Certainly, both transformations can be combined to have $\vartheta_k(z,s)=(h_k(z,s),\tau_k(z,s))$ under the condition that for every $\varepsilon>0$ there is only a finite number of indexes k satisfying $|h_k(z_k,s_k)|\geq\varepsilon$ and $\tau(z_k,s_k)\leq 1/\varepsilon$.

The special role of the time variable force a limitation on the previous transformation $(z_k,s_k)\mapsto(z_k,\tau(z_k,s_k))$, due essentially to the arrow of time, i.e., the change in the distributions of the jumps produced by stretching or shrinking the time-scale should preserve the order of jumps. Therefore, usually a *monotonic* assumption is required, i.e., if $s_i\leq s_j$ then the new jumps $\tau(s_i)\leq\tau(s_j)$, i.e., the relative order of jumps is preserved in time. Thus, eliminating a jump with the choice of $\tau_k=\infty$ forces the elimination of all successive jumps which is clearly not a desired action, a solution is to set $z_k=0$ and so, the point-function take values in $\mathbb{R}^m\times[0,\infty[$ but the integer measure is restricted to $\mathbb{R}^m_*\times[0,\infty[$. In this sense, a proper jumps occurs only when the size $z_k\neq 0$ and the time τ_k is finite. The stretching or shrinking of the time-scale under the monotonic assumption makes a clear sense whenever the jumps are ordered $s_1<s_2<\cdots<s_k<\cdots$, as is the case of a bounded integer measure.

3.6 Sequences of Probability

From the previous sections, the space of continuous functions $C([0,\infty[)$ with the locally uniform convergence and the space of cad-lag functions $D([0,\infty[)$ with the Skorokhod topology as in Definition 3.6 and metric (3.11) are both Polish spaces, i.e., complete separable metrizable spaces, and therefore, their Borel σ-algebras coincide with their cylindrical σ-algebras.

Recall that a sequence $\{P_n\}$ of probabilities on a Polish space Ω is tight if for every $\varepsilon>0$ there exists a compact set $K\subset\Omega$ such that $P_n(\Omega\smallsetminus K)<\varepsilon$, for

every n, see Definition 1.8. Moreover, Prohorov Theorem 1.9 states that any tight subset is pre-compact.

To recall that T in the expression for the modulus of continuity (3.1) and the cad-lag modulus of continuity (3.8) refers to the interval $[0,T]$, we use the notation $w_c(\omega,\delta,[0,1/\varepsilon])$ and $w(\omega,\delta,[0,1/\varepsilon])$ instead of $w_c(\omega,\delta,1/\varepsilon)$ and $w(\omega,\delta,1/\varepsilon)$ to have

Theorem 3.13. *Let $\{P_n\}$ be a sequence of probability measures on $C([0,\infty[)$ (respectively, on $D([0,\infty[)$) with its Borel σ-algebra. If for every $\varepsilon>0$ there exist $\delta>0$ such that (1) $P_n\{\omega : \sup_{t\in[0,1/\varepsilon]}|\omega(t)|\geq 1/\delta\}<\varepsilon$ and (2) $P_n\{\omega : w_c(\omega,\delta,[0,1/\varepsilon])\geq\varepsilon\}<\varepsilon$ (respectively, $P_n\{\omega : w(\omega,\delta,[0,1/\varepsilon])\geq\varepsilon\}<\varepsilon$), then there exists a weakly convergence subsequence.*

Proof. The arguments are rather simple, namely, based on Prohorov Theorem 1.9, for every given $\varepsilon>0$ we have to construct a compact set K_ε satisfying $P_n(K_\varepsilon)\geq 1-\varepsilon$, for every n.

Indeed, for every $\varepsilon>0$, $T>0$ and $k>0$ there exists $R=R(\varepsilon,T)$ and $r=r_k(\varepsilon,T)$ such that

$$P_n\Big\{\sup_{t\in[0,T]}|\omega(t)|\geq R\Big\}<\varepsilon 2^{-T-1},$$
$$P_n\big\{w(\omega,r_k,[0,T])|\geq 1/k\big\}<\varepsilon 2^{-T-k-1}.$$

Thus, if

$$K_{\varepsilon,T}=\Big\{\omega:\sup_{t\in[0,T]}|\omega(t)|<R,\ w(\omega,r_k,[0,T])|<1/k,\ \forall k\geq 1\Big\}$$

then

$$1-P_n(K_{\varepsilon,T})\leq P_n\Big\{\sup_{t\in[0,T]}|\omega(t)|\geq R\Big\}+$$
$$+\sum_{k=1}^{\infty}P_n\big\{w(\omega,r_k,[0,T])|\geq 1/k\big\}\leq\varepsilon 2^{-T}.$$

Hence we deduce that the compact set $K_\varepsilon=\bigcap_{T=1}^{\infty}K_{\varepsilon,T}$ satisfies $P_n(K_\varepsilon)\geq 1-\varepsilon$, for every n. □

For a family $\{P_i : i\in I\}$ of probability measures on the canonical sample space $D([0,\infty[)$, consider the following conditions:

(a) for every $\varepsilon>0$ there exist $\delta>0$ such that

$$P_n\{\omega:|\omega(0)|>1/\delta\}<\varepsilon,\quad\forall i\in I;$$

(b1) for every $\varepsilon>0$ and $0\leq t<s<t'\leq 1/\varepsilon$ there exist $\delta>0$ such that

$$P_n\{\omega:|\omega(t')-\omega(s)|\wedge|\omega(s)-\omega(t)|>\varepsilon\}\,\delta<(t'-t)^{1+\delta}\varepsilon^{-1/\delta},\quad\forall i\in I;$$

(b2) for every $\varepsilon > 0$ and $0 \le s < t \le 1/\varepsilon$ there exist $\delta > 0$ such that

$$P_n\{\omega : |\omega(t) - \omega(s)| > \varepsilon\}\, \delta < t^{1+\delta}\varepsilon^{-1/\delta}, \quad \forall i \in I;$$

(b) for every $\varepsilon > 0$ and $0 \le t < s \le 1/\varepsilon$ there exist $\delta > 0$ such that

$$P_n\{\omega : |\omega(t) - \omega(s)| > \varepsilon\}\, \delta < (s-t)^{1+\delta}\varepsilon^{-1/\delta}, \quad \forall i \in I.$$

Essentially with the technique of Theorem 3.13, we can show that (a), (b1) and (b2) imply that the family is tight. Note the extra condition (b2) (the oscillation at 0) need to replace the cad-lag modulus of continuity (3.8) with either w' given by (3.4) or w'' as in (3.9). The condition (b) is particular to (but used otherwise) probability measures in the space $C([0,\infty[)$, and we say that $\{P_i : i \in I\}$ is C-tight if (a) and (b) are satisfied. In this case, any limit point (in the weak convergence of measures topology) is a probability measure with support in the (closed) subspace $C([0,\infty[)$ of $D([0,\infty)$. It is also clear that in condition (a) we may use any fixed time T instead of the origin 0.

Most of the times, the condition (b1) and (b2) are deduced from the following estimate: there are $p, q > 0$ such that for every $T > 0$ we can find $C_T > 0$ satisfying

$$\mathbb{E}_n\big\{\big|\omega(s) - \omega(t)\big|^p \,\big|\, \mathcal{D}_t\big\} \le C_T(s-t)^{\frac{1}{2}+q}, \quad \forall 0 \le t < s \le T,\ \forall n,$$

where $\mathcal{D}_t$ is the σ-algebra defined in the previous section, and $\mathbb{E}_n\{\cdot \,|\, \mathcal{D}_t\}$ denotes the conditional expectation with respect to P_n. While

$$\mathbb{E}_n\big\{\big|\omega(s) - \omega(t)\big|^p\big\} \le C_T(s-t)^{1+q}, \quad \forall 0 \le t < s \le T,\ \forall n$$

is used to obtain condition (b). There are other ways of proving tightness of a family of probability, or criterium of compactness, but details are left for a more advance course.

As mentioned early, the reader is referred, for instance, to the books Billingsley [15, Chapter 3, pp. 109–153], Jacod and Shiryaev [69, Sections VI.1 and VI.2, pp. 288–310], Ethier and Kurtz [47, Section 3.5, pp. 116–154], Gikhman and Skorokhod [53, Chapter 4, pp. 144–173], and Pollard [105, Chapter V, pp. 89–121], among others.

3.7 Convergence of Processes

Since the concept of stochastic processes lead to the study of probability measures on a separable and complete metric space (also called *Polish space*), we continue the discussion initiated in Chapter 1 Section 1.3, where a processes is regarded as either a family of random variables and/or a probability (i.e., its distribution) on a Polish space.

A good discussion on this subject can be founded in Billingsley [15, Chapters 1,2 and 3, pp. 1–215], Ethier and Kurtz [47, Chapter 3, pp. 95–154] or Ikeda and Watanabe [62, Chapter 1, pp. 1–44]. We are going to state some of the

key elements, besides what was discussed in Chapter 1, Section 1.3, where the probability was regarded independently of the process itself. Actually, a key point is to focus in the sample space Ω and to use several probabilities as the distributions of processes under consideration.

First consider $\mathcal{P}(\Omega)$ the family of probability measures on $(\Omega, \mathcal{B})$, where $\mathcal{B} = \mathcal{B}(\Omega)$ is the Borel σ-algebra on the Polish space Ω. This is also referred to as the family of Borel probability measures on Ω. The Prohorov metric on $\mathcal{P}(\Omega)$ is defined by

$$d(P, Q) = \inf\{\varepsilon > 0 : P(A) \le Q(A^\varepsilon) + \varepsilon, \ \ \forall A \text{ closed in } \Omega\},$$

where $A^\varepsilon = \{\omega \in \Omega : \inf_{\omega' \in A} d_\Omega(\omega, \omega') < \varepsilon\}$, and $d_\Omega(\cdot, \cdot)$ is the metric on Ω. Thus $\mathcal{P}(\Omega)$ endowed with the above metric becomes a Polish space.

Denote by $C_b(\Omega)$ the space of real-valued continuous function on the Polish space (Ω, d_Ω) with the natural norm $\|f\| = \sup_\omega |f(\omega)|$. A sequence $\{P_n : n = 1, 2, \dots\}$ in $\mathcal{P}(\Omega)$ is said to *converge weakly* to P if

$$\lim_{n\to\infty} \int f(\omega)\, P_n(\mathrm{d}\omega) = \int f(\omega)\, P(\mathrm{d}\omega), \qquad \forall f \in C_b(\Omega).$$

Actually, even in a metric space Ω, and if the integral of f with respect to the probability P_n or P is written as $P_n(f)$ or $P(f)$ then, the above convergence is equivalent to any of the following conditions:

(1) $P_n(f) \to P(f)$, for every bounded Lipschitz function f;
(2) $\limsup_n P_n(f) \le P(f)$, for every upper semi-continuous function bounded from above;
(3) $\limsup_n P_n(f) \ge P(f)$, for every lower semi-continuous function bounded from below;
(4) $\limsup_n P_n(f) \le P(f)$, for every function of the form $f = \mathbb{1}_C$, with C a closed set;
(5) $\limsup_n P_n(f) \ge P(f)$, for every function of the form $f = \mathbb{1}_O$, with O an open set;
(6) $\limsup_n P_n(f) = P(f)$, for every function of the form $f = \mathbb{1}_B$, with B a Borel set with boundary ∂B having zero P-probability.

Moreover, it is sufficient to take only functions f which are uniformly continuous. Furthermore, if we know that the limit measure is a probability then it is enough to satisfy the convergence for uniformly continuous functions with a bounded support, even more, if the space Ω is locally compact, then it suffices to use continuous functions with a compact support. The important point here is that the convergence in the Prohorov metric is equivalent to the above weak convergence.

A classic result, so-called Skorokhod representation, gives some relation with the almost surely convergence.

Theorem 3.14 (Skorokhod)**.** *Let $\{P_n : n = 1, 2, \dots\}$ be a sequence of probability measures on a Polish space Ω which converge weakly to P. Then in some common probability space there exist random variables $X_n : n = 1, 2, \dots$ and X*

with values in Ω with distributions $P_n : n = 1, 2, \dots$ and P, respectively, such that $\lim_{n\to\infty} X_n = X$ almost surely. □

In the above the condition Polish space can be replaced by metric space, provided the support of the probability measure P is separable.

One of the typical applications of this representation is the fact that the weak convergence is preserved by Borel mapping which is almost continuous with respect to the limiting measure, i.e., suppose that X is a Borel measurable function form Ω into itself, where (Ω, d_Ω) is a separable and complete metric space, and $\{P_n : n = 1, 2, \dots\}$ is a sequence of probability measures on Ω which converges weakly to P; if $\{Q_n : n = 1, 2, \dots\}$ and Q are the image measures through the mapping X of $\{P_n : n = 1, 2, \dots\}$ and P then the sequence $\{Q_n : n = 1, 2, \dots\}$ converges weakly to the measure Q, provided X is P-almost surely continuous.

Another point is the characterization of pre-compacts or relatively compact sets (i.e., with a compact closure) set in $\mathcal{P}(\Omega)$ with the weak convergence i.e., endowed with the Prohorov metric.

Theorem 3.15 (Prohorov)**.** *A sequence of probability measures $\{P_n : n = 1, 2, \dots\}$ on a Polish space Ω has a weakly convergent subsequence if and only if the sequence is tight i.e, for any $\varepsilon, \delta > 0$ there exist $\omega_1, \dots, \omega_n$ in Ω such that*

$$P_n(\bigcup_{i=1}^{n} \{\omega : d_\Omega(\omega_i, \omega) \le \delta\}) \ge 1 - \varepsilon$$

for all $n = 1, 2, \dots$. □

Usually, a family of probability measures $\{P_\alpha\}$ on Ω is said to be *tight* if for any $\varepsilon > 0$ there exists a compact set $K \subset \Omega$ such that $P_\alpha(K) \ge 1 - \varepsilon$ for any index α. Since a set is pre-compact in Ω if and only if it is *totally bounded* this is equivalent to the above condition.

In view of the above characterization of the weak convergence of measures, it is important to understand the structure of compact sets in the particular spaces $C([0, \infty), E)$ and $D([0, \infty), E)$, where (E, d_E) is a Polish space, in particular $E = \mathbb{R}^d$. Classic results applies to say that pre-compact sets are equivalent to totally bounded and equi-continuous sets. Thus a family $\{\omega_\alpha\}$ of functions in $C([0, \infty), E)$ is relatively compact if and only if

(a) for any $\delta > 0$ and rational $r \ge 0$, there exist $x_1, \dots, x_n$ in E such that for any index α we have

$$\omega_\alpha(r) \in \bigcup_{i=1}^{n} \{x : d_E(x_i, x) \le \delta\}$$

(b) for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index α, and any t, s in $[0, T]$ we have $d_E(\omega_\alpha(t), \omega_\alpha(s)) < \delta$.

The fact that in (a) we require the condition to be satisfied only for rational

is convenient for later. Now, for the space $D([0,\infty),E)$ we need to use the modulus of continuity $w(\omega,\delta,T)$ defined by

$$w(\omega,\delta,T) := \inf_{\{t_i\}} \sup_i \sup\{d_E(\omega(t),\omega(s)) : t_{i-1} \le s < t < t_i\}$$

where $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \cdots < t_n = T$, with $t_i - t_{i-1} \ge \delta$ and $n \ge 1$. A shorter version of the modulus of continuity is given by the expression

$$w(\omega,\delta,T) := \sup_{0\le t<T-\delta} \sup_{t\le s\le t'\le t+\delta} \{d_E(\omega(t'),\omega(s)) \wedge d_E(\omega(s),\omega(t))\},$$

where $\wedge$ means the minimum between numbers. Therefore, we replace (b) by the condition

(b1) for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index α we have $w(\omega_\alpha,\delta,T) < \varepsilon$.

It is cleat that if $E = \mathbb{R}^d$ then (a) reduces to

(a1) for any rational $r \ge 0$, there $N > 0$ such that for any index α we have $|\omega_\alpha(r)| \le N$.

Theorem 3.16 (tight)**.** *Let $X_1, X_2, \dots$ be a sequence of random variables with values in $D([0,\infty),E)$, with E a Polish space and $P_1, P_2, \dots$ be its associated probability law on $D([0,\infty),E)$. Then the sequence $P_1, P_2, \dots$ is tight (hence relatively compact) in $D([0,\infty),E)$ if an only if the following two conditions hold:*

(a') *for any $\varepsilon,\delta > 0$ and rational $r \ge 0$, there exist $x_1,\dots,x_k$ in E such that for any index n we have*

$$P_n(X_n(r) \in \bigcup_{i=1}^{k} \{x : d_E(x_i,x) \le \delta\}) \ge 1 - \varepsilon,$$

(b') *for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index n we have*

$$P_n(w(X_n,\delta,T) \ge \varepsilon) < \varepsilon.$$

Moreover, if the sequence is tight, then it is weakly convergent if and only its finite-dimensional distributions converge. □

It is clear that some obvious modifications should be done for a sequence of random variables in the space $C([0,\infty),E)$, i.e., re-defining $w(\omega,\delta,T)$ as $\sup\{d_E(\omega(t),\omega(s)) : |t-s| < \delta,\ s,t \in [0,T]\}$. In the space $C([0,\infty),\mathbb{R}^d)$, condition (b') simply becomes:

(b') for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$P_n\big(\sup_{s,t\in[0,T],\ |t-s|<\delta} |X_n(t) - X_n(s)| \ge \varepsilon\big) < \varepsilon, \tag{3.17}$$

for any index n.

Naturally, the above result is useful to study cad-lag processes. It may be convenient to use Aldous' criterion in $D([0,\infty),\mathbb{R}^d)$, e.g., see Ethier and Kurtz [47, p. 137, Theorem 8.6, Chapter 3] or Liptser and Shiryayev [88, Section 6.3, pp. 515–519]. This is to replace condition (a') and (b') of the previous theorem with the following statement:

(a*) for any $\varepsilon > 0$ there exists $M > 0$ such that for any index n we have

$$P_n(|X_n(0)| \geq M) < \varepsilon, \tag{3.18}$$

(b*) for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index n we have

$$\sup_{0<s\leq\delta} P_n(|X_n(\tau_n + s) - X_n(\tau_n)| \geq \varepsilon) < \varepsilon, \tag{3.19}$$

for any stopping time (relative to X_n) τ_n satisfying $0 \leq \tau_n \leq T$. The key facts here are that the sup is outside of the *integral* and that s is a (deterministic) number, so that $\tau_n + s$ becomes an optional time with respect to process $X_n(\cdot - s)$. Moreover, (b*) is equivalent to the following condition:

(b") for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index n we have

$$P_n(|X_n(\theta_n) - X_n(\tau_n)| \geq \varepsilon) < \varepsilon, \tag{3.20}$$

for any stopping times (relative to X_n) θ_n and τ_n satisfying $0 \leq \theta_n \leq \tau_n \leq T$ and $\tau_n \leq \theta + \delta$.

If (P_n, X_n) is a sequence of homogeneous strong Markov processes in the canonical space $D([0,\infty),\mathbb{R}^d)$ with transition probability measure $P_n(x,t,\mathrm{d}y)$, and a sequence of stopping times τ_n, then the equality, with some $r > 0$,

$$\mathbb{E}\big\{\mathbb{E}\{|X_n(\tau_n + s) - X_n(\tau_n)|^r \,|\, X(\tau_n)\}\big\} = \\ = \int_{\mathbb{R}^d} \mathbb{E}\big\{|y - X(\tau_n)|^r P_n(X(\tau_n), s, \mathrm{d}y)\big\}$$

shows that Aldous' criterion conditions (b*) (3.19) is satisfied if

$$\lim_{s\to 0} \sup_n \int_{\mathbb{R}^d} |y - x|^r P_n(x, s, \mathrm{d}y) = \\ = \lim_{s\to 0} \sup_n \mathbb{E}\big\{|X_n(s) - X_n(0)|^r \,|\, X_n(0) = x\big\} = 0$$

which is a simple condition to verify. Moreover, the expression $|y - x|^r$ above could be replaced by $\gamma(|y - x|)$ with a strictly increasing continuous function $\gamma(\cdot)$ satisfying $\gamma(0) = 0$. For instance, the reader may check the book Bass [5, Chapter 34, pp 259–268].

The convergence of finite-dimensional distributions of a sequence $\{X_n : n \geq 1\}$ of $\mathbb{R}^d$-valued stochastic processes, means that for any finite number of times $t_1, \ldots, t_k$ we have

$$\lim_n \mathbb{E}\{h(X_n(t_1), \ldots, X_n(t_k))\} = \mathbb{E}\{h(X(t_1), \ldots, X(t_k))\},$$

for any continuous and bounded real-valued function h on $\mathbb{R}^k$. On the contrary to the convergence in $D([0,\infty), \mathbb{R}^d)$, no convergence condition on the paths is involved in this concept.

To control the modulus of continuity of a process $X(t)$, the following estimate is very useful. For any $\alpha, \beta > 0$ there exists a constant $C_0 = C_0(\alpha, \beta)$ such that

$$|f(t') - f(s')|^\alpha \leq C_0\, |t' - s'|^\beta \int_0^T \mathrm{d}t \int_0^T \frac{|f(t) - f(s)|^\alpha}{|t - s|^{2+\beta}} \mathrm{d}s, \tag{3.21}$$

for any continuous function f on $[0,T]$ and any t', s' in $[0,T]$, see Da Prato and Zabczyk [31, Theorem B.1.5, pp. 311–316] or Stroock and Varadhan [129, Theorem 2.1.3, pp. 47–49]. Therefore, if for some constants p, q, $C > 0$ a process $X(t,\omega)$ satisfies

$$\mathbb{E}\big\{|X(t) - X(s)|^p\big\} \leq C|t - s|^{1+q}, \quad \forall t, s \in [0,T], \tag{3.22}$$

then by taken $p = \alpha$ and $\beta = r$ with $0 < r < q$ we deduce that there is another constant $C_0 = C_0(p, q, C, r)$ such that

$$\mathbb{E}\Big\{ \sup_{t,s\in[0,T],\ |t-s|<\delta} \{|X(t) - X(s)|^p\}\Big\} \leq C_0\, \delta^r, \quad \forall \delta > 0. \tag{3.23}$$

Essentially, an estimate in L^∞ of the modulus of continuity is obtained based on an estimate in L^p. This is of particular interest for stochastic processes with continuous paths.

For cad-lag processes, a bound of the type: for every $0 \leq t \leq s \leq t + \delta \leq T$, and some positive constants C, p and q,

$$\mathbb{E}\big\{\big[|X(t+\delta) - X(s)| \wedge |X(s) - X(t)|\big]^p\big\} \leq C\delta^{1+q}, \quad \forall \delta > 0, \tag{3.24}$$

yields the estimate

$$\mathbb{E}\Big\{ \sup_{0\leq t\leq T-\delta} \sup_{t\leq s\leq s+\delta} \{|X(t+\delta) - X(s)| \wedge |X(s) - X(t)|\}^p\Big\} \leq C_0\, \delta^r, \tag{3.25}$$

for every $\delta > 0$, any $0 < r < q$ and another constant $C_0 = C_0(p, q, C, r)$. The reader may consult the books Billingsley [15, Chapter 3, pp. 109–153] or Ethier and Kurtz [47, Chapter 3, pp. 95–154] for a complete discussion.

Sometime we have to use the space $B(\Omega)$ of all bounded and Borel measurable functions from the Polish space Ω into $\mathbb{R}^n$. The weak topology we need is the so-called *boundedly and pointwise* convergence i.e., a sequence of functions $\{f_n : n = 1, 2, \ldots\}$ in $B(\Omega)$ converge boundedly and pointwise to f if

$\sup_{n,\omega} |f_n(x)| < \infty$ and $f_n(x) \longrightarrow f(\omega)$ for every $\omega \in \Omega$. The (sequentially) closure of a set M in this topology is referred to as the bp-closure of M. A typical application of the Monotone Class Theorem shows that the bp-closure of $C_b(\Omega)$ is the whole space $B(\Omega)$ i.e., it is bp-dense. Moreover, since Ω is separable, there exists a sequence $\{f_n : n = 1, 2, \dots\}$ of nonnegative continuous and bounded functions that span a bp-dense set in $B(\Omega)$. Note that this is not to say that any function in $B(\Omega)$ is a boundedly and pointwise limit of a sequence of function in $C_b(\Omega)$. Due to the probability measure, we prefer to use the Lebesgue space $L^\infty(\Omega, \mathcal{F}, P)$ instead of $B(\Omega)$, when ever is possible. The reader may consult the books Doob [38, Chapters VIII,...,X, pp. 123–177], Jacod and Shiryaev [69, Chapter VI, pp. 288–347], among other, for a complete discussion of convergence of measures and processes.

3.8 Existence of Probabilities

At this point, the reader may revise the some of the basic subjects treated in the book Malliavin [92]. In particular, a revision on measure theory, e.g., as in Kallenberg [71, Chapters 1 and 2, pp. 1–44], may be necessary.

Perhaps the Gaussian probability in $\mathbb{R}^n$ is the best well now situation, i.e., the probability space $(\mathbb{R}^n, \mathcal{B}^n, P_n)$, where $\mathcal{B}^n$ is the Borel σ-algebra in $\mathbb{R}^n$ and P_n is the probability measure given by

$$P_n(A) = (2\pi)^{-n/2} \int_A \exp\Big(-\frac{1}{2} \sum_{i=1}^n |x_i|^2 \Big) \mathrm{d}x, \quad \forall A \in \mathcal{B}^n,$$

standard normal distribution or Gaussian with mean 0 and variance 1. However, the extension of this probability to the space $\mathbb{R}^\infty$ of all sequences of real numbers (with the product topology) is not so trivial. If X_n denotes the projection from $\mathbb{R}^\infty$ into $\mathbb{R}^n$, i.e., $X_n(x) = (x_1, \dots, x_n)$ the first n coordinates of x, then it is not immediate to establish the existence of a probability P defined on the Borel σ-algebra $\mathcal{B}^\infty$ such that $P(A) = P_n(X_n(A))$ for any A in $\mathcal{B}^\infty$. In general, the two points of interest here are the σ-algebra generated by the cylindrical sets $\{X_n^{-1}(A) : A \in \mathcal{B}^n, n \geq 1\}$ and the σ-additivity of P. Therefore, on the probability space $(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$, we may look at $\{X_n : n \geq 1\}$ as a sequence of Gaussian random variables which generates the σ-algebra $\mathcal{B}^\infty$. It is well know that the Hermit polynomials provide an orthonormal basis for the Hilbert space $L^2(\mathbb{R}^1, \mathcal{B}^1, P_1)$, however some tedious notation and details are needed to deduce an orthonormal basis for $L^2(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$.

Now, our interest turns into the existence of probability measures, first in $\mathbb{R}^n$, next in separable Hilbert spaces and finally in Polish spaces, particularly in the space of tempered distributions. Thus, the discussion about the existence of a particular stochastic process with values in $\mathbb{R}^n$ becomes a discussion on the existence of probability measures on relatively large spaces, such as the Schwartz space of tempered distributions, where the Fourier transform can be used.

One way of constructing a probability measure is by prescribing its characteristic function (or its Fourier transform). In finite dimensional spaces we have

the classical Bochner's Theorem stated as follow:

Theorem 3.17. *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, $i = 1, \ldots, k$ 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 $\mathbb{R}^n$.* □

This is also known as Bochner-Khintchin's Theorem, for instance, a complete proof can be find in Gnedenko [54, Section 39, pp. 289–293] or Jacob [67, Vol 1, Theorem 3.5.7, pp. 108–109].

Next, the (Schwartz) space of rapidly decreasing and smooth functions $\mathcal{S}(\mathbb{R})$ and its dual space of tempered distributions $\mathcal{S}'(\mathbb{R})$ is identified (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 $\{\xi_n(x) : n \geq 1\}$, with

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

where p_n is the Hermite polynomial of order n) with the Fréchet space of rapidly decreasing sequences

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

This space 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

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

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,

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

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 (e.g., see Holden et al. [61, Appendix A, pp. 193–197]) 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 $|\Psi(a) - 1| < \varepsilon$, which yields

$$\int_{\mathbb{R}^\infty} \cos(\langle a', a\rangle)\, P(\mathrm{d}a') \geq 1 - \varepsilon - 2\delta^{-2}\|a\|_m^2, \quad \forall a \in s. \tag{3.26}$$

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

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

Now, take $b_k = (1+k^2)^{-m-1}$ to have $\sum_{k=0}^{n}(1+k^2)^m b_k = C < \infty$, 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 can state the following version of a Bochner-Minlos theorem: On the space of test functions $\mathcal{S}(\mathbb{R})$ we give a functional Ψ which is positive definite, continuous and satisfies $\Psi(0) = 1$, 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.

Certainly, this 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 3.18 (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*

$$\int_{\mathbb{R}^d}\big(|y|^2\wedge|y|\big)\pi(\mathrm{d}y)<\infty,\qquad \pi(\{0\})=0. \tag{3.27}$$

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

$$\begin{aligned}\mathbb{E}\big\{\exp\big[\mathtt{i}\langle\cdot,\varphi\rangle\big]\big\}&=\exp\Big(-\frac{1}{2}\int_{\mathbb{R}}|\sigma\varphi(t)|^2\mathrm{d}t\Big)\times\\&\times\exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}\big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)}-1-\mathtt{i}(\varphi(t),y)\big]\pi(\mathrm{d}y)\Big),\end{aligned} \tag{3.28}$$

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

$$\int_{\mathbb{R}^m}|y|^2\,\pi(\mathrm{d}y)<\infty, \tag{3.29}$$

then

$$\mathbb{E}\big\{\big|\langle\cdot,\varphi\rangle\big|^2\big\}=\int_{\mathbb{R}}\big|\sigma\varphi(t)\big|^2\mathrm{d}t+\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}\big|(\varphi(t),y)\big|^2\pi(\mathrm{d}y), \tag{3.30}$$

for any test function φ. □

Note that by replacing φ with $\lambda\varphi$, taking derivatives with respect to λ and setting $\lambda=0$ we deduce the isometry condition (3.30), 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 $\mathcal{S}(\mathbb{R};\mathbb{R}^d)$. Clearly, 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

$$\begin{aligned}\mathbb{E}\{\big\{h(\langle\cdot,\varphi_1\rangle,\dots,\langle\cdot,\varphi_d\rangle)\big\}&=\\&=(2\pi)^{-d/2}\int_{\mathbb{R}^d}\hat{h}(\xi)\Psi(\xi_1\varphi_1+\ldots+\xi_d\varphi_d)\mathrm{d}\xi,\end{aligned} \tag{3.31}$$

where Ψ is the characteristic function, i.e., the right-hand-side in (3.28).

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 (3.28), 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 (3.28) 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 $(\ell(t):t\ge0)$ is stationary with independent increments. The existence of a cad-lag version follows from the estimate

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

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

$$\int_{\mathbb{R}^d} \big(|y|^2 \wedge 1\big)\pi(\mathrm{d}y) < \infty, \qquad \pi(\{0\}) = 0. \tag{3.32}$$

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 (3.28). 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 (3.28) 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

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

replaces the second exponential in (3.28).

The theory of martingales (see Section 5.4 later on) 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, $\omega = (\omega_{n\text{↰}}, \omega_{\text{↱}m})$ where $\omega_{n\text{↰}}$ and $\omega_{\text{↱}m}$ 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 action-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 sure 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

$$X_t = x + \big(w_1(t), \ldots, w_n(t), q_1(t), \ldots, q_m(t)\big), \quad \forall\, t \geq 0,\ x \in \mathbb{R}^{n+m} \tag{3.33}$$

(also denoted by X^x_t) 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 (3.29), we may use the estimates

$$\begin{aligned}
&\mathbb{E}\big\{|w_i(t) - w_i(s)|^4\big\} = \mathbb{E}\big\{\big(w_i(t-s)\big)^4\big\} \leq C|t-s|^2,\\
&\mathbb{E}\big\{|q_j(s+r) - q_j(t)|^2 |q_j(t) - q_j(s)|^2\big\} =\\
&\qquad\qquad = \mathbb{E}\big\{\big(q_j(s+r-t)\big)^2\big\}\, \mathbb{E}\big\{\big(q_j(t-s)\big)^2\big\} \leq Cr^2,
\end{aligned}$$

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 (3.32) holds then we can obtain suitable estimates using the equality (3.31). We have then described a way of constructing these processes.

Actually, the only properties used in Lévy's Theorem 3.18 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 [30, 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 3.19 (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* $(S_\varepsilon h, h) \le 1$ *yields* $1 - \Re\{\Psi(h)\} \le \varepsilon$. □

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

$$\Psi(h_1, h_2) = \exp\Big(-\frac{1}{2} \int_{\mathbb{R}} |\sigma_1 h_1(t)|_0^2 \mathrm{d}t \Big) \times$$
$$\times \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_0} \big[\mathrm{e}^{\mathtt{i}(\sigma_2 h_2(t), \sigma_2 u)_0} - 1 - \mathtt{i}(\sigma_2 h_2(t), \sigma_2 u)_0 \big] \pi(\mathrm{d}u) \Big), \quad (3.34)$$

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

$$\int_{H_0} \big(|\sigma_2 u|_0^2 \wedge |\sigma_2 u|_0 \big) \pi(\mathrm{d}u) < \infty, \qquad \pi(\{0\}) = 0. \quad (3.35)$$

Under these assumptions the function Ψ is continuous on H, positive definite, $\Psi(0) = 1$ and the condition (d) of Theorem 3.19 is satisfied for a given $\varepsilon > 0$ with a trace class operator $S_\varepsilon : H \to H$ of the form

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

for any $k, \ell = 1, \dots,$ and for some $n = n(\varepsilon)$, where $\{e_j : j \ge 1\}$ is an orthonormal basis in Lebesgue space $L^2(\mathbb{R})$ and σ_i^* is the adjoint of σ_i, $i = 1, 2$, while $\{b_k : k \ge 1\}$ and $\{(b_k, b_\ell) : k, \ell \ge 1\}$ 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.,

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

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 3.20 (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

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

for every $i \ge 1$, then the sequence $\{\mu_i : i \ge 1\}$ regarded as Borel probability measure in $D([0,T],K_{-m})$ is tight, with $m \ge n$ 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

$$\mathbb{E}\big\{\exp\big(\mathtt{i}[(\varphi_1,\cdot)_1 + (\varphi_1,\cdot)_2]\big)\big\} = \exp\Big(-\frac{1}{2}\int_{\mathbb{R}^n} |\varphi_1(t)|_1^2 \mathrm{d}t\Big) \times$$
$$\times \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_2} \big[\mathrm{e}^{\mathtt{i}(\varphi_2(t),u)_2} - 1 - \mathtt{i}(\varphi_2(t),u)_2\big]\pi(\mathrm{d}u)\Big), \quad (3.36)$$

where π is a Radon measure on H_2 satisfying

$$\int_{H_2} \big(|u|_2^2 \wedge |u|_2\big)\pi(\mathrm{d}u) < \infty, \qquad \pi(\{0\}) = 0, \quad (3.37)$$

and $(\cdot,\cdot)_i$, $|\cdot|_i$ denote the inner product and the norm in H_i, $i = 1,2$. By comparison with (3.34) and (3.35) 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 [74, Chapters 1 and 2, pp, 1–83] for details on most of the preceding definitions.

If the probability to be constructed is not space-homogeneous (i.e., it is non-stationary) then the canonical process $(X_t^x : t \ge 0)$ does not define a Markov process under P. Thus, if for each x in $\mathbb{R}^d$ we have a $d\times d$ square matrix $\sigma(x)$ and a Radon measure $\pi(x,\mathrm{d}y)$ in $\mathbb{R}^d$ as before, then for every function ψ in $L^2(\mathbb{R},\mathbb{R}^d)$, we can construct (assuming some condition on the x-dependency of σ and π) a probability measure $Q(\psi,\cdot)$ on $\Omega = \mathcal{S}'(\mathbb{R};\mathbb{R}^d)$ such that its characteristic function satisfies

$$\int_\Omega \mathrm{e}^{\mathtt{i}\langle\omega,\varphi\rangle} Q(\psi,\mathrm{d}\omega) = \exp\Big(-\frac{1}{2}\int_{\mathbb{R}} |\sigma(\psi(t))\varphi(t)|^2 \mathrm{d}t +$$
$$+ \int_{\mathbb{R}} \mathrm{d}t \int_{\mathbb{R}^d} \big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)} - 1 - \mathtt{i}(\varphi(t),y)\big]\pi(\psi(t),\mathrm{d}y)\Big), \quad (3.38)$$

Next, the expected Markov process is the $\mathbb{R}^d$-valued canonical process

$$X^x(t,\omega) = (X_i^x(t,\omega) : i = 1,\dots,d), \quad X_i^x(t,\omega) := x_i + \langle\omega_i, \mathbb{1}_{(0,t]}\rangle$$

under the probability P^x, which is defined as the conditional probability

$$P^x\{\cdot \,|\, X^x\} = Q(X^x,\cdot).$$

Certainly, a drift and a killing terms can be added, and much more details are needed to complete this procedure. This involves the so called *pseudo-differential operators*, see the treatise by Jacob [67].

Related to the re-construction of probability is the following result (e.g., see Stroock and Varadhan [129, Theorem 1.3.5, pp. 34-36]). Let X be the canonical process in the canonical space either $C([0,\infty),E)$ or $D([0,\infty),E)$, where E is a Polish space. Assume $\{\tau_n : n \geq 0\}$ is a nondecreasing sequence of stopping times relative to the filtration $\{\mathcal{F}(t) : t \geq 0\}$, where $\mathcal{F}(t)$ and $\mathcal{F}(\infty)$ are the σ-algebras generated by $\{X(s) : 0 \leq s \leq t\}$ and $\{X(t) : t \geq 0\}$, respectively. Now, for each $n \geq 0$ let P_n be a probability measure defined on $\mathcal{F}(\tau_n)$. If $\lim_n P_n\{\tau_n \leq t\} = 0$ for every $t \geq 0$, and the probability P_{n+1} coincides with P_n on $\mathcal{F}(\tau_n)$ for any n, then there exists a probability measure P on $\mathcal{F}(\infty)$, which coincides with P_n on $\mathcal{F}(\tau_n)$ for every $n \geq 0$. Moreover, the same conclusion is true if $\mathcal{F}(t)$ is replaced by $\mathcal{F}(t+)$.

The reader interested in a guided tour to measure theoretic probability may take a look at the recent book by Pollard [106], and perhaps, later at Bichteler [14, Appendix A, pp. 363–469].

Chapter 4

Working on Probability Spaces

The purpose of this chapter is to supply the readers (who have taken a solid course in measure and integration theory but have little background in probability theory) with a *crash introduction* to stochastic processes. The focus is the *neighborhood* of Markov processes in continuous time. The first reading may be a little hard, since only references to proofs are given. The last section is only to complement the subject, some of the exercises are not so *simple exercises*, basically the reader should consult the references for the proofs if exercises are under consideration. Let us mention that a comprehensive introduction to probability, assuming *measure theory*, can be found in Stromberg [125] and in Stroock [126], among others. For instance, even without assuming measure theory, an introduction to probability can be found in Taylor [132], while an *analysis oriented* course on *diffusion processes* is given in Krylov [83]. Also Harlamov [58] and M. Iosifescu et al. [63] may be of a considerable help. An extensive classic study of the general theory of processes can be found in Dellacherie and Meyer [36], Gihman and Skorohod [52], Rao [109] and Sharpe [119]. For a complete discussion for foundation of probability, the reader may check the treatises De Finetti [35] and Loève[89], among many others. In a way, the material of this chapter could be regarded as the central point of a first course in stochastic processes.

4.1 Random Variables

Let $(\Omega, \mathcal{F})$ be a measurable space i.e., $\mathcal{F}$ is a σ-algebra of subsets in Ω. A *random variable* is a measurable mapping on $(\Omega, \mathcal{F})$, e.g. a real random variable x is a measurable function from $(\Omega, \mathcal{F})$ into $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B} = B(\mathbb{R})$ is the Borel σ-algebra of $\mathbb{R}$. Most of the information that we are interested in of a random variable x is contained in the σ-algebra generated by x i.e., $x^{-1}(\mathcal{B}) = \{F \in \mathcal{F} :$

$x(F) \in \mathcal{B}\}$. Thus if x is a characteristic (or indicator) function

$$x(\omega) = \begin{cases} 1 & \text{if } \omega \in F, \\ 0 & \text{if } \omega \in \Omega \setminus F, \end{cases}$$

for some F in $\mathcal{F}$, then $x^{-1}(\mathcal{B}) = \{\Omega, \emptyset, F, \Omega \setminus F\}$. If $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ are two measurable spaces, $\xi : \Omega \to \Omega'$ and $x : \Omega \to \mathbb{R}$ are two random variables, then x is $\sigma(\xi)$-measurable, i.e., $x^{-1}(\mathcal{B}) \subset \xi^{-1}(\mathcal{F}')$, if and only if there exists a measurable map $\eta : \Omega' \to \mathbb{R}$ such that $x(\omega) = \eta(\xi(\omega))$ for any ω in Ω. This is proved by means of a monotone class argument. Moreover, this remains true if $\mathbb{R}$ is replaced by a Polish space, i.e., a complete separable metric space.

A *stochastic process* is a collection of random variables indexed by some set e.g., a real valued stochastic process $X = \{X_t : t \in T\}$ is a family of measurable functions $X_t : \Omega \to \mathbb{R}$, with $t \in T$. Sometimes, the same process is denoted by $X = \{X(t) : t \in T\}$. Certainly, we can replace $\mathbb{R}$ with $\mathbb{R}^d$ in the previous discussion with almost not conceptual changes. Usually, when the random variables are indexed by a discrete set (countable set of isolated and totally ordered points) i.e. $\{\dots, -1, 0, 1, \dots\}$ or $\{1, 2, \dots\}$, we speak of a random sequence or a time series. In this context, we can view a time series as a random variable with values in $\mathbb{R}^\infty$, the set of real valued sequences $\{(x_1, x_2, \dots) : x_i \in \mathbb{R}, \ \forall i\}$. Here, we endowed $\mathbb{R}^\infty$ with the product topology and its associated Borel σ-algebra (e.g., Shiryayev [121]). A similar argument can be applied in general, but the discussion is more delicate. Thus, it is preferable to reserve the term *process* for uncountable index set T.

When the index set T is uncountable with a natural σ-algebra on it (for instance T is an interval), we restrict our attention to *measurable* stochastic process X i.e., we assume that the function $X : \Omega \times T \to \mathbb{R}$ is measurable. Moreover, if the index set T has a given topology and the stochastic process takes values in a topological space i.e., $\mathbb{R}^d$, then the following notions are necessary

Definition 4.1 (separable)**.** A d-dimensional *stochastic process* $\{X_t : t \in T\}$, $T \subset [0, +\infty)$ is *separable* if there exists a countable dense set of indexes $I \subset T$ (called separant) such that for any t in T and any ω in Ω 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 $X(t, \omega)$. □

For instance, the reader may want to take a look at the book by Meyer [99, Chapter IV] to realize the complexity of this notion of separability.

Unless otherwise stated, when referring to a stochastic process $\{X_t : t \in T\}$ in a measurable space $(\Omega, \mathcal{F})$, when T is a topological space, we mean a measurable and separable stochastic process, as understood from the context. Thus we denote by $\mathcal{L}^0(\Omega \times T, \mathbb{R}^d)$ the set of measurable stochastic processes with values in $\mathbb{R}^d$. Naturally, we can identify $X_t(\omega)$ with a measurable function in t, for each fixed ω, so that $\mathcal{L}^0(\Omega \times T, \mathbb{R}^d) = \mathcal{L}^0(\Omega, \mathcal{L}^0(T, \mathbb{R}^d))$ with the corresponding product σ-algebra. Thus we may look at a d-dimensional stochastic process as a random variable with values in $\mathcal{L}^0(T, \mathbb{R}^d)$. On the other hand, we may need to consider processes continuous in probability (see versions of processes) which are not expressible in terms of random variables.

Definition 4.2. A d-dimensional 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. □

Note that in the previous definition, the continuity is used as a global condition. Thus, if we denote by $\mathcal{C}^0(T, \mathbb{R}^d)$ the set of continuous functions, we may regard a d-dimensional stochastic process as a random variable with values in $\mathcal{C}^0(T, \mathbb{R}^d)$, provided a σ-algebra is defined on it. Similarly, we may define right (left) continuous and increasing (decreasing, locally bounded variation) processes.

When an order is given on the index set T, 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. Most of the time, the index set $T = [0, +\infty)$. In this case, for a given measurable and separable process $\{X_t : t \geq 0\}$ we associate a natural filtration $\{\mathcal{F}_t : t \geq 0\}$ as before. Certainly, X is *adapted* to the natural filtration i.e., the random variable X_t is $\mathcal{F}_t$-measurable for all $t \geq 0$. Also, X is *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, $T \subset \mathbb{R}^d$.

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.

This is essentially how far the analysis can go on measurable spaces. As soon as a probability measure space $(\Omega, \mathcal{F}, P)$ is given, any random variable is identified with its equivalence class. The same applies to processes when considered as random variables on function spaces, e.g., $\mathbb{R}^T$ or $C(T, \mathbb{R})$. In general, we may say that a measurable function from the sample space $(\Omega, \mathcal{F})$ into another measurable space $(E, \mathcal{E})$ is a random variable, and it is called a stochastic process if the value spaces has the form $(E^T, \mathcal{E}^T)$, for some set of indexes T (usually a subset of $\mathbb{R}$). Moreover, when a probability measure P is given on the measurable space $(\Omega, \mathcal{F})$ then random variables and stochastic processes are identified with their corresponding P-equivalence classes. For a given E-valued random variable x, the probability measure defined by $P_x(B) = P\{x^{-1}(B)\}$, B in $\mathcal{E}$, is called the distribution of x. However, for a given E-valued stochastic process X the family of probability on $\mathcal{E}^n$, $n \geq 1$, defined by $P_X(B_1 \times \cdots \times B_n) = P\{X(t_1, \omega) \in B_1, \ldots, X(t_n, \omega) \in B_n\}$, $B_1, \ldots, B_n$ in $\mathcal{E}$, $t_1, \ldots, t_n$ in T, is called the finite-dimensional distributions of X.

As long as the index set T is countable, no more detail is needed, however, for an uncountable index set T, e.g., $T = [0, \infty)$, we need to use the notion of *version* and *realization* of a stochastic process. Indeed, for a given (stochastic) process $\{X_t : t \in T\}$ on a probability space $(\Omega, \mathcal{F}, P)$ we say that $\{Y_t : t \in T\}$ is a *version* (or a modification) of the process $\{X_t : t \in T\}$ if $P(\{\omega : X_t(\omega) = Y_t(\omega)\}) = 1$, for any t in T. However, given a set of a priori properties that a process should satisfy (e.g., its finite-dimensional distributions, an assumption of continuity or measurability on its paths, or some other condition) then *realization* is a process, a probability space and any other items (such as a filtration) used to verify all required properties. Sometimes, we will refer to processes (not necessary defined on the same probability space) having the same finite-dimensional distribution or what is essentially the same (provided some regularity on the paths is assumed) having the same law in E^T or $C(T, E)$, as discussed later.

Only the case where the value set E is a complete separable metric space (Polish space), e.g., $E \subset \mathbb{R}^d$, endowed with the Borel σ-algebra $\mathcal{B}(E)$, and the set of index T is a totally ordered complete separable metric space, e.g., $T = [0, \infty)$, will be discussed herein. When the set of index T is uncountable, we impose some property (e.g., separability or continuity) on processes so that the value space E^T is replaced by better a space, e.g., E^I, I countable and dense in T, or $C(T, E)$ as discussed later.

Sometimes when dealing with extended real-valued random variables on a probability space $(\Omega, \mathcal{F}, P)$ we may need a definition of convergence in measure for random variables which may take values $\pm\infty$ with strictly positive probability. In this context we say that a sequence $\{x_n : n = 1, 2, \dots\}$ of random variables converges in measure to another random variable x if the sequence $\arctan(x_n)$ converges in measure to $\arctan(x)$ in the usual sense, equivalently, if

$$\lim_{n\to\infty} \mathbb{E}\{|\arctan(x_n) - \arctan(x)|\} = 0,$$

where $\mathbb{E}\{\cdot\}$ denote the *mathematical expectation*, i.e., the integral with respect to the probability measure P. The metric $d(x, y) = \mathbb{E}\{|\arctan(x) - \arctan(y)|\}$ on the space S of extended real-valued random variables make S a complete metric space, after the identification of two random variables whenever they are equal almost surely. Thus a measurable process $\{X_t : t \geq 0\}$ in the previous sense is (essentially) a Borel measurable mapping $t \mapsto X_t$ from $[0, +\infty)$ into S, we refer to Doob [39, pp. 407–410] for more details.

A typical generalization is to consider random variables with values in a *Polish space* (i.e, a complete and separable metric space), which is the analogous of stochastic processes if the Polish space is a function space. Stochastic processes are meant to model phenomenon which evolves in time in a random way. It is usually admitted that most often statistical experiments or physical considerations can only give information about the so-called finite-dimensional distributions of a process (note that two processes may have the same finite-dimensional distributions but having not the same probability space of reference). Therefore the choice of the Polish space becomes relevant for mathematical considerations.

For instance, consider the real-valued processes $X_t(\omega) = 1$ for every t in $[0,1]$, and $Y_t(\omega) = 1$ only when $t \neq \omega$ and $Y_t(\omega) = 0$ otherwise. It is clear that X is a continuous process while Y is (Borel) measurable but it is not separable. Moreover, if the probability measures (in which they are considered) have not atoms (each single event $\{\omega\}$) has zero probability) then these two processes have the same finite-dimensional distributions and from the phenomenological viewpoint they should be considered the same process. Mathematically, we prefers to take X.

Hence, in modeling a time-evolution random phenomenon, we are allowed to choose a *realization* of the process most suitable for our mathematical purpose. Questions like *is this process continuous?* really means *does there exist some process with the given finite-dimensional distributions whose paths are (almost sure) continuous?* or what is the same *is there a continuous realization of the process?*. This means that we can select the probability space $(\Omega, \mathcal{F}, P)$ and the map X among those satisfying the prescribed properties on the finite-dimensional distributions of the process. It will be clear by the end of this chapter, that there is a *canonical way* to performing this procedure of selecting a suitable realization such that the sample space Ω is a suitable Polish space and X is the *identity* as a random variable or the *coordinates mappings* if viewed as a stochastic process.

In what follows, we are going to denote *indistinctly* the notation $P(\{\cdot\})$, $P(\cdot)$ or $P\{\cdot\}$ for the probability measure, where the *dot* $\cdot$ represents a condition defining a set of events.

4.2 Typical Distributions

Let $(\Omega, \mathcal{F}, P)$ be a probability space i.e., P is a measure on $(\Omega, \mathcal{F})$ such that $P(\Omega) = 1$, called a probability measure. A measurable set (or a set in $\mathcal{F}$) is called an event. When a probability measure is involved, the previous concept of random variables becomes *equivalence classes* of random variables. For instance we may use the Lebesgue Banach spaces $L^p = L^p(\Omega, \mathcal{F}, P)$, for any $1 \leq p \leq \infty$. However, the study of stochastic processes is more delicate, since the family of random variable may not be countable.

As mentioned early, the *distribution* (or law) of a given random variable x is the probability measure P_x induced by x on $\mathcal{B}$ i.e., if x is a real random variable then its distribution is given by

$$P_x(B) = P(\{\omega : x(\omega) \in B\}), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Perhaps the three most important one-dimensional laws on $\mathbb{R}$ are the Gaussian (or normal) distribution, with parameters m and $r > 0$ $[N(m, r)]$, which has support on $\mathbb{R}$ and is given by

$$P_g(B) = \int_B (2\pi r^2)^{-1/2} \exp(-\tfrac{|x-m|^2}{2r^2}) \mathrm{d}x,$$

the Poisson distribution, with parameter $\lambda > 0$, which has support on $\mathbb{N}$ and is given by

$$P_p(B) = \exp(-\lambda) \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \mathbb{1}_{(n \in B)},$$

and the exponential distribution, with parameter $\alpha > 0$, which has support on $\mathbb{R}_0^+ = [0, +\infty)$ and is given by

$$P_e(B) = \int_{B \cap \mathbb{R}_0^+} \alpha \exp(-\alpha x) \mathrm{d}x.$$

Thus the mean and the variance are as follows

$$\int_{\mathbb{R}} x P_g(\mathrm{d}x) = m, \qquad \int_{\mathbb{R}} (x-m)^2 P_g(\mathrm{d}x) = r^2,$$
$$\int_{\mathbb{R}} x P_p(\mathrm{d}x) = \lambda, \qquad \int_{\mathbb{R}} (x-\lambda)^2 P_p(\mathrm{d}x) = \lambda,$$
$$\int_{\mathbb{R}} x P_e(\mathrm{d}x) = \alpha^{-1}, \qquad \int_{\mathbb{R}} [x-\alpha^{-1}]^2 P_e(\mathrm{d}x) = \alpha^{-2}.$$

The *characteristic* function (or Fourier transform) of a distribution (or probability law) P on $\mathbb{R}$ is the complex-value function

$$\widehat{P}(\xi) = \int_{\mathbb{R}} \mathrm{e}^{\mathtt{i}x\xi} P(\mathrm{d}x), \quad \forall \xi \in \mathbb{R},$$

with $\mathtt{i} = \sqrt{-1}$, and if the distribution P is on $\mathbb{R}_0^+$ then its *Laplace* transform is also defined

$$\widetilde{P}(\zeta) = \int_{\mathbb{R}_0^+} \mathrm{e}^{-x\zeta} P(\mathrm{d}x), \quad \forall \zeta \in \mathbb{R}_0^+.$$

For the previous distributions we have

$$\widehat{P_g}(\xi) = \exp\big(-\frac{1}{2} r\xi^2 + \mathtt{i}m\xi\big),$$
$$\widehat{P_p}(\xi) = \exp\big(\lambda(\mathrm{e}^{\mathtt{i}\xi} - 1)\big), \qquad \widetilde{P_p}(\zeta) = \exp\big(\lambda(\mathrm{e}^{-\zeta} - 1)\big),$$
$$\widehat{P_e}(\xi) = \frac{\lambda}{\lambda - \mathtt{i}\xi}, \qquad \widetilde{P_e}(\zeta) = \frac{\lambda}{\lambda + \zeta}.$$

There are others noted laws, such as the Cauchy distribution μ with parameters m and $c > 0$ and the Γ-distribution ν with parameters $c > 0$ and $\alpha > 0$ given by

$$\mu(B) = \pi^{-1} c \int_B \big[(x-m)^2 + c^2\big]^{-1} \mathrm{d}x, \quad \forall B \in \mathcal{B}(\mathbb{R}),$$
$$\nu(B) = \frac{\alpha^c}{\Gamma(c)} \int_{B \cap \mathbb{R}_0^+} x^{c-1} \mathrm{e}^{-\alpha x} \mathrm{d}x, \quad \forall B \in \mathcal{B}(\mathbb{R}_0^+),$$

with

$$\widehat{\mu}(\xi) = \exp(-c|\xi| + \mathtt{i}m\xi),$$
$$\widehat{\nu}(\xi) = (1 - \mathtt{i}\alpha^{-1}\xi)^{-c}, \qquad \widetilde{\nu}(\zeta) = (1 + \alpha^{-1}\zeta)^{-c}.$$

The Cauchy distribution μ does not have a mean value (i.e, μ does not integrate the function $|x|$) and the Γ-distribution has mean value equal to c/α. The exponential distribution is a particular case of the Γ-distribution, $c = 1$, and the Γ-distribution with $c = n/2$ and $\alpha = 1/2$ is referred to as the χ^2-distribution with n degrees of freedom. All these distributions are infinitely divisible, e.g., see Sato [116, Section 1.2, pp. 7-14].

Actually, for brevity we say *a family $\mathcal{A}$ of measurable sets is mutually independent relative to the probability P*, instead of saying a family $\mathcal{A}$ composed by measurable sets which are mutually independent relative to the probability P. However, in all what follows, we refer to *mutually independent* by saying only *independent*, i.e., we say a family of independent sets and a family of independent variables (or σ-algebras).

If $\mathcal{A}_i \subset \mathcal{F}$ is a family on a probability space $(\Omega, \mathcal{F}, P)$ indexed by $i \in I$, we define $\{\mathcal{A}_i : i \in I\}$ as independent if for any finite number of index $J \subset I$ and for any sets A_i in $\mathcal{A}_i$, $i \in J$, we have (2.2). It is clear that if $\mathcal{H}$ and $\mathcal{G}$ are two sub σ-algebras of $\mathcal{F}$, which are generated by the π-systems $\mathcal{H}_0$ and $\mathcal{G}_0$ (i.e., $\sigma(\mathcal{H}_0) = \mathcal{H}$ and $\sigma(\mathcal{G}_0) = \mathcal{G}$ (recall that a π-system means a collection of subsets closed or stable under finite intersections) then $\mathcal{H}$ and $\mathcal{G}$ are independent if and only if $\mathcal{H}_0$ and $\mathcal{G}_0$ are independent, i.e., if and only if $P(H \cap G) = P(H)P(G)$ for any H in $\mathcal{H}_0$ and G in $\mathcal{G}_0$, see Exercise 2.4.

Note that given a family $\mathcal{A}$ of three (or more) measurable sets, we may say that $\mathcal{A}$ is *pairwise* independent if any two subsets A_1 and A_2 of $\mathcal{A}$ are independent, i.e., $P(A_1 \cap A_2) = P(A_1)P(A_2)$. Clearly, this is distinct from the concept of *mutually* independent just defined. The same remark can be used for two or more families of either sub σ-algebras or random variables. On the other hand, two families $\mathcal{A}_1$ and $\mathcal{A}_2$ of measurable sets are (mutually or equivalently pairwise) independent $P(A_1 \cap A_2) = P(A_1)P(A_2)$ for any A_1 in $\mathcal{A}_1$ and A_2 in $\mathcal{A}_2$. Similarly, this definition can be extended to three or more families of measurable sets, where we need to distinguish between mutually and pairwise independent.

Note that if A and B are independent, i.e., $P(A \cap B) = P(A)P(B)$, then a simple calculation shows that $A' = \Omega \setminus A$ and B are also independent. As a consequence, if $\mathcal{F}_i$ denotes the σ-algebra generated by F_i, i.e., $\mathcal{F}_i = \{A_i, \Omega \setminus A_i, \emptyset, \Omega\}$, then a family of measurable sets (events) $\{A_i : i \in I\}$ is independent if and only if the family of σ-algebras $\{\mathcal{F}_i : i \in I\}$ is independent.

Thus, a sequence of independent random variables $\{x_i : i \in I\}$ is independent if and only if

$$P\Big(\bigcap_{j\in J}\{\omega : x_j(\omega) \in A_j\}\Big) = \prod_{j\in J} P\big(\{\omega : x_j(\omega) \in A_j\}\big)$$

for any measurable sets A_j and any finite subset J of I. In term of the characteristic functions, this is equivalent to

$$\mathbb{E}\{\exp[\mathtt{i}\sum_{j\in J}\alpha_j x_j]\} = \prod_{j\in J}\mathbb{E}\{\exp[\mathtt{i}\alpha_j x_j]\},$$

for any constants α_j and any finite subset J of I, where $\mathtt{i} = \sqrt{-1}$. There is a very close connection between the concepts of *independence* and *Cartesian product*. If x and y are two real valued random variables, we may look at (x, y) as a two-dimensional real valued random variable, then a direct comparison with the definition of independence shows that the fact that x and y are independent may be very simply expressed by the equation

$$P_{(x,y)} = P_x \times P_y,$$

i.e., the joint distribution of x, y is equal to the Cartesian product of the single distributions of x and y.

• *Remark* 4.3. If x is a Normal distributed random variable with parameters m and $r > 0$ then it characteristic function is given by

$$P\{\exp(\mathtt{i}\xi x)\} = \exp(-\frac{1}{2}r^2\xi^2 + \mathtt{i}m\xi).$$

Hence, if x_i, $i = 1, \dots, k$ are independent Normal distributed random variables with parameters m_i and $r_i > 0$ then any linear combination $x = c_1x_1 + \cdots + c_kx_k$, with c_i real numbers, is indeed a Normal distributed random variables with parameters $m = m_1 + \cdots + m_k$ and $r = \sqrt{r_1^2 + \cdots + r_k^2}$. Similarly, if x is a Poisson distributed random variable with parameter $\lambda > 0$ then it characteristic function is given by

$$P\{\exp(\mathtt{i}\xi x)\} = \exp\big(\lambda(\mathrm{e}^{\mathtt{i}\xi} - 1)\big).$$

Thus, if x_i, $i = 1, \dots, k$ are independent Poisson distributed random variables with parameters λ_i then the sum $x = x_1 + \cdots + x_k$ is indeed a Poisson distributed random variables with parameter $\lambda = \lambda_1 + \cdots + \lambda_k$. However, if x_i, $i = 1, \dots, k$ are independent exponentially distributed random variables with the same parameter λ, i.e., with characteristic function

$$\mathbb{E}\{\exp(\mathtt{i}\xi x_1)\} = \frac{\lambda}{\lambda - \mathtt{i}\xi} = \big(1 - \mathtt{i}\lambda^{-1}\xi\big)^{-1},$$

then the sum $x = x_1 + \cdots + x_k$ has a Gamma distribution with parameters λ and k, i.e.,

$$\mathbb{E}\{\exp(\mathtt{i}\xi x)\} = \big(1 - \mathtt{i}\lambda^{-1}\xi\big)^{-k} \quad \text{or} \quad P\{x \in \mathrm{d}t\} = \frac{\lambda^k t^{k-1}\mathrm{e}^{-\lambda t}}{(k-1)!}\,\mathrm{d}t.$$

On the other hand, the process of counting an independent identically exponentially distributed sequence $\{x_i\}$ of random variables with parameter λ, i.e., $n(t) = \sum_i \mathbb{1}_{x_i \le t}$, produces a family of random variables, indexed by $t \ge 0$, identically Poisson distributed with parameter $t\lambda$. □

• *Remark* 4.4. Certainly, there many other useful distributions, e.g., (1) the deterministic or delta or Dirac distribution, which is concentrated at one point say $x = x_0$, i.e., $P\{x = x_0\} = 1$ and $P\{\exp(\mathtt{i}\xi\cdot x)\} = \exp(\mathtt{i}\xi\cdot x_0)$, (2) the uniform or Lebesgue distribution, which is uniformly distributed on a region with finite volume, e.g., a random variable uniformly distributed over an interval $[a, b]$ has distribution $P\{\alpha < x \le \beta\} = (\min\{b, \beta\} - \max\{a, \alpha\})/(b - a)$, for every real numbers $\alpha \le \beta$, and (3) the compound Poisson distribution, which is frequently used and can be described as follows: if n is a Poisson distributed random variable with parameter $\lambda > 0$ which is independent of a sequence of independent identically distributed random variables $\{x_i\}$ with F as its common distribution satisfying $F(0) = 0$, then the random sum $x(\omega) = x_1(\omega) + \cdots + x_{n(\omega)}(\omega)$, complemented with the condition $x(\omega) = 0$ if $n(\omega) = 0$, has a compound Poisson distribution with parameters λ and F. Note that the condition that $F(0) = 0$ ensures that $x = 0$ only when $n = 0$. If F is a distribution in $\mathbb{R}^m_* = \mathbb{R}^m \setminus \{0\}$ then the k-convolution F^{*k} is the distribution of the independent sum $x_1 + \cdots + x_k$, and therefore, the *compound Poisson distribution* of a random variable x in $\mathbb{R}^m_*$ is given by

$$P\{x \in B\} = P_{cp}(B) = \exp(-\lambda) \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} F^{*k}(B),$$

with its characteristic function

$$P\{\exp(\mathtt{i}\xi \cdot x)\} = \widehat{P_{cp}}(\xi) = \exp\big(\lambda(\mathrm{e}^{\widehat{F}(\xi)} - 1)\big),$$

where $\widehat{F}$ is the characteristic function of the distribution F. It is interesting to remark that the two parameters λ and F can be combined to produce finite measure π, with $\lambda = \pi(\mathbb{R}^m_*)$ and $F = \pi/\lambda$, i.e., a compound Poisson random variable x with parameter π has characteristic function

$$P\{\exp(\mathtt{i}\xi \cdot x)\} = \exp\Big(\int_{\mathbb{R}^m_*} (\mathrm{e}^{\mathtt{i}x\cdot\xi} - 1)\pi(\mathrm{d}x)\Big),$$

which reduces to the (simple) Poisson distribution if F is a deterministic, e.g., if the finite-measure π is given by $\pi(B) = \lambda\delta_1(B) = \lambda\mathbb{1}_{1\in B}$, with B a Borel subset of $[0, \infty)$. It is also clear that a compound Poisson random variable x may or may not have first moment (or mean), but if it does, the expression $y = x - \mathbb{E}\{x\}$ produces what is called a *centered* or *compensated* Poisson distribution with parameter π and characteristic function

$$P\{\exp(\mathtt{i}\xi \cdot y)\} = \exp\Big(\int_{\mathbb{R}^m_*} \big(\mathrm{e}^{\mathtt{i}x\cdot\xi} - 1 - \mathtt{i}x \cdot \xi\big)\pi(\mathrm{d}x)\Big),$$

where the finite measure π must integrate the function $x \mapsto |x|$ and the random variable y has zero mean. At this point, let us mention that a distribution is infinitely divisible if and only if it is a limit of a sequence of compound Poisson distributions, e.g., see Prabhu [107, Chapter 4, pp.43–68]. □

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 [38], Feller [48] and Gnedenko [54], while an analytic view can be found in Dudley [41], Folland [49, Chapter 10], Halmos [57]), Stromberg [125] and Stroock [126].

• *Remark* 4.5 (conditional independence). In elementary probability theory we define the conditional probability of a event A (i.e., a measurable set) given another event C with $P(C) > 0$ as $P(A \,|\, C) = P(A \cap C)/P(C)$. A more sophisticate concept is the following: Given a measurable set C with $0 < P(C) < 1$, a family $\mathcal{A}$ of measurable sets is (mutually) conditional independent given C (relative to the probability P) if their elements are mutually conditional independent given C, i.e., if for any finite number of sets $A_1, \ldots, A_n$ in $\mathcal{A}$ we have

$$P\big(\bigcap_{i=1}^{n} A_i \,|\, C\big) = \prod_{i=1}^{n} P\big(A_i \,|\, C\big). \tag{4.1}$$

Now, a family of σ-algebras is (mutually) conditional independent given C, if for any finite number of σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$ in the family and any sets A_i in $\mathcal{F}_i$ we have (4.1). Similarly, a family of random variables is (mutually) conditional independent given C, if the family of their generated σ-algebras is (mutually) conditional independent given C. Moreover, if (4.1) holds for every C with $0 < P(C) < 1$ in a σ-algebra $\mathcal{C}$ then the sets $A_1, \ldots, A_n$ are mutually conditional independent given $\mathcal{C}$, and similarly, if this holds for every set A_i in $\mathcal{F}_i$ then the family of σ-algebras is (mutually) conditional independent given (the σ-algebra) $\mathcal{C}$. □

Note that if we allow $C = \Omega$ then we foldback to the case of complete independence Definition 2.3. It is also clear that condition (4.1) can be rewritten as

$$P\big(\bigcap_{j=1}^{n-1} A_{i_j} \,|\, A_{i_n},\, C\big) = \prod_{j=1}^{n-1} P\big(A_{i_n} \,|\, C\big),$$

for any permutation $\{i_1, \ldots, i_n\}$ of $\{1, \ldots, n\}$, provided $P(A_{i_n} \cap C) > 0$.

As in the case of complete independence, if $\mathcal{F}_i$ denotes the σ-algebra generated by F_i, i.e., $\mathcal{F}_i = \{A_i, \Omega \setminus A_i, \emptyset, \Omega\}$, then a family of events $\{A_i : i \in I\}$ is conditional independent given C if and only if the family of σ-algebras $\{\mathcal{F}_i : i \in I\}$ is conditional independent given C. However, if (4.1) condition holds for C then it does not necessarily hold for $\Omega \setminus C$, i.e., conditional independence with respect to an event C is not necessarily the same as conditional independence with respect to the σ-algebra generated by C, namely, $\mathcal{C} = \{\emptyset, \Omega, C, \Omega \setminus C\}$, see later Definition 2.8 on conditional independence.

Clearly, the conditional probability $P(\cdot\,|\,C)$ is itself a probability and thus, conditional independence given a set C is just independence under the conditional probability. Moreover, condition (4.1) can be rewritten in a less intuitive way as

$$P\big(\bigcap_{i=1}^{n} A_i \cap C\big)\,[P(C)]^{n-1} = \prod_{i=1}^{n} P\big(A_i \cap C\big),$$

which becomes $P(A_1 \cap A_2 \cap C)P(C) = P(A_1 \cap C)P(A_2 \cap C)$ when $n = 2$. In particular, the space $\Omega = A_1$ is conditional independent of any event $A = A_2$ given any event C with $P(C) > 0$. Also, if three events A, B and C are pairwise independent with $P(C) > 0$ then A and B are conditional independent given C, if and only if they are mutually independent according to Definition 2.3. Certainly, if $P(C) = 0$ the above equality could be used instead of condition (4.1), but it is trivially satisfied and the definition is meaningless. Some comments on this concept are given later, with the use of conditional expectation.

Going back to the three examples of distributions in $\mathbb{R}$, we can extend them to $\mathbb{R}^d$ as follows. Consider n independent identically distributed random variables $(\xi_1, \ldots, \xi_n)$, a linear transformation q from $\mathbb{R}^n$ into $\mathbb{R}^d$, and then for a given $m \in \mathbb{R}^d$, we look at the distribution of the random variable $y = m + Q(\xi_1, \ldots, \xi_n)$. Identifying the linear transformation Q with a canonical matrix, still denoted by Q, we deduce that if the common distribution of $(x_1, \ldots, x_n)$ is Gaussian, then

$$P_y(B) = \int_B p_n(x)\mathrm{d}x, \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

where

$$p_n(x) = [2\pi \det(QQ^*)]^{-d/2} \exp\Big(-\frac{[(x-m)^*(QQ^*)^{-1}(x-m)]^2}{2}\Big), \qquad (4.2)$$

and the * means the transpose (of a matrix), $\det(\cdot)$ is the determinant, and we have assumed that QQ^* is invertible. This is a Gaussian d-dimensional distribution. Similarly, d-dimensional Poisson (or exponential) distribution can be described.

Sums of independent random variables are studied with the purpose of generalizing and elaborating the *law of the large numbers*. Let $\{x_i : i \geq 1\}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, P)$, and let $\{s_n : n \geq 1\}$ be the sequence of partial sum, $s_n = x_1 + \cdots + x_n$. The first point is the Kolmogorov's *zero-one law*, namely, the series s_n converges almost surely or diverges almost surely (i.e., cannot converges for some ω and diverges for others ω). Next, if the two series of real numbers $\sum_i \mathbb{E}\{x_i\}$ (mean) and $\sum_i \mathbb{V}\{x_i\}$ (variance, $\mathbb{V}\{x\} = \mathbb{E}\big\{\big[x - \mathbb{E}\{x\}\big]^2\big\}$) converge then the series $\sum_i x_i$ converges almost surely. Another result, known as the *three series theorem*, affirms that $\sum_i x_i$ converges almost surely if and only if the following three series of real numbers $\sum_i \mathbb{E}\{x_i'\}$, $\sum_i \mathbb{V}\{x_i'\}$ and $\sum_i P\{x_i \neq x_i'\}$ are convergent, where

$x'_i = x_i$ if $-1 \le x_i \le 1$ and $x'_i = 0$ otherwise. There are several variants of these theorems, e.g., the strong law of the large number, namely, if $\mathbb{V}\{x_i\}$ is bounded then $\big[s_n - \mathbb{E}\{s_n\}\big]/n$ converges to 0 almost surely, or if x_i are integrable identically distributed then s_n/n converges to $\mathbb{E}\{x_1\}$ almost surely. Further in this analysis is the *central limit theorems* and law of the iterated logarithm, where we define the sequence of random variables $t_n = \big[s_n - \mathbb{E}\{s_n\}\big]/\sqrt{\mathbb{V}\{s_n\}}$ and give conditions under which the probability distributions of t_n converges to the Gauss or normal distribution $N(0,1)$. For instance, if the limit

$$\frac{1}{\mathbb{V}\{s_n\}} \sum_{i=1}^{n} \mathbb{E}\big\{|x_i - \mathbb{E}\{x_i\}|^2,\ |x_i - \mathbb{E}\{x_i\}| \ge \varepsilon\sqrt{\mathbb{V}\{s_n\}}\big\} \to 0$$

holds true for every $\varepsilon > 0$, then the probability distributions of t_n converges to the Gauss or normal $N(0,1)$, i.e.,

$$\lim_n P(a < t_n < b) = (2\pi)^{-1/2} \int_a^b e^{-x^2/2} \mathrm{d}x, \quad \forall b > a,$$

however we have

$$\limsup_n t_n = +\infty \quad \text{and} \quad \liminf_n t_n = -\infty$$

almost surely. This is used in the Gauss' theory of errors, namely, for every $n \ge 1$ let $\xi_{1,n}, \dots, \xi_{n,k(n)}$ be independent random variables and define $\sigma_n = \sum_{i=1}^{k(n)} \xi_{i,k(n)}$. If $\varepsilon_n := \sup_{i,\omega} |\xi_{i,k}(\omega)| \to 0$, $\mathbb{E}\{\sigma_n\} \to m$ and $\mathbb{V}\{s_n\} \to v$ then the probability distribution of σ_n converges to the Gauss or normal distribution $N(m,v)$. On the other hand, if the variables $x_{i,k}$ take only two values, i.e., assuming $P\{\xi_{i,k} = 1\} = p_{i,k}$ and $P\{\xi_{i,k} = 0\} = 1 - p_{i,k}$, and if $\bar{p}_n := \max_k p_{i,k}$ and $\sum_{i=1}^{k(n)} p_{i,k(n)} \to \lambda$ then the probability distribution of σ_n converges to the Poisson distribution with parameter λ, this last result is know as *Poisson's law of rare events*. Proofs of the above theorems can be found in several text books in probability, e.g, Breiman [20, Chapter 9, pp. 185–190] or Itô [64, Chapter 4, pp. 165–211].

It should be clear that given a probability space $(\Omega, \mathcal{F}, P)$, it is not possible to ensure the existence of (independent) random variables (or stochastic processes) 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 $n = 1, 2, \dots$ 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 $[0,1]$. 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 $i = 1, 2, \ldots$ then there exists a sequence $\{\xi_1, \xi_2, \ldots\}$ 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 $\mathcal{B}_i(S_i)$, $i = 1, 2, \ldots$, i.e., the distribution of ξ_i is exactly P_i, e.g., see Kallenberg [71, Theorem 3.19, pp. 55–57].

Let ξ be a random (vector) variable having a given (joint) density distribution p_ξ. Sometimes we are interested in computing

$$\mathbb{E}\{g(\xi)\} = \int g(x) p_\xi(x) \mathrm{d}x,$$

for some real-valued function g. In many situation, it is not analytically possible either to compute the above (multiple) integral exactly or even to numerically approximate it within a given accuracy. Another way to approximating $\mathbb{E}\{g(\xi)\}$ is by means of the co-called *Monte Carlo simulation* method. This goes as follows: start by generating a random (vector) variable ξ^1 having the (joint) density g, and then compute $\eta^1 := g(\xi^1)$. Now generate a second random (vector) variable ξ^2, independent of the first, and compute $\eta^2 := g(\xi^2)$. Keep on doing this, for a fix number of times n, to generate the independent and identically distributed random (vector) variables $\eta^i := g(\xi^i)$, $i = 1, \ldots, n$. As mentioned later, the strong law of large number applies and we find

$$\lim_n \frac{\eta^1 + \cdots + \eta^n}{n} = \mathbb{E}\{\eta^i\} = \mathbb{E}\{g(\xi)\}.$$

Clearly, the remaining problem is how to generate, or *simulate* random (vector) variables having a specific (joint) distribution. The first step in doing this is to be able to generate random variables from a uniform distribution on $(0, 1)$, i.e., a random variable u with valued in the interval $(0, 1)$ such that $P\{u \leq \lambda\} = \lambda$ for every λ in $[0, 1]$. One way of doing this would be to take 10 identical slips of paper, numbered $0, 1, \ldots, 9$, place them in a hat and then successively select n slips, with replacement, from the hat. The sequence of digits obtained (with a decimal point in from) can be regarded as the value of a uniform $(0, 1)$ random variable rounded off to the nearest 10^{-n}. This constitutes the so-called *random number tables*. Nowadays, digital computers simulate *pseudo random numbers* instead of the truly random numbers. Most of the random number generators start with an initial value ξ_0, called the seed, and then recursively compute values by specifying positive integers a and b and m, and then letting x_{n+1} be the remainder of $ax_n + b$ divided by m, i.e., $x_{n+1} := (ax_n + b) \mod (m)$. The quantity $u_n := x_n/m$ is taken as an approximation to a uniform $(0, 1)$ random variable. All other distributions are generated from uniform $(0, 1)$ random variables u. Indeed, the inverse transformation method is based on the fact that for any right-continuous distribution F, the random variable $\xi := F^{-1}(u)$ has distribution F. Note the definition of $F^{-1}(\lambda) := \inf\{s \in (0, 1) : F(s) = \lambda\}$, so that $t < F(\lambda)$ if and only if $F^{-1}(\lambda) < F(t)$.

The rejection method simulate a random variable η having density f on the basis of a random variable ξ having density g, it uses a two-step iteration as

follows: for two distributions f and g satisfying $f(x) \le cg(x)$ for every x and some constant c :

Step 1: simulate ξ having density g and simulate a random number u.

Step 2: if $u \le f(\xi)/[cg(\xi)]$ set $\eta = \xi$, otherwise return to Step 1.

This generates a random variable ξ having density f.

For a continuously differentiable distribution F, the *hazard rate function* of F is defined by

$$\lambda(t) := \frac{F'(t)}{F(t)}, \quad \forall t \ge 0.$$

The hazard rate method simulates a random variable ξ having $\lambda(t)$ as its hazard rate function, provided λ is a given nonnegative bounded function satisfying

$$\int_0^\infty \lambda(t)\mathrm{d}t = \infty.$$

This is defined as follows:

Step 1: select $r \ge \lambda(t)$ for every $t \ge 0$ and simulate independent random variables u^i, η^i, $i = 1, \dots,$ where u^i are $(0,1)$ uniform and η^i are exponential with rate r.

Step 2: stopping at $\tau := \min\left\{n : u_n r \le \lambda(\sum_{i\le n} \eta_i)\right\}$ define $\xi := \sum_{i\le\tau} \eta_i$.

It is proven that ξ has hazard rate function $\lambda(t)$. For instance, we refer to Ross [114], among others.

4.3 Filtrations and Optional Times

Let $(\Omega, \mathcal{F})$ be a measurable space and T be an partially order index set, with a first element called 0. Generally, the index set is the positive integers or a real interval, i.e. $[0,T]$, $0 < T \le +\infty$. Suppose we are given an increasing family of sub σ-algebras $\{\mathcal{F}(t) : t \in T\}$, i.e. $s \le t$ implies $\mathcal{F}(s) \subset \mathcal{F}(t)$. Sometimes the notation $\mathcal{F}_t = \mathcal{F}(t)$ is used. Define $\mathcal{F}^+(t) = \cap_{s>t}\mathcal{F}(s)$ for t in T to get another filtration with $\mathcal{F}(t) \subset \mathcal{F}^+(t)$ for any t in T. The filtration is said to be *right continuous* if $\mathcal{F}(\cdot) = \mathcal{F}^+(\cdot)$ and it is also denoted by $\mathcal{F}(t+)$. In particular, $\{\mathcal{F}^+(t) : t \in T\}$ is right continuous.

When a probability measure P is given on $(\Omega, \mathcal{F})$, the hypothesis that $\mathcal{F}(0)$ contains the null sets implies that the restriction of a given measure on each σ-algebra $\mathcal{F}(t)$ is *complete*, but the converse may be false. In most of the cases, we may assume that $\mathcal{F}(0)$ contains the null sets, at a cost of enlarging each σ-algebra to the σ-algebra generated by $\mathcal{F}(t)$ and the null sets. If the index set is a real interval $[0,T]$, or $[0,\infty)$, then it is possible without loss of generality to replace $\mathcal{F}(\cdot)$ by the right continuous filtration $\mathcal{F}^+(\cdot)$. If the index set does not have a last element, i.e. $[0,+\infty)$, then we add a last element denoted by $+\infty$ (or ∞) with $\mathcal{F}(+\infty)$ equal to the sub σ-algebra generated by all the $\mathcal{F}(t)$, $t \ge 0$. Thus we will refer to a filtration satisfying the *usual conditions* or a standard filtration when the filtration is completed and right continuous.

Definition 4.6 (filtration)**.** In a (complete) probability space $(\Omega, \mathcal{F}, P)$, a family of sub σ-algebras $\{\mathcal{F}(t) : t \in T\}$ is called a *filtration* if $s \leq t$ implies $\mathcal{F}(s) \subset \mathcal{F}(t)$, $\mathcal{F}(0)$ contains all sets of probability zero, and $\mathcal{F}(t) = \cap_{s>t}\mathcal{F}(s)$ i.e., unless explicitly stated we assume the usual conditions are satisfied. A family of random variables $\{X(t) : t \in T\}$ is said *adapted* to the filtration if $X(t)$ is $\mathcal{F}(t)$-measurable for any t in T. □

Given a stochastic process and a filtration we can talk about a stochastic process being adapted to a filtration, being progressively measurable, and so on. Several operations can be performed with processes and filtrations. For a family $\{X_\gamma(\cdot) : \gamma \in \Gamma\}$ of processes adapted to a common filtration $\mathcal{F}(\cdot)$ we may define the process *essential infimum* and *essential supremum*. For instance

$$X(t) = \operatorname*{ess\,sup}_{\gamma\in\Gamma} X_\gamma(t),$$

which can be taken adapted to the same common filtration $\mathcal{F}(\cdot)$. Similarly, the sample integral can be defined for a progressively measurable (see definition later on) integrable process $\{X(t), \mathcal{F}(t) : t \geq 0\}$. The resulting process

$$Y(t) = \int_0^t X(s,\omega)\mathrm{d}s$$

can be taken progressively measurable with respect to the same filtration $\mathcal{F}(\cdot)$.

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 4.7. An *optional time* (*stopping* or *Markov time*) τ with respect to a filtration $\mathcal{F}(\cdot)$ 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 [75, Section 1.2, pp. 6-11]. Since $\{\tau \leq t\} = \cup_{n\geq 1}\{\tau < t - 1/n\}$ 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 $\tau_1 \le \tau_2$, the *stochastic interval* $[\![\tau_1, \tau_2]\!]$, is defined by

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

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 (see Exercise 4.7).

(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 $\tau_1 \ge \tau$ 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 $\tau_1 \le \tau_2$, 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, $\{\tau_1 \le t\} \in \mathcal{F}(\tau_1 \wedge t)$.

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

(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 $n \ge n(\omega)$, 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 [39, pp. 419–423])

Theorem 4.8 (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 $\mathcal{F}(\cdot)$ satisfying the usual conditions on a probability space $(\Omega, \mathcal{F})$, Then the hitting, entry and exit times are optional times with respect to $\mathcal{F}(\cdot)$, 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 $t \ge 0$ 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. For instance, the reader may begin by taking a look at Revuz and Yor [111, Section I.4, pp. 41–48].

4.4 Versions of Processes

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 [68] or Williams [133], 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

$$P(\{\omega : X_t(\omega) = Y_t(\omega)\}) = 1, \quad \forall t \in T, \tag{4.3}$$

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. Moreover, a measure μ in the product space $(0,\infty)\times\Omega$ vanishes on every evanescent set if and only if it has the product form $\mu = \mathrm{d}\alpha(t,\omega)P(\mathrm{d}\omega)$ for some integrable nondecreasing process α. This is discussed in some detail later, in Chapter 4.

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. The following result (see Doob [38, Theorem 2.4, pp. 60], Billingsley [16, Section 7.38, pp. 551-563] or Neveu [102, 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 4.9 (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 (4.3) and the condition of separability in Definition 4.1, 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 $\mathbb{R}^d = [-\infty,+\infty]^d$. 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 [53, 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 [102, Section III.4, pp. 81–88].

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. Note that the concept of stochastic continuity (or continuity in probability) is not a sample path 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 [30, p. 72–75], Gikhman and Skorokhod [53, Section IV.3]) that

Theorem 4.10 (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 4.11 (continuity)**.** *Let $\{X_t : t \in [0,T]\}$ be a d-dimensional stochastic process in a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$E|X_t - X_s|^\alpha \leq C|t-s|^{1+\beta}, \quad \forall s,t \in [0,T], \tag{4.4}$$

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

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

The previous result is essentially based on the following arguments, e.g., Karatzas and Shreve [75, pp. 53–55]). Estimate (4.4) and the dyadic construction $\{X(k2^{-n}) : k = 0, 1, \ldots, 2^n,\ n = 1, 2, \ldots\}$ yields

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

$$\leq \sum_{k=1}^{2^n} P\{|X(k2^{-n}) - X((k-1)2^{-n})| \geq 2^{-\gamma}\} \leq C 2^{-n(\beta - \alpha\gamma)},$$

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 (4.4) 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., (4.4) 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 [53, Section IV.4], Wong [134, Proposition 4.3, p. 59] and its references)

Theorem 4.12 (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*

$$\mathbb{E}\{|X_{t+h} - X_s|^\alpha |X_s - X_t|^\alpha\} \le Ch^{1+\beta}, \quad \forall\, 0 \le t \le s \le t+h \le T, \tag{4.5}$$

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

Similarly, for processes of locally bounded variation we may replace the expression $|\cdot|$ in (4.4) 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 (4.4) holds for the metric $\mathrm{d}(X_t, X_s)$ instead of the Euclidean distance $|X_t - X_s|$, then the conclusions of Theorem 4.11 are valid with $\mathrm{d}(Y_t, Y_s)$ in lieu of $|Y_t - Y_s|$, e.g., see Durrett [43, 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.

If we can find 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 $\mathcal{B} = \Omega \bigcap \mathcal{B}^T(\mathbb{R})$ 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

$\{\omega \in \Omega : (\omega(s_1), \ldots, \omega(s_n)) \in B\}$ for any $B \in \mathcal{B}(\mathbb{R}^n)$, with $s = (s_1, \ldots, s_n)$, $n = 1, 2, \ldots$.

Note that the arguments in Theorems 4.10, 4.11 or 4.12 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., $P(\{\omega : X_t(\omega) = Y_t(\omega),\ \forall t \in T\}) = 1$.

Definition 4.13 (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. □

Clearly, the passage from general stochastic processes (i.e., a family of random variables X) to stochastic processes is very subtle (however very significant, since X becomes a random variable in some topological space). Technically, if we take the trivial choice $B = E^T$ then to label X a *stochastic process*, we need to know that X is also measurable for Borel σ-algebra $\mathcal{B}(E^T)$, which may be larger than $\mathcal{B}^T(E)$, and so not a priori satisfied. Note that we abandon the trivial choice $B = E^T$ because in several arguments, we need that the Borel σ-algebra $\mathcal{B}$ of the topological space B (the sample space, where all paths are) coincides with the cylindrical Borel σ-algebra in B, i.e., a B-valued function Y is $\mathcal{B}(B)$-measurable if and only if for each t, the E-valued function $Y(t)$ is $\mathcal{B}(E)$-measurable. Usually, the definition of the (topological) sample space B involves some topology on the index set T. Also, if the index set $T \subset \mathbb{R}^n$, $n \geq 2$ then the name *random field* is preferable. The notion of general stochastic processes is as general as possible, however the concept of stochastic processes imposes some *path regularity*. Depending on the interest and the objective of the study undertaken, there are other possible approaches, for instance, when $E = \mathbb{R}^d$ we may

consider (right- or left-)continuous processes, in probability or in mean-square, where processes are treated as continuous functions from T into $L^0(\Omega,\mathcal{F},P)$, the complete metric space of $\mathbb{R}^d$-valued random variables endowed with the convergence in probability (called stochastically continuous processes), or into the Hilbert space $L^2(\Omega,\mathcal{F},P)$ of square-integrable random variables (usually called second-order processes). This type of analysis is rarely used here, the interested reader may check the book Gikhman and Skorokhod [53], among others.

Starting from a general stochastic process X, usually, a two-step procedure is applied: first a *good* version $\bar{X}$ is chosen and then a *good* sample space B is found. For instance, given a $\mathbb{R}$-valued general stochastic process X with index set $T=[0,\infty)$, we can apply Theorem 4.9 to get a separable version of $\bar{X}$ (with extended values, i.e., in $[-\infty,+\infty]^T$). Thus if I is a denumerable separant subset of indexes then we may consider $\bar{X}$ as taking valued in $[-\infty,+\infty]^I$ (or the one-point compactification $\mathbb{R}^I\cup\{\infty\}$), which is not exactly a subset of $\mathbb{R}^T$, but it is essentially the space of all (extended) real-valued sequences. This point of view is not considered in this book. Alternatively, if we know that the initial general stochastic process X is right-(or left-)continuous in probability (so when X is also separable any dense set in T is a separant set) then we may apply Theorem 4.10 to get a measurable version. Still, this point of view is rarely used in this book. Finally, the two-step procedure is as follows: first, we assume that Theorems 4.11 or 4.12 is applicable to X, so that a continuous or at least cad-lag version $\bar{X}$ can be found (i.e., first $\bar{X}(t,\omega)=X(t,\omega)$ is defined for every ω but only for t in a countable dense subset of $[0,\infty)$ and then $\bar{X}(t,\omega)$ is extended for every t in $[0,\infty)$ but only for ω outside of a negligible set). Second, we take as B the space of continuous or cad-lag functions, with a suitable topology and we forget about the initial general stochastic process X, because X (technically its version $\bar{X}$) is considered now as a random variable with values in a complete separable metric space B endowed with its Borel σ-algebra $\mathcal{B}$. So, essentially, a version (or modification) of a process is allowed only once, and then all stochastic processes (with the same path regularity) indistinguishable of this good version are considered equals. In general we use the following result, e.g., Kallenberg [71, Lemma 3.24, pp. 58]): Let T be a set of index, and $\{X_t:t\in T\}$ and $\{Y_t:t\in T\}$ be family of random variables (perhaps defined on distinct probability spaces) taking values in some separable metric space E such that X_t and Y_t have the same finite-dimensional distribution. If the paths $t\mapsto Y_t$ lie in some Borel subset B of $\mathcal{B}^T(E)$ (σ-algebra generated by cylindrical Borel sets) then there exists a family of random variables $\{\tilde{X}_t:t\in T\}$ with paths in B such that $P(X_t=\tilde{X}_t)=0$ for every t in T, i.e., $\tilde{X}$ is a version of X.

Note that the initial probability space is irrelevant in the above context, we can always reduce to a canonical space of functions. Moreover, by taking the image measure through the map X if necessary, we can always reduce to the canonical process, i.e., the probability space becomes $(B,\mathcal{B},P_X)$ and the process is the random variable $\omega\mapsto X_t(\omega)=\omega(t)$ from B into itself. So that the *law* of a process carried all necessary information. If two or more processes are involved, then we have to deal with more that one probability measure on the sample space $(B,\mathcal{B})$. Thus only one process can be reduced to the canonical

process and the others can be viewed as probability measures or as measurable functions from B into itself. In most of the cases, the sample space B is a Polish (complete separable metric) space (better than E^T) where the motions of versions and equivalence classes coincide.

Clearly, most of the properties required for a stochastic process are stated relative to the probability P_X on the (Polish) sample space. However, for instance, when we state some property relative to the sample path of a process (such as integrability in one of the variables), the difficulty is the fact we refer either to the map $t \mapsto X(t,\omega)$ for a fixed ω or to the map $\omega \mapsto X(t,\omega)$ for any fixed t. Thus, sometime we are forced to comeback to the initial setting of general stochastic process as a family of E-valued random variables. Alternatively, we may define processes as functions from the *base space* $T \times \Omega$ into E, this is, we consider X as a function of two-variables $X = X(t,\omega)$, and immediately we restrict the attention to (joint) measurable functions, i.e., the so-called measurable processes. This approach yields delicate measurability problems when dealing with stochastic integration, as carefully discussed in Bichteler [14].

The following type of processes may be useful

Definition 4.14 (Gaussian)**.** A real valued process $\{X(t) : t \in T\}$ is a *Gaussian* process if for any finite sub-family $(t_1, \dots, t_n)$ of indexes in T, the random variable $(X(t_1), \dots, X(t_n))$ has a Gaussian n-dimensional distribution. Its *mean* is $m(t) := \mathbb{E}\{X(t)\}$ and its *covariance* is defined by $\Gamma(s,t) := \mathbb{E}\{[X(s) - m(s)][X(t) - m(t)]\}$, for any s, t in T. The process is called *centered* if $\mathbb{E}\{X(t)\} = 0$ for any t in T. □

An important property of the Gaussian processes is the fact that its covariance function is always semi-definite positive, i.e., for any $(t_1, \dots, t_n)$, any n, the matrix $\{\Gamma(t_i, t_j)\}$ is semi-definite positive. Moreover, any symmetric semi-definite positive function is the covariance of a centered Gaussian process, see Revuz and Yor [111, p. 36, Chapter 1].

Another important class of processes is the following

Definition 4.15 (stationary)**.** A E-valued process $\{X(t) : t \geq 0\}$ is called *stationary* if for every $t_1, \dots, t_n$ and t we have

$$P(\{X(t_1 + t) \in A_1, \dots, X(t_n + t) \in A_n\} = \\ = P(\{X(t_1) \in A_1, \dots, X(t_n) \in A_n\}),$$

for any Borel sets $A_1, \dots, A_n$ in E, i.e., its finite-dimensional distribution is invariant by a time translation. □

These processes play a central role in the study of *ergodicity* or *stability*, e.g., see the books Khasminskii [80] and Skorokhod [123].

4.5 Continuous 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 $1, \dots, n$. 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 cad-lag 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 $i = 1, \dots, n$, 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

$$P\{X(t) = j \mid X(r),\, r \le s\} = P\{X(t) = j \mid X(s)\}, \tag{4.6}$$

almost surely, for every $t > s$, $j = 1, \dots, n$. At this point, the reader may consult the classic book Doob [38, 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

$$\begin{aligned} &\sum_{j=1}^n p_{ij}(s,t) = 1,\ \forall t > s, \qquad \lim_{(t-s)\to 0} p_{ij}(s,t) = \delta_{ij},\ \forall t > s, \\ &p_{ij}(s,t) = \sum_{k=1}^n p_{ik}(s,r) p_{kj}(r,t),\ \forall t > r > s. \end{aligned} \tag{4.7}$$

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

$$\begin{aligned} &\partial_s p_{ij}(s,t) = \sum_{k=1}^n \rho_{ik}^+(s)\, p_{kj}(s,t), \quad \forall t > s,\ i, j, \\ &\rho_{ij}^+(s) = \lim_{r \to s} \partial_s p_{ij}(s,r) \quad \forall s,\ i, j, \end{aligned} \tag{4.8}$$

or the *forward* equation

$$\begin{aligned}\partial_t p_{ij}(s,t) &= \sum_{k=1}^{n} p_{ik}(s,t)\,\rho_{kj}^{-}(t), \quad \forall t > s,\ i,j,\\ \rho_{ij}^{-}(t) &= \lim_{r\to t} \partial_t p_{ij}(r,t) \quad \forall t,\ i,j,\end{aligned} \tag{4.9}$$

The quantities $\rho_{ij}^{+}(s)$ and $\rho_{ij}^{-}(s)$ are the characteristic of the process, referred to as *infinitesimal rate*. The initial condition of (4.7) 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

$$\rho(t,i,j) \geq 0, \quad \forall i \neq j, \qquad \rho(t,i,i) = -\sum_{j\neq i} \rho(t,i,j). \tag{4.10}$$

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

$$\begin{aligned}&\partial_s P(s,t) = -R(s)P(s,t), \quad \forall s < t,\\ &\partial_t P(s,t) = P(s,t)R(t), \quad \forall t > s,\\ &\lim_{t-s\to 0} P(s,t) = I, \quad \forall t > s.\end{aligned} \tag{4.11}$$

Conversely, given the integrable functions $\rho_{ij}(t)$, $i,j = 1,\dots,n$, $t \geq 0$ satisfying (4.10), we may solve the system of (non-homogeneous and linear) ordinary differential equations (4.8), (4.9) or (4.11) 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 [26], Yin and Zhang [136, Chapters 2 and 3, pp. 15–50].

Since $P(s,t)$ is continuous in $t > s \geq 0$ and satisfies the conditions in (4.7), 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 [22], Kemeny and Snell [79], Kleinrock [81], Nelson [101].

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 (as discussed later) of a Markov chain is given by

$$P\{X_0 \in A_0,\, X_1 \in A_1, \dots\, X_n \in A_n\} = \\ = \int_{A_0} \nu(\mathrm{d}x_0) \int_{A_1} P(x_0, \mathrm{d}x_1) \cdots \int_{A_n} P(x_{n-1}, \mathrm{d}x_n), \quad (4.12)$$

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 4.20 below) to obtain a Markov chain X_n for $n = 0, 1, \dots$ satisfying the above equation (4.12). 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

$$P^n(x, A) := \int_E P^{n-1}(x, \mathrm{d}y) P(y, A). \quad (4.13)$$

The reader may consult the book by Chung [26] and Shields [120], 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 $\mathbb{N} = \{0, 1, \dots\}$, 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.,

$$\begin{aligned} \partial_t p(t, n) &= -\lambda \big[p(t, n) - p(t, n-1)\big], \\ \partial_t p(t, 0) &= -\lambda p(t, 0), \end{aligned} \quad (4.14)$$

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 (4.14) becomes

$$\begin{aligned} \partial_t p(t,n) &= -\lambda_n p(t,n) + \lambda_{n-1} p(t,n-1), \\ \partial_t p(t,0) &= -\lambda p(t,0), \end{aligned} \tag{4.15}$$

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 (4.14) is modifies as follows

$$\begin{aligned} \partial_t p(t,n) &= -(\lambda_n + \mu_n) p(t,n) + \lambda_{n-1} p(t,n-1) + \mu_{n+1} p(t,n+1), \\ \partial_t p(t,0) &= -\lambda p(t,0) + \mu_1 p(t,1), \end{aligned} \tag{4.16}$$

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.

4.6 Markov Processes

There is an important class of processes known as Markov process 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 $\mathbb{R}^d$.

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 $f_1, \dots, f_n$, $g_1, \dots, g_n$, h, and times $s_1 \le \cdots \le s_n \le t \le t_1 \le \cdots \le t_n$ we have

$$\mathbb{E}\big\{h(X_t)\big(\prod_{i=1}^{n} f(X_{s_i})\big)\big(\prod_{i=1}^{n} g(X_{t_i})\big)\big\} = \\ = \mathbb{E}\big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n} f(X_{s_i}) \,|\, X_t)\}\prod_{i=1}^{n} g(X_{t_i})\big\}, \quad (4.17)$$

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

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

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 (4.17) is equivalent to (e.g., see Blumental and Getoor [17, Thm 1.3, pp. 12-14]) either

$$P(X_t \in B \,|\, X_{s_1}, \dots, X_{s_n}, X_s) = P(X_t \in B \,|\, X_s), \quad \text{a.s.}$$

for every $t > s \ge s_n > \cdots > s_1$, B in $\mathcal{B}(E)$, or

$$\mathbb{E}\{f(X_t) \,|\, X_{s_1}, \dots, X_{s_n}, X_s\} = \mathbb{E}\{f(X_t) \,|\, X_s\}, \quad \text{a.s.},$$

for every $t > s \ge s_n > \cdots > s_1$, and for any arbitrary bounded and continuous (actually, with compact support when E is locally compact) function f from E into $\mathbb{R}$.

Definition 4.16 (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 $\{X(s) : s \le t\}$ 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 are 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 $\{\omega : \tau(\omega) \le t\}$ 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 (4.17) 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 4.17 (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).

Remark that assuming that in $(\Omega.\mathcal{F}.P)$ regular conditional probabilities are used (their existence is implicitly assumed), and referring to the Markov property as stated in Definition 4.17, the equality is true except on a set of probability zero which may depend on t, s. Therefore, the path of the process should

have some regularity to completely identify the process in term of its finite-dimensional distributions.

To avoid deep difficulties, only cad-lag Markov processes are considered, 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. Note that 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. This means that the process $(X_t : t \in T)$ is necessary 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. In short, 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

$$P\{X_t \in A \mid X_s = x\} = P(s, x, t, A), \quad \forall s < t,\ x \in E,\ A \in \mathcal{B}(E),$$

for some function $P(s, x, t, A)$. Note that

$$P\{X_t \in A \mid X_s = x\} = \frac{P(\{X_t \in A,\ X_s = x\})}{P(\{X_s = x\})} = \\ = \frac{1}{P(\{X_s = x\})} \int_{\{X_s = x\}} P\{X_t \in A \mid X_s\}(\omega) P(\mathrm{d}\omega),$$

whenever $P(\{X_s = x\}) > 0$ and $P\{X_t \in A \mid X_s = x\} = 0$ when $P(\{X_s = x\}) = 0$, under the condition that a regular conditional probability exists.

Definition 4.18 (transition)**.** A *transition probability function* on a given measurable space $(E, \mathcal{E})$, 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 $\mathcal{E}$ 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 $(E, \mathcal{E})$,

(b) for each $s < t$ in T and A in $\mathcal{E}$ the function $x \mapsto P(s, x, t, A)$ is a measurable,

(c) for any s in T, x in E and A in $\mathcal{E}$ 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 $\mathcal{E}$ 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 $\mathcal{E} = \mathcal{B}(E)$, 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 4.17. 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

$$\begin{aligned} &\text{(a)} \ \lim_{t \to s} \sup_{x \in K} [1 - P(s,x,t,B(x,\varepsilon))] = 0, \\ &\text{(b)} \ \lim_{x \to \infty} \sup_{0 \le s < t \le 1/\varepsilon} P(s,x,t,K) = 0, \end{aligned} \tag{4.18}$$

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 (4.18), 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.

Perhaps, it could be convenient for the reader to check, for instance, Morimoto [100, Sec 2.6, pp. 73–76] to have a clean idea of how this Markov property is expressed in term of the solution of an ordinary differential equation driven by a Wiener process.

Theorem 4.19 (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,

$$P(X(\tau + t) \in A \,|\, \mathcal{F}_\tau^+) = P(t, X(\tau), A), \quad a.s.\ on\ \ \{\tau < \infty\},$$

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 [39, Theorems 8 and 9, pp. 556-560] or in Blumental and Getoor [17, 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 4.18 of transition function including condition (e) and (4.18) 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 $|\varphi(x)| \le \varepsilon$ for every x in $E \setminus K$), such that

(a) $0 \le P(t, s)\varphi \le 1$, for every φ in $C_0(E)$ with $0 \le \varphi \le 1$

(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 $\varsigma(\omega) := \{t \ge 0 : X(t) = \infty\}$, and assume that $X(t) = \infty$ for every $t \ge \varsigma$. 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 fact is 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 Blumental and Getoor [17], Dynkin [44, 45] or more recent books, such as, Çınlar [29], Chung [27], Dellacherie and Meyer [36], Ethier and Kurtz [47], Marcus and Rosen [94], Rogers and Williams [112], Taira [131], among many others.

4.7 Construction of Processes

A family of finite-dimensional distributions $\{P_s : s \in T^n,\ n = 1, 2, \dots\}$, on a Borel (usually open or closed) subset of $\mathbb{R}^d$, derived from a stochastic process forms a so-called *projective family*, i.e., the following (natural) *consistency conditions* conditions are satisfied:

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

(b) if $t = (t_1, \dots, t_n, t_{n+1})$ and $s = (t_1, \dots, t_n)$ and B in $\mathcal{B}(E^n)$ then $P_t(B \times E) = P_s(B)$.

If a total order is given on the index set T, it is enough to have the finite-dimensional distributions defined only for $(s_1, s_2, \dots, s_n)$ such that $s_1 < s_2 < \cdots < s_n$ and to satisfy only a consistency conditions of the form

(b') if $t = (t_1, \dots, t_n)$ and $s = (s_1, \dots, s_m)$ with $t_1 < \cdots < t_n < r < s_1 < \cdots < s_m$ 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 $n, m = 0, 1, \dots$.

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 [71], Karatzas and Shreve [75], Malliavin [93], Revuz and Yor [111], among others, for a comprehensive treatment).

Theorem 4.20 (Kolmogorov). *Let $\{P_s : s \in T^n,\ n = 1, 2, \dots\}$ 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 [39, 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 [112, Chapter 2, Sections 3 and 6].

We recall that a cylinder 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 $(t_1, t_2, \dots, t_n)$ and $B \in \mathcal{B}(E^n)$ such that $(\omega(t_1), \omega(t_2), \dots, \omega(t_n))$ belongs to B for any $i = 1, \dots, n$. The class of cylinder sets with $t_1, \dots, t_n$ fixed is equivalent to product σ-algebra in $E^{\{t_1,\dots,t_n\}} \simeq E^n$ 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 $T = \{1, 2, \dots\}$ 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. For instance,

combining the last two theorems, if a consistent family of distributions satisfies the a priori estimate (4.4), then we can redefine the probability measure P of the above theorem on the space $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$, of continuous functions.

There is a very important class of stochastic processes so-called *Lévy processes* which provides prototype examples, we refer to Bertoin [11] and Sato [116] for a comprehensive study.

Definition 4.21 (Lévy process)**.** An E-valued ($E \subset \mathbb{R}^d$) *Lévy (stochastic) process* is a couple (P_X, X) as in Definition 4.13 satisfying:

(a) its *increments are independent of the past* i.e., for any $t > s \geq 0$ the random variable $X_t - X_s$ is independent of X_s under P_X,

(b) it has *stationary increments* i.e., for any $t > s \geq 0$ the random variable $X_t - X_s$ has the same distribution as X_{t-s} under P_X,

(c) its paths are *continuous in probability* (referred to as *stochastically continuous*) i.e., for any $\varepsilon > 0$ and $s \geq 0$ we have

$$\lim_{t \to s} P(|X_t - X_s| \geq \varepsilon) = 0.$$

Usually the condition $P_X(X_0 = 0) = 0$ is added to normalize the process. □

Note that a process with independent increments, property (a), has its finite-dimensional distributions completely determined by the distributions of $(X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_k} - X_{t_{k-1}})$, for any $t_0 \leq t_1 \leq \cdots \leq t_k$ in $[0, +\infty)$. Moreover, if the process has stationary increments, property (b), then the distribution of the random variables $\{X_t : t \geq 0\}$ in E completely characterize its finite-dimensional distributions. Adding the continuity condition (c), the Lévy process is identified. Now, instead of looking at the distributions of X_t under P_X we may consider its *characteristic functions*, which is its Fourier transform i.e., for any $t \geq 0$

$$\varphi_t(\xi) = \mathbb{E}\{\mathrm{e}^{\mathrm{i}(\xi, X_t)}\} := \int_\Omega \mathrm{e}^{\mathrm{i}(\xi, X_t(\omega))} P_X(\mathrm{d}\omega),$$

where $(\cdot, \cdot)$ is the *dot* product in $\mathbb{R}^d$ and i is the imaginary unit.

It is not so hard to check that any characteristic function $\varphi(\xi)$, $\xi \in \mathbb{R}^d$ has the following properties:

(a) $\varphi(\cdot)$ is continuous from $\mathbb{R}^d$ into $\mathbb{C}$ and $\varphi(0) = 1$,

(b) $\varphi(\cdot)$ is *positive definite* i.e., for any $k = 1, 2, \dots$, $\zeta_1, \dots, \zeta_k$ in $\mathbb{C}$ and $\xi_1, \dots, \xi_k$ in $\mathbb{R}^d$ we have

$$\sum_{i,j=1}^{k} \varphi(\xi_i - \xi_j) \zeta_i \bar{\zeta}_j \geq 0,$$

we refer to Shiryayev [121, Section 2.12, pp. 272–294] for a more detail. So, this is satisfied by the family of characteristic functions $\{\varphi_t : t \geq 0\}$. Moreover,

since the Lévy process (P_X, X) has independent and stationary increments, see properties (a) and (b), the one-parameter functions φ_t satisfies $\varphi_{t+s} = \varphi_t \varphi_s$, for any $t \geq s \geq 0$. From the property (c) follows that the function $t \mapsto \varphi_t(\xi)$ is continuous for any $\xi \in \mathbb{R}^d$ from $[0, \infty)$ into $\mathbb{C}$. Thus the characteristic function are of the form $\varphi_t(\xi) = \exp[-t\psi(\xi)]$, for some non-negative and continuous function ψ with $\psi(0) = 0$. Actually the Lévy-Khintchine formula (see Protter [108, p. 32, Theorem 43] give a simple expression for the one-dimensional case

$$\psi(\xi) = \frac{\sigma^2}{2}\xi^2 - \mathtt{i}\alpha\xi + \int_{|x|\geq 1} (1 - \mathrm{e}^{i\xi x})\nu(\mathrm{d}x) + \int_{|x|<1} (1 - \mathrm{e}^{\mathtt{i}\xi x} + i\xi x)\nu(\mathrm{d}x),$$

where ν (called Lévy measure) is a Radon measure on $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$ satisfying

$$\nu(\{|x| \geq 1\}) + \int_{|x|<1} |x|^2 \nu(\mathrm{d}x) < +\infty,$$

and the constants σ, α are the characteristic parameters of the process i.e., uniquely determine the Lévy process. This generalizes to $\mathbb{R}^d$, where now σ^2 is a positive semi-definite quadratic $d \times d$ matrix, α, x and ξ are in $\mathbb{R}^d$ and the dot product is used, see Bertoin [11, Theorem I.1, pp. 13–15]. Moreover, the canonical filtration $\{\mathcal{F}(t) : t \geq 0\}$ associated with a Lévy process (P_X, X) is right-continuous, i.e. it satisfies the usual conditions, see Definition 4.6.

The converse can be established as follows. First, by means of the classic Bochner theorem (e.g. see Pallu de la Barrière [104, p. 157, Theorem 1] that for any non-negative and continuous function ψ with $\psi(0) = 0$ such that $\xi \mapsto \exp[-t\psi(\xi)]$ is positive definite for any $t \geq 0$, is indeed the characteristic function of some distribution), there exist a family of probability distribution $P(t, \mathrm{d}x)$ for any $t \geq 0$ whose characteristic function is precisely $\exp[-t\psi(\xi)]$. Now, based on the properties (a) and (b) of a Lévy process, the finite-dimensional distributions are determined by

$$\begin{aligned} &P(X(t_1) \in B_1\,,\, X(t_2) \in B_2\,,\, \dots\,,\, X(t_n) \in B_n) = \\ &= P(X(t_1) \in B_1\,,\, X(t_i) - X(t_{i-1}) \in B_i - B_{i-1}\,,\, 2 \leq i \leq n) = \\ &= P(t_1, B_1)P(t_2 - t_1, B_2 - B_1) \cdots P(t_n - t_{n-1}, B_n - B_{n-1})\,, \end{aligned}$$

for any $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$. Hence, Kolmogorov construction may be applied. The fact that the family of characteristic functions $\exp[-t\psi(\xi)]$ are continuous in t, locally uniformly in ξ implies that the process is continuous in probability i.e., condition (c) is satisfied. Therefore, if ν, σ and α are given then there exists a unique Lévy process with the prescribed characteristic parameters. Moreover, as proved in Protter [108, p. 21, Theorem 30] any Lévy process has a version which is *cad-lag* i.e., the paths $t \mapsto X_t(\omega)$ are right-continuous functions having left-hand limits from $[0, \infty)$ into E for any $\omega \in \Omega \setminus N$ with $P_X(N) = 0$. In other words, the Kolmogorov construction is valid on the space of cad-lag functions.

The transition function $P(t, x, A)$ of a Lévy process has the form of *convolutions semigroup*, i.e., a family of probability measures on $\mathbb{R}^d$ such that

(a) $\nu_s * \nu_t = \nu_{s+t}$ for any s, t

(b) $\nu_t \to \delta_0$ in the weak topology,

and

$$P(t, x, A) := \int_{\mathbb{R}^d} \mathbb{1}_A(x + y)\nu_t(\mathrm{d}y),$$

see Revuz and Yor [111, Propositon 2.18, pp. 96-97].

As seen in the classic Kolmogorov construction, the finite-dimensional distributions characterize the statistics of a process, but not its sample properties. If we are only interested in the statistics of a process, we may adopt another viewpoint to identify a process. Instead of given its finite-dimensional distributions, we may prescribe some time-evolution or dynamics.

Suppose that a transition (probability) function $P(s, x, t, A)$ (see Definition 4.18) and an initial (i.e. for $t = 0$) distribution ν are given. Then we can define an *absolute probability function* by

$$P_0(t, A) = \int P(0, x, t, A)\, \nu(\mathrm{d}x), \quad \forall A \in \mathcal{F}$$

and a family of finite-dimensional distributions for $t_1 < \cdots < t_n$ as follows

$$\begin{aligned} &P_{t_1,\ldots,t_n}(A_1 \times A_2 \times \cdots \times A_n) = \\ &\quad = \int_{A_1} P_0(t_1, \mathrm{d}x_1) \int_{A_2} P(t_1, x_1, t_2, \mathrm{d}x_2) \cdots \int_{A_n} P(t_{n-1}, x_{n-1}, t_n, \mathrm{d}x_n), \end{aligned}$$

for any $A_1, A_2, \ldots, A_n$ in $\mathcal{F}$. It is not hard to check that the consistency condition (b') is satisfied.

Thus for a given transition function on $(E, \mathcal{F})$, with $E \subset \mathbb{R}^d$, we can use Kolmogorov construction to get a canonical process (i.e. a probability measure) on $(E^T, \mathcal{B}(E^T))$. Moreover, if the continuity condition (4.4) is satisfied, which is verifiable in term of the transition function and the initial probability i.e.,

$$\int_E |x - y|^\alpha P(s, x, t, \mathrm{d}y) \le C|t - s|^{1+\beta}, \quad \forall s, t \in [0, T],\ \forall x \in E,$$

for some positive constants α, β and C, then the construction can be performed over the space $C(T, E)$, with the local uniform convergence topology and the induced Borel σ-algebra. Similarly, the cad-lag condition (4.5) becomes

$$\int_E |x - y|^\alpha P(s, x, t, \mathrm{d}y) \le C|t - s|^{\frac{1}{2}+\beta}, \quad \forall s, t \in [0, T],\ \forall x \in E,$$

after some simplification.

If we denote by (P_X, X) the process constructed as above on the product space $(E^T, \mathcal{B}(E^T))$, with its canonical filtration $\{\mathcal{F}(t) : t \ge 0\}$, then we have the following property (derived from the Chapman-Kolmogorov identity)

$$P_X(X(t) \in \cdot \mid X(s)) = P(s, X(s), t, \cdot), \quad \forall t \ge s \tag{4.19}$$

So that $P(s,x,t,A)$ represents the conditional probability that the state of the process belongs to A at time t knowing that initially it was x at time s. This is referred to as the *Markov property*.

Let us state an useful result relative to the construction of standard Markov process, actually the so-called Markov-Feller processes. There are several way of constructing a *good* Markov process from a given Markov (or sub-Markov) transition as in Definition 4.18 conditions (a), (b), (c) and (d). For instance we refer to Blumental and Getoor [17, Chapter 1, Theorem 9.4, pp. 46-51], Dynkin [45, Chapter III, Theorems 3.5, 3.6, pp. 81-92], Ethier and Kurtz [47, Chapter 4, Theorem 1.1, pp. 157-169], among others. We summarize these results in a non-homogeneous form as follows:

Theorem 4.22 (realization)**.** *Let $P(s,x,t,\mathrm{d}y)$ be a transition probability function satisfying properties* (a), (b), (c), (d) *of Definition 4.18 on a Polish space E. Then, for any given probability measure ν on E and any initial time $s \geq 0$, there exist a unique Markov process $(P, X_t, \mathcal{F}_t; t \geq s)$ satisfying condition (4.19) and such that X_s has ν as its distribution. If the transition probability function either* (1) *is Feller, i.e., the function $(s,x) \mapsto P(s,x,t,f)$ is continuous from $[0,t] \times E$ into $\mathbb{R}$ for any t in $(0,\infty)$ and any bounded continuous function f from E into $\mathbb{R}$, or* (2) *satisfies condition (4.18), then the Markov process $(P, X_t, \mathcal{F}_t; t \geq s)$ has cad-lag paths and right-continuous filtration. Moreover, if the function $(x,t) \mapsto P(s,x,t,f)$ is continuous from $[s,\infty) \times E$ into $\mathbb{R}$, for any s in $[0,\infty)$ and any bounded continuous function f from E into $\mathbb{R}$, then $(X_t, \mathcal{F}_t : t \geq s)$ is quasi-left-continuous, i.e., for any increasing sequence of stopping time $\{\tau_n : n = 1,2,\ldots\}$ almost surely strictly convergent to τ, $P(\tau_n \leq \tau_{n+1} < \tau < \infty, \tau > s) = 1$, then X_{τ_n} converges to X_τ a.s.* (or equivalently the σ-algebra $\mathcal{F}_\tau$ is the minimal σ-algebra containing the sequence of σ-algebra $\{\mathcal{F}_{\tau_n} : n = 1,2,\ldots\}$). Furthermore, if

$$\lim_{t \to s} \frac{1}{t-s} \sup_{x \in K} \{1 - P(s,x,t,B(x,\varepsilon))\} = 0,$$

for every $s \geq 0$ and x in E, then the Markov process $(P, X_t, \mathcal{F}_t; t \geq 0)$ has continuous paths. □

There are several ways of expressing the continuity condition (Feller's property) assumed in the previous theorem. Sometime, a transition probability $P(s,x,t,\mathrm{d}y)$ is called *stochastically continuous* if it satisfies

$$\lim_{t-s \to 0} P(s,x,t,\{y \in E : |y-x| < \delta\}) = 1, \quad \forall x \in E,$$

see Dynkin [45, Chapter 2]. This is equivalent to the continuity in probability of a realization of the Markov process and the Feller character used above. Moreover, a Markov process satisfying the above regularity on its transition function is called a *Markov-Feller process*. Actually, with the aid of the sample space $D([0,\infty),E)$ discussed later, we see that the so-called *realization* of the Markov-Feller process given in the previous theorem can be regarded as a probability measure P on $D([0,\infty),E)$, with $X_t(\omega) = \omega(t)$ the canonical process

and $(\mathcal{F}_t : t \geq 0)$ its canonical filtration. Moreover, it satisfies the strong Markov property. At this point, a reading of Chapter 4 in Marcus and Rosen [94] could be of some help.

We may look at stochastic processes with complex values. In this context, we have the so-called second-order process, which are well adapted to dynamical systems.

Definition 4.23 (second-order process)**.** A (complete) probability measure P_X on $(\Omega, \mathcal{F})$, where $\Omega \subset \mathbb{C}^T$ and $\Omega \bigcap \mathcal{B}(\mathbb{C}^T) \subset \mathcal{F}$, together with a measurable mapping X from Ω into itself satisfying:

(a) $\mathbb{E}_X\{|X_t|^2\} < \infty$, for all t in T,

(b) $t \mapsto X_t$ from T into $L^2(P_X)$ is continuous

will be called a mean square continuous second-order process (P_X, X). □

This definition is independent of the sample path of the process, i.e., any version of the process should satisfy the conditions (a) and (b) above. Thus, as the case of processes continuous in probability, second-order processes cannot be regarded as random variables with values in some appropriate (sample) function space.

For a second-order process we can define the *mean* and *covariance* function:

$$\mu(t) := \mathbb{E}_X\{X_t\} \qquad R(t,s) := \mathbb{E}_X\{(X_t - \mu(t))\overline{(X_s - \mu(s))}\},$$

where the *over-bar* denotes the complex conjugate. Several properties can be discussed for this class of processes. For instance, we can mention that a second-order process is continuous in mean square if and only if its covariance function $R(t,s)$ is continuous. On the other hand, any continuous second-order process (P_X, X) with $T = [a,b]$ has a version of the form

$$X_t(\omega) = \sum_{n=1}^{\infty} \sqrt{\lambda_n}\varphi_n(t)y_n(\omega),$$

where the convergence is in $L^2(P_X)$, uniformly for t in $[a,b]$, and $\{\varphi_n : n = 1,2,\dots\}$ are the orthonormal eigenfunctions and $\{\lambda_n : n = 1,2,\dots\}$ are the eigenvalues of

$$\int_a^b R(t,s)\varphi(s)\mathrm{d}s = \lambda\varphi(t), \qquad a \leq t \leq b$$

in $L^2(]a,b[)$ and $\{y_n : n = 1,2,\dots\}$ is an orthonormal system in $L^2(P_X)$ satisfying

$$y_x(\omega) = \sqrt{\lambda_n}\int_a^b \varphi_n(t)X_t(\omega)\mathrm{d}t.$$

We refer to Shiryayev [121, Section 2.10, pp. 260–272] for the discrete case, and to Wong [134, Chapter 3, pp. 74–138] for more details.

Related with the second-order processes are the so-called *processes with orthogonal increment*, which are defined as a (complete) probability measure P_X on $(\Omega, \mathcal{F})$, where $\Omega \subset \mathbb{C}^T$ and $\Omega \bigcap \mathcal{B}(\mathbb{C}^T) \subset \mathcal{F}$, together with a measurable mapping X from Ω into itself satisfying:

(a) $\mathbb{E}_X\{|X_t - X_s|^2\} < \infty$, for all t, s in T,

(b) $\mathbb{E}_X\{X_t\, X_s\} = 0$, for all t, s in T,

It can be proved that to each such a process corresponds a monotone nondecreasing functions F_X, uniquely determined up to an additive constant, satisfying

$$\mathbb{E}_X\{|X_t - X_s|^2\} = F_X(t) - F_X(s), \quad \forall t > s.$$

Moreover, the right-hand $X(t+)$ and the left-hand $X(t-)$ limits exists in the $L^2(P_X)$ sense, for every t in T, and $X(t-) = X(t) = X(t+)$, except for a countable set in T. Thus, processes with orthogonal increments can be normalized to be cad-lag in mean square sense. We refer to Doob [38, Chapter IX, pp. 425–451].

4.8 Examples of Markov processes

Two typical examples can be given, (a) the standard Wiener process (or Brownian motion) $(w(t), t \geq 0)$ and (b) the standard Poisson process $(p(t), t \geq 0)$. Both are the prototype of Lévy processes (see Definition 4.21). For the Wiener process, we take the Gauss kernel

$$p(t,x) = (2\pi t)^{-d/2} \exp(-\frac{|x|^2}{2t}), \quad t > 0,\ x \in \mathbb{R}^d \tag{4.20}$$

and consider the transition function

$$P_w(s,x,t,A) = \int_A p(t-s, y-x)\mathrm{d}y, \quad \forall t > s,\ A \in \mathcal{B}(\mathbb{R}^d),$$

the initial probability P_0 as the Dirac measure. Kolmogorov construction and path regularity apply in this case to generate a probability measure P_w in the space $C([0,+\infty), \mathbb{R}^d)$, called *Wiener measure*. Under P_w, the canonical process is a standard Wiener process. Certainly there several ways to construct a Brownian motion and a critical point is to show continuity of its paths. In general, a one-dimension standard Brownian motion is defined as a real valued stochastic process $\{B(t) : t \geq 0\}$ satisfying:

(a) $B(0) = 0$ and for $0 \leq s < t < \infty$, the difference $B(t) - B(s)$ is a normally distributed random variable with mean zero and variance $t - s$, i.e., for every α in $\mathbb{R}$,

$$P\big(B(t) - B(s) > \alpha\big) = \int_\alpha^{+\infty} \big(2\pi(t-s)\big)^{-1/2} \exp\big(-\frac{|x|^2}{2(t-s)}\big)\mathrm{d}x,$$

(b) for $0 \le t_0 < t_1 < \cdots < t_n$, the family $\{B(t_0), B(t_k) - B(t_{k-1}),\ k = 1, \dots, n\}$ is a set of independent random variables.

The above two properties characterize a *Brownian motion* (or Wiener process) as an indistinguishable stochastic process. Because it is well know that (e.g. Chung [27, p. 145]) any Brownian motion has a version with continuous path, we work always with a continuous version, actually with the Wiener measure. A d-dimensional version is constructed by taking d independent one-dimensional Brownian motions. Note that the characteristic function is given by

$$\mathbb{E}\{\exp \mathtt{i}(\xi, B(t))\} = \exp\Big(-t\frac{|\xi|^2}{2}\Big), \qquad \forall t \ge 0,$$

with $\mathtt{i} = \sqrt{-1}$. There are many properties on the paths of a Wiener process $\{w(t) : t \ge 0\}$, we mention some of them (see Itô and McKean [66], Krylov [83, pp. 36-38])

(a) for any c non zero constant, $w(t+c) - w(c)$, $c^{-1}w(c^2t)$, $\forall t \ge 0$ and $tw(1/t)$, $\forall t > 0$ (and $= 0$ for $t = 0$) are also Wiener process;

(b) for any constants α in $(0, 1/2)$ and T in $(0, \infty)$ there exists a random variable $C(\omega)$ such that $\mathbb{E}\{|C|^p\} < \infty$ for any positive and finite p and

$$|w(t,\omega) - w(s,\omega)| \le C(\omega)|t-s|^\alpha, \quad \forall t, s \in [0,T],\ \omega \in \Omega.$$

In particular, $|w(t,\omega)| \le C(\omega)t^\alpha$ for $t \in [0,T]$ and $|w(t,\omega)| \le C(\omega)t^{1-\alpha}$ for $t > T$.

(c) for any ω in $\Omega \setminus N$, with $P(N) = 0$ we have

$$\limsup_{s \downarrow 0} \frac{|w(t+s) - w(t)|}{\sqrt{-2s \ln s}} = 1, \quad \forall t > 0,$$

$$\limsup_{t \downarrow 0} \frac{w(t)}{\sqrt{2t \ln(-\ln t)}} = 1, \qquad \liminf_{t \downarrow 0} \frac{w(t)}{\sqrt{2t \ln(-\ln t)}} = -1,$$

which are referred as the *law of the iterated logarithm.*

(d) let $\{t_{0,n}, \dots, t_{k,n} : n = 1, 2, \dots\}$ be a sequence of partition of $[a,b] \subset [0,\infty)$ with mesh going to zero, then

$$\lim_{n \to \infty} \sum_{i=1}^{k} [w(t_i) - w(t_{i-1})]^2 = b - a, \quad a.s.$$

As a consequence, almost surely the sample paths $t \mapsto w(t,\omega)$ of a standard Wiener process are of unbounded variation on any interval. It is possible to construct a Wiener process based on a complete system of orthonormal function in L^2. For instance (Krylov [83, pp. 32]), let $\{\xi_n : n = 0, 1, \dots\}$ be a sequence of independent (standard) normal distributed random variables. Define the process

$$w_t^k(\omega) = \frac{1}{\sqrt{\pi}} t\, \xi_0(\omega) + \sqrt{\frac{2}{\pi}} \sum_{n=1}^{N(k)} \xi_n(\omega) \frac{1}{n} \sin(nt).$$

Then there exists increasing sequence of positive integers $\{N(k) : k = 1, 2, \dots\}$ such that the sequence of processes w_t^k converges uniformly for $t \in [0, \pi]$ and $\omega \in \Omega \setminus N$, $P(N) = 0$ to a Wiener process. Another typical expression is the locally uniformly convergent series

$$w_t(\omega) = \frac{\sqrt{2}}{\pi} \sum_{n=0}^{\infty} \xi_n \frac{\sin\big(t\pi(n + 1/2)\big)}{n + 1/2},$$

e.g., see Knight [82, Chapter 1].

Related to a Wiener process (or Brownian motion) is a Brownian motion with drift, defined by $x(t) := \sigma w(t) + bt$, for some constants σ and b, where each random variable $x(t)$ has a Gaussian distribution. Similarly, a *geometric* Wiener process is defined by

$$x(t) := \exp[\sigma w(t) + bt], \quad t \geq 0,$$

where now $x(t)$ has a log-normal distribution. On the other hand,

$$x(t) := w(t) + \sup_{0 \leq s \leq t} \{\max[w(s), 0]\}, \quad t \geq 0$$

defines a *reflected* Brownian motion on $[0, \infty)$, and $x(t) := w(t) - tw(1)$, $0 \leq t \leq 1$ defines a so-called *Brownian bridge*.

For a one-dimensional standard Poisson process with parameter $c > 0$, we define the transition function

$$P_p(s, x, t, A) = \exp\big[-c(t - s)\big] \sum_{k=0}^{\infty} \frac{[c(t - s)]^k}{k!} \mathbb{1}_A(x + k) \tag{4.21}$$

and we apply Kolmogorov construction. Here, the continuity condition (4.4) is not satisfied, but the process is continuous in probability (see property (c) of Definition 4.21). A Poisson process $\{p(t) : t \geq 0\}$, with parameter $c > 0$, is characterized by the following properties:

(a) $p(0) = 0$ and $0 \leq s < t < \infty$, the difference $p(t) - p(s)$ is a Poisson random variable with mean $c(t - s)$ i.e.,

$$P\big(p(t) - p(s) = n\big) = \big[c(t - s)\big]^n \exp\big[-c(t - s)\big]/n!, \quad n = 0, 1, \dots;$$

(b) for $0 \leq t_0 < t_1 < \cdots < t_n$, the family $\{p(t_0), p(t_k) - p(t_{k-1}),\ k = 1, \dots, n\}$ is a set of independent random variables.

Any Poisson process has a version with right continuous (and left hand limits) paths (see Chung [27, Theorem 3, p. 29]). Almost surely, the paths of a Poisson process are constant except for upward jumps of size one, of which there are finitely many in each bounded interval, but infinitely many in $[0, \infty]$. The times between successive jumps are independent and exponentially distributed with parameter c. Thus, if τ_n is the time between the n and the $(n + 1)$ jumps, then $P(\tau_n > t) = \exp(-ct)$ for each $t \geq 0$. Actually, based on this last property,

a Poisson process can be constructed as follows. Let $\{\tau_n : n = 1, 2, \dots\}$ be a sequence of independent exponentially distributed (with parameter c) random variables. Define the *counting process* as

$$p(t) = \begin{cases} 0 & \text{if } t < \tau_1, \\ n & \text{if } \sum_{i=1}^{n} \tau_i \le t < \sum_{i=1}^{n+1} \tau_i \end{cases}$$

with values in $\{0, 1, 2, \dots\}$. It can be proved (see Protter [108, Chapter 1]) that $\{p(t) : t \ge 0\}$ is a Poisson process. Its characteristic function is given by

$$\mathbb{E}\{\exp(\mathtt{i}\xi p(t))\} = \exp\big[-tc(1 - \mathrm{e}^{\mathtt{i}\xi})\big], \qquad \forall t \ge 0.$$

Similarly, a *Compound-Poisson* process is given by

$$p(t) = \begin{cases} 0 & \text{if } t < \tau_1, \\ Z_n & \text{if } \sum_{i=1}^{n} \tau_i \le t < \sum_{i=1}^{n+1} \tau_i, \end{cases}$$

where $Z_n := \zeta_1 + \zeta_2 + \cdots + \zeta_n$, and $\{\zeta_n : n = 1, 2, \dots\}$ is a sequence of independent identically distributed (with distribution law ν) random variables, independent of the sequence $\tau_1, \dots, \tau_n$. Its transition function is given by the expression of the form

$$P_{cp}(s, x, t, A) = \exp\big[-c(t-s)\big] \sum_{k=0}^{\infty} \frac{[c(t-s)]^k}{k!} \nu^k(x + A), \tag{4.22}$$

where $\nu^k = \nu * \nu^{k-1}$ is the k measure-convolutions of ν, i.e., the distribution of the random variable $Z_k = \zeta_1 + \zeta_2 + \cdots + \zeta_k$. In general, if ν is a distribution in $\mathbb{R}^d$ with $\nu(\{0\}) = 0$ then a compound Poisson process $p_{cp}(t)$ in $\mathbb{R}^d$ has the characteristic function

$$\mathbb{E}\{\exp(\mathtt{i}\xi \cdot p_{cp}(t))\} = \exp\big[-tc(1 - \hat{\nu}(\xi))\big], \qquad \forall\, t \ge 0,$$

where $\hat{\nu}$ is the characteristic transform of ν, i.e.,

$$\hat{\nu}(\xi) = \int_{\mathbb{R}^d} \exp(\mathtt{i}\xi \cdot x)\nu(\mathrm{d}x), \qquad \forall\, \xi \in \mathbb{R}^d.$$

In particular, if $d = 1$ and $\nu = \delta_1$ then we get an standard Poisson process as above, e.g., see Sato [116, p.18].

A more general viewpoint is to define a (standard or homogeneous) *Poisson measure* (or Poisson point process) $\{p(t, \cdot) : t \ge 0\}$ with Lévy (characteristic or intensity) measure $m(\cdot)$ by the conditions:

(a) $m(\cdot)$ is a Radon measure on $\mathbb{R}^d_* := \mathbb{R}^d \setminus \{0\}$, i.e., $m(B) < \infty$ for any compact subset B of $\mathbb{R}^d_*$;

(b) $\{p(t, B) : t \ge 0\}$ is a Poisson process with parameter $m(B)$, for any compact subset B in $\mathbb{R}^d_*$ (with $p(t, B) := 0$ if $m(B) = 0$);

(c) if $B_1, B_2, \dots, B_n$ are disjoint Borel set in $\mathbb{R}^d_*$ then the Poisson processes $p(\cdot, B_1), p(\cdot, B_2), \dots, p(\cdot, B_n)$ are independent.

Note that the Lévy measure $m(\cdot)$ may have a singularity at the origin (at most) of order 2, i.e.,

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

Clearly from (b) follows

$$\mathbb{E}\{p(t, B)\} = t m(B), \quad \forall t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d_*).$$

Also, Poisson measures can be represented by means of Dirac point mass (or atoms), i.e.,

$$p(t, B) = \sum_{0 \leq s \leq t} \mathbb{1}_{e(s) \in B}, \quad \forall t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d_*),$$

where $\{e(t) : t \geq 0\}$ is a *Poisson point process* with characteristic measure $m(\cdot)$, i.e., $\{e(t) : t \geq 0\}$ is a cad-lag (independent of each other) process taking values in $\mathbb{R}^d$, which is equal to 0 except for a countable number of times and its counting process $p(t, B)$ is a Poisson process. Equivalently, a (random) sequence $\{(e_n, \tau_n) : n \geq 1\}$ in $\mathbb{R}^d_* \times [0, \infty)$ such that its counting process

$$p(t, B) = \sum_n \mathbb{1}_{e_n \in B}\, \mathbb{1}_{0 \leq \tau_n \leq t}, \quad \forall t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d_*)$$

is a Poisson process. If $m(\cdot)$ is bounded, i.e., $m(\mathbb{R}^d_*) < \infty$ then the following expression

$$P_p(s, x, t, A) = \exp\big[-m(\mathbb{R}^d_*)(t-s)\big] \sum_{k=0}^{\infty} \frac{(t-s)^k}{k!} m^k(x + A),$$
$$m^k(B) = \int_{\mathbb{R}^d_*} m^{k-1}(z + B) m(\mathrm{d}z), \quad k = 1, 2, \dots,$$

gives the corresponding transition function for the Poisson measure process. Comparing with (4.22), we see that compound Poisson processes are particular cases of Poisson measure process. We refer to Bensoussan and Lions [9, Chapter 3, Section 5] Bertoin [11], Bremaud [21], Jacod and Shiryaev [69], Protter [108] and Sato [116] for a systematic discussion of the above statements.

Other less typical processes (but particular cases of the above) are (c) the Cauchy process $(c(t), t \geq 0)$ and (d) the deterministic process $(d(t), t \geq 0)$. For the one-dimensional Cauchy process, we define the transition function

$$P_c(s, x, t, A) = \int_A \frac{\pi(t-s)\mathrm{d}y}{(t-s)^2 + (y-x)^2}, \tag{4.23}$$

and we apply Kolmogorov construction. It can be proved that the continuity condition (4.4) is not satisfied. Only a cad-lag version of this process can be constructed. This process can be thought of as the *trace* on the real line of trajectories of a two-dimensional Brownian motion, which moves by jumps.

For the deterministic process, we use an ordinary differential equation $[0,T]\times\mathbb{R}^d$, e.g. an initial valued problem

$$\dot{y}_{sx} = f(t, y_{sx}), \quad t > s, \quad y_{sx}(s) = x, \tag{4.24}$$

to define the transition function

$$p_d(s,x,t,A) = \mathbb{1}_A\big(y_{sx}(t)\big).$$

Certainly, the only deterministic model comparable with the previous examples is the case $y_{sx}(t) = (t-s)a + x$, for some constant a, which is homogeneous in time and space.

In this case we may use Kolmogorov construction with the continuity condition (4.4) satisfied. However, a direct approach show that the probability measure P_d constructed in the space $C([0,T],\mathbb{R}^d)$ is simply the Dirac measure with respect to $y_{sx}(t)$, i.e. for all $A \in \mathcal{B}(C)$

$$P_d(A) = \begin{cases} 1 & \text{if } y_{0x}(\cdot) \in A, \\ 0 & \text{otherwise,} \end{cases}$$

where the initial probability is equal to δ_x.

Note that the Poisson process has values in a countable set and it is a typical example of the so-called (e) Markov pure jump process. Its (time homogeneous) transition density function, denoted by $p(t,i,j)$, should satisfy the Chapman-Kolmogorov identity

$$p(t-s,i,j) = \sum_k p(r-s,i,k)\, p(t-r,k,j), \quad \forall t > r > s,\ i,j.$$

Hence, differentiate either in s or in t and let either $r-s$ or $t-r$ approaches 0 to deduce either the *backward* equation

$$\begin{aligned} \partial_s p(t-s,i,j) &= \sum_k \rho^+(s,i,k)\, p(t-s,k,j), \quad \forall t > s,\ i,j, \\ \rho^+(s,i,j) &= \lim_{r\to s} \partial_s p(r-s,i,j) \quad \forall s,\ i,j, \end{aligned} \tag{4.25}$$

or the *forward* equation

$$\begin{aligned} \partial_t p(t-s,i,j) &= \sum_k p(t-s,i,k)\, \rho^-(t,k,j), \quad \forall t > s,\ i,j, \\ \rho^-(t,i,j) &= \lim_{r\to t} \partial_t p(t-r,i,j) \quad \forall t,\ i,j, \end{aligned} \tag{4.26}$$

The quantities $\rho^+(s,i,j)$ and $\rho^-(s,i,j)$ are the characteristic of the process, and clearly

$$\begin{aligned} &\rho^-(s,i,j) = -\rho^+(t,i,j) := \rho(i,j), \quad \text{independent of } s,t, \\ &\rho(i,j) \geq 0, \quad \forall i \neq j, \qquad \rho(i,i) = -\sum_{j\neq i} \rho(i,j). \end{aligned} \tag{4.27}$$

As in the deterministic example, instead of prescribing the transition function we give an infinity array of real numbers $\{\rho(i,j) : i,j = 1,\dots\}$ satisfying the above equation (4.27). These functions represent the *infinitesimal rate* at which i will jumps to j at time s. We can construct the transition function by solving the forward equation (4.26) with an initial condition, which is a linear system of ordinary differential equations if the i,j ranges over a finite set. We have to solve for $p(t,i,j)$ the equation

$$\begin{aligned} \dot p(t,i,j) &= \sum_k p(t,i,k)\,\rho(k,j), \quad \forall t>0,\ i,j,\\ p(0,i,j) &= \delta_{i,j}, \end{aligned} \tag{4.28}$$

where the dot means derivative in t and $\delta_{i,j} = 1$ if $i = j$ and zero otherwise. It can be proved that this equation (4.28) admits a unique solution which provides a transition function $p(t,i,j)$ as above, see Lamperti [86, Section 6.2, pp. 114-117]. The cad-lag condition (4.5) is satisfied, so that by means of Kolmogorov construction a *realization* of the Markov pure jump process exists in $D([0,T],\mathbb{R})$. Actually, a discrete arguments will provide a realization in $\{1,2,\dots\}^{[0,T]}$. For instance we refer to Chung [26] for a exhaustive treatment.

A so-called (Gaussian) *white noise* is generally understood in engineering literature as a family of independent centered Gaussian random variables $\{\xi(t) : t \ge 0\}$, i.e., mean $\mathbb{E}\{\xi(t)\} = 0$ and a covariance $R(t-s) = \mathbb{E}\{\xi(t)\xi(s)\}$ with constant spectral density, so that

$$\mathbb{E}\{\exp[-\mathtt{i}\xi(t)]\} = \frac{1}{2\pi}, \quad \forall t \ge 0.$$

Such a process has a spectrum on which all frequencies participate with the same intensity, hence a *white* spectrum (in analogy with the *white light* in optics, which contains all frequencies of visible light, uniformly). However, such a process does not exist in the traditional sense because we should have $R(t-s) = \delta(t-s)$, where δ is the Dirac's delta *generalized* function. All this can be mathematical interpreted in the Schwartz distribution sense.

Definition 4.24 (generalized process)**.** Denote by $\mathcal{S}(]0,\infty[)$ the space of rapidly decreasing infinite differentiable functions and by $\mathcal{S}'(]0,\infty[)$ its dual space, the space of temperate distributions on $]0,\infty[$. In the same way that a stochastic process can be considered as a random variable with values in $\mathbb{R}^{[0,\infty]}$, a *generalized stochastic process* is random variable with values in the Schwartz distribution space $(\mathcal{S}'(]0,\infty[), \mathcal{B}(\mathcal{S}'))$. □

Note that $\mathcal{S}(]0,\infty[)$ and $\mathcal{S}'(]0,\infty[)$ are reflexive and Fréchet (locally convex, complete and metrizable) spaces, e.g., see Schwartz [118], but only with a partial order. In particular, the concept of *a process adapted to a filtration* and *stopping time* need to be reviewed with generalized processes. Clearly, we may use $\mathcal{S}(\mathbb{R}^d)$ and $\mathcal{S}'(\mathbb{R}^d)$ instead of $\mathcal{S}(]0,\infty[)$ and $\mathcal{S}'(]0,\infty[)$.

Thus, a process $\{x(t) : t \geq 0\}$ can be regarded as a generalized process Φ, where

$$\langle \Phi_x(\omega), \varphi \rangle := \int_0^\infty x(t,\omega)\, \varphi(t) \mathrm{d}t, \quad \forall \varphi \in \mathcal{S}(]0,\infty[),\ \omega \in \Omega,$$

but it is clear that we loss the complete order on the index set. Also, a generalized (stochastic) process Φ is said to be Gaussian if for arbitrary linearly independent functions $\varphi_1, \varphi_2, \ldots, \varphi_n$ the n-dimensional random variable $(\Phi(\varphi_1), \Phi(\varphi_2), \ldots,\ \Phi(\varphi_n))$ is normally distributed. As in the classical sense, a generalized Gaussian process is uniquely determined by its means and covariance. An important advantage of a generalized process is the fact that its derivative always exists and is itself a generalized stochastic process. In particular if we start with a Wiener process $\{w(t) : t \geq 0\}$ consider the generalized derivative of its associated generalized process Φ_w i.e.

$$\langle \dot{\Phi}_w(\omega), \varphi \rangle := -\langle \Phi_w(\omega), \dot{\varphi} \rangle \quad \forall \varphi \in \mathcal{S}(]0,\infty[),\ \omega \in \Omega.$$

we obtain a Gaussian white noise (generalized) process. Similarly, from a Poisson process $\{p(t) : t \geq 0\}$ we get a Poisson white noise (generalized) process as the derivative of the generalized process associated with $\{\bar{p}(t) := p(t) - \mathbb{E}\{p(t)\} : t \geq 0\}$. For instance we refer to Arnold [2, pp. 50–56], Kallianpur and Karandikar [73] and Pallu de la Barrière [104, Chapter 7]. A comprehensive analysis on Markov (and Gausssian) processes can be found in the recent book by Marcus and Rosen [94].

Chapter 5

Working with Stochastic Processes

Contrary to previous Chapter 4, the purpose of this chapter is to learn a little bit more on Markov processes (and a little 'neighborhood' of them) based on typical examples, but with a certain emphasis in the arguments used. This material could be adapted, perhaps, to a series of 'extensive lectures' on stochastic processes, with a clear focus on Markov processes.

For instance, the interested reader may want to check the book by Revuz and Yor [111, Chapters 0-III, pp. 1–118] to find a carefully written introduction with full details on various questions that are almost ignored in what follows. Moreover, there many excellent books on this subject, with various orientations, that the reader should search to find the most suitable one in each particular situation.

A general stochastic process with values in a Polish space (separable complete metric space) E, usually locally compact (e.g., closed subset of $\mathbb{R}^d$ or a one-point compactification of an open), on a probability space $(\Omega, \mathcal{F}, P)$ is a family of random variables (measurable mappings, almost surely defined), denoted by either $X = X(t)$ or $X = X_t$, from Ω into E indexed by a set T (usually a subset of real numbers). We suppress the qualifier general when a suitable regularity on the paths is imposed, i.e., at least separable and measurable. From the phenomenological viewpoint, a stochastic process is identified by prescribed properties given on the family of finite-dimensional distributions. A priori, a stochastic process can be viewed as a random variable X with values in the product topological space E^T, which is not a Polish space if E and T are uncountable. Mathematically, we add some regularity conditions (such as continuity) on the paths of the stochastic process ($X(\cdot, \omega)$, for each ω), so that we can consider a much smaller subset of E^T, with a structure of Polish space (non-locally compact), where the process actually lives. Usually, these Polish spaces are either $C([0, \infty), E)$, the space of E-valued continuous functions, or $D([0, \infty), E)$, the space of cad-lag functions as described in section 1.12. Ac-

tually, we use always the cad-lag regular form of a stochastic process, that is, given a stochastic process $\{Y(t) : t \in T\}$ with an uncountable (index) subset $T \subset [0,\infty)$, the *cad-lag regularisation* (usually this is also a version of $Y(t)$) is the stochastic process X given by $X(t,\omega) = \lim_{s\in\mathbb{Q}, s\to t} Y(t,\omega)$ and defined for any t in T, where the limits from the right and from the left (on the rational or dyadic-rational numbers $\mathbb{Q}$) exist finitely. Therefore, we say that the stochastic process Y is *regularisable* if the cad-lag regularisation is defined for any t in T. It can be proved that a process is regularisable if and only if the process and its (rational) up-crossings are locally bounded, moreover, the set of regularisable stochastic processes is measurable, see Rogers and Williams [112, Chapter 2, Section 5, pp. 163–166].

We are going to deal only with stochastic processes admitting a cad-lag version, unless otherwise stated. Thus for the initial (nominal or reference) stochastic process, we take a canonical realization, i.e., a complete probability space $(\Omega, \mathcal{F}, P)$ and a random variable X with values in the Polish space $D([0,\infty), E)$. Usually, we may want to set $\Omega = D([0,\infty), E)$, $\mathcal{F} = \mathcal{B}(\Omega)$, $X(t) = \omega(t)$, $X_-(t) = \omega(t-)$, and use the (complete) σ-algebras $\{\mathcal{F}(t),\, t \in T\}$ and $\{\mathcal{F}_-(t),\, t \in T\}$ generated by the family of E-valued random variables $\{X(s),\, s \le t,\, s \in T\}$ and $\{X_-(s),\, s \le t,\, s \in T\}$, respectively. Most of the times, $E \subset \mathbb{R}^d$ and P is the (unique) Wiener measure on $D([0,\infty), \mathbb{R}^d)$ with support in $C([0,\infty), \mathbb{R}^d)$. Other processes Y, Z, ... are regarded as random variables with values in the (non-locally compact) Polish space $D([0,\infty), E)$ or with values in $C([0,\infty), E)$ if possible. Sometimes, these processes are regarded also as a maps from $\Omega \times T$ into E. For instance, an increasing process Y will have also the property (on the paths) that $t \mapsto Y(t,\omega)$ is increasing, for every ω in $\Omega \setminus N$, with $P(N) = 0$. Thus, unless otherwise stated, a *E-valued stochastic process* is a measurable function from a complete probability space $(\Omega, \mathcal{F}, P)$ into the canonical space $D([0,\infty), E)$, i.e., a $D([0,\infty), E)$-valued random variable, which imposes the cad-lag regularity on paths. In any case, it is implicitly assumed that we have choose a probability space $(\Omega, \mathcal{F}, P)$ where regular conditional probability exists.

Our main interest is (strong) Markov processes admitting a cad-lag realization and having a (continuous) transition function. These stochastic processes are identified by (1) a prescribed (continuous) transition function, or (2) a given (continuous) Markov-Feller semigroup, which in turn can be obtained by its infinitesimal generator. With the previous introduction, the purpose of this chapter is to discuss (in some detail) certain topics in *stochastic analysis* as a tool to describe (or generate) Markov processes. Certainly, it is implicity assume a minimum understanding of probability, e.g., see Bremaud [21, Appendix A1, pp. 255–295].

5.1 Discrete Time

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 complete probability space $(\Omega, \mathcal{F}, P)$ and an increasing sequence (so-called *filtration*) of sub σ-algebras $(\mathcal{F}_n : n = 0, 1, \ldots)$, $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n$, for all $n = 1, 2, \ldots$, such that $\mathcal{F}_0$ contain all null sets. 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 $\mathcal{F}_\infty = \mathcal{F}$. 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 $n \geq 0$.

Similar to Kolmogorov's extension theorem (see Theorem 4.20) we can mention the following construction of the direct product of probability spaces (e.g., Halmos [57, 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 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. Sometimes, cylindrical sets (or cylinder sets) are defined as $C_n = F_n \times \prod_{k=n+1}^\infty \Omega_k$, with F_n in the (finite) product σ-algebra $\prod_{k=1}^n \mathcal{F}_k$. Even if the representation is not unique (e.g., replace F_n with $F_n \times \Omega_{n+1}$ and begin the infinite product at $n+2$ instead of $n+1$), under this later definition the class of all cylindrical sets is an algebra of sets in Ω. Indeed, this follows from the properties (a) the complement of cylindrical set corresponding to n, F_n is the cylindrical set corresponding to n, F_n^c, with F_n^c being the complement of F_n in $\prod_{k=1}^n \Omega_k$ and (b) the union of two cylindrical sets (corresponding to n, F_n and to n, F_n') is the cylindrical set corresponding to $n, F_n \cup F_n'$.

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

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

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

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 $\mathbb{R}^d$.

Thus, 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.

The reader is also referred to Tulcea's theorem (e.g., Neveu [102, Section V.1, pp. 153–159], Shiryayev [121, Section II.9, Theorem 2, pp. 243–250]), 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

$$C_n = \prod_{i=1}^n F_i \times \prod_{i=n+1}^\infty \Omega_i, \quad \text{with} \quad F_i \in \mathcal{F}_i, \ \forall i, \quad n = 1, 2, \dots \tag{5.1}$$

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 (5.1). 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, we assume 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:

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

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

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

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

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 (5.1) 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 $\cup_n\mathcal{F}^n$.

Theorem 5.1 (Tulcea)**.** *Under the above notation, the function $P_n(\omega,F)=P(\omega_1,\dots,\omega_n;F)$, with $\omega=(\omega_1,\dots,\omega_n,\dots)$, 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 $\mathcal{F}^n$.* □

Proof. Only a brief idea is given. The central point is show the σ-additivity of P_n on the algebra $\cup_n\mathcal{F}^n$ 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 $\cup_n\mathcal{F}^n$ such that $\cap_k A_k=\emptyset$ 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. □

First let us discuss Kolmogorov's extension theorem (see Theorem 4.20) in a general product space $\Omega=\prod_{t\in T}\Omega_t$, $\mathcal{F}=\prod_{t\in T}\mathcal{F}_t$. We assume given a family of finite-dimensional distributions $\{P_s : s\in T^n,\ n=1,2,\dots\}$ on $\Omega^s=\Omega_{t_1}\times\dots\times\Omega_{t_n}$, with $s=(t_1,\dots,t_n)$ which satisfies the consistency conditions, namely

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

(b) if $t=(t_1,\dots,t_n,t_{n+1})$ and $s=(t_1,\dots,t_n)$ and B in $\mathcal{F}^s=\mathcal{F}_{t_1}\times\dots\times\mathcal{F}_{t_n}$ then $P_t(B\times\Omega_{t_{n+1}})=P_s(B)$.

If a total order is given on the index set T, it is enough to have the finite-dimensional distributions defined only for $(s_1, s_2, \ldots, s_n)$ such that $s_1 < s_2 < \cdots < s_n$ and to satisfy only a consistency conditions of the form

(b') if $t = (t_1, \ldots, t_n)$ and $s = (s_1, \ldots, s_m)$ with $t_1 < \cdots < t_n < r < s_1 < \cdots < s_m$ and $A \times B$ in $\mathcal{F}^t \times \mathcal{F}^s$ then $P_{(t,r,s)}(A \times \Omega_r \times B) = P_{(t,s)}(A \times B)$, for any $n, m = 0, 1, \ldots$.

Consistency along is not sufficient to ensure the existence of a probability P defined on $(\Omega, \mathcal{F})$ such that P_s be the restriction (or image trough the projection) of P over $(\Omega^s, \mathcal{F}^s)$. Some sort of topology is necessary on Ω_t so that P_s results inner regular (e.g., see Doob [39, pp. 403, 777], Neveu [102, Section III.3, pp. 74–81]), for instance, if Ω_t is a Lusin space (i.e., Ω_t is homeomorphic to a Borel subset of a compact metrizable space) and $\mathcal{F}_t = \mathcal{B}(\Omega_t)$ its Borel σ-algebra then every probability measure is inner regular. Under these conditions, the construction of the measure P is possible.

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 $s = (t_1, \ldots, t_n)$, has the form

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

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.

A very useful and well know result is the following

Lemma 5.2 (Borel-Cantelli)**.** *Let $(F_1, F_2, \ldots)$ a sequence of measurable sets in a probability space $\Omega, \mathcal{F}, P$). (1) If $\sum_n P(F_n) < \infty$ then $P(\cap_n \cup_{k \geq n} F_k) = 0$. (2) If $(F_1, F_2, \ldots)$ is independent and $\sum_n P(F_n) = \infty$ then $P(\cap_n \cup_{k \geq n} F_k) = 1$.* □

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.

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

Definition 5.3 (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

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

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 [78].

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 $0 \leq s_1 < t_1 < \cdots < s_k < t_k \leq N$, $X_{s_i} < a$ and $X_{t_i} > b$, for any $i = 1, 2, \ldots k$. Then for any super-martingale the estimate

$$(b - a)\, \mathbb{E}\{U_N(X, [a, b])\} \leq \mathbb{E}\{(X_N - a)^-\} \tag{5.2}$$

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 $n \geq 2$,

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

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

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

for each ω, the estimate (5.2) follows.

The Doob's super-martingale convergence states that for a super martingale $(X_n : n = 0, 1, \ldots)$ 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 $(X_n : n = 0, 1, \ldots)$ is uniformly integrable, and in this case we have $\mathbb{E}\{X_\infty \,|\, \mathcal{F}_n\} \leq X_n$, 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 (5.2) 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 $(X_n : n = 0, 1, \ldots)$ is a nonnegative sub-martingale bounded in L^p then Doob's L^p inequality reads as follows

$$\|\sup_n X_n\|_p \leq p' \sup_n \|X_n\|_p, \qquad \text{with } \; 1/p + 1/p' = 1, \tag{5.3}$$

where $\|\cdot\|_p$ denotes the in $L^p = L^p(\Omega, \mathcal{F}, P)$. However, $p' = \infty$ for $p = 1$, this inequality becomes

$$\|\sup_n X_n\|_1 \leq \frac{\mathrm{e}}{\mathrm{e}-1} \sup_n \|X_n \ln^+ X_n\|_1, \tag{5.4}$$

where $\ln^+(\cdot)$ is the positive part of $\ln(\cdot)$. Note that $(p')^p \leq 4$ for every $p \geq 2$. Indeed, if the set Ω_n^r of all ω where $\sup_{k\leq n} X_k \geq r$ 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 $\Omega^r_{n,0} = \{X_0 \geq r\}$, then we have $X_k \geq r$ on $\Omega^r_{n,k}$, 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 (5.3) for sub-martingale bounded in L^p, with $p > 1$, first check the claim that for any two nonnegative random variables x and y,

$$\text{if} \quad r\,P(y \geq r) \leq \mathbb{E}\{x \mathbb{1}_{y \geq r}\} \quad \text{then} \quad \mathbb{E}\{y^p\} \leq (p')^p \mathbb{E}\{x^p\}. \tag{5.5}$$

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 (5.5). Next, choose $y = \sup_n X_n$ and $x = X_n$ to conclude.

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

$$X_n = X_0 + M_n + A_n, \ \forall n, \quad \text{and} \quad M_0 = A_0 = 0. \tag{5.6}$$

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

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

and $(M_n : n = 1, \dots)$ with $M_n = X_n - X_0 - A_n$ to obtain the decomposition (5.6). 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 $V = (V_n : n = 1, 2, \ldots)$,

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

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 (5.3), $p = 2$, 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

$$c_p\, \mathbb{E}\{([M]_n)^{p/2}\} \leq \mathbb{E}\{\sup_{k \leq n} |M_k|^p\} \leq C_p\, \mathbb{E}\{([M]_n)^{p/2}\}, \tag{5.7}$$

valid for any $n \geq 1$ and $p > 0$ and some constants $C_p > c_p > 0$ independent of the martingale $(M_n : n = 0, 1, \ldots)$. Even for $p = 1$, we may use $C_1 = 3$ in the right-hand side of (5.7). Moreover, the L^2-martingale $(M_n : n = 0, 1, \ldots)$ may be only a *local martingale* (i.e., there exists a sequence of stopping times $\eta = (\eta_k : k = 0, 1, \ldots)$ such that $M^{\eta,k} = (M_n^{\eta,k} : n = 0, 1, \ldots)$, 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 $X = (X_n : n = 0, 1, \ldots)$ be a sub-martingale with respect to $(\mathcal{F}_n : n = 0, 1, \ldots)$ 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 $A = (A_n : n = 0, 1, \ldots)$ and $M = (M_n : n = 0, 1, \ldots)$ the predictable and martingale sequences given in the decomposition (5.6), $X_n = X_0 + M_n + A_n$, 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

$$\mathbb{E}\{M_\theta \,|\, \mathcal{F}_\eta\} = M_\eta, \quad \text{a.s.} \quad \text{and} \quad \mathbb{E}\{X_\theta \,|\, \mathcal{F}_\eta\} \geq X_\eta, \quad \text{a.s.} \tag{5.8}$$

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 [21], Chung [26], Dellacherie and Meyer [36, Chapters I–IV], Doob [38, 40], Karlin and Taylor [76, 77], Nelson [101], Neveu [103], Williams [133], among others.

5.2 Filtered Spaces

We have seen that the study of stochastic processes leads to the discussion on probability measures in separable complete metric spaces, i.e., in Polish spaces (recall that a countably product of Polish spaces is a Polish space with the product topology and that any Borel set of a Polish space is a continuous image of the product Polish space $\mathbb{N}^\infty$, sequences of nonnegative integer numbers). Natural models for stochastic dynamical systems are the Markov processes. Since past, present and future information are represented by sets of events, a systematic analysis of families of σ-algebras is necessary. Thus, the starting point is a (complete) probability space with either a filtration or a process with its natural filtration.

Remark that all properties concerning a Filtered space are not preserved accross any possible version of a given process, e.g., if X and Y are versions of the same process (namely, $P\{X(t) \neq Y(t)\} = 0$ for every $t \geq 0$) then X may be adapted to the filtration but not Y. In this way, for a given process X we can take any version, but as soon as a filtration is involved, the given version kept. However, changing the processes in an evanescent set is fine, since the filtration is complete.

Definition 5.4. Given a *filtered space* $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$, this is a (usually complete) probability space $(\Omega, \mathcal{F}, P)$ and a (standard) filtration $\mathbb{F} = (\mathcal{F}(t) : t \geq 0)$, i.e., $\mathcal{F}(0)$ contains all P-negligible (or P-null) sets in $\mathcal{F}$ (complete) and $\mathcal{F}(t) = \bigcap_{\varepsilon>0} \mathcal{F}(t+\varepsilon)$ (right-continuous). Sometimes, this is also called *stochastic basis*, see Jacod and Shiryaev [69, Chapter 1].
(a) The σ-algebra $\mathcal{M}$ 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 $\mathcal{F}(t) \times \mathcal{B}([0,t])$ for every $t \geq 0$.
(b) The σ-algebra $\mathcal{O}$ of *optional* or *well-measurable* sets is generated by sets of the form $F_0 \times \{0\}$ and $F \times [s,t)$, where $F_0 \in \mathcal{F}(0)$ and $F \in \mathcal{F}(s)$ for $s < t$ in $(0,\infty)$.
(c) The σ-algebra $\mathcal{P}$ of *predictable* sets is generated by sets of the form $F_0 \times \{0\}$ and $F \times (s,t]$, where $F_0 \in \mathcal{F}(0)$ and $F \in \mathcal{F}(s)$ for $s < t$ in $(0,\infty)$. □

Note that sometimes the variables t and ω may be exchanged so that the σ-algebras $\mathcal{M}$, $\mathcal{O}$ and $\mathcal{P}$ are regarded as defined on $[0,\infty)\times\Omega$ instead of $\Omega\times[0,\infty)$. As long as no confusion arrives, we will ignore this fact.

It may be convenient to use the notation $\mathbb{F} = (\mathcal{F}(t) : t \geq 0)$ for a filtration and assume that $\mathcal{F}$ is the minimal σ-algebra containing all sets belonging to $\mathcal{F}(t)$ for some $t \geq 0$, so that $(\Omega, \mathbb{F}, P)$ denotes a stochastic basis. If a given

filtration $(\mathcal{F}_0(t) : t \geq 0)$ does not satisfy the usual conditions of completeness and right-continuity then its *usual augmentation* $(\mathcal{F}(t) : t \geq 0)$ is defined as the minimal filtration satisfying $\mathcal{F}_0(t) \subset \mathcal{F}(t)$ for all t, plus the usual conditions. The σ-algebra $\mathcal{F}(t)$ can be constructed in two steps, first $\mathcal{F}_0(t)$ is completed with all null sets to a new σ-algebra $\bar{\mathcal{F}}_0(t)$ and second $\bar{\mathcal{F}}_0(t)$ is made right-continuous by defining $\mathcal{F}(t) = \bigcap_{s>t} \bar{\mathcal{F}}_0(s)$. Thus, for any F in $\mathcal{F}(t)$ there exists a F_0 in $\mathcal{F}_0(t+) = \bigcap_{s>t} \mathcal{F}_0(s)$ such that $F \Delta F_0 = (F \setminus F_0) \cup (F_0 \setminus F)$ is a null set. Completing the family of increasing σ-algebras to become a filtration (satisfying the usual condition) is a routing task, however this is an important issue when dealing with the *strong Markov property* as discussed later on. The passage from $\mathcal{F}_0(t)$ to $\mathcal{F}_0(t+)$ is a very technical matter that we have to deal when the realization or simulation of a processes is studied. For instance, it can be easily proved that the completion of history of a Lévy process (or of a *right-constant* process) is actually right-continuous, e.g., see Bremaud [21, Appendix A2.3, pp. 303–311], Davis [34, Appendix A2, Theorem A2.2, pp. 259–261], Protter [108, Section 1.4, Theorem 31, pp. 22–23].

Note that the three σ-algebras defined in term of the filtration $(\mathcal{F}(t) : t \geq 0)$ are all on $\Omega \times [0, \infty)$ and not on Ω alone. We have $\mathcal{P} \subset \mathcal{O} \subset \mathcal{M}$ and it can be proved that the predictable class $\mathcal{P}$ (optional class $\mathcal{O}$, resp.) is the minimal σ-algebra for which adapted left-continuous (right-continuous, resp.) processes are measurable as function from $\Omega \times [0, \infty)$ into the state space (e.g., $E \subset \mathbb{R}^d$).

Recall that a random variable τ with values in $[0, \infty]$ is called a *stopping time* (or optional time) if sets of the form $\{\omega : \tau(\omega) \leq t\}$ are measurable with respect to $\mathcal{F}(t)$ for every $t \geq 0$. This is equivalent to imposing that the stochastic interval $[\![\tau, \infty[\![$ is optional, see Definition 4.7. Thus, $\mathcal{P}$ ($\mathcal{O}$, resp.) are generated by stochastic interval of the form $[\![0, \tau]\!]$ ($[\![0, \tau[\![$, resp.) where τ is any stopping time. Filtration satisfying the usual condition (right-continuity and completeness) are necessary to be able to identify a stopping time with it equivalent class, as explained below.

Assume that a right-continuous filtration $\mathbb{F} = \{F(t) : t \geq 0\}$ is given. If $\mathcal{O}$ is an open set of $\mathbb{R}^d$ and X is a cad-lag $\mathbb{F}$-adapted process with values in $\mathbb{R}^d$, then the *hitting time* $\tau_{\mathcal{O}}$ of an open set $\mathcal{O}$ is a stopping time, where

$$\tau_{\mathcal{O}} = \inf \big\{ t > 0 \, : \, X(t) \in \mathcal{O} \big\},$$

and $\tau_{\mathcal{O}} = +\infty$ if $X(t) \in \mathbb{R}^d \setminus \mathcal{O}$ for every $t \geq 0$. Indeed the relation $\tau_{\mathcal{O}}(\omega) < t$ if and only if $X(s, \omega) \in \mathcal{O}$ for some rational number in $[0, t)$ shows that the event $\{\tau_{\mathcal{O}} < t\}$ belongs to $\mathcal{F}(t)$ and so $\{\tau_{\mathcal{O}} \leq t\}$ is in $\mathcal{F}(t+)$. Similarly, if C is a closed set of $\mathbb{R}^d$ then the *contact time* $\tilde{\tau}_C$ of a closed set C is also an stopping time, where

$$\tilde{\tau}_C = \inf \big\{ t \geq 0 \, : \, \text{either } X(t) \in C \text{ or } X(t-) \in C \big\},$$

with $X(0-) = X(0)$. Indeed, use the fact that $\tilde{\tau}_C(\omega) \leq t$ if and only if the infimum over all rational numbers s in $[0, t]$ of the distance from $X(s, \omega)$ to C is zero. However, if K is a compact set of $\mathbb{R}^d$ then to check that the *entry time*

$\bar{\tau}_K$ of a compact set K is also an stopping time, where

$$\bar{\tau}_K = \inf\big\{t \geq 0 : X(t) \in K\big\},$$

is far more delicate. The argument uses ordinal numbers and involves the assumption of P-completion $\bar{\mathbb{F}}$ for the filtration $\mathbb{F}$, i.e., $\bar{\mathbb{F}}$ is the minimal right-continuous filtration such that $\mathcal{F}(0)$ contains all P-nulls sets in $\bar{\mathcal{F}}$, the P-completion of $\mathcal{F}$. In this case, for any $\bar{\mathbb{F}}$-stopping time T there exists a $\mathbb{F}$-stopping time S such that $P\{T = S\} = 1$ and $\bar{F}(T)$ is the smaller σ-algebra containing $\mathcal{F}(S+)$ all P-null sets. Furthermore, by means of the so-called *analytic sets* (i.e., continuous or Borel images of Borel sets in a Polish space), a deeper result shows that the hitting time of any Borel set is indeed a stopping time, e.g., see Doob [39, pp. 419–423] or Rogers and Williams [112, Sections II.73–76].

A process X with values in $E \subset \mathbb{R}^d$ is progressively measurable (resp. optional or predictable) if the map $(t,\omega) \mapsto X(t,\omega)$ or equivalently $(t,\omega) \mapsto \big(t, X(t,\omega)\big)$ is measurable with respect to $\mathcal{M}$ (resp. $\mathcal{O}$ or $\mathcal{P}$). In particular, if $(t,x) \mapsto a(t,x)$ is a Borel function and X is progressively measurable (respectively, optional or predictable) then so is the map $(t,\omega) \mapsto a\big(t, X(t,\omega)\big)$. On the other hand, a stopping time τ is called *predictable* if there exists an *announcing sequence* of stopping times $\{\tau_n : n = 1, 2, \dots\}$, i.e., τ_n increases to τ and $P(\tau_n < \tau \text{ if } \tau > 0) = 1$; sometimes the condition $P(\tau > 0) = 1$ is also requested. It is not hard to show that τ is a predictable time if and only if the stochastic interval $[\![\tau, \infty[\![$ is predictable. Note that $\tau + \varepsilon$ is a predictable (stopping) time for any stopping time τ and any constant $\varepsilon > 0$. Moreover, if τ and θ are predictable times then all stochastic intervals that have τ, θ, 0 or ∞ as endpoints are predictable sets. There are many interesting measurability question on these points, e.g., see Bichteler [14, Section 3.5]

Based on the alternative way of generating the σ-algebras $\mathcal{O}$ and $\mathcal{P}$ (as mentioned above), we deduce that a right-continuous (resp. left-continuous) progressively measurable process is optional (resp. predictable). When working with cad-lag (continuous) processes, the difference between the progressively measurable and optional (predictable) processes have almost no importance. Recall that given a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ a stochastic process X is called *adapted* if the random variable $\omega \mapsto X(t,\omega)$ is $\mathcal{F}(t)$-measurable for any $t \geq 0$. Thus, any adapted cad-lag process is progressively measurable.

The concept of a *predictable* (also called *previsible*) or *optional* process implies that of *adapted process* (to a given filtration) in a way suggested by the name. Denote by $\mathcal{P}_{\mathcal{R}}$ the family of subsets of $\Omega \times [0,\infty)$ containing all sets of the form $F_0 \times \{0\}$ and $F \times (s,t]$, where $F_0 \in \mathcal{F}(0)$ and $F \in \mathcal{F}(s)$ for $s < t$ in $(0,\infty)$, is called the *class of predictable rectangles*. Sometimes, the sets $F_0 \times \{0\}$ need special consideration and some authors prefer to remove these type of sets from the definition of the σ-algebra $\mathcal{P}$. As it was defined, the σ-algebra $\mathcal{P}$ of subsets of $[0,\infty) \times \Omega$ generated by all predictable rectangles is called the *predictable* σ-algebra associated with the filtration $\{\mathcal{F}(t) : t \geq 0\}$. Another equivalent way of generating the predictable σ-algebra is to define $\mathcal{P}$ as generated by the stochastic intervals of the form $[\![0, \tau]\!]$, for stopping times τ with respect to the given filtration $\{\mathcal{F}(t) : t \geq 0\}$.

Sometimes, the filtration $\{\mathcal{F}(t) : t \geq 0\}$ is *quasi-left continuous*, namely, $\mathcal{F}(\tau-) = \mathcal{F}(\tau)$ for any *predictable* stopping time τ. This is the case of a (continuous) Markov-Feller process (see Rogers and Williams [112, Chapter 6, Theorem 18.2, pp. 346–347]).

It should be clear by now that filtered spaces are a fundamental feature of the theory of stochastic processes and definitions of our central object, Markov processes, will involve a filtration. Heuristically speaking, the σ-algebra $\mathcal{F}(t)$ is the collection of event that may occur before or at the time t (i.e., the set of possible *pasts* up to time t). In what follows, unless otherwise stated, we assume that filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ is given, sometime denoted by $(\Omega, \mathbb{F}, P)$ and also called stochastic basis, with $\mathbb{F} = (\mathcal{F}(t) : t \geq 0)$, $\mathcal{F} = \sigma\{\mathcal{F}(t) : t \geq 0\}$, and the three σ-algebras $\mathcal{P} \subset \mathcal{O} \subset \mathcal{M}$, called predictable, optional and progressively measurable, respectively, as in Definition 5.4. Moreover, if Ω is also a Polish space then $\mathcal{F}$ contains the Borel σ-algebra $\mathcal{B}(\Omega)$, actually, $\mathcal{F} = \mathcal{B}(\Omega)$ in most of the cases.

On a given measurable space $(\Omega, \mathcal{F})$ we may define its *universal completion* $\mathcal{F}^0 = \bigcap_P \mathcal{F}^P$, where $\mathcal{F}^P$ is the completion of $\mathcal{F}$ relative to P, and the intersection is over all probability measures P on $\mathcal{F}$. This is to say that $A \in \mathcal{F}^0$ if and only if for every P there exist B and N in $\mathcal{F}$ such that $B \setminus N \subset A \subset B \cup N$ and $P(N) = 0$ (since B and N may depend on P, clearly this does not necessarily imply that $P(N) = 0$ for every P). Thus, a universally complete measurable space satisfies $\mathcal{F} = \mathcal{F}^0$. The concept of universally measurable is particularly interesting when dealing with measures in a Polish space Ω, where $\mathcal{F} = \mathcal{B}(\Omega)$ is its Borel σ-algebra, and then any subset of Ω belonging to $\mathcal{F}^0$ is called universally measurable. In this context, it is clear that a Borel set is universally measurable. Also, it can be proved that any analytic set is universally measurable, and on any uncountable Polish space there exists a analytic set (with not analytic complement) which is not a Borel set, e.g., see Dudley [41, Section 13.2].

A detailed discussion on the strong Markov property involves a *measured filtration* $(\Omega, \mathbb{F}, \mathbb{P})$, i.e., besides the filtration $(\Omega, \mathbb{F})$, we have a class of probability measures $\mathbb{P}$ on $(\Omega, \mathcal{F})$, with $\mathcal{F} = \mathcal{F}(\infty)$. Then, the $\mathbb{P}$-universal completion $\mathbb{F}^0$ (i.e., the family $\mathbb{P}$ is implicitly understood in the notation) of the filtration $\mathbb{F}$ is defined by adding all null sets, i.e.,

$$\mathcal{F}^0(t) = \bigcap_{\varepsilon>0} \bigcap_{P\in\mathbb{P}} \sigma\big(\mathcal{F}(t+\varepsilon), \mathcal{N}^P\big), \quad \forall t \geq 0,$$

where $\mathcal{N}^P$ denotes the σ-algebra of $(P, \mathcal{F})$-null sets (i.e., all the subsets of some set N in $\mathcal{F}$ with $P(N) = 0$). Hence the new filtration $\mathbb{F}^0$ is right-continuous and universally complete (i.e., the universal completion of $\mathcal{F}^0(t)$ is again $\mathcal{F}^0(t)$, for every t) by construction but not necessarily complete with respect to a particular probability P in the class $\mathbb{P}$. Furthermore, sometimes not all $(P, \mathcal{F})$-null sets are necessary, and completion arguments are reviewed. Countable unions of sets N with $P(N) = 0$ and the property of being $\mathcal{F}(t)$-measurable (for some finite t) are called P-*nearly empty* sets. Thus, a set N in $\mathcal{F} = \mathcal{F}(\infty)$ with $P(N) = 0$

which is not in $\mathcal{F}(t)$, for every t finite, may not be nearly empty. Then, a measured filtration is called $\mathbb{P}$-*regular* if $\mathcal{F}(t) = \mathcal{F}^{\mathbb{P}}(t)$ for every $t \geq 0$, where $\mathcal{F}^{\mathbb{P}}(t) = \bigcap_{P\in\mathbb{P}} \mathcal{F}^P(t)$, with $\mathcal{F}^P(t)$ the σ-algebra of all subsets A of Ω such that for every P in the class $\mathbb{P}$, the symmetric difference $(A \setminus A_P) \cup (A_P \setminus A)$ is P-nearly empty for some A_P in $\mathcal{F}(t)$. Note that $\mathcal{F}^P(t)$ contains the completion of $\mathcal{F}(t)$ relative to the restriction of P to $\mathcal{F}(t)$ (so it is universally complete), but it could be smaller than $\sigma\big(\mathcal{F}(t), \mathcal{N}^P\big)$. Moreover, filtration $\{\mathcal{F}^{\mathbb{P}}(t+) : t \geq 0\}$ is also $\mathbb{P}$-regular, and called the $\mathbb{P}$-*natural enlargement* of $\mathbb{F}$. Essentially, as long as we work with a right-continuous regular filtration, the technical points about measurability are resolved, this is usually referred to as the natural conditions. For instance, see the book Bichteler [14, Section 1.3] for a comprehensive study.

When a process is viewed as a function of two-variables, $(t, \omega) \mapsto X(t, \omega)$, properties like continuity or monotonicity refers to the path (i.e., to the function $t \mapsto X(t, \omega)$ for a fixed ω, which are global property on the variable t only), while properties like integrable or bounded may refer to either one of the variables or even to both variables. However, the qualifier integrable usually refers to the random variable $\omega \mapsto X(t, \omega)$, for a fixed $t \geq 0$. As discussed later, when a filtration is given, the term *locally* will apply to any property of a process involving both variables with a precise meaning, e.g., a process X is locally bounded if there is a (increasing) sequence $(\tau_n : n \geq 1)$ of stopping times satisfying $P(\tau_n \uparrow \infty) = 1$ such that the stopped process $X_n(t) = X(t \wedge \tau_n)$ is bounded. In all statements and procedures, processes are considered equals if their paths differ in a set of measure zero (i.e., they are indistinguishable one of each other), but sometimes we may select a particular element of its equivalent class to perform a specific construction. All these terminologies become clear from the context of the discussion.

5.3 Bounded Variation

Let us consider real-valued (or $\mathbb{R}^d$-valued) processes $(X(t) : t \geq 0)$ in a probability space we $(\Omega, \mathcal{F}, P)$. If a filtration $\mathbb{F} = (\mathcal{F}(t) : t \geq 0)$ is also given then the term *adapted* (to $\mathbb{F}$) is implicitly assumed (although sometimes is explicitly mentioned) and the qualifier *raw* is to be used to explicitly assume not necessarily adapted.

5.3.1 Pathwise Properties

An *(monotone) increasing* [*(monotone) decreasing*, resp.] stochastic process $(X(t) : t \geq 0)$ is such that the function $t \mapsto X(t, \omega)$ is increasing [decreasing, resp.] for every ω, except perhaps in a null set. Because an increasing function has left and right-hand limits at each points, it is convenient to normalize the process to be right-continuous. Thus an increasing process is a random variable X (almost surely defined) with values in the sample space $D([0, \infty), \mathbb{R})$ such that $X(t) \geq X(s)$ for every $t \geq s$. Also, vector valued process (i.e., in $D([0, \infty), \mathbb{R}^d)$) can be considered.

Similarly, a stochastic process $X = (X(t) : t \geq 0)$ is said to be of *locally bounded variation* in $[0, \infty)$ (or simplifying, of *finite variation*) if it is a random variable X with values in the sample space $D([0, \infty), \mathbb{R})$ and its variation process $\{\text{var}(X, [0, t]) : t \geq 0\}$ is finite,

$$\text{var}(X, [0, t]) = \sup\{\sum_{i=1}^{n} |X(t_i) - X(t_{i-1})| \, : \, 0 = t_0 < t_1 < \cdots < t_n = t\},$$

where the supremum is taken over all partitions of the interval $[0, t]$, var is referred to as the *variation operator*. Clearly, as long as the process is cad-lag, we may only consider some countable family of partitions, e.g., partitions with $t_i = i2^{-n}$ for $i = 0, 1, \ldots, 2^n$. It can be defined the positive $\{\text{var}^+(X, [0, t]) : t \geq 0\}$ and the negative $\{\text{var}^-(X, [0, t]) : t \geq 0\}$ variation processes exchanging the absolute value $|\cdot|$ with the positive part $[\cdot]^+$ and the negative part $[\cdot]^-$ of a real number in the above definition. Note that because X is cad-lag, the supremum can be taken over partitions with end points $t_1 < \cdots < t_{n-1}$ in a countable dense set so that the functions $X \mapsto \text{var}(X, \cdot)$, $X \mapsto \text{var}^+(X, \cdot)$ and $X \mapsto \text{var}^-(X, \cdot)$ are measurable from $D([0, \infty), \mathbb{R})$ or $C([0, \infty), \mathbb{R})$ into itself.

It can be checked that, e.g., Gordon [55, Chapters 4 and 6],

$$\begin{aligned} &\text{var}(X, [0, t]) = \text{var}^+(X, [0, t]) + \text{var}^-(X, [0, t]) \quad \text{and} \\ &X_t - X_0 = \text{var}^+(X, [0, t]) - \text{var}^-(X, [0, t]) \quad \forall t \in [0, \infty), \end{aligned}$$

and that the three variation processes

$$\{\text{var}(X, [0, t]) : t \geq 0\}, \ \{\text{var}^+(X, [0, t]) : t \geq 0\}, \ \{\text{var}^-(X, [0, t]) : t \geq 0\}$$

are (monotone) increasing (and cad-lag); and they are adapted, optional or predictable if the initial process X is so. Thus we can look at a locally bounded variation process X as two random variables $\text{var}^+(X, [0, t])$ and $\text{var}^-(X, [0, t])$ with values in the sample space $D = D([0, \infty), \mathbb{R})$, i.e., a probability measure P on D with the Borel σ-algebra $\mathcal{B}(D)$ and two increasing and measurable maps $\text{var}^+(X, [0, t])$ and $\text{var}^-(X, [0, t])$ from D into itself. Note that $\text{var}^+(X, [0, t])$ and $\text{var}^-(X, [0, t])$ are minimal in the sense that if X is of bounded variation and $X = Y - Z$ with each Y and Z monotone increasing then $\text{var}^+(X, [0, t]) - \text{var}^+(X, [0, s]) \leq Y_t - Y_s$ and $\text{var}^-(X, [0, t]) - \text{var}^-(X, [0, s]) \leq Z_t - Z_s$, for every $t \geq s$. This is the so-called Jordan decomposition. On the other hand, given a (cad-lag) bounded variation process X, its *continuous part* is defined as $X^c(t) = X(t) - X^{jp}(t)$, where the *jump part* is defined by $X^{jp}(t) = \sum_{0<s\leq t} \delta X(s)$, with $\delta X(s) = X(s) - X(s-)$. It is clear that, by rearranging the jumps, we can rewrite the jumps part as $X^{jp}(t) = \sum_n [X(\tau_n) - X(\tau_n-)] \mathbb{1}_{\tau_n \leq t}$, where the series converges absolutely for any t and the random times τ_n are stopping times if the process X^{jp} is adapted, see Sato [116, Lemma 21.8, Chapter 4, pp. 138–140]. Moreover, since the continuous part $X^c(t)$ is still of bounded variation, it is differentiable almost everywhere and we have $X^c(t) = X^{ac}(t) + X^{sc}(t)$, where $X^{ac}(t) = \int_0^t \dot{X}^c(s) \mathrm{d}s$ is called the *absolutely continuous part* and $X^{sc}(t)$

is the *singular continuous part*. Thus, any bounded variation process X can be written as a unique sum $X^{ac} + X^{sc} + X^{jp}$ called Lebesgue decomposition.

On the other hand, for any cad-lag process X and any $\varepsilon > 0$ we can define the ε-jumps process as

$$X^{\varepsilon j}(t) = \sum_{0<s\le t} \delta X(t) \mathbb{1}_{|\delta X(t)|\ge\varepsilon}, \quad \forall t > 0,$$

and the ε-almost continuous process $X^{\varepsilon c}(t) = X(t) - X^{\varepsilon j}(t)$. However, the continuous part, i.e., $\lim_{\varepsilon\to 0} X^{\varepsilon c}(t)$ may not be defined in general. Certainly, this would be $X^c(t)$ when X has locally bounded variation. However, the above limit will make sense in the L^2 topology for *square integrable local martingales* and defined for every process which is a local martingale, as discussed later.

A monotone increasing or a locally bounded variation process X induces a Borel (positive or signed) measure on $[0, \infty)$ by setting

$$\mu(\{0\}) = X(0, \omega) \quad \text{and} \quad \mu(]a, b]) = X(b, \omega) - X(a, \omega), \quad 0 < a < b,$$

for each sample path $X(\cdot, \omega)$, which is referred to as the Lebesgue-Stieltjes measure. Conversely, if a (Borel) locally finite signed measure μ on $[0, \infty)$ is defined for ω which is *weakly measurable* i.e. $\omega \mapsto \mu(A, \omega)$ is measurable for each Borel subset A of $[0, \infty)$ then we can define process of bounded variation as $X(t, \omega) = \mu([0, t], \omega)$, for any $t \ge 0$. Based on Fubini's theorem, it can be proved that given two processes X and Y of locally bounded variation (cad-lag) we have the integration-by-part formula

$$X(b)Y(b) - X(a)Y(a) = \int_{(a,b]} X(t-)\mathrm{d}Y(t) + \\ + \int_{(a,b]} Y(t-)\mathrm{d}X(t) + \sum_{a<t\le b} \delta X(t)\, \delta Y(t).$$

When the integrands $Y(t-)$ and $X(t-)$ are left-continuous and the integrator $X(t)$ and $Y(t)$ are right-continuous as above, the integral can be regarded in the Riemann-Stieltjes sense, where $X(t-)$ is the left-hand limit at t. Also, the last two terms may be grouped and considered as an integral in the sense of Lebesgue-Stieltjes, i.e.,

$$\int_{(a,b]} Y(t)\mathrm{d}X(t) = \int_{(a,b]} Y(t-)\mathrm{d}X(t) + \sum_{a<t\le b} \delta X(t)\, \delta Y(t),$$

where $\delta X(t) = X(t) - X(t-)$ is the jump at t. Moreover, given a locally bounded variation (cad-lag) process Y, the equation

$$X(t) = 1 + \int_{(0,t]} X(t-)\mathrm{d}Y(t), \ \forall t \ge 0,$$

has a unique solution X, in the class of locally bounded variation process (cad-lag), which is explicitly given by the formula

$$X(t) = \exp\left[Y^c(t) - Y^c(0)\right] \prod_{0<s\le t} (1 + \delta Y(s)),$$

where $Y^c(t) = Y(t) - \sum_{0<s\le t} \delta Y(s)$ is the continuous part of the process $(Y(t) : t \ge 0)$, the (infinite) product is the exponential of the absolutely convergence series $\sum_{0<s\le t} \ln(1 + \delta Y(s))$, and clearly, if $Y(\tau) - Y(\tau-) = -1$ for some $\tau > 0$ then $X(t) = 0$ for any $t \ge \tau$, see Shiryayev [121, pp. 204–208], Doob [40, pp. 160–166], Chung and Williams [28, pp. 4–6].

An *elementary process* (or piecewise constant over stochastic intervals) is a stochastic process of the form

$$Y(t) = \sum_{i=1}^{n} Y_{i-1} \mathbb{1}_{(\tau_{i-1},\tau_i]}(t), \qquad t \ge 0, \tag{5.9}$$

where $0 = \tau_0 \le \tau_1 \le \ldots \le \tau_n$ are stopping times and Y_{i-1} is a $\mathcal{F}(\tau_{i-1})$ measurable random variable for any i. It is called *bounded* if all Y_{i-1} are bounded random variables. The set of (bounded) elementary processes form a subalgebra (i.e., closed by natural addition and multiplication) of predictable sets. Note that processes of the form (5.9) are left-continuous with right-hand limits, i.e., the (right-hand limit) process $Y(t+)$ is cad-lag, and satisfies $Y(0) = 0$.

If Y is an elementary process and X is a locally bounded variation process (cad-lag) then we may take Y as an *integrand* and X as an *integrator* to construct the integral process for $t \ge 0$ by

$$Z(t) = \int_{(0,t]} Y(s)\mathrm{d}X(s) = \sum_{i=1}^{n} Y_{i-1}[X(t \wedge \tau_i) - X(t \wedge \tau_{i-1})]. \tag{5.10}$$

This integral extends to all predictable processes Y in either Riemann-Stieltjes or Lebesgue-Stieltjes sense. In particular, the above integral makes sense for any bounded adapted cag-lad (left-continuous with right-hand limits) Y. Actually, if the Lebesgue-Stieltjes integral is used then $\mathrm{d}X(t)$ means $\mathrm{d}\mu_X$ (the signed measure induced by the cad-lag bounded variation process X) integration over the interval $(0, t]$, but if the Riemann-Stieltjes integral is used then we mean

$$\int_{(0,t]} Y(s)\mathrm{d}X(s) = \lim_{\alpha\to 0^+,\, \beta\to t^+} \int_{\alpha}^{\beta} Y(s)\mathrm{d}V(s) = \int_{]0,t]} Y(s)\mu_X(\mathrm{d}s).$$

Clearly, the Lebesgue-Stieltjes integral makes also sense when the integrand Y is not necessarily cag-lad. Moreover, from the integration by part formula we deduce the following property. If a function f is continuously differentiable from $\mathbb{R}^d$ into $\mathbb{R}$ and $X = (X_1, \ldots, X_d)$ is a stochastic process with values in $\mathbb{R}^d$ where

each components X_i is a process of locally bounded variation then we have

$$f(X(t)) - f(X(0)) = \int_{(0,t]} \nabla f(X(s-))\mathrm{d}X(s) + \\ + \sum_{0<s\le t} \big[f(X(s)) - f(X(s-) - \nabla f(X(s-))\,\delta X(s))\big], \quad (5.11)$$

If the initial process X is adapted then the integral (5.10) defines an adapted cad-lag stochastic process $(Z(t) : t \ge 0)$ of locally bounded variation, with

$$\mathrm{var}(Z, [0,t]) = \int_{(0,t]} |Y(s)|\,\mathrm{var}(X, \mathrm{d}s).$$

Actually, as long as the above integral is finite (in Lebesgue sense) with a predictable processes $Y(s)$ and locally bounded variation process $X(s)$, the integral (5.10) defines a process $Z(t)$ with locally bounded variation.

Sometimes it is necessary to make a *time change* in Stieltjes integrals. For a given increasing cad-lag process A with values in $[0,\infty]$ consider

$$A^{-1}(s) = \inf\{t \ge 0 : A(t) > s\}, \quad \forall s \ge 0, \quad (5.12)$$

with $A^{-1}(s) = 0$ for $s \le \sup\{t \ge 0 : A(t) = 0\}$, and $A^{-1}(s) = +\infty$ if $A(t) \le s$ for all $t \ge 0$. This define a cad-lag process $(A^{-1}(s) : s \ge 0)$ with the properties

$$A^{-1}(s-) = \inf\{t \ge 0 : A(t) \ge s\}, \quad \forall s \ge 0,$$
$$A[A^{-1}(s)] \ge s, \quad \forall s \ge 0,$$
$$A(t) = \inf\{s \ge 0 : A^{-1}(s) > t\}, \quad \forall t \ge 0.$$

If A is continuous then A^{-1} may not be continuous (when A is not strictly increasing). The following change of variables formula can be obtained. For any nonnegative Borel measurable function f on $[0,\infty)$ we have

$$\int_{[0,\infty)} f(t)\mathrm{d}A(t) = \int_0^\infty f(A^{-1}(s))\,\mathbb{1}_{A^{-1}(s)<\infty}\mathrm{d}s,$$
$$\int_{[u(a),u(b)]} f(t)\mathrm{d}A(t) = \int_{[a,b]} f(u(t))\mathrm{d}A(u(t)),$$

for any continuous non-decreasing process u on the bounded interval $[a,b]$.

A typical example is a real-valued Poisson process $X(t)$ with parameter $c > 0$, which is a process of bounded variation and a jumps process of the form $X(t) = \sum_n \mathbb{1}_{t\ge\theta_n}$, where $\theta_n = \tau_1 + \cdots + \tau_n$ and $(\tau_1, \tau_2, \ldots)$ is a sequence of independent exponentially distributed (with parameter c) random variables. If Y is a (cad-lag) bounded adapted process with respect to X then the following

(cad-lag) adapted processes are defined by the Riemann-Stieltjes integrals

$$M(t) = \int_{]0,t]} Y(s-)\mathrm{d}X(s) - c\int_0^t Y(s)\mathrm{d}s, \quad \forall t \geq 0,$$
$$N(t) = M^2(t) - c\int_0^t Y^2(s)\mathrm{d}s, \quad \forall t \geq 0,$$
$$E(t) = \exp\Big\{ \int_{]0,t]} Y(s-)\mathrm{d}X(s) + c\int_0^t [1 - \mathrm{e}^{Y(s)}]\mathrm{d}s\Big\}, \ \forall t \geq 0.$$

Taking the left-hand limit $Y(s-)$ in the integral with respect to X is essential to make the Riemann-Stieltjes integral meaningful. It will be seen later that these three processes M, N, E are martingales and the above integral will be called stochastic integral when the process Y is predictable with respect to X.

Note that all arguments made above for locally bounded variation process are of a pathwise character, without any assumption of integrability in Ω.

5.3.2 Integrable Finite Variation

No specific difference was made in the previous pathwise discussion regarding path with bounded variation within any bounded time-interval and within the half (or whole) real line, i.e., bounded variation paths (without any other qualitication) 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 realted 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 [69, 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.

A simple application of the change of time (5.12), i.e., the following expression for a cad-lag increasing process A,

$$\mathbb{E}\Big\{\int_0^T X(t)\mathrm{d}A(t)\Big\} = \int_0^\infty \mathbb{E}\big\{X\big(A^{-1}(s)\big)\mathbb{1}_{A^{-1}(s)<\infty}\big\}\mathrm{d}s,$$

proves that for any two nonnegative measurable processes (non necessarily adapted) X and Y satisfying $\mathbb{E}\big\{X(\tau)\mathbb{1}_{\tau<\infty}\big\} = \mathbb{E}\big\{Y(\tau)\mathbb{1}_{\tau<\infty}\big\}$, for every stopping time τ, we have

$$\mathbb{E}\Big\{\int_0^r X(t)\mathrm{d}A(t)\Big\} = \mathbb{E}\Big\{\int_0^r Y(t)\mathrm{d}A(t)\Big\}, \quad \forall r \in (0,\infty].$$

Now, if $\mathcal{F}_\tau$ denotes the σ-algebra associated with a stopping time (see Definition 4.7) then the condition

$$\mathbb{E}\big\{X(\tau)\mathbb{1}_{\tau<\infty}\big|\mathcal{F}(\tau)\big\} = \mathbb{E}\big\{Y(\tau)\mathbb{1}_{\tau<\infty}\big|\mathcal{F}(\tau)\big\}, \quad \text{a.s.}$$

implies

$$\mathbb{E}\Big\{\int_\tau^\infty X(t)\mathrm{d}A(t)\big|\mathcal{F}(\tau)\Big\} = \mathbb{E}\Big\{\int_\tau^\infty Y(t)\mathrm{d}A(t)\big|\mathcal{F}(\tau)\Big\}, \tag{5.13}$$

almost surely, which is used later with martingale.

On the other hand, we can verify that if A and B are two cad-lag increasing processes (non necessarily adapted) such that

$$\mathbb{E}\big\{A(t) - A(s)\big|\mathcal{F}(s)\big\} = \mathbb{E}\big\{B(t) - B(s)\big|\mathcal{F}(s)\big\}, \quad \text{a.s.},$$

for every (extended) real numbers $0 \le s \le t \le \infty$, then we have

$$\mathbb{E}\Big\{\int_0^T X(t-)\mathrm{d}A(t)\big|\mathcal{F}(\tau)\Big\} = \mathbb{E}\Big\{\int_0^T X(t-)\mathrm{d}B(t)\big|\mathcal{F}(\tau)\Big\}, \tag{5.14}$$

for every $T \ge 0$ and for every nonnegative cad-lag adapted process X.

Let us go 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

$$\mu_A(]a,b] \times F) = \mathbb{E}\Big\{ \mathbb{1}_F \int_{]a,b]} \mathrm{d}A(t) \Big\}, \quad \forall b > a \geq 0,\ F \in \mathcal{F}. \tag{5.15}$$

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 [36, Section VI.2, pp. 113–164], Jacod and Shiryaev [69, Section 1.3, pp. 27–32], Rogers and Williams [112, Sections VI.19–21, pp. 347–352], and Elliott [46], Protter [108], among others, to complement the above remarks and following theorem–definition

Definition 5.5 (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 (5.15). 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 $A(t+) - A(t-) \le k+1$ and $\tau^{i,j} \le k$, 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 $r \ge 0$ we set

$$\tau_r = \inf\{t \ge 0 : A(t) \ge r\} \quad \text{and} \quad \theta_r = \inf\{t \ge 0 : A(t) > r\},$$

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 [14, Section 2.4, pp. 69–71].

As mentioned above, another measure associated with a process X is the so-called *jumps measure*, which is a random measure on $[0, \infty) \times \mathbb{R}_*$, with $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$ with integer values and defined, for each ω, by

$$\nu(]a, b] \times B) = \#\{t : a < t \le b,\ X(t) - X(t-) \in B\},$$

for every $b > a \ge 0$ and B in $\mathcal{B}(\mathbb{R}_*)$, i.e., ν is the number of jumps of the process X in the time interval $]a, b]$ which belongs to the set B. Typically, if $X = P$ is a Poisson measure process then the compensator of ν is indeed the (deterministic) Lévy measure m. Clearly, the above integer-valued random measure is defined even if the process X is not of locally bounded variation, only the cad-lag property is used.

On the other hand, we may define the quadratic variation of X over a partition $\pi = (t_0 < t_1 < \cdots < t_n)$ is given by

$$\mathrm{var}_2(X, \pi) = \sum_{i=1}^n |X(t_i) - X(t_{i-1})|^2,$$

and then quadratic variation operator as

$$\begin{aligned} &\mathrm{var}_2(X, [0, t]) = \lim_{r \to 0} \mathrm{var}_2(X, [0, t], r), \\ &\mathrm{var}_2(X, [0, t], r) = \sup\{\mathrm{var}_2(X, \pi)\}, \quad \forall t > 0, \end{aligned} \tag{5.16}$$

where the supremum is taken over all partitions $\pi_t = (0 = t_0 < \cdots < t_n = t)$ of the interval $[0, t]$, with $t_i - t_{i-1} \le r$. It is easy to imagine a process with only jumps such that $\mathrm{var}_2(X, [0, t]) < \infty$ but $\mathrm{var}(X, [0, t]) = \infty$ for any $t > 0$, i.e., the

sum of *small* jumps at the origin is infinite but the sum of the square converges. Moreover, if the process X is continuous with bounded variation then the estimate $\mathrm{var}_2(X,\pi) \le w(X,\pi)\,\mathrm{var}(X,\pi)$ shows that necessarily $\mathrm{var}_2(X,\pi) \to 0$ as the mesh of the partition $\delta(\pi) = \max_i\{t_i - t_{i-1}\}$ vanishes, where $w(X,\pi)$ is the modulus of continuity of X on π, namely,

$$w(X,\pi) = \sup_i \sup\big\{|X(t) - X(s)| : t, s \in [t_{i-1}, t_i]\big\}.$$

However, we may construct a continuous process X with unbounded variation and with the above vanishing quadratic variation property. Furthermore, for a process X with vanishing quadratic variation we can setup and define the Riemann-Stieltjes integral to show that

$$\int_a^b [X(t)]^m \,\mathrm{d}X(t) = \frac{1}{m+1}\Big[[X(b)]^{m+1} - [X(a)]^{m+1}\Big],$$

for every $b > a \ge 0$. For instance, if $X = W$ is a Wiener process then $\mathbb{E}\{\mathrm{var}_2(W,\pi)\} = t_n - t_0$ and the relevance of the quadratic variation is clear when a pathwise analysis is not available. As seen later, this is a typical behavior for martingale processes. We refer the interested reader to Doob [40, Chapters X–XI, pp. 157–204] for a neat analytic approach.

The technique to treat cad-lag processes is essentially as follows. On one hand, the pathwise study is efficient for cad-lag process with local bounded variation paths. This includes (1) continuous process with local bounded variation path and (2) jump processes X with jumps of local bounded variation, i.e., $\sum_{s\le t} |\delta X(s)| < \infty$ for every $t > 0$. For every cad-lag jump process X, there is only a finite number of jumps that are larger than any deterministic constant r, so the number of larger jumps is finite, i.e., $\sum_{s\le t} \mathbb{1}_{\{|\delta X(s)|\ge r\}} < \infty$. Thus,

$$X(t) = \lim_{r\to 0} \sum_{s\le t} \mathbb{1}_{\{|\delta X(s)|\ge r\}} \delta X(s), \quad \forall t > 0,$$

but the series may not converge absolutely. In particular, we look at jumps processes with jumps satisfying $\sum_{s\le t} |\delta X(s)|^2 < \infty$ for every $t > 0$, which does not necessarily have local bounded variation paths. For these processes, the pathwise arguments are not longer valid. By imposing a local integrability (with respect to the path, i.e., $\mathbb{E}\{\sum_{s\le\tau_k} |\delta X(s)|^2\} < \infty$), for some increasing sequence of stopping time $\{\tau_1, \tau_2, \ldots\}$ with $\tau_k \to \infty$, the compensator/martingale theory can be used. This is part of the stochastic integral theory, where jump processes are better viewed as random measure. Continuous martingale processes with no local bounded variation paths are also studied with non-pathwise technique. A more pure analytic point of view is the use of the so-called *orthogonal random measure*.

Some arguments use a enumeration of the jumps, certainly, they are denumerable but to have them in a ordered way, we need to use ordinal numbers. An intuitive feeling is the fact that we can count through countable ordinals

(where each nonempty subset has a first element) as follows:

$$1, 2, 3, \dots, \infty, \infty+1, \infty+2, \dots, 2\infty, 2\infty+1, \dots \\ 3\infty, 3\infty+1, \dots, \infty^2, \dots, \infty^3, \dots, \infty^\infty, \infty^\infty+1, \dots$$

where ∞ is the first infinite ordinal. Each countable ordinal is either a successor $\alpha+1$ of some countable ordinal α or a limit ordinal $\beta = \sup\{\alpha : \alpha < \beta\}$, which is the supremum of ordinals less than it. For instance, to count the jumps of a cad-lag process X, first we set $\tau_0 = 0$, $a_0 = X(0)$ and given an ordinal i with successor $i+1$ we define

$$\tau_{i+1} = \inf\big\{t \geq \tau_i : X(t) \neq X(t-)\big\}, \qquad a_{i+1} = X(\tau_{i+1}) - X(\tau_{i+1}-),$$

while, given a limit ordinal i we define

$$\tau_i = \sup_{j<i} \tau_j, \qquad a_i = X(\tau_i) - X(\tau_i-).$$

Thus, for each countable ordinal i we have defined τ_i and a_i such that $i \leq j$ implies $\tau_i \leq \tau_j$, where τ_i may be infinite for some ω. Because there is a countable number of jumps, we have $\sup_i \tau_i(\omega) = \infty$ and so, for every t and ω there is a first (necessary countable) ordinal such that for $\kappa = \kappa(t,\omega)$ we have $\tau_i \wedge t = \tau_{i+1} \wedge t$, for every $i \geq \kappa$. This means that all the jumps of X within $[0,t]$ are listed with τ_i and a_i for $i \leq \kappa$, the problem is that the possible values of the $\kappa(t,\omega)$ is uncountable (in much the same way that the number of finite ordinals is infinite), so that τ_i may not be an stopping time for some limit ordinal i. However, τ_i is almost surely equal to a stopping time. Indeed, set

$$c_i = \mathbb{E}\{\exp(-\tau_i)\}, \qquad c = \inf_i c_i,$$

where the infimum is taken over all countable ordinals. Thus, there exists a sequence of (countable) ordinals $\{\iota(n) : n \geq 1\}$ independent of ω such that $c_{\iota(n)} \to c$ as $n \to \infty$. If $\iota(\infty)$ is the countable ordinal $\lim_n \iota(n)$ we have $c_{\iota(\infty)} = c$ and the stopping time $\sup\big\{\tau_{\iota(n)} : n\big\}$ is equal to $\tau_{\iota(\infty)}$ almost surely. Hence, each τ_i in the above construction can be chosen to be a stopping time. On the other hand, to construct a (purely) jump process from its jumps, we need some assumptions on the cad-lag process, e.g., locally bounded variation. Indeed, by induction procedure, we may define the sum or series, starting from $S_0 = a_0$, we set $S_{i+1} = a_{i+1} + S_i$ if i has a successor and $S_i = a_i + \sum_{j<i} S_j$ (which converges absolutely) if i is a limit ordinal. Hence, the process $S_{\kappa(t,\omega)}(\omega)$ or equivalently $\sum_{i \leq \kappa(t,\omega)} a_i(\omega)$ is optional. This same argument applies to semi-martingales, as seen in the next section.

5.4 Martingales

Related to the Markov processes with values in $\mathbb{R}^d$, is the concept of (vector) *martingales*. Moreover, the martingale property can be extended to processes with values in Hilbert, Banach or co-nuclear (the strong dual of a nuclear

space) spaces, e.g., see Kallianpur and Xiong [74, Chapter 3, pp. 85–126] and Métivier [98].

Definition 5.6 (general martingale)**.** A (general) martingale with states in $E \subset \mathbb{R}^d$ is a (complete) probability measure P on $(\Omega, \mathcal{F})$, together with a measurable mapping M (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 *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 is the t-component of M. If the family of σ-algebras $(\mathcal{F}_t : t \in T)$ is not mentioned, then it is assumed $(\mathcal{F}_t : t \in T)$ is the history $(\mathcal{H}_t : t \in T)$ of the process $(M_t : t \in T)$, i.e., $\mathcal{H}_t$ is generated by the random variables $\{M_s : s \le t\}$ and the null sets. Moreover, we say that the martingale is *cad-lag* if $(\mathcal{F}_t : t \in T)$ is a filtration satisfying the usual conditions and except on a set of P-probability zero, the paths of $(M_t : t \in T)$ are cad-lag. The martingale is *continuous* if their paths are continuous. Furthermore, if $d = 1$, i.e., with values in $\mathbb{R}$, we may define also *super-* or *sub*-martingale by replacing the equal sign by either $\le$ or $\ge$ in the above condition. □

In most of the cases considered here, the index T is a bounded real interval or $[0, \infty)$, and the probability P is fixed, so that a (good) particular member of the P-equivalence class is used and called (super- or sub-)martingale. As usually, the conditional expectation operator identifies an equivalence class of processes satisfying the above condition and so another condition on the sample path is needed to make the above martingale condition workable in continuous time, e.g., a minimal condition would be a separable martingale and a more reasonable condition is right-continuity in probability. It is clear that if $(M_t : t \ge 0)$ is a cad-lag martingale relative to (or with respect to) $(\mathcal{F}_t : t \ge 0)$ then it is also a cad-lag martingale relative to its canonical (or natural) filtration $(\mathcal{H}_t : t \ge 0)$, the history of the process, see Definitions 4.16 and 4.17 on Markov processes. Certainly, if $(M_t : t \ge 0)$ is a super-martingale then $(-M_t : t \ge 0)$ is a sub-martingale.

When the filtration is the history of the process, the second condition in the martingale property of the above Definition 5.6 can be rephrased as follows

$$\mathbb{E}\Big\{ M(t) \prod_{i=1}^{n} h_i\big(M(s_i)\big) \Big\} = \mathbb{E}\Big\{ M(s) \prod_{i=1}^{n} h_i\big(M(s_i)\big) \Big\} \tag{5.17}$$

for any integer n, for every $0 \le s_1 < s_2 \le \cdots \le s_n \le s < t$, any (real-valued) Borel and bounded (in $\mathbb{R}^d$) functions h_i, $i = 1, \ldots, n$. Moreover, if the process $(M_t : t \ge 0)$ is right-continuous in L^1, i.e.,

$$\lim_{t \downarrow s} \mathbb{E}\{|M_t - M_s|\} = 0, \quad \forall s \ge 0,$$

then, applying the martingale property (5.17) to continuous functions h_i and $s_i + \varepsilon_i$, $s + \varepsilon$, with $0 < \varepsilon_i \le \varepsilon$, we deduce another expression of the martingale

property, namely,

$$\mathbb{E}\Big\{ M(t) \prod_{i=1}^{n} h_i\big(M(s_i+0)\big) \Big\} = \mathbb{E}\Big\{ M(s) \prod_{i=1}^{n} h_i\big(M(s_i+0)\big) \Big\} \tag{5.18}$$

for any integer n, for every $0 \leq s_1 < s_2 \leq \cdots \leq s_n \leq s < t$, any (real-valued) continuous and bounded (in $\mathbb{R}^d$) functions h_i, $i = 1, \ldots, n$. Note that relation (5.18) represents the second condition in the martingale property of Definition 5.6, where $(\mathcal{F}_t : t \geq 0)$ is the smallest filtration satisfying the usual conditions which makes the process $(M_t : t \geq 0)$ adapted. This proves that if $(M_t : t \geq 0)$ is a right-continuous (actually it suffices that it be right-continuous in probability) martingale with respect to $(\mathcal{F}_t : t \geq 0)$ then $(M_t : t \geq 0)$ is also a martingale relative to the (possible larger) right-continuous filtration $(\bar{\mathcal{F}}_t : t \geq 0)$, with $\bar{\mathcal{F}}_t = \cap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$. Clearly, if $(M_t : t \geq 0)$ and $(N_t : t \geq 0)$ are two sub-martingales (or super-martingales, respectively) relative to the same filtration $(\mathcal{F}_t : t \geq 0)$ then the new process $(M_t \vee N_t : t \geq 0)$ (or $(M_t \wedge N_t : t \geq 0)$, respectively) is also a sub-martingale (or super-martingale, respectively).

It is clear that the martingale condition does not distinguish modifications of the process, so it may be possible to have a (general) martingale which paths are not necessarily cad-lag, with a filtration which is not necessarily right-continuous (or completed). Thus, the assumption of a setup with a filtration satisfying the usual conditions is not at all granted, completion with null sets is a rather technical condition, but the right-continuity is essential to the well behavior and mathematically workable study of (sub-/super-) martingale processes. This is illustrated by the Doob's regularization result, which uses the following concept.

Let D be a countable dense set in $\mathbb{R}$ (e.g., the rational numbers) and x be a function from D into $\mathbb{R}$. The function x is called *regularisable* if the right-hand and left-hand limits exist finitely within D for every real value, i.e., for every t in $\mathbb{R}$ there exist real values $x(t+)$ and $x(t-)$ such that for every $\varepsilon > 0$ there is a $\delta > 0$ (possible depending on ε, $x(\cdot)$ and t) such that $0 < s - t < \delta$ implies $|x(s) - x(t+)| < \varepsilon$ and $0 < t - s < \delta$ implies $|x(s) - x(t-)| < \varepsilon$. Clearly, when the initial function x is defined in an interval I of $\mathbb{R}$, first the function x is restricted to the set $I \cap D$ and then the above definition is applied for t in the interval I. In most cases, the countable dense set D can be arbitrary chosen or easily understood from the context. Usually $I = [0, \infty)$ and so $x(0-)$ is either not defined or set equal to 0 for the sake of completeness. If a function x is regularisable (within D) then x_+ and x_- (right and left limits) denote the new functions obtained as the pointwise limits.

An interesting point is that a function $x : D \cap [0, \infty) \to \mathbb{R}$ is regularisable if and only if for every integer N and any compact interval $[a, b]$ with $a < b$ in D, the following expressions are finite,

$$\begin{aligned} \|x\|_{\infty, D\cap[a,b]} &= \sup\big\{ |x(s)| \,:\, s \in D \cap [a,b] \big\}, \\ U_N(x, [a,b]) &= \sup\big\{ k \,:\, 0 \leq s_1 < r_1 < s_2 < \ldots < s_k < r_k \leq N, \\ &\qquad\qquad , x(s_i) < a,\ x(r_i) > b,\ s_i, r_i \in D,\ \forall i \big\}. \end{aligned}$$

As in the discrete case, $U_N(x,[a,b])$ is called the upcrossings of the interval $[a,b]$ by time N.

Theorem 5.7. *Let $M = \{M_t : t \geq 0\}$ be a real-valued family of random variables in a probability space $(\Omega, \mathcal{F}, P)$ which satisfies the (super- or sub-) martingale property relative to an increasing family of σ-algebras $\{\mathcal{F}(t) : t \geq 0\}$, i.e.,*

$$\mathbb{E}\{|M(t)|\} < \infty, \ \forall t, \qquad \mathbb{E}\{M(t) \mid \mathcal{F}_s\} = M_s, \quad a.s. \ \ \forall t > s \geq 0,$$

with $=$ replaced by $\leq$ or $\geq$ when (super- or sub-) is used. Then M is regularisable except in a set of probability zero, and the processes $M_+ = \{M(t+) : t \geq 0\}$ and $M_- = \{M(t-) : t \geq 0\}$ are cad-lag (super- or sub) martingales relative to $\{\mathcal{F}(t+) : t \geq 0\}$ and $\{\mathcal{F}(t-) : t \geq 0\}$, respectively. Moreover, if the function $t \mapsto \mathbb{E}\{M(t)\}$ is right-continuous (resp., left-continuous) then M_+ (resp., M_-) is a version of M. □

For a complete detail on the proof see, e.g., Dellacherie and Meyer [36, Section VI.1] or Rogers and Williams [112, Section II.5, Subsections 65–67, pp. 169–174]. Clearly, the above results include the following statement. Let $\{\mathcal{F}(t) : t \geq 0\}$ be a right-continuous and complete the filtration, and assume that the function $t \mapsto \mathbb{E}\{M(t)\}$ is right-continuous (e.g., this mean right-continuity holds if M is a martingale). Then M_+ is a version of M, which is a cad-lag (super- or sub-) martingale relative to the filtration $\{\mathcal{F}(t) : t \geq 0\}$. Moreover, if M is separable then so is M_+ and therefore M_+ is indistinguishable from M, i.e., M itself is a cad-lag (super- or sub-) martingale.

An integrable process with independent increments and zero mean is not always a typical example of martingale, some regularity on the path is needed. For instance, if $\{w(t) : t \geq 0\}$ is a standard Wiener process in $\mathbb{R}^d$ then it is also a continuous martingale, and if $\{p(t) : t \geq 0\}$ is a standard Poisson process then $(M_t : t \geq 0)$, with $M_t = p(t) - \mathbb{E}\{p(t)\}$, is a cad-lag martingale. In general, we will see that if $(X_t : t \geq 0)$ is a cad-lag Markov process with infinitesimal generator A (see then the stochastic process

$$M_t = \varphi(X_t) - \int_0^t A\varphi(X_s)\mathrm{d}s, \ \forall t \geq 0$$

is a cad-lag martingale, for any (smooth) function φ in the domain $\mathcal{D}(A)$ of the infinitesimal generator A. In fact, as seen later, this is a characterization of the Markov processes in terms of the so-called *martingale problem*. On the other hand, the concept of martingale is a sort of complementary definition with respect to bounded variation processes, in the sense that the only continuous martingale of bounded variation is the trivial or constant process.

To study martingales we begin with either (1) a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$ satisfying the usual conditions and we look at cad-lag stochastic processes $(X_t : t \geq 0)$ as random variables with valued in the canonical space $D([0,\infty), \mathbb{R}^d)$ or (2) a canonical space $D = D([0,\infty), \mathbb{R}^d)$, with its Borel σ-algebra $\mathcal{B}$, the

canonical process ($X_t = \omega(t)$, $t \geq 0$) and it associated the filtration ($\mathcal{F}_t : t \geq 0$) and we look for probability measures on D. Thus, a cad-lag martingale is viewed as a random variable with values in the canonical space D, identified with its equivalence class, namely, all processes which are indistinguishable (or equivalent) of it, and as long as we use *cad-lag* (or *separable*) processes this agree with the notion of version (or modification).

We rephrase the above martingale concept

Definition 5.8 (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 $\{\mathcal{F}(t) : t \geq 0\}$ 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 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 5.9. *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

$$r\, P\big(\sup_{s\leq t} X(s) \geq r\big) \leq \mathbb{E}\{X(t) \mathbb{1}_{\sup_{s\leq t} X(s)\geq r}\} \leq \mathbb{E}\{X(t)\}, \tag{5.19}$$

and therefore

$$\|\sup_{s\leq t} X(s)\|_p \leq p' \, \|X(t)\|_p, \quad \text{ with } \ 1/p + 1/p' = 1, \tag{5.20}$$

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

$$\|\sup_{s\le t} X(s)\|_1 \le \frac{\mathrm{e}}{\mathrm{e}-1}\,\|X(t)\ln^+ X(t)\|_1, \tag{5.21}$$

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

Note that (5.13) implies that for any positive cad-lag martingale M, which is written as $M(t) = \mathbb{E}\{M(\infty)|\mathcal{F}(t)\}$ if M is uniformly integrable, and any local integrable increasing process A with $A(0) = 0$, we have

$$\mathbb{E}\big\{X(t)A(t)\big\} = \mathbb{E}\Big\{\int_0^t Y(s)\mathrm{d}A(s)\Big\}, \quad \forall t \in (0,\infty),$$

and even for $t = \infty$ if M is uniformly integrable.

Now, based on (5.14), an local integrable increasing process A with $A(0) = 0$ is called *natural* if

$$\mathbb{E}\Big\{\int_0^T M(t)\mathrm{d}A(t)\Big\} = \mathbb{E}\Big\{\int_0^T M(t-)\mathrm{d}A(t)\Big\}, \quad \forall T \in \mathbb{R}, \tag{5.22}$$

for every nonnegative, bounded and cad-lag continuous martingale M. Since the process $Y(t) = Y(t)\mathbb{1}_{t<\tau} + Y(\tau)\mathbb{1}_{s\ge\tau}$ is a martingale for any stopping time τ, we deduce that (5.22) is equivalent to either

$$\mathbb{E}\Big\{\int_0^\infty M(t)\mathrm{d}A(t)\Big\} = \mathbb{E}\Big\{\int_0^\infty M(t-)\mathrm{d}A(t)\Big\}, \tag{5.23}$$

or

$$\mathbb{E}\Big\{\int_0^\tau M(t)\mathrm{d}A(t)\Big\} = \mathbb{E}\Big\{\int_0^\tau M(t-)\mathrm{d}A(t)\Big\}, \quad \forall \text{ stopping time } \tau,$$

and that the increasing process $B(t) = B(t)\mathbb{1}_{t<\tau} + B(\tau)\mathbb{1}_{s\ge\tau}$ is also natural. Finally, if A is an integrable increasing natural process then (5.23) holds for any uniformly integrable cad-lag martingale M.

With all these properties in place, we can check that if X is a cad-lag sub martingale and A and B are two cad-lag increasing natural processes such that $A(0) = B(0) = 0$ and $X - A$ and $X - B$ are (cad-lag) martingale then $A(t) = B(t)$, almost surely, for every $t \ge 0$.

As we can check later, this notion of (cad-lag) increasing natural process agrees with the more general condition of (cad-lag) increasing predictable process. For further details, the reader may want to take a look at certain points in the book by Meyer [99].

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

Theorem 5.10 (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.* □

For instance, a comprehensive proof of this fundamental results can be found Rogers and Williams [112, 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 predictable *compensator* as in Definition 5.5. Certainly, this can be extended to integrable bounded variation processes, by using the positive and negative variation.

Therefore, the previous convergence Theorem 5.9 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 M. Note that μ_A on $[0,\infty)\times\Omega$ associated with the increasing process A, as defined by (5.15), satisfies

$$\mu_A(\llbracket\tau,\theta\rrbracket) = \mathbb{E}\{A_\theta - A_\tau\} = \mathbb{E}\{X_\theta - X_\tau\},$$

for any stopping times $\tau \le \theta$ and where the the stochastic interval $\llbracket\tau,\theta\rrbracket$ is viewed as the subset $\{(\omega,t) : \tau(\omega) < t \le \theta(\omega)\}$ of $[0,\infty)\times\Omega$. This is one of the key elements used in the definition of the process A, i.e., the fact that for any given sub-martingale X of class (D) we can construct a unique bounded (positive) measure on $[0,\infty)\times\Omega$ defined by $\mu_A(]\tau,\theta]) = \mathbb{E}\{X_\theta - X_\tau\}$. Actually, it is also established in the Doléans' proof of the above decomposition, that for a quasi-leftcontinuous (or regular) sub-martingale X (i.e., $\mathbb{E}\{X(\tau)\} = \mathbb{E}\{X(\tau-)\}$ for any predictable stopping time τ or equivalently A is continuous) the predictable (or conditional) variation for a partition $\pi = (t_0 < t_1 < \cdots < t_n)$,

$$^{\mathrm{p}}\mathrm{var}(X,\pi) = \sum_{i=1}^{n} \mathbb{E}\{X(t_i) - X(t_{i-1}) \,|\, \mathcal{F}(t_{i-1})\}, \tag{5.24}$$

which is equal to $^{\mathrm{p}}\mathrm{var}(A,\pi)$, has the following property: for any $\varepsilon > 0$ there is a $\delta > 0$ such that $\mathbb{E}\{|^{\mathrm{p}}\mathrm{var}(X,\pi_t) - A(t)|\} \le \varepsilon$ for any partition $\pi = \pi_t$ with $t_n = t$ and $t_i - t_{i-1} \le \delta$, for every $i = 1,2,\ldots,n$.

Let us denote by $M^2(\Omega,P,\mathcal{F},\mathcal{F}(t) : t \ge 0)$ the space of square-integrable martingales M null at time zero, i.e., besides the martingale conditions in Definition 5.8 we impose $M(0) = 0$ and $\sup_{t\ge0}\mathbb{E}\{|M(t)|^2\} < \infty$. 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 $M^2(\Omega,P,\mathcal{F},\mathcal{F}(t) : t \ge 0)$ 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 $M^*(t) = \sup_{s\le t}|M(s)|$ and its limit $M^*(\infty) = \sup_{t\ge0}|M(t)|$ then we have

$$\mathbb{E}\{|M^*(\infty)|^2\} \le 4 \sup_{t\ge0} \mathbb{E}\{|M(t)|\} = 4\,\mathbb{E}\{|M(\infty)|^2\},$$

after using Doob's estimate (5.20) with $p = 2$. Thus, $M^2(\Omega,P,\mathcal{F},\mathcal{F}(t) : t \ge 0)$ 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 5.10 applies to get a unique predictable (continuous) increasing process $\langle M \rangle$, referred to as the *predictable quadratic variation* process. 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

$$[M](t) = \langle M_c \rangle(t) + \sum_{s \leq t} (M_d(s) - M_d(s-))^2, \quad \forall t > 0. \tag{5.25}$$

Note that $[M_c] = \langle M_c \rangle$ and $M_d(t) - M_d(t-) = M(t) - M(t-)$, for any $t > 0$, but that $\langle M \rangle$ is the predictable (dual) projection of the increasing process $[M]$, as defined in Section 5.3. We re-state these facts for a further reference

Theorem 5.11 (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 $\langle M \rangle = 0$.* □

Also, the optional quadratic variation can be defined by means of the stochastic integral and for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\mathbb{E}\Big\{ \sup_{0<t\leq 1/\varepsilon} |\mathrm{var}_2(M, \pi_t) - [M](t)| \Big\} \leq \varepsilon,$$

for any partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ with $0 < t_i - t_{i-1} \leq \delta$, for every $i = 1, 2, \ldots, n$, where $\mathrm{var}_2(M, \pi_t)$ is the optional quadratic variation

operator

$$\mathrm{var}_2(M, \pi_t) = \sum_{i=1}^{n} \mathbb{E}\{|M(t_i) - M(t_{i-1})|^2\}, \tag{5.26}$$

and t_i in the partition may be stopping times. This previous limit could be used as definition of $[M]$, and then we may define the predictable variation process $\langle M \rangle$ as the compensator of the optional quadratic variation $[M]$.

If the filtration can be chosen quasi-left continuous (i.e., satisfying $\mathcal{F}(\tau-) = \mathcal{F}(\tau)$ for every predictable stopping time τ) or equivalently if the predictable variation process $\langle M \rangle$ is continuous then for any $\varepsilon > 0$ there is a $\delta > 0$ such that $\mathbb{E}\{|{}^{\mathrm{p}}\mathrm{var}_2(M, \pi_t) - \langle M \rangle(t)|\} \le \varepsilon$ for any partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ with $0 < t_i - t_{i-1} \le \delta$, for every $i = 1, 2, \ldots, n$, where $\mathrm{var}_2(M, \pi_t)$ is the optional quadratic variation operator and ${}^{\mathrm{p}}\mathrm{var}_2(M, \pi_t)$ is the predictable quadratic variation operator defined by (5.24), i.e., ${}^{\mathrm{p}}\mathrm{var}_2(M, \pi_t)$ converges in L^1 to $\langle M \rangle$ as the mesh goes to zero.

These are key results in the study of martingales and foundation of the stochastic integrals for continuous martingales. To understand the convergence in the L^1-norm of the predictable quadratic variation as defined in Theorem (5.11), first we realize that the predictable quadratic variation operator on M is equal to the predictable variation operator on $\langle M \rangle$, i.e., ${}^{\mathrm{p}}\mathrm{var}_2(M, \pi) = {}^{\mathrm{p}}\mathrm{var}(\langle M \rangle, \pi)$, as defined by (5.24). Setting $A^k(s) = \min\{\langle M \rangle(s), k\}$, for $s \ge 0$, for a given partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ we consider the (finite) sequence of bounded random variables

$$x_i = \mathbb{E}\{A^k(t_i) - A^k(t_{i-1}) \,|\, \mathcal{F}(t_{i-1})\} - [A^k(t_i) - A^k(t_{i-1})],$$

for $i = 1, 2, \ldots, t_n$, which are orthogonal in $L^2(\Omega, \mathcal{F}, P)$. Based on the elementary bound $(a-b)^2 \le 2a^2 + 2b^2$ and Jensen's inequality we obtain

$$\mathbb{E}\{x_i^2\} \le 4\, \mathbb{E}\{[A^k(t_i) - A^k(t_{i-1})]^2\},$$

which yields

$$\mathbb{E}\{[{}^{\mathrm{p}}\mathrm{var}(A^k, \pi_t) - A^k(t)]^2\} = \sum_{i=1}^{n} \mathbb{E}\{x_i^2\} \le$$

$$\le 4 \sum_{i=1}^{n} \mathbb{E}\{[A^k(t_i) - A^k(t_{i-1})]^2\} \le 4\, \mathbb{E}\{\rho(A^k, [0,t], \delta)\, A^k(t)\},$$

where $\rho(A^k, [0,t], \delta)$ is the modulus of continuity of A^k, i.e.,

$$\rho(A^k, [0,T], \delta) = \sup\{A^k(t) - A^k(s) \,:\, 0 \le s < t \le T\},$$

and $0 < t_i - t_{i-1} \le \delta$, $i \ge 1$. Since

$$\mathbb{E}\{|{}^{\mathrm{p}}\mathrm{var}_2(M, \pi_t) - \langle M \rangle(t)|\} \le$$

$$\le \mathbb{E}\{|{}^{\mathrm{p}}\mathrm{var}(A^k, \pi_t) - A^k(t)|\} + \mathbb{E}\{\langle M \rangle(t) - A^k(t)\},$$

we obtain the L^1 convergence stated in quadratic variation theorem above, when we assume that $\langle M\rangle$ is continuous, which is equivalent to the quasi-left continuity of the filtration.

Moreover, if M is a continuous martingale with $M(t)=0$ then

$$^{\mathrm{p}}\mathrm{var}_2(M,\pi_t) \le {}^{\mathrm{p}}\mathrm{var}(M,\pi_t)\,\rho(M,[0,t],\delta),$$

for any partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ with $0 < t_i - t_{i-1} \le \delta$, $i = 1, 2, \ldots$. Thus, the predictable square variation process $\langle M\rangle$ vanishes, when M has almost surely path of bounded variation, i.e., M is the null process.

Similarly, the predictable ℓ-variation ($\ell > 2$) of any ℓ-integrable martingale M null at time zero is defined by

$$^{\mathrm{p}}\mathrm{var}_\ell(M,\pi) = \sum_{i=1}^{n} \mathbb{E}\{|M(t_i) - M(t_{i-1})|^\ell \,|\, \mathcal{F}(t_{i-1})\}, \tag{5.27}$$

with $\pi = (t_0 < t_1 < \cdots < t_n)$. Therefore, the inequality

$$^{\mathrm{p}}\mathrm{var}_\ell(M,\pi_t) \le {}^{\mathrm{p}}\mathrm{var}_2(M,\pi_t)\,[\rho(\langle M\rangle,[0,t],\delta)]^{\ell/2-1},$$

for any partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ with $0 < t_i - t_{i-1} \le \delta$, $i = 1, 2, \ldots$ and $\ell > 2$, proves that $^{\mathrm{p}}\mathrm{var}_\ell(M,\pi_t) \to 0$ almost surely as the mesh of the partition δ vanishes. These two facts about the convergence of the predictable quadratic variation and ℓ-variation ($\ell > 2$) are essential to define the stochastic integral.

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

$$\begin{aligned} \langle M,N\rangle &= \big(\langle M+N\rangle - \langle M-N\rangle\big)/4, \\ [M,N] &= \big([M+N] - [M-N]\big)/4, \end{aligned} \tag{5.28}$$

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 [36, Chapters V–VIII], Jacod and Shiryaev [69], Karatzas and Shreve [75], Neveu [103], Revuz and Yor [111], among others.

Given a cad-lag integrable process X we can associate the so-called Föllmer *finitely additive* measure on the ring $\mathcal{R}$ generated by all predictable rectangles by the expression

$$\begin{aligned} &\lambda_X(]s,t]\times F) = \mathbb{E}\{(X_t - X_s)\,\mathbb{1}_F\}, \quad \forall t > s \ge 0,\ F \in \mathcal{F}_s, \\ &\lambda_X(\{0\}\times F) = 0 \quad \forall F \in \mathcal{F}_0. \end{aligned} \tag{5.29}$$

The *variation* of λ_X for any A in $\mathcal{R}$ is defined by

$$|\lambda_X|(A) = \sup\Big\{\sum_{i=1}^{n} |\lambda_X(A_i)|\Big\}, \tag{5.30}$$

where the supremum is taken over all finite partition of A, i.e., $A = \cup_{i=1}^{n} A_i$, with A_i in $\mathcal{R}$ and $A_i \cap A_j = \emptyset$ if $i \neq j$. Replacing the absolute value $|\cdot|$ in the above definition (5.30) with the positive or negative part, the positive or negative variation λ_X^+ or λ_X^- is also defined. It satisfies $\lambda_X^+ + \lambda_X^- = |\lambda_X|$ and $\lambda_X^+ - \lambda_X^- = \lambda_X$. It is easily seen that the three variations $|\lambda_X|$, λ_X^+ and λ_X^- are finitely additive measure.

The process X is called a *quasi-martingale* if its variation $|\lambda_X|([0,t] \times \Omega)$ given by (5.30) is finite for every $t > 0$. An interesting point is that the sum of a bounded martingale and an adapted (cad-lag) integrable process or the difference of two non-negative super (or sub) martingale is a quasi-martingale, and indeed (see Protter [108, Theorem II.3.8, pp. 96-97]) any quasi-martingale admits a decomposition as a difference of two positive right continuous super (or sub) martingales. The notion of quasi-martingale can be easily generalized to the multi-dimensional case (even with values in a Banach space). It can be proved that X is a quasi-martingale if and only if

$$\sup \Big\{ \sum_{i=1}^{n} \big| \mathbb{E}\{X(t_i) - X(t_{i-1}) \,|\, \mathcal{F}(t_{i-1})\} \big| \Big\} < \infty,$$

where the supremum is taken over all partition of the form $0 = t_0 < t_i < \cdots < t_n = t$, any $n \geq 1$.

If the initial process X has integrable bounded variation then λ_X can be extended to a σ-additive measure on the predictable σ-algebra $\mathcal{P}$. Conversely, the finitely additive measure λ_X on $\mathcal{R}$ can be extended to σ-additive measure on $\mathcal{P}$ if X is a quasi-martingale of the class (D), i.e., such that the family of random variables $\{X(\tau) : \tau \text{ is a stopping time}\}$ is uniformly integrable. The interested reader may consult the book by Métivier [98].

5.5 Semi-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 5.12 (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 $\{X(t) : t \geq 0\}$. □

In some cases, it may be some ambiguity regarding the above definition, e.g., when we refer to a *local* uniform integrable martingale or to a uniform integrable *local* martingale, fortunately, we can prove that all cases used here are exactly the same. One of the reasons for using the above localization is the following construction: if $(\tau_n : n \geq 1)$ is a reducing sequence of stopping times for a local martingale X defined on $[0, \tau)$, $\tau_n \to \tau$ then define $\tau_0 = 0$ and

$$\gamma(t) = \begin{cases} t - k + 1 & \text{if } \tau_{k-1} + k - 1 \leq t < \tau_k + k - 1, \\ \tau_k & \text{if } \tau_k + k - 1 \leq t < \tau_k + k, \end{cases}$$

which yields $\gamma(t) = t$ on $[0, \tau_1]$, $\gamma(t) = \tau_1$ on $[\tau_1, \tau_1 + 1]$, $\gamma(t) = t - 1$ on $[\tau_1 + 1, \tau_2 + 1]$, $\gamma(t) = \tau_2$ on $[\tau_2 + 1, \tau_2 + 2]$, $\gamma(t) = t - 2$ on $[\tau_2 + 2, \tau_3 + 2]$, $\gamma(t) = \tau_3$ on $[\tau_3 + 2, \tau_3 + 3]$, etc. Strictly speaking γ compresses the time interval $[0, \infty)$ onto $[0, \tau)$ and $X^\gamma = (X(\gamma(t)) : t \geq 0)$ is well defined, actually X^γ is a martingale relative to $(\mathcal{F}(\gamma(t)) : t \geq 0)$. This construction is a key element to extend previous results on (super-/sub-) martingales to *local* (super-/sub-) martingales, where integrability is no more an issue. Actually, by means of the Doob's optional sampling theorem if (super-/sub-) martingale X relative to the filtration $(\mathcal{F}(t) : t \geq 0)$ then $X^\gamma = (X(\gamma(t)) : t \geq 0)$ is a (super-/sub-) martingale relative to the filtration $(\mathcal{F}(\gamma(t)) : t \geq 0)$ for any family of stopping times such that $P(\gamma(s) \leq \gamma(t) < \infty) = 1$ for any $0 \leq s \leq t$, for instance, see the books by Ikeda and Watanabe [62, pp. 32–34] or Durrett [43, pp. 38–42] for more details and comments.

A very important point in the localization principle is the fact that when a property is localized, we are not given away only the integrability $\mathbb{E}\{|X(t)|\} < \infty$, for any $t \geq 0$, more is included. For instance, there are continuous non-negative super-martingales which are local martingales but not martingales, a typical example is a $M(t) = 1/|W(t)|$, where W is a Wiener process in $\mathbb{R}^3$ with $|W(0)| = 1$. Indeed, by means of the (Gaussian) density of W we may show that $\mathbb{E}\{|M(t)|\} < \infty$ and because $1/|x|$ is a fundamental solution for the Laplace equation we can complete the argument. Note that $\tau_n = \inf\{t \geq 0 : |M(t)| > n\}$ is a reducing sequence for M but the family of random variables $\{M(t \wedge \tau_n) : n = 1, 2, \ldots\}$ is not uniformly integrable. On the other hand, a local (super-/sub-) martingale X satisfying

$$\mathbb{E}\{\sup_{s \leq t} |X(s)|\} < \infty, \quad \forall t,$$

is indeed a (super-/sub-) martingale, note the sup inside the mathematical expectation.

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 5.10 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.

Note 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

$$[M](t) = \langle M_c \rangle(t) + \sum_{s \le t} [M(s) - M(s-)]^2, \quad \forall t \ge 0, \tag{5.31}$$

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). An important point here is the fact that the square of the jumps are locally integrable, i.e., the process $\sqrt{[M]}$ is locally integrable and therefore

$$\sum_{s \le t} [M(s) - M(s-)]^2 < \infty, \quad \forall t > 0, \tag{5.32}$$

almost surely. This follows from the use the compensator of Definition 5.5 and two facts: (1) for any cad-lag process there is only a finite number of jumps greater than a positive constant, i.e., $|M(s) - M(s-)| > \varepsilon$, almost surely, and (2) any local martingale with jumps bounded by a constant is locally square-integrable.

On the other hand, given a local martingale M and a real number κ there exists two local martingales V_κ and N_κ such that $M = V_\kappa + N_\kappa$, where V is a locally bounded (or finite) variation process and the jumps of N are bounded by κ. Thus, a local martingale is the sum of a local square-integrable martingale and a locally finite variation process.

It is also clear that we can write $M = M_c + M_d$, where M_c is the continuous local martingale part and M_d is the so-called purely discontinuous local martingale part, so that $M(t) - M(t-) = M_d(t) - M_d(t-)$ for any $t > 0$. Beside the defining fact that any square-integrable purely discontinuous martingale is orthogonal to any square-integrable integrable continuous martingale, we may define purely discontinuous martingales as locally uniform L^2-limits of local martingales with local finite variation, i.e., M_d is purely discontinuous if and only if there exists a sequence $\{X_n : n \ge 1\}$ of locally integrable finite variation processes of the form

$$X_n(t) = M_d(0) + A_n(t) - A_n^p(t), \qquad A_n(t) = \sum_{0 < s \le t} \big(X_n(s) - X_n(s-)\big),$$

where A_n^p is the compensator of A_n as in Definition 5.5, such that

$$\mathbb{E}\{\sup_{0\le t\le T}|X_n(t)-M_d(t)|^2\}\to 0,$$

for any constant $T>0$, e.g., see Kallenberg [71, Theorem 26.14, pp. 527–529].

On the other hand, 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 (5.28) for any two local martingales M and N. For instance we refer to Rogers and Williams [112, 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 (5.20), 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

$$\mathbb{E}\{\sup_{0\le t\le T}|M(t)|^p\}\le C_p\,\mathbb{E}\{\big(\langle M\rangle(T)\big)^{p/2}\},\tag{5.33}$$

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 (5.33) follows. This involves Lenglart–Robolledo inequality, see Liptser and Shiryayev [88, 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

$$c_p\,\mathbb{E}\{\big([M](T)\big)^{p/2}\} \le \mathbb{E}\{\sup_{t\le T}|M(t)|^p\} \le C_p\,\mathbb{E}\{\big([M](T)\big)^{p/2}\}, \tag{5.34}$$

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.

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 (5.33) and (5.34) 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 [71, Theorem 26.12, pp. 524–526], Liptser and Shiryayev [88, Sections 1.5–1.6, pp. 70–84] or Dellacherie and Meyer [36, 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 [111, Section IV.4, pp. 160–171] for continuous local martingales.

Now, combining bounded variation processes with martingales processes and localization arguments, we are led to the following definition.

Definition 5.13 (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 paths. □

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 [71], Protter [108].

Theorem 5.14. *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 (5.31) 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 [88, 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 [69, 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. It is clear that $\mathbb{R}^d$-valued processes with independent increments are completely described by their characteristic functions, namely

$$\hat{X}(\lambda, t) = \mathbb{E}\{\exp(\mathtt{i}\, \lambda \cdot X(t))\}, \quad \forall t \geq 0,\ \lambda \in \mathbb{R}^d,$$

which is a complex-valued cad-lag function. It can be proved that a process X with independent increments is a general semi-martingale if and only if the (deterministic) function $t \mapsto \hat{X}(\lambda, t)$ has locally bounded variation for any λ in $\mathbb{R}^d$. Moreover, a process X with independent increments has the form $X = Y + A$, where Y is a general semi-martingale with independent increments and A is a deterministic cad-lag process (or function) from $[0, \infty)$ into R_+^d with $A(0) = 0$. On the other hand, if X is an integrable (cad-lag!) process with independent increments, i.e., $\mathbb{E}\{|X(t)|\} < \infty$ for every $t \geq 0$, and $(\mathcal{F}(t) : t \geq 0)$ is its natural filtration then

$$\mathbb{E}\{X(t) \,|\, \mathcal{F}(s)\} = \mathbb{E}\{X(t) - X(s)\} + X(s), \quad \text{a.s.},$$

for any $t \geq s \geq 0$. Hence, X is a (super-/sub-) martingale if and only if $\mathbb{E}\{X(t) - X(s)\} = 0$ $(\leq 0 / \geq 0)$ for any $t \geq s \geq 0$.

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 (5.34), another very useful estimate is the Lenglart's inequality: 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

$$P\big\{\sup_{t\le\tau}|X_t|\ge\varepsilon\big\}\le\frac{1}{\varepsilon}\Big[\eta+\mathbb{E}\big\{\sup_{t\le\tau}(A_t-A_{t-})\big\}\Big]+P\big\{A_\tau\ge\eta\big\},\tag{5.35}$$

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 (5.35). 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 (5.35) is when the predictable compensator process $\langle M\rangle$ is continuous, and therefore $[M] = \langle M\rangle$, so that

$$P\big\{\sup_{t\le\tau}|M_t|\ge\varepsilon\big\}\le\frac{\eta}{\varepsilon^2}+P\big\{\langle M\rangle_\tau\ge\eta\big\},\quad\forall\varepsilon,\eta>0,\tag{5.36}$$

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

For a comprehensive treatment with proofs and comments, the reader is referred to the books by Dellacherie and Meyer [36, Chapters V–VIII], Liptser and Shiryayev [88, Chapters 2–4, pp. 85–360]. Rogers and Williams [112, Section II.5, pp. 163–200], among others. A treatment of semi-martingale directly related with stochastic integral can be found in Protter [108].

Let us insist on the following concept, which characterize a large class of Markov processes suitable for modelling.

Definition 5.15 (quasi-left continuous)**.** As mentioned previously, a filtration $(\mathcal{F}(t) :\ge 0)$ of a probability space $(\Omega, \mathcal{F}, P)$ is called *quasi-left continuous* or *regular* if for any increasing sequence of stopping time $\{\tau_n : n = 1, 2, \ldots\}$ almost surely strictly convergent to τ, $P(\tau_n \le \tau_{n+1} < \tau < \infty, \tau > 0) = 1$, the σ-algebra $\mathcal{F}(\tau)$ is the minimal σ-algebra containing the sequence of σ-algebra $\{\mathcal{F}(\tau_n) : n = 1, 2, \ldots\}$. This is equivalent to the condition $\mathcal{F}(\tau) = \mathcal{F}(\tau-)$ for any predictable stopping time, recall that a stopping time τ is predictable if there exists an announcing sequence of stopping times $\{\tau_n : n = 1, 2, \ldots\}$, i.e.,

τ_n increases to τ and $P(\tau_n < \tau, \tau > 0) = 1$. A (cad-lag) integrable stochastic process X adapted to a filtration $\{\mathcal{F}(t) : t \geq 0\}$ is called *quasi-left continuous* or *regular* if $X(\tau_n)$ converges to $X(\tau)$ almost surely for any announcing sequence of stopping times τ_n convergent to τ. □

It can be proved (e.g., Rogers and Williams [112, Chapter VI, Theorems 18.1-2, pp. 346-347]) that a filtration $\{\mathcal{F}(t) : t \geq 0\}$ is quasi-left continuous if and only if every uniformly integrable martingale M relative to $\{\mathcal{F}(t) : t \geq 0\}$ satisfies $M(\tau) = M(\tau-)$ for any predictable stopping time and that any *Markov-Feller process* (also called Feller-Dynkin process) is regular with respect to its natural filtration, the discussion goes as follows.

Let E be a locally compact Polish (i.e., complete separable metric) space (usually, E is an open or closed subset of $\mathbb{R}^d$). A *Markov-Feller process* with states in E possesses a Feller semigroup $\{P(t) : t \geq 0\}$ in $C_0(E)$, with infinitesimal generator A with domain $D(A) \subset C_0(E)$. Its transition function $P(t, x, \mathrm{d}y)$ can be defined on a compact base space $\dot{E}$, the one-point compactification of E, by $P(t, x, \{\infty\}) = 1 - P(t, x, E)$, so that $P(t, x, \dot{E}) = 1$. For any initial distribution on $\dot{E}$, we denote by P the (complete) probability measure induced by the transition function $P(t, x, \mathrm{d}y)$ on the canonical space $D([0, \infty), \dot{E})$ with its Borel σ-algebra $\mathcal{B}$, its canonical process $X(t) = \omega(t)$ and its filtration $\{\mathcal{F}(t) : t \geq 0\}$, see Definition 4.18 and Theorem 4.22. Note that the probability measure P and the completion necessary to generate the filtration $\{\mathcal{F}(t) : t \geq 0\}$ depend on the initial distribution $r \to P\{X(0) \leq r\}$. All these elements constitute a standard *realization* of a Markov-Feller or Feller-Dynkin process with state in E (strictly speaking in $\dot{E}$).

Now, for a function f in $C_0(\dot{E})$ and $\lambda > 0$ the resolvent operator is given by

$$R_\lambda f(x) = \int_0^\infty \mathrm{e}^{-\lambda t}\, P(t) f(x)\, \mathrm{d}t = \int_0^\infty \mathrm{d}t \int_E \mathrm{e}^{-\lambda t}\, f(y)\, P(t, x, \mathrm{d}y),$$

and satisfies

$$R_\lambda f(x) = \mathbb{E}\{\xi \,|\, X(0) = x\}, \qquad \text{with } \xi = \int_0^\infty \mathrm{e}^{-\lambda t}\, f(X(t))\mathrm{d}t.$$

Denoting by $\mathbb{E}^x\{\cdot\}$ the (conditional) expectation with respect to the probability measure P with the Dirac measure at x as the initial distribution and applying Markov property, we find that

$$\mathbb{E}^x\{\xi \,|\, \mathcal{F}(t)\} = \int_0^t \mathrm{e}^{-\lambda s}\, f(X(s))\mathrm{d}s + \mathrm{e}^{-\lambda t} R_\lambda f(X(t)), \quad \text{a.s.} \quad \forall t > 0,$$

which proves that the right-hand side is a uniformly integrable martingale. Thus the Optional-Stopping Theorem 5.9, part (b), yields

$$\mathbb{E}^x\Big\{ \int_0^\tau \mathrm{e}^{-\lambda t}\, f(X(t))\mathrm{d}t \Big\} + \mathbb{E}^x\{\mathrm{e}^{-\lambda \tau} R_\lambda f(X(\tau))\} = R_\lambda f(x).$$

Hence, if $f = (\lambda - A)\varphi$ for some φ in the domain $D(A)$ of the infinitesimal generator A we deduce that the process $Y(t) = Y(t, \varphi, \lambda)$ given by

$$Y(t) = \mathrm{e}^{-\lambda t}\,\varphi(X(t)) + \varphi(x) + \int_0^t \mathrm{e}^{-\lambda s}\,(\lambda - A)\varphi(X(s))\mathrm{d}s,$$

is a uniformly integrable martingale relative to $(P^x, \mathcal{F}(t) : t \geq 0)$. Therefore, the following identity, so-called *Dynkin's formula*

$$\mathbb{E}\{\varphi(X_\theta)\} = \mathbb{E}\{\varphi(X_\tau)\} + \mathbb{E}\{\int_\tau^\theta A\varphi(X(t))\,\mathrm{d}t\} \tag{5.37}$$

holds for any function φ in $D(A)$ and any stopping time satisfying $P\{\tau \leq \theta < \infty\} = 1$. Moreover, the filtration $\{\mathcal{F}(t) : t \geq 0\}$ is quasi-left continuous, i.e. $\mathcal{F}(\tau) = \mathcal{F}(\tau-)$ and $X(\tau) = X(\tau-)$ (almost surely) for any predictable stopping time τ.

For instance, we refer to Rogers and Williams [112, Chapter III, pp. 227–349], Dellacherie and Meyer [36, Chapters XI–XVI], Dynkin [45], among others. Also, the reader interested in a comprehensive study on the theory of martingales may consult the books He et al. [59] or Liptser and Shiryayev [88].

5.6 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.,

$$\mathbb{E}\{f(X(t) \mid \mathcal{F}(s)\} = \mathbb{E}\{f(X(t) \mid X(s)\}, \tag{5.38}$$

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 (5.38), 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 $0 \le s \le t$, 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 $0 \le s \le t$, B in $\mathcal{B}(E)$ the function $x \mapsto p(s, x, t, B)$ is Borel measurable,

(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

$$p(s, x, t, B) = \int_E p(r, y, t, B) p(s, x, r, \mathrm{d}y), \tag{5.39}$$

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

$$\begin{aligned} P_{t_0,t_1,\ldots,t_n}(B_0 \times B_1 \times \cdots \times B_n) = \\ = \int_{B_0} \mu(\mathrm{d}x_0) \int_{B_1} p(t_0, x_0, t_1, \mathrm{d}x_1) \int_{B_2} p(t_1, x_1, t_2, \mathrm{d}x_2) \cdots \\ \cdots \int_{B_{n-1}} p(t_{n-2}, x_{n-2}, t_{n-1}, \mathrm{d}x_{n-1})\, p(t_{n-1}, x_{n-1}, t_n, B_n) \end{aligned} \tag{5.40}$$

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 (5.39), then the above relation (5.40) 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 $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P_x, t \geq 0, x \in E)$ 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 (5.38) as

$$P\{X(s+t) \in B \mid \mathcal{F}(s)\} = p(X(s), t, B), \tag{5.41}$$

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.,

$$P\{X(T+t) \in B \mid \mathcal{F}(T)\} \mathbb{1}_{T<\infty} = p(X(T), t, B) \mathbb{1}_{T<\infty}, \tag{5.42}$$

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 [39, Theorems 8 and 9, pp. 556-560], see Theorem 4.19, for conditions ensuring the right-continuity of the filtration and the strong Markov property. In the statement (5.42), 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 Blumental and Getoor [17], Dynkin [45] or more recent books, e.g., Davis [34], Rogers and Williams [112]. For instance, Morimoto [100, Sec 2.6, pp. 73–76] is a good reading to grasp the meaning of Markov property in a simple way when applied to ordinary stochastic differential equations driven by a Wiener process.

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.,

$$P\{\theta_n \le t\} = \frac{\lambda^n}{(n-1)!} \int_0^t s^{n-1} \mathrm{e}^{-\lambda x} \mathrm{d}s, \quad \forall\, t \ge 0,$$

and the *counting process* defined by

$$p(t, \omega) = \sum_{n=1}^{\infty} \mathbb{1}_{\theta_n(\omega) \le t}, \quad \forall\, t \ge 0 \tag{5.43}$$

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 $n = 0, 1, \ldots$, and for any $0 \le t_0 < t_1 < \cdots < t_n$ the family $\{p(t_0), p(t_k) - p(t_{k-1}) : k = 1, 2, \ldots, n\}$ is a set of independent random variables. The parameter λ is usually called *jump rate*.

In a *compound* Poisson process the construction (5.43) is modified as follows

$$p_c(t, \omega) = \sum_{k=1}^{\infty} \eta_n(\omega) \mathbb{1}_{\theta_n(\omega) \le t}, \quad \forall\, t \ge 0, \tag{5.44}$$

where $\{\eta_n : n = 1, 2, \ldots\}$ 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

$$\rho_c(t, B, \omega) = \sum_{k=1}^{\infty} \mathbb{1}_{\theta_k(\omega) \le t} \mathbb{1}_{\eta_k(\omega) \in B}, \quad \forall\, t \ge 0,\ B \in \mathcal{B}(\mathbb{R}^d), \tag{5.45}$$

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

$$Ih(x) = \lambda \int_{\mathbb{R}} \big[h(x+y) - h(x)\big]\, \nu(\mathrm{d}y), \quad \forall\, x \in \mathbb{R}^d, \tag{5.46}$$

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 (5.45) 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 (5.46) 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

$$M(x, B) = \lambda(x)\, \ell(\{\zeta \in [0,1] : x + j(x,\zeta) \in B\}), \tag{5.47}$$

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

$$\Theta(x, u) = \inf\big\{t \geq 0 : \exp[-t\,\lambda(x)] \leq u\big\}, \tag{5.48}$$

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

$$\begin{aligned} \theta_k &= \theta_{k-1} + \Theta(x_{k-1}, U_k), \\ x_k &= x_{k-1} + j(x_{k-1}, Z_k) \end{aligned}$$

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

$$\rho(t, B, \omega) = \sum_{k=1}^{\infty} \mathbb{1}_{\theta_k(\omega) \leq t} \mathbb{1}_{x_k(\omega) \in B}, \quad \forall\, t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d). \tag{5.49}$$

The integral operator becomes

$$Ih(x) = \lambda(x) \int_{[0,1]} \big[h(x + j(x,\zeta)) - h(x)\big]\, \ell(\mathrm{d}\zeta), \quad \forall\, x \in \mathbb{R}^d, \tag{5.50}$$

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.

5.7 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 (5.39) 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 [34, 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 $t > 0$ 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 $P(t)f \in \mathcal{D}(A)$ 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

$$\partial_t u(t) = Au(t), \quad u(0) = f, \tag{5.51}$$

which is an abstract version of the so-called *Kolmogorov backward* equation. The semigroup is determined by (5.51) 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

$$\mathbb{E}_x\{f(X(t)\} = f(x) + \mathbb{E}_x\Big\{ \int_0^t Af(X(s) \mathrm{d}s \Big\}, \tag{5.52}$$

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

$$M_f(t) = f(X(t)) - f(X(0)) - \int_0^t Af(X(s))\mathrm{d}s. \tag{5.53}$$

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.,

$$\mathbb{E}_x\{M_f(t) \mid \mathcal{F}(s)\} = M_f(s), \quad \forall t \geq s \geq 0.$$

Thus, a natural extension of the domain $\mathcal{D}(A)$ of the (strong) infinitesimal generator is as follows.

Definition 5.16 (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 (5.52) 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 [34, Chapter 1], Ethier and Kurtz [47, 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 (5.43), 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 $(p_c(t) : t \geq 0)$ (5.43), A is the integral operator I given by (5.46). 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 (5.48) and (5.49), the full description of the extended domain $\mathcal{D}(\bar{A})$, with $A = I$ as in (5.50), 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 $L^1_{\text{loc}}(\rho)$, 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 $L^1_{\text{loc}}(\rho)$. This is particular case of Davis [34, Theorem 26.14, pp. 69–74].

5.8 Poisson Processes and Queues

In a practical way, a stochastic process is a mathematical model of a probabilistic experiment that generates a sequence of numerical values as it evolves in time. Each numerical value in the sequence is modelled by a random variable, so a stochastic process is simply a (finite or infinite) sequence of random variables. However, the properties of the evolution in time become essential when the focus is on the dependencies in the sequence of values generated by the process. Typically, arrival-type or outcome-type processes occur very frequently ("arrival" of such as message receptions at a receiver, job completions in a manufacturing cell, customer purchases at a store, trials of coin tosses, etc),

where the focus is on modeling the "inter-arrival" (times between successive arrivals) are independent variables. These processes become Markov processes as the dimension is increased. In Markov processes, the experiments that evolve in time exhibit a very special type of dependence: the next value depends on the past values only through the current value, the present.

Clearly, the way how time is measured is critical. Essentially, there are only two ways, "discrete time" (where a unit time is identified and used, i.e., integer numbers are the model) and "continuous time" (where the time goes continuously, i.e., real numbers are the model). For instance, if the arrivals occur in discrete time and the inter-arrival times are geometrically distributed, this is the *Bernoulli process* described as a sequence $\{x_1, x_2, \dots\}$ of independent random variables x_i with $P\{x_i = 1\} = p$ (arrival occurs in the i trial with probability $0 < p < 1$) and $P\{x_i = 0\} = p$ (arrival does not occur in the i trial). Here, arrival also means success in the outcome under consideration. Standard calculations show that if $S_n = x_1 + x_2 + \cdots + x_n$ denotes the number of arrivals in n independent trials then

$$P\{S_n = k\} = \binom{n}{k} p^k (1-p)^k, \quad k = 0, 1, \dots, n,$$

with a mean $\mathbb{E}\{S_n\} = np$ and a variance $\mathbb{E}\{S_n - np\}^2 = np(1-p)$, i.e., a binomial distribution with parameters p and n. Similarly, if $T = \inf\{i \geq 0 : x_i = 1\}$ denotes the number of trials up to (and including) the first arrival then

$$P\{T = k\} = (1-p)^{k-1} p, \quad k = 1, 2, \dots,$$

with mean $\mathbb{E}\{T\} = 1/p$ and a variance $\mathbb{E}\{S_n - 1/p\}^2 = (1-p)/p^2$, i.e., a geometric distribution with parameter p. This yields the *memoryless* fact that the sequence of random variables $\{x_{n+1}, x_{n+2}, \dots\}$ (the future after n) is also a Bernoulli process, which is independent of $\{x_1, \dots, x_n\}$. Also, the *fresh-start property* holds, i.e., for a given n define $T_n = \inf\{i \geq n : x_i = 1\}$ then $T_n - n$ has a geometric distribution with parameter p, and is independent of the random variables $\{x_1, \dots, x_n\}$.

The equivalent of this in continuous time is the Poisson process, where the inter-arrival times are exponentially distributed, i.e., given a sequence of independent identically exponentially distributed random variables $\{\tau_1, \tau_2, \dots\}$, the counting process

$$p(t) = \begin{cases} 0 & \text{if } t < \tau_1, \\ n & \text{if } \sum_{i=1}^{n} \tau_i \leq t < \sum_{i=1}^{n+1} \tau_i \end{cases}$$

with values in $\{0, 1, 2, \dots\}$, is called a Poisson process. A realization of this process requires some properties on the probability space.

Perhaps the simplest example of an uncountable probability space is the unit interval with the Lebesgue measure $([0,1], \mathcal{L}, \ell)$, where $\mathcal{L}$ is the Lebesgue σ-algebra. The real-valued random variable $U(\omega) = \omega$ satisfies $\ell(U \leq r) =$

$(r \wedge 1) \vee 0$ for every r in $\mathbb{R}^d$, which is referred to as the *uniform distribution* on $[0,1]$. More general the Hilbert cube $\Omega = [0,1]^{\{1,2,\dots\}}$, i.e., the space of sequences $\omega = (\omega_1, \omega_2, \dots)$ with values in $[0,1]$, endowed with the product σ-algebra $\mathcal{F} = \mathcal{L}^{\{1,2,\dots\}}$ and the product measure $P = \ell^{\{1,2,\dots\}}$, provides a canonical space for a sequence of independent random variables $(U_1, U_2, \dots)$, each having uniform distribution on $[0,1]$, defined by $U_i(\omega) = \omega_i$, for every $\omega = (\omega_1, \omega_2, \dots)$. In theory, almost all statistical simulation is based on this probability space. Random number generator in computers produce sequences of numbers which are statistically indistinguishable (as much as possible) from samples $(U_1, U_2, \dots)$. Random variables with other distributions are then produced by well-known transformations. For instance, given a distribution F in the real line, i.e., a function $F : \mathbb{R} \to [0,1]$ monotonically increasing and right-continuous with $F(-\infty) = 0$ and $F(+\infty) = 1$, its inverse function defined by $F^{-1}(\rho) = \inf\{r : F(r) \geq \rho\}$ satisfies $F^{-1}(\rho) \leq r$ if and only if $\rho \leq F(r)$. Thus, if U is a random variable uniformly distributed in $[0,1]$ then $V = F^{-1}(U)$ satisfies $P(V \leq r) = F(r)$, i.e., F is the distribution of V.

Given a Borel subset E of $\mathbb{R}^d$, it is possible to construct a one-to-one Borel function $\phi : E \to [0,1]$ such that $\phi(E)$ is a Borel subset of $[0,1]$ and $\phi^{-1} : \phi(E) \to E$ is Borel measurable. From this we deduce that for any measure μ on a Borel subset E of R^d there exists a measurable function $\Upsilon : [0,1] \to E$ such that $\ell(\Upsilon^{-1}(B)) = \mu(B)$ for every B in $\mathcal{B}(E)$, the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ restricted to E. Indeed, setting $F(r) = \mu(\phi^{-1}([0,r]))$ and $F^{-1}(\rho) = \inf\{r : F(r) \geq \rho\}$ as above we may take $\Upsilon(\rho) = \phi^{-1}(F^{-1}(\rho))$ for any $F^{-1}(\rho)$ belongs to $\phi(E)$ and $\Upsilon(\rho) = 0$ otherwise.

One of the advantages of stochastic modeling is that calculations are greatly facilitated if the model is formulated as a Markov process, so that general methods for computing distributions and expectations (based on the Dynkin formula and the Kolmogorov backward equation) are available. If the randomness is in the form of point events then the prototype is the Poisson process. A non-negative real random variable T is exponentially distributed if its survivor function $F(t) = F_T(t) = P(T > t)$, for every $t \geq 0$, has the form $F(t) = \mathrm{e}^{-\lambda t}$, for some constant $\lambda > 0$. The mean and the standard deviation of T are both equal to $1/\lambda$. The *memoryless* property of the exponential distribution relative to the conditional distribution, i.e.,

$$P(T > t + s \mid T > s) = \frac{F(t+s)}{F(s)} = F(t) = \mathrm{e}^{-\lambda t},$$

make T a prototype of a (Markov) stopping time. Thus the conditional distribution of the remaining time (i.e., given $T > s$) is just the same as the unconditional distribution of T, regardless of the elapsed time s. Another way of expressing this is in terms of the *hazard rate*, which is by definition a function $h(t)$ satisfying

$$\lim_{\delta \to 0} \frac{P(T \in]s, s+\delta] \mid T > s) - h(s)}{\delta} = 0,$$

i.e., $h(s)\delta$ expresses, to first order, the probability that T occurs 'now' given that it has not occurred 'so far'. In the exponential case we have $P(T \in]s, s+\delta] \mid T > s) = 1 - \mathrm{e}^{-\lambda\delta}$, so that the hazard rate is constant, $h(t) = \lambda$. For a non-negative random variable with a general density function ψ the hazard rate is given by

$$h(s) = \frac{\psi(s)}{\Psi(s)}, \quad \forall s \in [0, c[, \quad \text{with}$$
$$\Psi(s) = \int_s^\infty \psi(r)\mathrm{d}r, \quad c = \inf\{r : \Psi(r) = 0\},$$

where $\Psi(s)$ is the corresponding survivor function. In fact, there is a one-to-one correspondence between h and Ψ based on the ordinary differential equation $\dot{\Psi} = -h\Psi$ and the initial condition $\Psi(0) = 1$. Thus the exponential is the only distribution with constant hazard rate.

Let us construct a sequence of independent identically distributed (exponential with parameter $\lambda > 0$) random variables $(\tau_1, \tau_2, \ldots)$ in the canonical Hilbert cube $(\Omega, \mathcal{F}, P)$. Let $(U_1, U_2, \ldots)$ be the canonical sequence of independent random variables each having uniform distribution on $[0, 1]$ as above. Then setting $\Psi(t) = \mathrm{e}^{-\lambda t}$ and $\Psi^{-1}(u) = \inf\{t \geq 0 : \Psi(t) \leq u\}$, for every $u > 0$, we define $\tau_i = \Psi^{-1}(U_i(\omega)) = -\ln[U_i(\omega)]/\lambda$, for every $i \geq 1$, which satisfies $P(\tau_i > t) = \Psi(t)$, i.e., exponentially (with parameter λ) distributed and independent.

Now define $\theta_0 = 0$, $\theta_n = \tau_1 + \tau_2 + \ldots + \tau_n$, which has $\Gamma(\lambda, n)$, i.e., $P(\theta_n \in \mathrm{d}t) = \big(\lambda^n s^{(n-1)}/(n-1)!\big)\mathrm{e}^{-\lambda t}\mathrm{d}t$, and

$$N(t) = \sum_{i=1}^{\infty} \mathbb{1}_{t \geq \theta_i}, \qquad \text{i.e.,} \quad N(t) = n \quad \text{if} \quad \theta_n \leq t < \theta_{n+1}.$$

The sample functions of $(N(t) : t \geq 0)$ are right-continuous step functions with jumps of height 1 at each τ_i, in particular it is cad-lag, belonging to the canonical sample space $D([0, \infty[)$. The random variable $N(t)$ has a Poisson distribution $P(N(t) = n) = \mathrm{e}^{-\lambda t}(\lambda t)^n/n!$, with mean $\mathbb{E}\{N(t)\} = \lambda t$. Denote by $(\mathcal{F}_t : t \geq 0)$ it natural filtration, i.e., $\mathcal{F}_t$ is the σ-algebra generated by the random variables $\{N(s) : 0 \leq s \leq t\}$. Fix $t > 0$ and denote by θ^t the last jump time before t, i.e, $\theta^t(\omega) = \theta_n(\omega)$, with $\theta_n(\omega) \leq t < \theta_{n+1}(\omega)$ and $n = n^t(\omega)$. In view of the memoryless property of the exponential, if $\tau_1^* = \theta_{n+1} - t$ and $\tau_i^* = \tau_{n+i}$, for $i \geq 2$, with $n = n^t(\omega)$, then the conditional distribution of τ_1^* given $\mathcal{F}_t$ (or equivalently, given that $\tau_{n+1} > t - \theta_n$) is exponential, $P(\tau_1^* > s \mid \mathcal{F}_t) = \mathrm{e}^{-\lambda t}$, and so the sequence $\{\tau_1^*, \tau_2^*, \ldots\}$ is independent identically distributed (exponential with parameter $\lambda > 0$). It follows that $N^*(s) = N(t+s) - N(t)$, $s \geq 0$ is a Poisson process independent of $\mathcal{F}_t$, i.e., the process 'restart' at time t. In particular, it has independent increments, i.e., $N(t_2) - N(t_1)$ and $N(t_4) - N(t_3)$ are independent variables for any $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$. This implies that $(N(t) : t \geq 0)$ is a Markov process, indeed, for any bounded and Borel measurable function and

$t > s \geq 0$ we have

$$\mathbb{E}\{f(N(t)) \mid \mathcal{F}_s\} = \mathrm{e}^{\lambda(t-s)} \sum_{k=0}^{\infty} f(k + N_s) \frac{[\lambda(t-s)]^k}{k!},$$

since $N(t) - N(s)$ is Poisson distributed with mean $\lambda(t-s)$. The Poisson process may be considered as a Markov process in either the integer numbers $E = \{0, \pm 1, \pm 2, \ldots\}$ or the non-negative integer numbers $E = \{0, 1, 2, \ldots\}$. The process $i + N(t)$ yields probability measure P_i and the transition function is

$$p(i,t,j) = \begin{cases} \mathrm{e}^{\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}, & \text{if } j \geq i, \\ 0 & \text{otherwise,} \end{cases}$$

for any i, j in E. This defines a semigroup $P(t) = \mathbb{E}\{f(x + N(t))\}$ on the space $B(E)$ of real (Borel) bounded functions on E. The infinitesimal generator A is

$$Af(x) = \lim_{t \to 0} \frac{\mathbb{E}\{f(x + N(t))\}}{t} = \lambda[f(x+1) - f(x)], \quad \forall x \in E,$$

where the domain $\mathcal{D}(A)$ of the strong infinitesimal generator is the space functions f for which the above limit exists uniformly in x. Consider the process

$$M_x^f(t) = f(x + N(t)) - f(x) - \\ - \lambda \int_0^t [f(x + N(r) + 1) - f(x + N(r))] \mathrm{d}r, \quad t \geq 0$$

for any f in $B(E)$ and x in E. In view of the independent increment property and the fact that $N(t) - N(s)$ is Poisson distributed, we have for $t > s$

$$\mathbb{E}\{f(x + N(t)) - f(x + N(s)) \mid \mathcal{F}_s\} = \\ = \mathrm{e}^{-\lambda(t-s)} \sum_{k=0}^{\infty} [f(k + x) - f(x)] \frac{[\lambda(t-s)]^k}{k!}$$

and

$$\mathbb{E}\{\lambda \int_s^t [f(x + N(r) + 1) - f(x + N(r))] \mathrm{d}r\} = \\ = \lambda \sum_{k=0}^{\infty} [f(x + k + 1) - f(x + k)] \int_s^t \mathrm{e}^{-\lambda r} \frac{(\lambda r)^k}{k!} \mathrm{d}r,$$

which yields $\mathbb{E}\{M_x^f(t) - M_x^f(s) \mid \mathcal{F}_s\} = 0$, i.e., $(M_x^f(t) : t \geq 0)$ is a martingale. Actually, this calculation remains valid for any function (because E is countable, all functions are Borel measurable) such that $\mathbb{E}\{|f(x + N(t))|\} < \infty$, for every x in E and $t \geq 0$. By the optional sampling theorem, the process $M_x^{f,n}(t) =$

$M_x^f(t \wedge n \wedge \theta_n)$ is also a martingale, since $n \wedge \theta_n$ is a bounded stopping time. However, the process $(M_x^{f,n}(t) : t \geq 0)$ involves only the values of f on the finite set $\{x, x+1, \ldots, x+n\}$. Therefore, the process $(M_x^f(t) : t \geq 0)$ is a local martingale for any function. Thus the domain $\mathcal{D}(\bar{A})$ of the extended infinitesimal generator $\bar{A}$ consists of all functions $f : E \to \mathbb{R}$ with not restriction at all.

The *renewal process* is closely related to the Poisson process. It is a point process $(N(t) : t \geq 0)$ defined in a similar way to the Poisson process but with the inter-arrival time τ_i now being a sequence of independent identically distributed random variables with some density function ψ on $[0, \infty)$, not necessarily exponential. The process clearly 'restarts' at each 'renewal time' θ_i and the well-known renewal equation

$$m(t) = \mathbb{E}\{N(t)\}, \qquad m(0) = 0,$$
$$m(t) = \int_0^t [1 + m(t-r)]\psi(r)\mathrm{d}r, \quad \forall t \geq 0,$$

which can be solved by the Laplace transform methods.

The sequences of inter-arrival time $\{\tau_1, \tau_2, \ldots\}$ is now constructed as follows. Then setting $\Psi(t) = \int_t^\infty \psi(r)\mathrm{d}r$ and $\Psi^{-1}(u) = \inf\{t \geq 0 : \Psi(t) \leq u\}$, for every $u > 0$, we define $\tau_i = \Psi^{-1}(U_i(\omega))$, for every $i \geq 1$, which satisfies $P(\tau_i > t) = \Psi(t)$, i.e., independent identically distributed with density ψ.

When τ_i is not exponentially distributed, the memoryless property does not hold and the conditional distribution of the residual time $\tau_1^* = \theta_{n+1} - t$ (as defined above for the Poisson process) given $\mathcal{F}_t$ depends on the time $t - \theta_n$ since the last jump. Therefore, the renewal process itself is not a Markov process, if we add a new variable $S(t) = t - \theta_n$, the time since the last jump where $\theta_n \leq t < \theta_{n+1}$, then the new two-component process $X = \{(N(t), S(t)) : t \geq 0\}$ is a Markov process on $E = \{0, 1, 2, \ldots\} \times [0, \infty)$. Its evolution can be simulate as follows. For a fixed (n, s) and with $\psi(t)$, $\Psi(t)$ as above, we set first

$$\lambda(r) = \begin{cases} \dfrac{\psi(r)}{\Psi(r)} & \text{if } 0 \leq r < c_\Psi, \\ 0 & \text{otherwise,} \end{cases}$$

with $c_\Psi = \inf\{r : \Psi(r) = 0\}$, and then

$$\Psi(s,t) = \exp\Big(-\int_s^t \lambda(r)\mathrm{d}r\Big), \qquad \forall t \geq s \geq 0.$$

Note that $\Psi(0,t) = \Psi(t)$, for every $t \geq 0$. Thus, we re-define

$$\Psi^{-1}(s,u) = \inf\{t \geq 0 : \Psi(s,t) \leq u\}, \quad \tau_1(\omega) = \theta_1(\omega) = \Psi^{-1}(s, U_1(\omega)),$$

with the convention that $\Psi^{-1}(s,u) = +\infty$ if $\Psi(s,t) > u$ for every $t \geq 0$. The sample path $X(t,\omega)$ is then $(n, s+t)$ if $0 \leq t < \theta_1(\omega)$ and $(n+1, 0)$ if

$t = \theta_1(\omega) < \infty$. Next, if $\theta_1(\omega) < \infty$ we restart with the initial state $(n+1, 0)$ and the same recipe. This is $\tau_2(\omega) = \Psi^{-1}(0, U_2(\omega))$, $\theta_2(\omega) = \theta_1(\omega) + \tau_2(\omega)$ and

$$X(t,\omega) = \begin{cases} (n+1, t-\theta_1(\omega)) & \text{if } \theta_1(\omega) \le t < \theta_2(\omega), \\ (n+2, 0) & \text{if } t = \theta_2(\omega) < \infty, \end{cases}$$

and so on. The key point is that this construction generalizes to a much more general situation.

The formal expression of the infinitesimal generator is

$$Af(n,s) = \lim_{t\to 0} \frac{\mathbb{E}\{f(n+N(t), s+S(t)\} - f(n,s)}{t} = \\ = \partial_s f(n,s) + \lambda(s)[f(n+1,0) - f(n,s)],$$

where the hazard rate $\lambda(t) = \psi(t)/\Psi(t)$ and ∂_s means the partial derivative in the second variable, i.e., in s. The domain $\mathcal{D}(A)$ of the strong infinitesimal generator should include conditions to ensure that the above limit exists uniformly in (n, s), in particular $f(n, s)$ should be differentiable in s. However, the domain $\mathcal{D}(\bar{A})$ of the extended infinitesimal generator would only impose that $s \mapsto f(n, t)$ be absolutely continuous.

Another typical example is a *single-server queue*. Customers arrive at a queue at random times $\{\theta_1 \le \theta_2 \le \cdots\}$ which require a service time $\{\varsigma_1, \varsigma_2, \ldots\}$, measured in units of time for processing. The total service load presented up to time t is $L(t) = L_0 + \sum_i \varsigma_i \mathbb{1}_{\theta_i \le t}$, where $L_0 \ge 0$ is the service load existing at time 0. The virtual waiting time $V(t)$ is the unique solution of the equation

$$V(t) = L(t) - \int_0^t \mathbb{1}_{V(r)>0} \mathrm{d}r, \quad \forall t \ge 0,$$

and represents the time a customer arriving at time t waits for service to begin, or equivalently, the amount of unprocessed load at time t.

A similar way to describe a queueing system is by means of the relation $Q(t) = Q_0 + A(t) - D(t)$, where $(A(t) : t \ge 0)$ and $(D(t) : t \ge 0)$ are non-explosive point processes without common jumps, i.e., $A(t) = n$ for $\theta_n \le t < \theta_{n+1}$ and $D(t) = n$ for $\vartheta_n \le t < \vartheta_{n+1}$, $\theta_0 = \vartheta_0 = 0$, $\theta_n, \vartheta_n \to \infty$ as $n \to \infty$ and $P(\theta_i = \vartheta_j) = 0$, for every i, j. The random variable Q_0 is the initial state and the state process satisfies $Q(t) \ge 0$, for any $t \ge 0$, which is interpreted as the number of customer waiting in line or being attended by the server (i.e., in the system). The processes A and D are called arrival and departure processes. Thus, if A and B are two nonexplosive point processes without common jumps then to achieve the condition $Q(t) \ge 0$ we set $Y(t) = Q_0 + A(t) - B(t)$ and $m(t) = \min\{Y(r) \wedge 0 : r \in [0, t]\}$. Hence, a simple queueing system $Q(t) + Q_0 + A(t) - D(t)$ can be constructed with $Q(t) = Y(t) - m(t)$, $D(t) = \int_0^t \mathbb{1}_{Q(r-)>0} dB(r)$, where also $m(t) = \int_0^t \mathbb{1}_{Q(r-)=0} dB(r)$.

There is a conventional classification $A/B/n$ of queueing systems, where A refers to the arrival process (i.e., statistics of the increasing sequence of random

variables $\{\theta_1, \theta_2, \ldots\}$), B to the service process (i.e., statistics of the sequence of random variables $\{\varsigma_1, \varsigma_2, \ldots\}$) and n is the number of servers. For instance, consider a $M/G/1$ queue, i.e., the letter M (for Markov) means that arrival are independent and exponential, i.e., from a Poisson process, and G (for general) means that the service time independent identically distributed with some arbitrary distribution on $(0, \infty)$.

A variable ν indicates whether the queue is busy $\nu = 1$ or empty $\nu = 0$. This means that ν vanishes, $\nu = 0$, if and only if the virtual waiting time vanishes, $v = 0$. Thus, starting from a time $t_0 \geq 0$ with $\nu(t_0) = 1$ and $V(t_0) = v$, the process $v(t) = V(t)$ decreases at unit rate until it hits zero, say at time t_1. Then $\nu(t_1)$ becomes zero, $\nu(t_1) = 0$, and $v(t) = 0$ until a new arrival $t_2 > t_1$ which takes an exponential time, and (ν, v) jumps to $(1, \vartheta)$, i.e., $\nu(t_2-) = 0$, $\nu(t_2+) = 1$, $v(t_2-) = 0$ and $v(t_2+) = \varsigma$. The state (ν, v) has a Markov evolution on the set $E = \{(0,0)\} \cup \{1\} \times (0, \infty)$, which is normalized to be a cad-lag process. Roughly speaking, if the initial state is $(1, v)$ then after a short time δ the state becomes $(1, v - \delta)$ with probability $(1 - \lambda\delta)$, while with probability $\lambda\delta$ the Markov process jumps to the new state $(1, v + \varsigma - \delta)$, where ς has the distribution on $(0, \infty)$ of the services time, namely F_ϑ, and λ is the parameter of the exponential distribution of the arrival times. With this in mind, the infinitesimal generator has the expression

$$Af(1, v) = -\partial_v f(1, v) + \lambda \int_{(0,\infty)} [f(1, v + z) - f(1, v)] F_\vartheta(\mathrm{d}z),$$

$$Af(0, 0) = \lambda \int_{(0,\infty)} [f(1, z) - f(0, 0)] F_\vartheta(\mathrm{d}z),$$

for any v in $(0, \infty)$. It is clear that this formula of the infinitesimal generator A does not include the fact that the process jumps from $(1, 0)$ to $(0, 0)$, immediately after hitting $(1, 0)$. This is the *boundary conditions*

$$f(1, 0) = \int_{[0,\infty)} f(1, z) F_\vartheta(\mathrm{d}z)$$

added to the strong domain $\mathcal{D}(A)$ or extended $\mathcal{D}(\bar{A})$.

A construction of the Markov process starting at $x = (\nu, v)$ is described in the canonical Hilbert cube $(\Omega, \mathcal{F}, P)$, where $(U_1, U_2, \ldots)$ is sequence of independent random variables each having uniform distribution on $[0, 1]$. First we set

$$Q(x, B) = \int_{[v,\infty)} \mathbb{1}_B(z) F_\varsigma(\mathrm{d}z),$$

$$\mathtt{X}(x, t) = \begin{cases} (\nu, v - t) & \text{if } \nu = 1, \\ (\nu, v) & \text{otherwise,} \end{cases}$$

and

$$\mathtt{T}(x) = \begin{cases} v & \text{if } \nu = 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Given an initial state $x = (\nu, v)$ in $E = \{(0,0)\} \cup \{1\} \times (0, \infty)$ we define

$$\Psi(x,t) = \mathbb{1}_{t<\mathtt{T}(x)} \mathrm{e}^{-\lambda t},$$

the survivor function of the first jump time θ_1 of the process and its (generalized) inverse

$$\Psi^{-1}(x,u) = \inf\{t \geq 0 : \Psi(x,t) \leq u\}, \quad \tau_1(\omega) = \theta_1(\omega) = \Psi^{-1}(x, U_1(\omega)),$$

with the convention that $\Psi^{-1}(x,u) = +\infty$ if $\Psi(x,t) > u$ for every $t \geq 0$. This yields $P(\theta_1 > t) = \Psi(x,t)$. As mentioned above, we are working in the canonical Hilbert cube and there exist a measurable function Υ from $E \times [0,1]$ into $\mathring{E} = \{1\} \times (0, \infty)$ such that $\ell(\{u : \Upsilon(x,u) \in B\}) = Q(x,B)$, for every B in $\mathcal{B}(E)$, where ℓ is the Lebesgue measure on $[0,1]$. The sample path $X(t,\omega)$ is defined up to the first jump as follows:

$$\begin{aligned} &X(t,\omega) = \mathtt{X}(x,t), \quad \text{if} \quad 0 \leq t < \theta_1(\omega), \\ &X(\theta_1(\omega),\omega) = \Upsilon(\mathtt{X}(x,\theta_1(\omega)),\omega), U_2(\omega)), \quad \text{if} \quad \theta_1(\omega) < \infty. \end{aligned}$$

Note that when $\nu = 1$, as long as $t < \theta_1(\omega)$ we have $v - t > 0$. If $t = \theta_1(\omega) < \infty$ then $v - t \geq 0$. On the other hand, when $\nu = 0$ we have $\theta_1(\omega) < \infty$ and $X(t,\omega) = (0,v)$ for every $t < \theta_1(\omega)$. In any case, the definition of Υ ensure that $X(\theta_1(\omega),\omega)$ belongs to $\mathring{E} = \{1\} \times (0,\infty)$. Now, if $\theta_1(\omega) < \infty$ the process restarts from $X(\theta_1(\omega),\omega)$ according to the same recipe. Thus, if $\theta_1(\omega) < \infty$ we define

$$\tau_2(\omega) = \Psi^{-1}(X(\theta_1(\omega),\omega), U_3(\omega)), \qquad \theta_2(\omega) = \theta_1(\omega) + \tau_2(\omega)$$

and the sample path $X(t,\omega)$ up to the next jump is given by

$$\begin{aligned} &X(t,\omega) = \mathtt{X}(x, t - \theta_1(\omega)), \quad \text{if} \quad \theta_1(\omega) \leq t < \theta_2(\omega), \\ &X(\theta_2(\omega),\omega) = \Upsilon(\mathtt{X}(x,\tau_2(\omega)),\omega), U_4(\omega)), \quad \text{if} \quad \theta_2(\omega) < \infty, \end{aligned}$$

and so on.

This procedure define the sample path $X(t,\omega)$ if the sequence $\theta_k(\omega) \to \infty$. Hence, a common assumption is to impose that

$$\mathbb{E}\{N(t)\} < \infty, \quad \text{with} \quad N(t) = \sum_k \mathbb{1}_{t \geq \theta_k},$$

which yield some condition on the distribution $F_\varsigma(t)$ of the sequence $\{\varsigma_1, \varsigma_2, \ldots\}$ associated with the service time. Since

$$P(t < \tau_i < \infty) = \mathbb{1}_{t<\mathtt{T}(x)} \mathrm{e}^{-\lambda t}$$

a condition on the type $P(\varsigma > \varepsilon) = 1$ for some positive ε, on the service time distribution ensures the required assumption.

If the arrival process is a renewal process instead of a Poisson process then we need one more variable to have a Markov process, the time since the last jump

s, so that the state is $x = (\nu, v, s)$ in $E = \{(0,0)\} \times [0,\infty) \cup \{1\} \times (0,\infty) \times [0,\infty)$ and an *intensity* or hazard rate $\lambda(t) = \psi(t)/\Psi(t)$, where ψ is the density of the arrival times. The previous simulation can be adapted, where $s + t$ is the evolution in last variable, which is reset to zero each time a new arrival occurs.

It is clear that the above technique can be used for more general situation, e.g., a $G/G/n$ queue system and many other stochastic models.

Another type of common jump process is the so-called *doubly stochastic Poisson process* or *conditional Poisson process* $(N(t) : t \geq 0)$ with stochastic intensity $\lambda = \lambda(t,\omega) \geq 0$ relative (i.e., adapted) to the filtration $(\mathcal{F}_t : t \geq 0)$ on a probability space $(\Omega, \mathcal{F}, P)$, which is defined by the condition

$$\mathbb{E}\big\{ \mathrm{e}^{i\zeta[N(t)-N(s)]} \mid \mathcal{F}_s \big\} = \exp\Big[\big(\mathrm{e}^{iu} - 1\big) \int_s^t \lambda(r) \mathrm{d}r \Big],$$

for every $t \geq s \geq 0$, where $\lambda(r)$ is an $\mathcal{F}_0$-measurable and almost surely integrable function in $[0,t]$. Usually, the intensity takes the form $\lambda(t,\omega) = \lambda(t, Y(t,\omega))$, where the process $(Y(t) : t \geq 0)$ is $\mathcal{F}_0$-measurable and the function $(t,y) \mapsto \lambda(t,m)$ is a nonnegative Borel measurable with some appropriate integrability conditions. This means that on one hand we have a Poisson process with intensity $\lambda(t,y)$ where the parameter y is replaced by an independent process $y = Y(t)$ and the σ-algebra $\mathcal{F}_0$ is augmented with the σ-algebra generated by $(Y(t) : t \geq 0)$ to meet the $\mathcal{F}_0$-measurability condition. A conditional Poisson process $(N(t) : t \geq 0)$ is characterized by the property

$$\mathbb{E}\Big\{ \int_0^\infty X(t) \mathrm{d}N(t) \Big\} = \mathbb{E}\Big\{ \int_0^\infty X(t) \lambda(t) \mathrm{d}t \Big\},$$

valid for any $\mathcal{F}_t$-predictable process $(X(t) : t \geq 0)$, c.f., Bremaud [21, Chapter 2, Theorem T4]. Thus its predictable jumps compensator is the integral process $\langle N \rangle(t) = \int_0^t \lambda(r,\omega) \mathrm{d}r$. Conditional Poisson processes are in general not Markov processes, however, if the intensity function $\lambda(t,\omega) = \lambda(Y(t))$, where $(Y(t) : t \geq 0)$ is a $\mathcal{F}_0$-measurable Markov process, then the couple (N, Y) becomes a Markov process with a suitable filtration, actually a compound Poisson process.

To end this section let us mention the so-called *multivariate point processes* which are defined by means of two sequences $\{\theta_0, \theta_1, \theta_2, \ldots\}$ and $\{\zeta_1, \zeta_2, \ldots\}$ of random variables with values in $[0,\infty]$ and $\{1, 2, \ldots, d\}$, respectively, and satisfying $\theta_0 = 0$, if $\theta_n < \infty$ then $\theta_n < \theta_{n+1}$, and called *nonexplosive* when $\theta_\infty = \lim_n \theta_n = \infty$. The sample path is defined by the *d-counting* process $(N_i(t) : t \geq 0)$, $i = 1, 2, \ldots, d$

$$N_i(t) = \sum_{n=1}^{\infty} \mathbb{1}_{\theta_n \leq t} \, \mathbb{1}_{\zeta_n = i}, \forall t \geq 0.$$

Both the d-vector process $(N(t) : t \geq 0)$ with nonnegative integer-values components as above and the double sequence $\{\theta_n, \zeta_n : n = 1, 2, \ldots\}$ are called d-variate point process. Note each component $(N_i(t) : t \geq 0)$ is a (simple or univariate) point process and that only one component jumps at a given time,

i.e., there is not common jumps among all the processes $(N_i(t) : t \geq 0)$, for $i = 1, 2, \ldots, d$. For instance, if the double sequence $\{\theta_n, \zeta_n : n = 1, 2, \ldots\}$ is independent identically distributed, θ_n $\Gamma(\lambda, n)$ distributed and ζ_n such that $P(\zeta_n = 1) = p$, $P(\zeta_n = 2) = 1 - p$, with some constant p in $(0, 1)$, then interpreting $N_1(t)$ as the births and $N_2(t)$ as the deaths up to time t of a given population, the expression $N_1(t) - N_2(t)$ is a birth-and-death process.

For instance, the reader is refereed to the books by Borovkov [19, Chapters 5 and 6, pp. 155–193], Bremaud [21, 22], Daley and Vere-Jones [33, 32], Davis [34, Chapter 1], Revuz and Yor [111, Section XII.1, pp. 471–480], among others.

5.9 Piecewise Deterministic Processes

Non-diffusion stochastic models called piecewise-deterministic Markov processes (PDP) is proposed as a general framework for studying problems involving non-diffusion continuous-time dynamical systems whose deterministic motion is punctuated by random jumps. A great number of applications in engineering systems, operations research, management science, economic and applied probability show the importance of these systems. Queuing systems, investment planning, stochastic scheduling and inventory systems are some examples, we refer to Davis [34] for a comprehensive study.

5.9.1 Vector Fields and ODE

Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be a globally Lipschitz continuous function, i.e., there exists a constant $M > 0$ such that $|g(x) - g(x')| \leq M|x - x'|$ for every x, x' in $\mathbb{R}^d$. It is well-know that the ordinary differential equation (ODE) relative to g, i.e., the initial value problem (IVP) $\dot{x}(t) = g(x(t))$, $x(0) = x$, has a unique solution defined for every x in $\mathbb{R}^d$. We denote by $\mathtt{X}(x, t)$ its solution, i.e.,

$$\partial_t \mathtt{X}(x, t) = g(\mathtt{X}(x, t)), \quad \forall t \in \mathbb{R}, \qquad \mathtt{X}(x, 0) = x, \quad \forall x \in \mathbb{R}^d,$$

which has the properties:

(1) the map $\mathtt{X}_t : x \mapsto \mathtt{X}(x, t)$ is Lipschitz continuous (uniformly in t), one-to-one and onto, indeed, $\mathtt{X}^{-1}(x, t) = \mathtt{X}(x, -t)$, for every x in $\mathbb{R}^d$ and t in $\mathbb{R}$,

(2) the family $\{\mathtt{X}_t : t \in \mathbb{R}\}$ is a group, i.e., $\mathtt{X}_t \circ \mathtt{X}_s = \mathtt{X}_{t+s}$, or more explicitly $\mathtt{X}(x, t + s) = \mathtt{X}(\mathtt{X}(x, s), t)$, for every x in $\mathbb{R}^d$ and t, s in $\mathbb{R}$.

This is referred to as an *homogeneous Lipschitz flow* in $\mathbb{R}^d$.

If f is a real valued continuously differentiable function, i.e., f in $C^1(\mathbb{R}^d)$, then we may consider $\mathtt{X}$ as a first-order differential operator, $\mathtt{G} : C^1(\mathbb{R}^d) \to C^0(\mathbb{R}^d)$, defined by

$$\mathtt{G}f(x) = \sum_{i=1}^{d} g_i(x) \partial^i f(x), \quad \forall x \in \mathbb{R}^d,$$

where $\partial^i f$ means the first partial derivative with respect to the variable x_i. By means of the chain rule we deduce that $t \mapsto x(t)$ is a solution of the ODE relative to g if and only if

$$\mathtt{G}f(x(t)) = \sum_{i=1}^{d} g_i(x(t))\partial^i f(x(t)), \quad \forall t \in \mathbb{R},, \quad \forall f \in C^1(\mathbb{R}^d),$$

which is a coordinates-free form of the differential equation, the operator $\mathtt{G}$ (and the function g) is know as the *vector field* associated with the flow $\mathtt{X} = \{\mathtt{X}_t : t \in \mathbb{R}^d\}$.

If the function g is continuously differentiable then the function homeomorphism $x \mapsto \mathtt{X}(x,t)$ is indeed a diffeomorphism and it satisfies the linear system of ODEs

$$\partial_t \mathtt{X}_{i,j}(x,t) = \sum_{k=1}^{d} g_{i,k}[\mathtt{X}(x,t)]\, \mathtt{X}_{k,j}(x,t),$$

for every $i, j = 1, \dots, d$, x in $\mathbb{R}^d$ and $t \in \mathbb{R}$, where the subscript i, j as in $\mathtt{X}_{i,j}$ denotes the the first partial derivative in the x_j variable of the i component of $\mathtt{X}(x,t)$, i.e., $\partial_j \mathtt{X}_i(x,t)$.

This analysis can be extended to non-flat manifolds and the assumption on g can be weakened. Of particular interest for us is the case where $\mathbb{R}^d$ is replaced by a finite intersection of nonempty domain D having a representation of the form

$$D = \{x \in \mathbb{R}^d : \phi(x) < 0\}, \quad \phi \in C^1(\mathbb{R}^d), \quad |\nabla\phi(x)| \geq 1,\ \forall x \in \partial D, \quad (5.54)$$

which implies that D is an open set with an outward unit normal vector given by $\nabla\phi(x)/|\nabla\phi(x)|$ on the boundary ∂D. The function g defining the vector field $\mathtt{G}$ and the flow $\mathtt{X}$ is usually assumed locally Lipschitz continuous in $\bar{D}$ and with linear growth when D is unbounded. Local uniqueness and existence of solution to the IVP is ensure by the local Lipschitz continuity, and so the solution is extended to its maximum interval of existence with bounds (which may be infinite) $t^{\wedge}_{\bar{D},x} < t^{\vee}_{\bar{D},x}$. If $t^{\vee}_{\bar{D},x} < \infty$ (or $t^{\wedge}_{\bar{D},x} < \infty$) then we assume that there exist $t < t^{\vee}_{\bar{D}}$ ($t > t^{\wedge}_{\bar{D}}$) such that $\mathtt{X}(x,t)$ does not belongs to $\bar{D}$. This *non-explosive* condition effectively rule out the case where an explosion occurs in closure $\bar{D}$. The linear growth condition implies this non-explosion assumption, but it is not necessary. Thus under this non-explosion condition we can define the first exit time from any Borel subset E of $\bar{D}$ as follows

$$\tau_E(x) = \inf\{t \geq 0 : \mathtt{X}(x,t) \notin E\}, \quad \forall x \in \bar{D},$$

with the convention that $\tau_E(x) = \infty$ if $\mathtt{X}(x,t)$ belongs to E for every $t \geq 0$. It is clear that $\tau_E(x) = 0$ for any x in $\bar{D} \setminus E$ and that $x \mapsto \tau_E(x)$ is a measurable $[0,\infty]$-valued function. Also, if $g(x) \cdot \nu(x) > 0$, with $\nu(x) = \nabla\phi(x)/|\nabla\phi(x)|$ the exterior unit normal vector at x in ∂D, then $\tau_{\bar{D}}(x) = 0$. On the other hand, if

$g(x) \cdot \nu(x) \leq 0$ for every x in ∂D then $\tau_{\bar{D}}(x) > 0$ for any x in ∂D, i.e., $\mathtt{X}(t,x)$ belongs to $\bar{D}$ for every x in $\bar{D}$ and $t \geq 0$.

It is convenient to divide the flow $\mathtt{X} = \{\mathtt{X}_t : t \in \mathbb{R}^d\}$ into two flows, a forward flow $\mathtt{X} = \{\mathtt{X}_t : t \geq 0\}$ and a backward flow $\mathtt{X} = \{\mathtt{X}_t : t \leq 0\}$.

Theorem 5.17. *Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be a locally Lipschitz continuous function which yields a forward flow $\mathtt{X} = \{\mathtt{X}_t : t \geq 0\}$ without explosions in $\bar{D}$ as above, so that the first exist times τ_D from the open set D and $\tau_{\bar{D}}$ from the closed set $\bar{D}$. Then the functions $x \mapsto \tau_D(x)$ and $x \mapsto \tau_{\bar{D}}(x)$ are lower and upper semi-continuous, respectively. If $\partial^0 D = \{x \in \partial D : \tau_{\bar{D}}(x) = 0\}$ then $\tau_{\bar{D}}(\cdot)$ is continuous if and only if $\partial^0 D$ is closed.*

Proof. Take x and s such that $\tau_{\bar{D}}(x) < s$. Then there exists s' such that $\mathtt{X}(x,s')$ does not belongs to $\bar{D}$. If $x_n \to x$, by the continuity of $\mathtt{X}(\cdot,s')$ and because $\mathbb{R}^d \setminus \bar{D}$ is open, there exists N such that $\mathtt{X}(x_n,s')$ does not belongs to $\bar{D}$, for any $n \geq N$. Thus $\tau_{\bar{D}}(x_n) \leq s'$. This proves that $\limsup_n \tau_{\bar{D}}(x_n) \leq \tau_{\bar{D}}(x)$, i.e., $x \mapsto \tau_{\bar{D}}(x)$ is upper semi-continuous.

Similarly, take x and s such that $\tau_D(x) > s$. Then the closed set $\{\mathtt{X}(x,t) : 0 \leq t \leq s\}$ is contained in the open set D. If $x_n \to x$, again by continuity, there exists N such that $\{\mathtt{X}(x_n,t) : 0 \leq t \leq s\}$ contained in D, for any $n \geq N$. Thus $\tau_D(x_n) \geq s$. This proves that $\liminf_n \tau_D(x_n) \geq \tau_D(x)$, i.e., $x \mapsto \tau_D(x)$ is lower semi-continuous.

If $\tau_{\bar{D}}(\cdot)$ is continuous then it is clear that $\partial^0 D$ is closed. On the other hand, take x in $\bar{D}$ such that $\tau_{\bar{D}}(x) < \infty$. Since the composition property yields $\mathtt{X}(x, t + \tau_{\bar{D}}(x)) = \mathtt{X}(\mathtt{X}(x, \tau_{\bar{D}}(x)), t)$, we deduce that $\mathtt{X}(x, \tau_{\bar{D}}(x))$ must belongs to $\partial^0 D$. Hence the forward flow exists necessarily though $\partial^0 D$. If $\tau_{\bar{D}}(x) > s$ then the closed set $\{\mathtt{X}(x,t) : 0 \leq t \leq s\}$ has a positive distance to the closed set $\partial^0 D$. As in the case of $\tau_D(\cdot)$ we deduce that $\tau_{\bar{D}}(\cdot)$ is lower semi-continuous. □

We state for further reference the following concept.

Definition 5.18 (locally Lipschitz continuous forward flow)**.** Let E be a set in $\mathbb{R}^d$ having the following property, either

(1) E is the union of an open set $\mathring{E}$ in $\mathbb{R}^d$ and a relative open part $\partial E \setminus \partial^0 E$ of its boundary ∂E (which is the *non-active boundary*, so that the active boundary $\partial^0 E$ is closed); the interior set $\mathring{E}$ is a finite intersection of nonempty domain D having a representation of the form (5.54), or

(2) after a permutation of coordinates the set E has the form $E = E_1 \times E_2$, where E_1 is as in (1) but relative to $\mathbb{R}^{d_1}$ with $d_1 < d$ and E_2 is a single point in $\mathbb{R}^{d-d_1}$ and the corresponding vector field g has only d_1 non-zero components, i.e., g can be considered as a function from $\mathbb{R}^d$ into $\mathbb{R}^{d_1}$.

If g is a locally Lipschitz function from the closure $\bar{E}$ into $\mathbb{R}^d$ then the following ODE $\dot{x}(t) = g(x(t))$, for any $t > 0$, can be uniquely solved for any given initial condition $x(0) = x$ in $\mathbb{R}^d$ and locally define flow $(x,t) \mapsto \mathtt{X}(x,t)$ as the solution of the above IVP on the maximal interval of existence $[0, t_x^\vee[$, i.e., for any x in $\mathbb{R}^d$ the solution $x(t) = \mathtt{X}(x,t)$ is defined for every $0 \leq t < t_x^\vee \leq +\infty$. We say that $\mathtt{X} = \{\mathtt{X}(x,t) : t \geq 0,\ x \in E\}$ is a *locally Lipschitz forward flow* associated to

the vector field g on $E \subset \mathbb{R}^d$ or to first-order differential operator $\mathtt{G}$ on $C^1(\mathbb{R}^d)$, with *active boundary* $\partial^0 E$ if for any x in E we have either *(1)* $t_x^\vee = +\infty$ and $\mathtt{X}(x,t)$ belongs to E for any $t \geq 0$ or *(2)* $\mathtt{X}(x,t)$ belongs to $\partial^0 E$ for some t in $]0, t_x^\vee[$. The forward flow $\mathtt{X} = \{\mathtt{X}(x,t) : t \geq 0\}$ is locally Lipschitz continuous in x in E, locally uniformly in t and the first exit time $\tau_E(x)$ from E is denoted by $\mathtt{T}(x)$. We may use the notation $\mathtt{X}^E$ and $\mathtt{T}^E$ to emphasize the dependency on the domain E. □

If the boundary ∂E of E (or E_1 if necessary) is smooth (e.g., $E = \{x \in \mathbb{R}^d : \phi(x) < 0\}$ as above) so that the outward normal vector $\nu(x)$ at x can be defined then every x in ∂E satisfying $g(x) \cdot \nu(x) > 0$ belongs to $\partial^0 E$ (active boundary) while and $g(x) \cdot \nu(x) < 0$ implies that x in the non-active boundary. The fact that we assume $\partial^0 E$ closed or $\partial E \setminus \partial^0 E$ relatively open ensure that the $\partial^0 E$ is the closure of x in ∂E satisfying $g(x) \cdot \nu(x) > 0$. Moreover, a key property is the continuity of the first exit time from E, i.e.,

$$\mathtt{T}(x) = \inf\{t > 0 : \mathtt{X}(x,t) \notin E\} = \inf\{t > 0 : \mathtt{X}(x,t) \in \partial^0 E\},$$

which is defined for any x in E, with the convention that $\mathtt{T}(x) = \infty$ if $\mathtt{X}(x,t)$ belongs to E for every $t > 0$. This means that the sets $\{x : \mathtt{T}(x) < \infty\}$ and $\{x : \mathtt{T}(x) = \infty\}$ are disjointed (one of them may be empty), and that the function $x \mapsto \mathtt{T}(x)$ is continuous on $\{x : \mathtt{T}(x) < \infty\}$. Note that the part of the boundary $\partial E \setminus \partial^0 E$ plays no role, and it is actually ignored. Two interesting cases are when either (1) $\partial^0 E = \emptyset$ so that E is closed, $E = \bar{E}$, or (2) $\partial^0 E = \partial E$ so that E is relative open $E = \mathring{E}_1 \times E_2$, where E_2 is a single point in $\mathbb{R}^{d-d_1}$. Moreover, the simplest situation is when $E = \mathbb{R}^{d_1} \times E_2$ and g is locally Lipschitz in $\mathbb{R}^d$ with a linear growth.

Given a locally Lipschitz continuous forward flow $\mathtt{X} = \{\mathtt{X}(x,t) : t \geq 0,\, x \in E\}$ with (closed!) active boundary $\partial^0 E$ then

$$P(t)f(x) = f(\mathtt{X}(x, t \wedge \mathtt{T}(x))), \quad \forall t \geq 0,\ x \in E,$$

defines a *semigroup* on the $C_*(E)$, real-valued bounded continuous functions on E vanishing on $\partial^0 E$. If f is a C^1 function then

$$P(t)f(x) - f(x) = \int_0^t P(s)\mathtt{G}f(x)\mathrm{d}s = \int_0^t \mathtt{G}P(s)f(x)\mathrm{d}s,$$

for every $t \geq 0$ and x in E. Thus the extended infinitesimal generator of the semigroup $\{P(t) : t \geq 0\}$, denoted by $\bar{\mathtt{G}}$ has the domain $\mathcal{D}(\bar{\mathtt{G}})$ which are all real-valued measurable functions f on E vanishing on $\partial^0 E$ such that the function $t \mapsto P(t)f(x)$ is absolutely continuous on $[0,T]$ for every positive real number $T \leq \mathtt{T}(x)$ and for every x in E. In this case, the expression $P(s)\mathtt{G}f(x)$ is only defined almost every where in s relative to the Lebesgue measure. The function f is only continuous along the flow, not necessarily continuous in other directions.

In same situations, it is important to single out the time variable so that weaker assumptions can be made. For instance, we may call $\mathtt{X} = \{\mathtt{X}(s,x,t) : t \geq s \geq 0,\, x \in E\}$ a nonhomogeneous locally Lipschitz forward flow associated

with the time-variant vector fields $\mathtt{G} = \{\mathtt{G}(t) : t \geq 0\}$ the solution of the IVP $\dot{x}(t) = g(x(t), t)$, $x(s) = x$. Typical assumptions are (1) the function $x \mapsto g(x, t)$ is locally Lipschitz continuous with linear growth, uniformly in t, i.e., for any compact subset K of $\bar{E}$ and any $T > 0$ there exists a constant M such that $|g(x, t) - g(x', t)| \leq M|x - x'|$ for every x, x' in K, and t in $[0, T]$, and if E is unbounded then there exist a constant C such that $|g(x, t)| \leq C(1 + |x|)$, for any x in $\bar{E}$ and $t \geq 0$, and the function $t \mapsto g(x, t)$ is (Borel) measurable. Under this assumptions, the IVP has only absolutely continuous solutions and the flow have the *composition* property $\mathtt{X}_{r,t} \circ \mathtt{X}_{s,r} = \mathtt{X}_{s,t}$ or equivalent $\mathtt{X}(s, x, t) = \mathtt{X}(s, \mathtt{X}(r, x, t), r)$, for every $t \geq r \geq s$. Differentiating with respect to the initial data, we deduce the well-known relations between the flow and the vector field, namely $\partial_s \mathtt{X}(s, x, t) = -\mathtt{G}(s)\mathtt{X}(s, x, t)$, i.e.,

$$\partial_s \mathtt{X}_i(s, x, t) = -g_j(x, s)\, \partial_j \mathtt{X}_i(s, x, t), \quad \forall t \geq s \geq 0,\ x \in E,$$

and $\partial_t \mathtt{X}(s, x, t) = -\mathtt{G}^*(t)\mathtt{X}(s, x, t)$, i.e.,

$$\partial_t \mathtt{X}_i(s, x, t) = \partial_j \big[g_j(x, t)\, \mathtt{X}_i(s, x, t)\big], \quad \forall t \geq s \geq 0,\ x \in E,$$

which are the deterministic equivalent of Kolmogorov backward and forward equations. Also, we refer the reader to the book by Ladde and Lakshmikantham [85] for a complete treatment of ODE with random coefficients.

5.9.2 Definition of PDP

Again, this is a generalization of the Poisson process in the direction of strong Markov processes as in a queueing system. Essentially, this stochastic process has deterministic evolutions between two consecutive jumps, instead of being constant as a Poisson process. The deterministic dynamic are characterized by a family vector fields, namely $g(n, x)$ defined for every x in $E \subset \mathbb{R}^d$ and each n in a countable set N. The jump mechanism has a jump rate $\lambda(n, x)$ and a transition distribution $Q(n, x, \cdot)$. Note that the jump rate determines *when to jump*, so a vanishing jump rate (i.e., $\lambda = 0$) means not jump at that particular position, while an infinite jump rate (i.e., $\lambda = +\infty$) translates into a instantaneous jump. The transition distribution rate $Q(n, x, \cdot)$ determines *where to jump*, also called transition probability measures of jumps.

Let $\mathtt{X} = \{\mathtt{X}(n, x, t) : t \geq 0,\ x \in E,\ n = 0, 1, \ldots\}$ be a family (indexed by n) of locally Lipschitz forward flows associated to vector fields $g(n, \cdot)$ on $E \subset \mathbb{R}^d$ or to first-order differential operator $\mathtt{G}(n, \cdot)$ on $C^1(\mathbb{R}^d)$, with active boundary $\partial_n^0 E$ (which may depend on n) and first exit from E time $\mathtt{T}(n, x)$ (or first hitting time to $\partial^0 E_n$), see Definition 5.18. Occasionally, we may use $\mathtt{X}_n(x, t) = \mathtt{X}(n, x, t)$ to emphasize the countable index n. Note that the dimension d is fixed and generally large, since $g(n, \cdot)$ may have several zero components which change with n. Clearly, the case where only finitely many n are used may be defined as a *module* operation in the variable n. Moreover, it may be useful to allow the set E to depend on n, i.e., $E_n \subset \mathbb{R}^{d_n}$, but we chose to disregard this case for now.

There are two type of jumps: (1) *interior jumps*, which are produced while inside region $E \setminus \partial_n^0 E$ and (2) *boundary jumps*, which are produced while on the active boundary $\partial_n^0 E$. Note that a point x belongs to the active boundary $\partial_n^0 E$ if and only if $\mathtt{T}(n,x) = 0$. The forward flow $\mathtt{X}(n,x,t)$ is defined for t in $[0, \mathtt{T}(n,x)]$ and for every (n,x) the backward flow $\mathtt{X}(n,x,-t)$ belongs to the inside $E \setminus \partial_n^0 E$ for t sufficiently small. Also, the functions $x \to \mathtt{T}(n,x)$ and $(x,t) \to \mathtt{X}(n,x,t)$ are continuous for every n. The interior jumps have the same exogenous origin as the one produced in the Poisson process, but the boundary jumps are forced or imposed by the continuous dynamic of the forward flow $\mathtt{X}$.

The state space of this piecewise deterministic process is $N \times E$, where N is a countable set (possible finite). Sufficient conditions on g to construct a locally Lipschitz forward flow have been discussed in the previous section, and the assumption of the jump are the following:

(1) the map $(n,x) \mapsto \lambda(n,x)$ is a Borel measurable function from $N \times E$ into $[0,\infty]$ and for each (n,x) there exists $\varepsilon = \varepsilon(n,x) > 0$ such that $t \mapsto \lambda(n, \mathtt{X}(n,x,t))$ is integrable in the interval $[0,\varepsilon)$,

(2) for each B in $\mathcal{B}(N \times E)$ the map $(n,x) \mapsto Q(n,x,B)$ is measurable, and for each (n,x) in $N \times E$ the map $B \mapsto Q(n,x,B)$ is a probability measure on $N \times E$ satisfying $Q(n,x,\{(n,x)\}) = 0$.

Note that the integrability condition of (1) ensure that after any jumps, we do have a continuous evolution following the forward flow for a positive time, while the last condition of (2) ensure a state discontinuity at every jump time. Piecewise deterministic processes viewed as Markov processes have state in $N \times E$, a discrete (piecewise constant in t) component $\{n(t) : t \geq 0\}$ in N and a continuous (piecewise continuous in t) component $\{x(t) : t \geq 0\}$ in $E \subset \mathbb{R}^d$. These conditions are mainly necessary to make sense to the jump mechanism, however, we need another condition to forbid the accumulation of boundary jumps.

Perhaps, the most typical situation in *hybrid system modelling*, including most of the queueing systems, is the case of finitely many n, i.e., the state space is $\{0,1,2,\ldots,K\} \times E$, and the locally Lipschitz forward flows $\{\mathtt{X}_n(x,t) : t \geq 0,\ x \in E\}$ are indexed by $n = 0,1,\ldots,N$. The particular case when with only one n, namely $n = 0$, is essentially different from the others, the discrete component is useless, and we may work directly on E. This is, we do have a Markov process in E which generalize the Poisson process, without adding a discrete component. Even in this simple situation, we do not have a Feller process. The active boundary introduces instantaneous predictable jumps, producing a *deterministic* discontinuity. Thus, unless there is not active boundary, a piecewise deterministic process is not a Feller process, but we do have a strong Markov process. According to the definition and assumptions on the locally Lipschitz forward flow $\mathtt{X}_n$ in Definition 5.18, the active boundary $\partial_n^0 E$ is closed and contains all reachable points from the inside $E \setminus \partial_n^0 E$, i.e., defining

$$\mathtt{T}(n,x) = \inf\{t > 0 : \mathtt{X}(n,x,t) \notin E\} = \inf\{t > 0 : \mathtt{X}(n,x,t) \in \partial_n^0 E\}, \quad (5.55)$$

which is defined for any x in E, with the convention that $\mathtt{T}(n,x) = \infty$ if $\mathtt{X}(n,x,t)$

belongs to E for every $t > 0$. The sets $\{x : \mathtt{T}(n,x) < \infty\}$ and $\{x : \mathtt{T}(n,x) = \infty\}$ are disjointed (one of them may be empty), and that the function $x \mapsto \tau(n,x)$ is continuous on $\{x : \mathtt{T}(n,x) < \infty\}$. To effectively rule out the accumulation of boundary jumps, see Davis [34, Proposition 24.6, pp 60–61], we may assume λ is bounded and that there exists $\varepsilon > 0$ such that

$$Q(n,x,\{(n,x) : \mathtt{T}(n,x) \geq \varepsilon\}) = 1, \quad \forall\,(n,x),\ x \in \partial_n^0 E, \tag{5.56}$$

which include the particular case of an empty active boundary, i.e.,

$$\mathtt{T}(n,x) = \infty, \quad \forall\,(n,x) \in N \times E.$$

Let us discuss a realization (or construction) of piecewise deterministic processes are described above, similarly to Section 5.8, but to emphasize the two components (discrete and continuous), we use the notation (n,x) instead of x. Thus the Markov process with sample path $(n(t),x(t) : t \geq 0)$, starting form a fixed initial point $(n(0),x(0))$, is realized in the canonical Hilbert cube $(\Omega,\mathcal{F},P)$, where $(U_1,U_2,\ldots)$ is sequence of independent random variables each having uniform distribution on $[0,1]$. We suppose given the characteristics g, λ and Q on the state space $N \times E$ which determine the flow $\mathtt{X}$, $\mathtt{T}$ as in Definition 5.18 satisfying (5.55). The jump mechanism satisfies the conditions (1) and (2) mentioned above and assumption (5.56).

First we define

$$\Psi(n,x,t) = \mathbb{1}_{t<\mathtt{T}(n,x)} \exp\Big[-\int_0^t \lambda(n,\mathtt{X}(n,x,s))\mathrm{d}s\Big], \tag{5.57}$$

the survivor function of jumps times and its (generalized) inverse

$$\Psi^{-1}(n,x,u) = \inf\{t \geq 0 : \Psi(n,x,t) \leq u\}, \tag{5.58}$$

with the convention that $\Psi^{-1}(n,x,u) = +\infty$ if $\Psi(n,x,t) > u$ for every $t \geq 0$. Note that $P\{\Psi^{-1}(n,x,U_k) > t\} = \Psi(n,x,t)$, for any k. In other words, always $0 \leq \Psi^{-1}(n,x,U_k) \leq \mathtt{T}(n,x)$, $P\{\Psi^{-1}(n,x,U_k) = \mathtt{T}(n,x) < \infty\} \leq P\{U_k = 0\} = 0$, and if the times $\mathtt{T}(n,x) = \infty$ then $\{\Psi^{-1}(n,x,U_k)\}$ is an independent identically distributed random variables in $[0,\infty[$, where its common distribution has $s \mapsto \lambda(n,\mathtt{X}(n,x,s))$ as intensity. This random variable $\Psi^{-1}(n,x,U_k)$ represents the waiting time for the next jump, while in the path $s \mapsto \mathtt{X}(n,x,s)$. Secondly, as mentioned above, we are working in the canonical Hilbert cube and there exists a measurable function satisfying

$$\begin{aligned} &\Upsilon : N \times E \times [0,1] \longrightarrow \{(n,x) : \mathtt{T}(n,x) \geq \varepsilon\} \subset N \times E, \\ &\ell(\{u : \Upsilon(n,x,u) \in B\}) = Q(n,x,B), \quad \forall\, B \in \mathcal{B}(N \times E), \end{aligned} \tag{5.59}$$

where ε is as in (5.56) and ℓ is the Lebesgue measure on $[0,1]$.

Now the sample path $\big(n(t,\omega),x(t,\omega) : t \geq 0\big)$ is defined by induction as follows. Given θ_{k-1}, n_{k-1} and x_{k-1}, with $k = 1,2,\ldots,$ we set

$$\theta_k(\omega) = \theta_{k-1}(\omega) + \Psi^{-1}(n_{k-1}(\omega),x_{k-1}(\omega),U_{2k-1}(\omega)), \tag{5.60}$$

and if $\theta_{k-1}(\omega) \le t < \theta_k(\omega)$ then

$$n(t,\omega) = n_{k-1}(\omega), \qquad x(t,\omega) = \mathtt{X}(n_{k-1}(\omega), x_{k-1}(\omega), t), \tag{5.61}$$

and if $\theta_k(\omega) < \infty$ then

$$(n_k, x_k) = \Upsilon\big(n_{k-1}(\omega), x(\theta_k(\omega)-,\omega), U_{2k}(\omega)\big), \tag{5.62}$$

where $x(s-,\omega)$ is the left-hand limit at a time s. If $\theta_{k-1}(\omega) = \theta_k(\omega)$ then we skip (5.61), define $x(\theta_k(\omega)-,\omega) = x_{k-1}(\omega)$ and go to (5.62). Therefore, if $\theta_k(\omega) = \infty$ then we have define the sample path for every time $t \ge \theta_{k-1}$, otherwise we have define the sample path in the stochastic interval $[\![\theta_{k-1}, \theta_k[\![$ as well as θ_k, n_k and x_k, and we can iterate (5.60),...,(5.62) with the initial $\theta_0 = 0$.

To actually see that the sample path is defined for every time, we will show that

$$\lim_{k\to\infty} \theta_k = \infty, \quad \text{a.s..} \tag{5.63}$$

Indeed, define the *counting jump process* $N_t = k$ if $\theta_{k-1} \le t < \theta_k$ with $N_0 = 0$ and $k \ge 1$. By means of (5.57) and the fact that λ is bounded ($\lambda \le c$), we can construct a Poisson process $(N_t^p : t \ge 0)$ with rate c, that dominate the counting process $(N_t : t \ge 0)$, i.e, $N_t \le N_t^p$. Thus $\mathbb{E}\{N_t\} \le \mathbb{E}\{N_t^p\} = ct$, which implies the condition (5.63).

We will only consider the class of piecewise deterministic processes satisfying the condition (5.63) is satisfied, or even more if $\mathbb{E}\{N_t\} < \infty$ for every , besides the assumptions made on the vector fields $\mathtt{X}_n$, $\mathtt{T}_n$ and (1) and (2) above on the jump mechanism.

Since $(n(t) : t \ge 0)$ is piecewise constant, $(x(t) : t \ge 0)$ is piecewise continuous and both are right-continuous, we take pass to the canonical sample space $\mathcal{D}([0,\infty), N \times E)$ and define the integer-valued measure process

$$\rho(t, B, \omega) = \sum_{k=1}^{\infty} \mathbb{1}_{\theta_k(\omega)\le t} \mathbb{1}_{x_k(\omega)\in B}, \quad \forall\, t \ge 0,\ B \in \mathcal{B}(\mathbb{R}^d) \tag{5.64}$$

associated with the piecewise deterministic process constructed above. When necessary, we may write $\rho_{nx}(t,\omega,B)$ to indicate dependency on the initial state $n_0 = n$, $x_0 = x$ at time $\theta_0 = 0$.

It is proved in Davis [34, Sections 25, 26, Theorems 24.3, 25.5, 26.14] that the filtration (history) generated by the piecewise deterministic processes is right-continuous, that $\big(n(t,\omega), x(t,\omega) : t \ge 0\big)$ is an homogeneous strong Markov process in the canonical sample space $\mathcal{D}([0,\infty), N \times E)$. Also, the extended infinitesimal generator has complete description as follows.

The expression of the extended infinitesimal generator

$$\begin{aligned} \bar{A}f(n,x) &= \mathtt{G}_n f(n,x) + \mathtt{I}f(n,x), \\ \mathtt{G}_n h(x) &= \sum_{i=1}^{d} g_i(n,x)\partial_i h(x), \\ \mathtt{I}f(n,x) &= \lambda(n,x) \int_{N\times E} [f(\eta,\xi) - f(n,x)]\, Q(n,x,\mathrm{d}\eta \times \mathrm{d}\xi) \end{aligned} \tag{5.65}$$

where the first-order differential operator $\mathtt{G}_n$ is acting only on the continuous variable x, while the integral operator $\mathtt{I}$ may involves both variable n and x.

To full describe the extended domain $\mathcal{D}(\bar{A})$ we need the following concept. Now we say that a process $\{h(n,x,t,\omega) : n \in N,\, x \in \mathbb{R}^d,\, t \geq 0,\}$ belongs to $L^1(\rho)$, with $\rho = \rho(t,\omega,B)$ given by (5.64) if

$$\mathbb{E}\Big\{ \int_{N\times E\times\mathbb{R}^+} h \mathrm{d}\rho \Big\} = \mathbb{E}\Big\{ \sum_{i=1}^{\infty} h(n_k, x_k, \theta_k, \cdot) \Big\} < \infty. \tag{5.66}$$

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

$$\mathbb{E}\Big\{ \sum_{i=1}^{\infty} h(n_k, x_k, \theta_k \wedge \sigma_n, \cdot) \Big\} < \infty, \quad \forall\, n,$$

i.e., $h(n,x,t,\cdot)\, \mathbb{1}_{t<\sigma_n}$ belongs to $L^1(\rho)$.

A measurable function $f\colon N \times E \to \mathbb{R}$ belongs to $\mathcal{D}(\bar{A})$ if and only if several conditions are met:

(1) we have $f(n, \mathtt{X}(n,x,-t)) \to f(n,x)$ as $t \to 0$ and

$$f(n,x) = \int_{N\times E} f(\eta,\xi)\, Q(n,x,\mathrm{d}\eta \times \mathrm{d}\xi)$$

for every n in N, x in E such that $\mathtt{T}(n,x) = 0$,

(2) the function $t \to f(n, \mathtt{X}(n,x,t))$ is absolutely continuous on $[0, \mathtt{T}(n,x) \wedge T[$, for every $T > 0$, n in N and x in E,

(3) for every n in N and x in E, the process

$$h(n,x,t) = f(n,x) - f(n(t-,\omega), x(t-,\omega)), \quad \forall\, t > 0,$$

with $h(n,x,0) = 0$, belongs to $L^1_{\mathrm{loc}}(\rho_{nx})$.

Property (1) is called *boundary condition* since $\mathtt{T}(n,x) = 0$ if and only if x belongs to the active boundary $\partial^0_n E$, which is mainly related to the discrete variable n. Condition (2) involves only the continuous variable x and provided a weak sense to the differential operator $\mathtt{G}f(n(t),x(t))$ as the derivative (almost every t, the discrete variable n and ω are regarded as parameters) of the function $t \to f(n, \mathtt{X}(n,x,t))$. Property (3) can be re-written as

$$\mathbb{E}\Big\{ \sum_{k=1}^{n} |f(n_k,x_k) - f(n_{k-1}, \mathtt{X}(n_{k-1}, x_{k-1}, \theta_k)| \mathbb{1}_{\theta_k} \leq n \Big\} < \infty,$$

for every $n \geq 1$, by taken $\sigma_n = \theta_n \wedge n$, which is certainly verified if f is bounded.

To complete this discussion let as mention that when $\mathtt{T}(n,x) = \infty$ for every (n,x) and the jump rate $\lambda(n,x)$ is bounded then a piecewise deterministic process is a Feller process. In general, a piecewise deterministic process is a Borel right process. This means that (a) the state space $N \times E$ is topological

homeomorphic to a Borel subset of a compact metric space, (b) the semigroup $P(t)h = \mathbb{E}\{h(n(t), x(t)\}$ maps the bounded Borel functions into itself, (c) the sample paths $t \to (n(t), x(t))$ are right continuous almost surely, (d) if f is an α-excessive function for $\{P(t) : t \geq 0\}$ then the function $t \to f(n(t), x(t))$ is right continuous almost surely. Recall that a non-negative function f is called α-excessive (with $\alpha \geq 0$) if $\exp(-t\alpha)P(t)f \leq f$ for every $t \geq 0$ and $\exp(-t\alpha)P(t)f \to f$ as $t \to 0$.

As mentioned early, a comprehensive study on piecewise deterministic process can be found in Davis [34].

5.10 Lévy Processes

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 5.19. 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 $\varepsilon > 0$.

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

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

Several properties relating characteristic functions are known, e.g., if $\hat{\mu}$ is integrable in $\mathbb{R}^d$ then μ is absolutely continuous with respect to the Lebesgue measure, with a bounded continuous density g given by the inversion formula

$$g(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} \mathrm{e}^{-\mathrm{i}\, x\cdot y}\, \hat{\mu}(y)\mathrm{d}y.$$

In particular $y \mapsto \hat{\mu}(y)$ is uniformly continuous from $\mathbb{R}^d$ into the complex plane $\mathbb{C}$, $\hat{\mu}(0) = 1$ and $\hat{\mu}$ is *positive definite*, i.e., for any $k = 1, 2, \ldots$, $\zeta_1, \ldots, \zeta_k$ in $\mathbb{C}$ and $x_1, \ldots, x_k$ in $\mathbb{R}^d$ we have $\sum_{i,j=1}^k \hat{\mu}(x_i - x_j)\zeta_i\bar{\zeta}_j \geq 0$. Moreover, Bochner Theorem tell us that the converse, i.e., any complex-valued continuous function φ in R^d with $\varphi(0) = 1$ and positive definite is the characteristic function of a distribution, e.g., Da Prato and Zabczyk [30, Theorem I.2.3, pp. 48–52] for a proof valid in separable Hilbert spaces.

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 R^d-valued random variable $X(1)$, i.e.,

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

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 [126, 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 $\mathbb{E}\{\mathrm{e}^{\mathrm{i}\, y\cdot X(t)}\} = \exp[-t\phi(y)]$ 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 $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$ 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 [84, Section 5.2, pp. 137–144], for the one-dimensional case.

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 $c\, t$, for every $t \geq 0$. 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 $\mathbb{E}\{\mathrm{e}^{\mathtt{i}\, y \cdot X(t)}\} = \exp[-t\, c(\hat{\gamma}(y) - 1)]$, for any $t \geq 0$ and y in $\mathbb{R}^d$. 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 γ and n), the expressions

$$X(t) = \sum_{n=1}^{\infty} \zeta_n \mathbb{1}_{t \geq \theta_n}, \quad \text{with} \quad \delta X(t) = X(t) - X(t-)$$

$$\delta X(\theta_n) = \zeta_n, \quad \text{and} \quad \delta X(t) = 0 \ \text{ if } t \neq \theta_n,\ \forall n, \ \text{or equivalently}$$

$$X(t) = \zeta_1 + \zeta_2 + \cdots + \zeta_n \ \text{ if } \ \sum_{i=1}^{n} \tau_i = \theta_n \leq t < \theta_{n+1} = \sum_{i=1}^{n+1} \tau_i,$$

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

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

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 5.19, 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 $\mathcal{B}(\mathbb{R}^d)$, $i = 1, 2, \ldots, k$ the integer-valued random variables

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

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 m of any Lévy process with increasing path satisfies

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

e.g., see books Bertoin [11, Chapter III, pp. 71-102], Itô [65, Section 1.11] and Sato [116, Chapter 6, pp. 197-236].

Another interesting case is the so-called *symmetric* Lévy processes where the characteristic exponent (also called Lévy exponent) $\phi(y)$ (defined early) satisfies

$$\int_{\mathbb{R}^d} \Big|\frac{1}{1 + \phi(y)}\Big|\, \mathrm{d}y < \infty,$$

which implies that $\phi(y)$ is a positive real-valued even function. Moreover, the only possible case occurs when the dimension $d = 1$, and actually, $\phi(y)$ takes the form

$$\phi(y) = \frac{1}{2} Q y^2 + 2 \int_{\mathbb{R}^d_*} \big[1 - \cos(xy)\big] m(\mathrm{d}x), \quad \forall y \in \mathbb{R}$$

for some nonnegative constant Q and some measure m on $\mathbb{R}$ which integrates the function $(1 \wedge x^2)$. In this one-dimensional case, the resolvent (measure)

$$\int_{\mathbb{R}} f(y)\, R(\lambda, \mathrm{d}y) = \int_0^\infty \mathrm{e}^{-\lambda t} \mathbb{E}\big\{ f\big(X(t) + x\big) \big\} \mathrm{d}t,$$

has the density

$$r(\lambda, y) = \frac{1}{\pi} \int_0^\infty \frac{\cos(xy)}{\lambda + \phi(y)}\, \mathrm{d}t, \quad \forall y \in \mathbb{R},$$

while the transition densities are given by

$$p(t, x) = \frac{1}{\pi} \int_{\mathbb{R}} \mathrm{e}^{-xy}\, \mathrm{e}^{-t\phi(y)} \mathrm{d}y, \quad \forall t > 0,\, x \in \mathbb{R},$$

see the book by Marcus and Rosen [94, Section 4.1, pp. 135-144] for details. The interested reader, may consult the book by Applebaum [1], which discuss Lévy process at a very accessible level.

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 $\mathbb{E}\{\mathrm{e}^{y \cdot X(t)}\} = \exp\left[-t(|\sigma y|^2/2 - i\, b)\right]$, 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

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

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

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

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 $(k-1)\pi \leq t < k\pi$, for $k \geq 1$.) 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 [31, Theorem B.1.5, pp. 311–316]. For more details on the construct arguments, see, e.g., Friedman [50] or Krylov [83].

We are ready to state the general existence result

Theorem 5.20 (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 $(X(t) = \omega(t) : t \geq 0)$ is a Lévy process with characteristic *(g, Q, m), i.e.,*

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

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

Proof. Only some details are given. First, consider the case where $Q = 0$, which corresponds to *Poisson measures* and *point processes*, a step further from the compound Poisson processes. Essentially, a point process is the jumps process constructed from a cad-lag process. Poisson measures are particular case of integer-valued measures, which are the distribution of the jumps of cad-lag processes. More extensive comments can be are given, here we recall a couple of arguments used to construct a Poisson measure. Let m be a Radon measure in $\mathbb{R}^d_*$ (which integrates the function $|x|^2 \wedge 1$ is used later) and write $m = \sum_k m_k$, where $m_k(B) = m(B \cap R_k)$, $\mathbb{R}^d_* = \cup_k R_k$, $m(R_k) < \infty$ and $R_k \cap R_\ell = \emptyset$ if $k \neq \ell$. To each m_k we may associate a compound Poisson and point processes

$$Y_k(t) = \sum_{n=1}^{\infty} \zeta_{n,k}\,\mathbb{1}_{t\geq\theta_{n,k}} \qquad \text{or} \qquad Y_k(t) = Z_{n,k} \ \text{ if } \ \theta_{n-1,k} < t \leq \theta_{n,k},$$
$$\delta Y_k(t) = Y_k(t) - Y_k(t-) = \zeta_{n,k}\,\mathbb{1}_{t=\theta_{n,k}}, \quad \forall t \geq 0,$$

where $\theta_{n,k} = \tau_{1,n,k} + \tau_{2,n,k} + \cdots + \tau_{n,n,k}$, $\{\tau_{i,n,k} : i = 1, \dots, n,\ n = 1, 2, \dots\}$ is a sequence of independent exponentially distributed (with parameter $m(R_k) = c_k$) random variables, and $Z_{n,k} = \zeta_{1,k} + \zeta_{2,k} + \cdots + \zeta_{n,k}$, $\{\zeta_{n,k} : n = 1, 2, \dots\}$ is another sequence of independent identically distributed (with distribution law m_k/c_k) random variables, the family $\{\tau_{i,n,k},\ \zeta_{n,k} : i = 1, \dots, n,\ n, k \geq 1\}$ is independent. Since the processes $\{Y_k : k \geq 1\}$ are independent, the characteristic function of the point process $Y = \sum_k Y_k$ is the product of those of Y_k, which should reconstruct the measure m. The independence property and the diffuse character (non atoms) of the exponential distribution ensure that there are no simultaneous jumps among the $\{Y_k : k \geq 1\}$. Hence, the jump process $\delta Y = \sum_k \delta Y_k$ is indeed a Poisson point process with characteristic measure m, i.e.,

$$\nu(]s,t] \times B) = \sum_{n,k=1}^{\infty} \mathbb{1}_{s<\theta_{n,k}\leq t}\, \mathbb{1}_{\zeta_{n,k}\in B}, \quad \forall t > s \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d_*)$$

is a Poisson random measure with intensity measure $\mathbb{E}\{\nu(]s,t] \times B)\} = (t-s)m(B)$. In general, we cannot re-order the jumps in a increasing manner as those of a compound Poisson process. Next, some *martingale* estimates are necessary to establish good behavior of the process Y. First the exponential formula, for any complex-valued Borel function f on $\mathbb{R}^d_*$

$$\text{if} \quad \int_{R^d_*} |1 - \mathrm{e}^{f(x)}|\, m(\mathrm{d}x) < \infty \quad \text{then} \quad \forall t \geq 0 \quad \text{we have}$$
$$\mathbb{E}\Big\{ \exp\Big[\sum_{0\leq s\leq t} f(\delta Y(s)) \Big] \Big\} = \exp\Big\{ -t \int_{R^d_*} [1 - \mathrm{e}^{f(x)}]\, m(\mathrm{d}x) \Big\}.$$

Secondly, the Doob's maximal inequality for the compensated jumps

$$\mathbb{E}\Big\{ \sup_{0\leq t\leq T} \Big| \sum_{0\leq s\leq t} f(\delta Y(s)) - t \int_{R^d_*} f(x)\, m(\mathrm{d}x) \Big|^2 \Big\} \leq 4T \int_{R^d_*} |f(x)|^2\, m(\mathrm{d}x),$$

valid for any real valued Borel function f on $\mathbb{R}^d_*$ and any $T > 0$, see Sections 5.3 and 5.4 for more detail and references.

After *compensate* the small jumps, this Poisson measure ν and its associated jump process δY yield a Lévy process with characteristic $(0, 0, m)$. Indeed, define

$$X_1(t) = \int_{]0,t]\times\{|x|\geq 1\}} x\, \nu(\mathrm{d}s \times \mathrm{d}x) = \sum_{s\leq t} \delta Y(s)\, \mathbb{1}_{|\delta Y(s)|\geq 1},$$

and

$$X_2^{\varepsilon}(t) = \int_{]0,t]\times\{\varepsilon\leq|x|<1\}} x\, \nu(\mathrm{d}s \times \mathrm{d}x) - \mathbb{E}\Big\{ \int_{]0,t]\times\{\varepsilon\leq|x|<1\}} x\, \nu(\mathrm{d}s \times \mathrm{d}x) \Big\}$$
$$= \sum_{s\leq t} \delta Y(s)\, \mathbb{1}_{\varepsilon<|\delta Y(s)|\leq 1} - t \int_{\varepsilon\leq|x|<1} x\, m(\mathrm{d}x),$$

for $t, \varepsilon > 0$, which are two compound Poisson processes with characteristic exponents

$$\phi_1(y) = \int_{|x|\geq 1} \big[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x}\big]\, m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d,$$
$$\phi_2^\varepsilon(y) = \int_{\varepsilon\leq|x|<1} \big[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x} + \mathtt{i}\, y\cdot x\big]\, m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d,$$

respectively, after using the above exponential formula. In view of the martingale inequality

$$\mathbb{E}\big\{\sup_{t\leq T} |X_2^\varepsilon(t) - X_2^\delta(t)|^2\big\} \leq 4T \int_{\delta\leq|x|<\varepsilon} |x|^2\, m(\mathrm{d}x),$$

for every $T > 0$, $\varepsilon > \delta > 0$, and because the intensity measure m integrates the function $|x|^2 \wedge 1$, the family of processes $\{X_2^\varepsilon : \varepsilon > 0\}$ converges to a process X_2 and

$$\mathbb{E}\big\{\sup_{t\leq T} |X_2^\varepsilon(t) - X_2(t)|^2\big\} \leq 4T \int_{|x|<\varepsilon} |x|^2\, m(\mathrm{d}x),$$

for every $T > 0$ and $\varepsilon > \delta > 0$. This cad-lag process X_2 has stationary independent increments, so a Lévy process with characteristic exponent

$$\phi_2(y) = \int_{|x|<1} \big[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x} + \mathtt{i}\, y\cdot x\big]\, m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d.$$

Therefore, $X_1 + X_2$ is a Lévy process with characteristic $(0, 0, m)$.

Now to conclude, consider given a drift g in $\mathbb{R}^d$ and a co-variance Q (non-negative-definite $d \times d$ matrix). If B is a standard Wiener process independent of (X_1, X_2) (i.e., of the previous construction) then define $X_3 = (\sqrt{Q}\, B(t) - g\, t : t \geq 0)$, which is a Lévy process with characteristic exponent

$$\phi_3(y) = \mathtt{i}\, g\cdot x + \frac{1}{2}\, Qy\cdot y, \quad \forall y \in \mathbb{R}^d.$$

Finally, $X = X_1 + X_2 + X_3$ is a Lévy process with the desired characteristic (g, Q, m). The converse follows from Lévy-Khintchine formula for infinitely divisible distributions.

An important point to remark is that the above construction shows that any Lévy process is a Wiener process plus the limit of a sequence of compound Poisson processes. Also note that any infinitely divisible probability measure on $\mathbb{R}^d$ can be viewed as the distribution of a Lévy process evaluated at time 1. □

It is perhaps relevant to remark that even if any Lévy process can be expressed as a limit of compound Poisson processes, the structure of a typical graph of Levy process eludes us. For instance, almost surely, the jumping times $J = \{t : X(t,\omega) \neq X(t-,\omega)\}$ are countable, and (a) if the Levy measure satisfies $m(\mathbb{R}^d) = \infty$ then J is dense in $[0,\infty)$ while (b) if $m(\mathbb{R}^d) < \infty$ then J

can be written as an increasing sequence $\{\tau_k : k \geq 1\}$, $\tau_k \leq \tau_{k+1}$, of independent random variables having exponential distributions with mean $1/m(\mathbb{R}^d)$, see Sato [116, Theorem 21.3, pp. 136–137].

Note that for a given Lévy process X with the characteristic (g, Q, m) we can define $\delta X(t) = X(t) - X(t-)$ and the integer-valued (random) measure ν_X associated with the jumps δX of X, and its martingale measure $\mu_X = \nu_X - m$, since $m = \nu_X^p$ is the predictable jump compensator (which is actually deterministic). To make sense to discontinuous (purely jumps) part $X_d(t)$ of X, which is the *compensated* sum of all jumps $\sum_{s\leq t} \delta X(s)$, and therefore define its continuous part as $X_c = X - X_d$ we proceed essentially as above. For $t, \varepsilon > 0$, consider

$$X_1(t) = \int_{]0,t]\times\{|x|\geq 1\}} x\, \nu_X(\mathrm{d}s \times \mathrm{d}x) = \sum_{s\leq t} \delta X(s)\, \mathbb{1}_{|\delta X(s)|\geq 1},$$

and

$$X_2^\varepsilon(t) = \int_{]0,t]\times\{\varepsilon\leq|x|<1\}} x\, \mu_X(\mathrm{d}s \times \mathrm{d}x) = \\ = \sum_{s\leq t} \delta X(s)\, \mathbb{1}_{\varepsilon<|\delta X(s)|\leq 1} - t \int_{\varepsilon\leq|x|<1} x\, m(\mathrm{d}x)$$

are compound Poisson process, and using the canonical semi-martingale decomposition, the limit process $X_d = X_1 + \lim_{\varepsilon\to 0} X_2^\varepsilon$ is a Lévy process with characteristic exponent

$$\phi_\delta(y) = \int_{\mathbb{R}^d_*} \left[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x} + i\, y\cdot x\,\right] \mathbb{1}_{|x|<1}\, m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d,$$

which is called the discontinuous (purely jumps) part of X. If the jumps are not of bounded variation, the series $\sum_{s\leq t} \delta X(s)$ is meaningless, unless it is compensated with the *small jumps* (we used jumps greater than 1, but it suffices greater than some positive constant). Sometimes we *ignore* large jumps (by assuming that the Lévy measure m integrates z at infinite), and so the Lévy process corresponding to the characteristic exponent

$$\phi_d(y) = \int_{\mathbb{R}^d_*} \left[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x} + i\, y\cdot x\,\right] m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d.$$

is uniquely determined could be used as X_d.

Since X is quasi-left continuous, we have $\delta X = \delta X_d$. Due to the independence of increments, $X(t) - X(t-)$ results independent of $X(t-) = X_c(t)$, i.e., the processes X_c and X_d are independent and the characteristic exponent of X_c must be

$$\phi_c(y) = \mathtt{i}\, g\cdot x + \frac{1}{2}\, Qy\cdot y, \quad \forall y \in \mathbb{R}^d.$$

Thus X_c is a Wiener process and the characteristic (g, Q, m) can also be found as

$$g = -\mathbb{E}\{X_c(1)\}, \qquad Q = \mathbb{E}\{[X_c(1) + g]^*[X_c(1) + g]\},$$
$$m(B) = \mathbb{E}\Big\{ \sum_{0<t\le 1} \mathbb{1}_{\delta X(t)\in B}\Big\}, \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

where $\delta X(t) = X(t) - X(t-)$, $X_c = X - X_d$ and $(\cdot)^*$ is the transpose operator.

On the other hand, if A is a $n \times d$ matrix then $AX = (AX(t) : t \ge 0)$ is a n dimensional Lévy process with characteristic

$$g = Ag + \int_{\mathbb{R}^d_*} Ax[\mathbb{1}_{|Ax|<1} - \mathbb{1}_{|x|<1}]\, m(\mathrm{d}x),$$
$$Q_A = AQA^*, \qquad m_A = mA^{-1},$$

where $mA^{-1}(B) = m(\{x : Ax \in B\})$.

If $Q = 0$ and the Lévy measure m integrates the function $|x| \wedge 1$ then the characteristic exponent may be re-written in a simpler way, as

$$\phi(y) = -\mathtt{i}\, g \cdot x + \int_{\mathbb{R}^d_*} \left[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x}\right] m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d,$$

where g is now referred to as the *drift coefficient*. In this case, the Lévy process has locally bounded variation, not necessarily integrable, unless m integrates the function $|x|$. Certainly, if X has a finite Lévy measure m on $\mathbb{R}^d_*$ then X is a compound Poisson process plus a drift.

Now we take a look at the resolvent operators associated with Lévy processes. Let $\{P(t) : t \ge 0\}$ be the semigroup associated with a Lévy process, i.e.,

$$P(t) : C_0(\mathbb{R}^d) \to C_0(\mathbb{R}^d), \qquad P(t)f(x) = \mathbb{E}\{f(X(t) + x)\}, \tag{5.67}$$

where $C_0(\mathbb{R}^d)$ is the Banach space of continuous functions vanishing at infinity. Then, the family $\{R(\lambda) : \lambda > 0\}$ of linear and bounded operators from $C_0(\mathbb{R}^d)$ into itself and the family of $\{R(\lambda, dy) : \lambda > 0\}$ of finite measures on $\mathbb{R}^d$, defined by

$$R(\lambda)f(x) = \int_0^\infty \mathrm{e}^{-\lambda\, t}\, P(t)f(x)\mathrm{d}t \quad \text{and}$$
$$\int_{R^d} f(y)R(\lambda, dy) = \mathbb{E}\Big\{ \int_0^\infty \mathrm{e}^{-\lambda\, t}\, f(X(t))\mathrm{d}t\Big\}, \tag{5.68}$$

which satisfies

$$R(\lambda)f(x) = \mathbb{E}\Big\{ \int_0^\infty \mathrm{e}^{-\lambda\, t}\, f(X(t) + x)\mathrm{d}t\Big\} \quad \text{and}$$
$$\int_{R^d} f(y)R(\lambda, dy) = R(\lambda)f(0),$$

are called the *resolvent operators* and the *resolvent kernel* associated with the Lévy process X. It is also clear that $R(\lambda)$ is a convolution operator, i.e.,

$$\begin{aligned} &\text{if} && \int_{R^d} f(y)\check{R}(\lambda, dy) = \mathbb{E}\Big\{ \int_0^\infty \mathrm{e}^{-\lambda t} f(-X(t))\mathrm{d}t \Big\} \\ &\text{then} && (\check{R}(\lambda,\cdot) \star f)(x) = \int_{R^d} f(x-y)\check{R}(\lambda, dy) = R(\lambda)f(x). \end{aligned}$$

The resolvent operators describe the distribution of the Lévy process evaluated at independent exponential times, i.e., if $\tau = \tau(\lambda)$ is an independent (of X) random variable having an exponential law with parameter $\lambda > 0$, then $\mathbb{E}\{f(X(\tau)+x)\} = \lambda\, R(\lambda)f(x)$.

The semigroup property yields the identity

$$\begin{aligned} &R(\lambda) - R(\mu) = (\mu-\lambda)R(\lambda)R(\mu), \quad \forall \lambda,\mu > 0 \quad \text{or} \\ &\check{R}(\lambda,\cdot) - \check{R}(\mu,\cdot) = (\mu-\lambda)\check{R}(\lambda,\cdot) \star \check{R}(\mu,\cdot) \end{aligned} \tag{5.69}$$

so-called *resolvent equation*. Thus the image of $C_0(\mathbb{R}^d)$ under $R(\lambda)$, denoted by $\mathcal{D}$, does not depend on $\lambda > 0$. Since

$$\lambda R(\lambda)f(x) - f(x) = \int_0^\infty \mathrm{e}^{-s}[P(s/\lambda)f(x) - f(x)]\mathrm{d}s$$

we deduce that $\lambda R(\lambda)f \to f$ in $C_0(\mathbb{R}^d)$ as $\lambda \to \infty$. Therefore, $\mathcal{D}$ is a dense subspace of $C_0(\mathbb{R}^d)$. Moreover, if $R(\lambda)f = R(\lambda)g$ for some $\lambda > 0$, the resolvent equation shows that $R(\lambda)f = R(\lambda)g$ for any $\lambda > 0$ and then, as $\lambda \to \infty$ we deduce $f = g$, i.e., $R(\lambda)$ is a one-to-one mapping from $C_0(\mathbb{R}^d)$ onto $\mathcal{D}$. The infinitesimal generator A from $\mathcal{D}$ into $C_0(\mathbb{R}^d)$ is defined by the relation

$$R(\lambda)(\lambda I - A) = I \quad \text{or equivalently} \quad A = \lambda I - [R(\lambda)]^{-1}, \tag{5.70}$$

where I is the identity mapping, and $\mathcal{D} = \mathcal{D}_A$ is called the *domain* of the infinitesimal generator A.

The Fourier transform for f in $L^1(\mathbb{R}^d)$, namely

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\, x\cdot\xi} f(x)\mathrm{d}x, \quad \forall \xi \in \mathbb{R}^d,$$

yields simple expressions of these operators in term of the characteristic exponent

$$\phi(\xi) = -\ln(\mathbb{E}\{\mathrm{e}^{\mathrm{i}\, X(1)\cdot\xi}\})$$

of the Lévy process X. We have for any f in $L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$

$$\begin{aligned} &\widehat{P(t)f}(\xi) = \mathrm{e}^{-t\phi(-\xi)}\,\widehat{f}(\xi), \quad \forall t \geq 0,\ \xi \in \mathbb{R}^d, \\ &\widehat{R(\lambda)f}(\xi) = [\lambda + \phi(-\xi)]^{-1}\,\widehat{f}(\xi), \quad \forall \lambda > 0,\ \xi \in \mathbb{R}^d, \end{aligned} \tag{5.71}$$

and, for any f in $\mathcal{D}_A$ such that Af belongs to $L^1(\mathbb{R}^d)$

$$\widehat{Af}(\xi) = -\phi(-\xi)\widehat{f}(\xi), \quad \forall \xi \in \mathbb{R}^d. \tag{5.72}$$

Hence, Lévy-Khintchine formula and the inversion of Fourier transform yield the following expression for the infinitesimal generator of Lévy processes

$$Af(x) = -g \cdot \nabla f + \frac{1}{2}\nabla \cdot Q\nabla f + \\ + \int_{\mathbb{R}^d_*} \left[f(\cdot + y) - f - \mathbb{1}_{|y|<1}\, y \cdot \nabla f \right] m(dy), \tag{5.73}$$

for any smooth function f, e.g., twice-continuously differentiable and bounded function f.

Let us mention a result from Blumental and Getoor [17]. If the resolvent kernel $R(\lambda, x + dy)$ is absolutely continuous (with respect to the Lebesgue measure) for some $\lambda > 0$ and some x in R^d then it is absolutely continuous for every $\lambda > 0$ and every x in R^d. Moreover this is equivalent to the so-called *strong Feller property* of the resolvent operators, namely, for any $\lambda > 0$ and f in $L^\infty(R^d)$ the function $x \mapsto R(\lambda)f(x)$ is continuous.

To conclude this section we briefly discuss the so-called *local time* associated with a Levy process, full details can be found in the book Bertoin [11, Chapter V, pp. 125–154]. Let $X(t)$ a one dimensional Lévy process, then its characteristic function is given by

$$\mathbb{E}\{\exp(\mathtt{i}\xi X(t))\} = e^{-t\psi(\xi)}$$

where ψ is characterized by

$$\psi(\xi) = -\mathtt{i}\gamma\xi + \tfrac{1}{2}\sigma^2\xi^2 - \int_{\mathbb{R}} \left(e^{\mathtt{i}\xi y} - 1 - \mathtt{i}\xi y \mathbb{1}_{\{|y|<1\}} \right) \mathrm{d}m(y)$$

where $\mathrm{d}m(y)$ is the Lévy measure. For instance, the choice $\gamma = 0$, $\sigma = 1$ and $\mathrm{d}m = 0$ yields the Brownian motion, while $\gamma = 0$ and $\sigma = 0$ produces a pure jump processes. In particular, $m = \delta_1$ corresponds to the Poisson process and for $\mathrm{d}m(y) = \frac{1}{|y|^{1+\alpha}}\mathrm{d}y$, with $0 < \alpha < 2$, we get the so-called *α-stable Lévy processes* with $\psi(\xi) = \frac{2}{c}|\xi|^\alpha$, where

$$|\xi|^\alpha = -\frac{c}{2}\int_{\mathbb{R}} \left(e^{\mathtt{i}\xi y} - 1 - \mathtt{i}\xi y \mathbb{1}_{\{|y|<1\}} \right) m(\mathrm{d}y), \qquad \frac{1}{c} = \int_0^\infty \frac{1 - \cos s}{s^{1+\alpha}} \mathrm{d}s,$$

i.e., $c = \frac{2}{\pi}\Gamma(1+\alpha)\sin(\frac{\pi\alpha}{2})$.

For any $t > 0$, the *occupation measure* $\mu(t, \mathrm{d}x)$ on the time interval $[0, t]$ of the Lévy process X is defined as

$$\mu(t, B, \omega) = \int_0^t \mathbb{1}_{\{X(s,\omega)\in B\}} \mathrm{d}s,$$

for every Borel subset of $\mathbb{R}$.

Theorem 5.21 (occupation measure)**.** *For any $t \geq 0$, the occupation measure $\mu(t, \mathrm{d}x)$ is absolutely continuous with respect to the Lebesgue measure with a density in $L^2(\mathrm{d}y \times P)$ if and only if*

$$\int_{\mathbb{R}} \Re\Big\{\frac{1}{1+\psi(y)}\Big\}\mathrm{d}y < \infty. \tag{5.74}$$

Moreover, if the above condition fails, then $\mu(t, \mathrm{d}x)$ is singular with respect to the Lebesgue measure for any $t > 0$ and with probability 1. □

Brownian motions and α-stable Lévy processes with $1 < \alpha < 2$ satisfy condition (5.74) and therefore theirs occupation measures have densities with respect to the Lebesgue measure. While Poisson processes and α-stable Lévy processes with $0 < \alpha \leq 1$ do not satisfy (5.74).

Now, if condition (5.74) is satisfied then for every $t > 0$ and any x in $\mathbb{R}$ we can define

$$\ell(t, x) = \limsup_{\epsilon \to 0} \tfrac{1}{2\epsilon} \int_0^t \mathbb{1}_{\{|X(s)-x|<\epsilon\}}\mathrm{d}s,$$

which is referred to as the *local time* at the level x and time t for the Lévy process X. It is clear that $\{\ell(t, x) : x \in \mathbb{R}\}$ serves as a $\mathcal{F}(t)$-measurable version of the density of $\mu(t, \mathrm{d}x)$. Note that for every x, the process $\ell(\cdot, x)$ is (cad-lag) nondecreasing, which may increase only when $X = x$. Thus, ℓ is jointly measurable.

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 5.19 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 [116, Theorem 11.5, p. 65], which is actually indistinguishable if the initial Lévy or additive process was a separable process.

Proposition 5.22. *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 $\{y(s) : 0 \leq s \leq t\}$. 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 $\mathcal{F}_t(y) = \cap_{s>t}\mathcal{F}_s(y)$, for any $t \geq 0$.*

Proof. Set $\mathcal{F}_t^+(y) = \cap_{s>t}\mathcal{F}_s(y)$ and $\mathcal{F}_\infty^0(y) = \vee_{t\geq 0}\mathcal{F}_t^0(y)$. 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

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

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

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

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

i.e., we have

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

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

• *Remark* 5.23. 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 $s > t \geq 0$. 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 $\mathcal{F}(t+) = \cap_{\varepsilon>0}\mathcal{F}(t+\varepsilon)$. 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 5.22. This proof is inspired by Letta [87], based on a personal communication. □

The reader is referred to the books by Bremaud [21], Elliott [46], Protter [108]), and the comprehensive works by Bertoin [11, Chapters O and I, pp. 1–42] and Sato [116, Chapters 1 and 2, pp. 1–68].

5.11 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

$$P\{X(t) \in B \mid X(r),\, r \leq s\} = P\{X(t) \in B \mid X(s)\}, \tag{5.75}$$

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*,

$$P(s, x, t, B) = \int_{\mathbb{R}^d} P(s, x, r, \mathrm{d}y)\, P(r, y, t, B), \tag{5.76}$$

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

$$\Phi(t, s) = \Phi(t, r)\Phi(r, s) \ \text{ in } \ B(\mathbb{R}^d), \quad \forall t > r > s > 0, \tag{5.77}$$

and in the case of time-homogeneous Markov processes, the resolvent operators satisfy the so-called *resolvent equation*, namely

$$R(\lambda) - R(\nu) = (\nu - \lambda)R(\lambda)R(\nu) \ \text{ in } \ B(\mathbb{R}^d), \quad \forall \lambda, \nu > 0. \tag{5.78}$$

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 (5.75) 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 (5.76) we have

Definition 5.24 (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

$$P(t, s)\varphi(x) = \int_{\mathbb{R}^d} \varphi(y) P(s, x, t, \mathrm{d}y) = P(s, x, t, \varphi), \tag{5.79}$$

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 [129, 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 (5.79) with $\Phi(t,s) = P(t,s)$. In the time-homogeneous case, this relates with the resolvent operators by

$$R(\lambda)\varphi(x) = \int_0^\infty \mathrm{e}^{-t\,\lambda}\,\Phi(t)\varphi(x)\mathrm{d}t = \\ = \int_0^\infty \mathrm{e}^{-t\,\lambda}\mathrm{d}t \int_{\mathbb{R}^d} \varphi(y)P(t,x,\mathrm{d}y),\ \ \forall x \in \mathbb{R}^d, \quad (5.80)$$

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

$$P\{X(t) \in B \mid X(r),\, r \le s\} = P(s, X(s), t, B), \quad (5.81)$$

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 (5.75) 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

$$P\{X(t+\tau) \in B \mid X(r+\tau),\, r \le 0\} = P(\tau, t, X(\tau), B), \quad (5.82)$$

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 (5.82).

Only in very particular cases the transition function is explicitly known, such as a Wiener or a Poisson process, see (4.20) or (4.21). 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 $\{\Phi(t) : t \ge 0\}$ 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 5.24 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 Appendix B 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 5.24, 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 (5.81) 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 (5.81) 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 5.24 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

$$\alpha(\varepsilon,T,\delta,K) = \sup\big\{1 - P\big(s,x,t,\{y : |y-x| \leq \varepsilon\}\big) : \\ : \forall x \in K,\ s,t \in [0,T],\ 0 < t-s \leq \delta\big\}, \quad (5.83)$$

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

$$\lim_{\delta \to 0} \alpha(\varepsilon,T,\delta,\mathbb{R}^d) = 0, \quad \forall \varepsilon, T > 0, \quad (5.84)$$

This condition (5.84) is satisfied for a Feller transition.

The following result addresses the construction of standard Markov processes

Theorem 5.25. *Let $P(s,x,t,B)$ be a transition probability function satisfying (5.84). 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*

$$\lim_{\delta \to 0} \frac{\alpha(\varepsilon,T,\delta,\mathbb{R}^d)}{\delta} = 0, \quad \forall \varepsilon, T > 0, \quad (5.85)$$

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.,

$$P_{sx}\{X(\theta) \in B \mid \mathcal{F}(\tau)\} = P(\tau, X(\tau), \theta, B),\ \ \forall B \in \mathcal{B}(\mathbb{R}^d), \quad (5.86)$$

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 [36, Section XIV.24, pp. 169–172] or Sato [116, 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

$$\mathbb{E}\{Z\,\mathbb{1}_{B^X(n,4\varepsilon,R)}\}\leq\mathbb{E}\{Z\}\,[2\alpha(\varepsilon,T,\delta,\mathbb{R}^d)]^n, \tag{5.87}$$

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 (5.83). 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 (5.87) 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 (5.84), 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 (5.85) 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

$$\mathbb{E}\{R(n, \varepsilon, \cdot)\} \le n\, \alpha(\varepsilon, \ell, \ell/n).$$

Hence, condition (5.85) 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 $|X^*(t,\omega) - X^*(t-,\omega)| \le \varepsilon$, 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 (5.86) 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 $\{\tau_n : n \ge 1\}$ 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 (5.82) 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 (5.84) is replaced by

$$\begin{aligned} &\text{(a)} \quad \lim_{|x|\to\infty} \sup_{0\le s<t\le T} P(s,x,t,K) = 0, \\ &\text{(b)} \quad \lim_{\delta\to 0} \alpha(\varepsilon,T,\delta,K) = 0, \quad \forall \varepsilon, T > 0, \end{aligned} \tag{5.88}$$

for any compact subset K of $\mathbb{R}^d$, and assumption (5.85) can be substituted by

$$\lim_{\delta\to 0} \frac{\alpha(\varepsilon,T,\delta,K)}{\delta} = 0, \quad \forall \varepsilon, T > 0, \quad \text{any compact } K \subset \mathbb{R}^d, \tag{5.89}$$

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 [130, 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 5.19, 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 $\mathcal{B}(R^d)$ so that X is a (stochastically continuous) Markov process in 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 5.25 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, see Definition 5.15.

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 5.10. 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 $\mathcal{B}(\mathbb{R}^d)$,

$$\mu(B) = \pi^{-(d+1)/2}\Gamma(\tfrac{d+1}{2})\, c \int_B \left(|x-\gamma|^2 + c^2\right)^{-(d+1)/2} \mathrm{d}x,$$
$$\text{and} \quad \hat{\mu}(y) = \mathrm{e}^{-c|y| + \mathrm{i}\,\gamma\cdot y}, \quad \forall y \in \mathbb{R}^d, \quad (5.90)$$

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 [116, Theorem 11.7, pp. 63-64]. For instance, for dimension $d = 1$, the characteristic function is

$$\mathbb{E}\{\mathrm{e}^{\mathrm{i}\,y\cdot X(t)}\} = \mathrm{e}^{-t\,y^2/2}, \quad \forall t \geq 0,\ y \in \mathbb{R}^d,$$

and a simple calculation shows that condition (5.85) of Theorem 5.25 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 Blumental and Getoor [17], Dellacherie and Meyer [36, Chapters XI–XVI], Ethier and Kurtz [47], Sato [116, Chapter 1 and 2, pp. 1–68], among others.

5.12 Hunt and Standard Processes

In the modern theory of Markov processes, the emphasis is put on Markov transition functions $p(t, x, A)$, where $t \geq 0$, $x \in E$, a locally compact Hausdorff space, and A is any element of the Borel σ-algebra in E, as described in the previous Section 5.11 for the case $E = \mathbb{R}^d$. Thus, starting from a Markov transition function (or its Laplace transform, the resolvent), the actual construction of a Markov process having the prescribed transition functions is known a *realization* of the Markov process. Certainly, Kolmogorov construction and path regularity is the natural approach. Moreover, the strong Markov property is a highly desired. Therefore, continuous time Markov processes are usually constructed in the canonical cad-lag sample space $\Omega = D([0,\infty), E)$, i.e., we construct a probability measure P on Ω such that the canonical process $X_t(\omega) = \omega(t)$ is the desired Markov process. Furthermore, to simplify the notation, only the time-homogenous Markov process is considered, since by adding one dimension to the space (i.e., using $E \times [0,\infty)$ instead of E) all assumptions can be transported to the time-dependent case. Clearly, it is not necessary to have an explicit form for the transition function (or resolvent). Only a number of properties are involved,

which can be obtained from a given semigroup. The semigroup is described in term of its infinitesimal generator or its Dirichlet form.

Essentially based on super-median functions and super-martingales arguments, the general theory of processes (e.g. Dellacherie and Meyer [36, Section XIV.24, pp. 169–172]) shows that a cad-lag realization can be constructed for any Markov transition function, i.e., satisfying (a), (b) and (d) of Definition 5.24, i.e.,

(a) for each $t > 0$ and x in E the function $B \mapsto P(t, x, B)$ is a probability measure on $(E, \mathcal{B}(E))$,

(b) for each B in $\mathcal{B}(E)$ the function $(t, x) \mapsto P(t, x, B)$ is a measurable,

(c) for any x in E and B in $\mathcal{B}(E)$ we have

$$\lim_{t \to 0} P(t, x, B) = \delta_x(A),$$

i.e., the limit is equal to 1 if x belongs to A, and 0 otherwise,

(d) for each $s, t > 0$, x in E and B in $\mathcal{B}(E)$ we have

$$P(s + t, x, B) = \int_E P(s, x, \mathrm{d}y)\, P(t, y, B),$$

i.e., the Chapman-Kolmogorov identity holds.

Condition (c) is not actually necessary, but certainly is a natural complement to (d), which becomes necessary for the sub-Markov transitions, see below. Note that the Chapman-Kolmogorov identity combined with the inequality $0 \leq P(t, x, B) \leq 1$ show that the limit in (c) is monotone decreasing.

Besides the Feller property, there is another key property (necessary to build a nice theory), the so called *quasi-left continuity*. Moreover, the one-point compactification $\bar{E} = E \cup \{\infty\}$ (where the symbol ∞ does not belong to E) of a locally compact Hausdorff space E (the state space) and the concept of a *coffin state* and a *lifetime* are necessary (recall that for a locally compact space E, a point ∞ is adjoined to E as the point at infinity if K is not compact, and as an isolated point if K is compact). Typically, given a stopping time ς and an adapted E-valued cad-lag process Y on a filtered probability space $(\Omega, \mathbb{F}, P)$, $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$, we define a new process X as follows

$$X(t) = \begin{cases} Y(t) & \text{if } t < \varsigma, \\ \infty & \text{if } t \geq \varsigma, \end{cases}$$

which is $\bar{E}$-valued with lifetime ς. Clearly, the process Y needs only to be defined on the semi-open stochastic interval $[\![0, \varsigma[\![$. Now, a cad-lag (up to its lifetime) process X with values in $\bar{E}$ has the lifetime $\varsigma = \inf\{t \geq 0 : X(t) = \infty\}$ if $X(t) = \infty$ for every $t \geq \varsigma$. It is important to observe that X with values in $\bar{E}$ may not be "fully" cad-lag, the limit as $t \to \varsigma$ does not necessarily exist. Thus, the canonical cad-lag space $D([0, \infty), E)$ cannot be used with $\bar{E}$ instead of E, this requires some adjustment, via the so-called Ray resolvent. This means that the coffin state and one-point compactification state space are just convenient

notations, the key elements are the lifetime ς and the E-valued process X, which is cad-lag only up to ς, i.e., the pathwise left-hand limit $X(\varsigma-)$ may not exist, even on $\varsigma < \infty$.

Definition 5.26 (quasi-left continuity)**.** Let ς be a stopping time on a filtered probability space $(\Omega, \mathbb{F}, P)$, $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$. A cad-lag adapted process X with valued in E is called *quasi-left continuous* on $[\![0, \varsigma[\![$ if $X(\tau_n) \to X(\tau)$ almost surely on $\{\tau < \varsigma\}$, for any increasing sequence $\{\tau_n\}$ of stopping times converging to τ, with $\tau_n < \tau$ almost surely. □

Note that for a (cad-lag) quasi left-continuous we have $P\{X(t) = X(t-)\} = 0$ for every t in $[\![0, \varsigma[\![$ and the set of probability zero, where X_{τ_n} fails to converge, depends on the stopping time τ and the sequence $\{\tau_n\}$. Hence, in general the process is not continuous on the left, because it may not be possible to find a common set of probability zero for all times, i.e., a continuous modification of X does not necessarily exist.

Recall also that a stopping time τ is predictable if and only if that there exists an increasing sequence of stopping times τ_n with $\tau_n < \tau$ almost surely and converging to τ. Also, a stopping time τ is called *totally inaccessible* if for any predictable stopping time θ we have that $P(\tau = \theta < \infty) = 0$.

It is proved in Jacod and Shiryaev [69, Propositions 1.32, 2.26], that the random set of discontinuities of any cad-lag process X, i.e., the set of random jumps $J_X = \{(t, \omega), X(t_-, \omega) \neq X(t, \omega)\}$ has the form $J_X = \bigcup_n [\![\tau_n]\!]$ where τ_n is a sequence of stopping times, which is called a sequence that exhausts the jumps of X. Moreover, the following statements are equivalent:

(1) the process X is quasi left-continuous on $[\![0, \infty[\![$,

(2) we have $X(\tau^-) = X(\tau)$ almost surely on $\{\tau < \infty\}$ for every predictable time τ,

(3) there exists a sequence of totally inaccessible stopping times that exhausts the jumps of X.

5.12.1 Digression on Markov processes

As mentioned early and mainly for the notation simplicity, details are given only for time-homogeneous Markov processes.

Definition 5.27 (Markov processes)**.** A cad-lag *Markov process* with values in a Hausdorff space E is composed by the following elements:

(1) the canonical cad-lag space $D = D([0, \infty), E)$ endowed with σ-algebra $\mathcal{F}$, the σ-algebra generated by the canonical process $X_t(\omega) = \omega(t)$,

(2) a right-continuous increasing family of σ-algebras $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$ on $(D, \mathcal{F})$, i.e., a not necessarily completed filtration,

(3) a family of cad-lag processes (P^μ, X) or (P, X^μ) (depending on what is to be emphasized) indexed by μ, i.e., P^μ is a probability measure on $(D, \mathcal{F})$ and X^μ a measurable function from D into itself, such that the E-valued random variable $X(t)$ is $\mathcal{F}_t$-measurable, for every $t \geq 0$, and μ is the initial distribution

on the state space E, i.e., $P\{X(0) \in B\} = \mu(B)$, for every B in $\mathcal{B}(E)$,
(4) a transition function p on E, i.e., (a),..., (d) above are satisfied,
All these elements come together in the Markov property

$$P(X(t) \in B \mid \mathcal{F}_s) = p(t-s, X(s), B), \quad \forall t \geq s \geq 0, \ \forall B \in \mathcal{B}(E), \tag{5.91}$$

which is an almost surely equality. □

Observe that the left hand side in the Markov property is only defined almost surely, while the right hand side is well defined everywhere. This means that for every μ there exists a subset $N = N^\mu$ of the canonical space D with $P^\mu(N) = 0$ such that equality (5.91) holds outside of N.

The above definition may take place in an arbitrary probability space $(\Omega, \mathcal{F})$, but since the processes are cad-lag, the canonical sample can be used to concretize arguments. Moreover, with the notation (P^μ, X) the probability measure is emphasized with $X(t,\omega) = \omega(t)$ the canonical process, while (P, X^μ) suggests a fixed probability with a selectable process X^μ.

The indexation in term of distributions μ on the state space is not actually necessary, a smaller class suffices. Indeed, if the Markov process can be constructed for any $\mu = \delta_x$ with x in E and some measurability conditions are imposed on the mapping $x \mapsto (P_x, X)$ or $x \mapsto (P, X_x)$, where P_x and X_x correspond to the initial distribution δ_x, e.g., $P_x\{X(0) = x\} = 1$ or $P\{X_x(0) = x\} = 1$. Clearly, we have $P^\mu = P_x \, \mu(\mathrm{d}x)$.

The filtration $\mathbb{F}$ needs to be *universally completed* with respect to the probability measures P^μ or P_x. This is to enlarge each $\mathcal{F}_0$ (and consequently $\mathcal{F}_t$ for every $t > 0$) with sets which have measure zero relative to each P^μ, i.e., first define $\mathcal{F}_t^\mu$ by completing $\mathcal{F}_t$ with respect to P^μ, next set $\mathcal{F}_t^0 = \cap_\mu \mathcal{F}_t^\mu$ and finally, take $\bar{\mathcal{F}}_t = \cap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}^0$ to make it right-continuous if necessary, i.e., $\bar{\mathbb{F}} = (\bar{\mathcal{F}}_t : t \geq 0)$. In most of the cases, the initial filtration $\mathbb{F}$ is just the history of the process X, which may depends on μ if we insist in the (P, X^μ) setting. So, by using the embedding in the canonical space (P^μ, X) with $X(t,\omega) = \omega(t)$ the canonical processes, the initial filtration is fixed and $\bar{\mathbb{F}}$ is usefully.

For this universally completed filtration $\bar{\mathbb{F}}$, we consider an equality similar to (5.91), namely, for every almost surely finite stopping time τ relative to $\bar{\mathbb{F}}$ assume

$$P(X(\tau + t) \in B \mid \bar{\mathcal{F}}_\tau) = p(t, X(\tau), B), \quad \forall t \geq 0, \ \forall B \in \mathcal{B}(E), \tag{5.92}$$

which is again an almost surely equality, and where $\bar{\mathcal{F}}_\tau$ is the σ-algebra generated by the stopping time τ. This is referred as the *strong Markov property* and a Markov process satisfying this condition is called a *strong Markov process*. Certainly, condition (5.92) can be relative to the initial filtration $\mathbb{F}$, but some technical reasons lead to the universally completed filtration $\bar{\mathbb{F}}$.

For the canonical cad-lag space, the so-called *shift operator* $\vartheta_t(\omega) = \omega(\cdot + t)$ is viewed as mapping D into itself, and satisfying

$$X_s \circ \vartheta_t = X_t \circ \vartheta_s = X_{s+t}, \quad \forall t, s \geq 0,$$

where the composition $X_s \circ \vartheta_t(\omega) = X_s(\vartheta_t(\omega))$ is used. In the abstract setting the shift map ϑ is postulate with the above properties. The strong Markov property takes the form

$$P(X_t \circ \vartheta_\tau \in B \mid \bar{\mathcal{F}}_\tau) = p(t, X_\tau, B), \quad \forall t \geq 0,\ \forall B \in \mathcal{B}(E),$$

for every almost surely finite stopping time τ relative to $\bar{\mathbb{F}}$.

In all the above, it is clear that the transition function is the main element in the definition of Markov processes. The construction of a Markov transition function is quite delicate and several ways are known. Starting from a infinitesimal generator or a Dirichlet form, functional analysis and in particular the semigroup theory are used to obtain suitable transition functions. Its probabilistic counterpart starts with simple (or known) Markov processes and via some transformations a transition function is obtain, the key arguments are stochastic differential equations and its generalizations. In some cases, the transition function obtained is not quite a Markov transition function, it is what is called a *sub-Markov transition function*, i.e., all conditions are satisfied, except for (a) where $p(t, x, \cdot)$ is only a sub-probability, i.e., $B \mapsto p(t, x, B)$ is a measure with $p(t, x, E) \leq 1$. In this case, the condition (c) is more important.

Some restrictions on the state space E are necessary to deal with sub-Markov transition function, namely, E is now a locally compact Hausdorff. This is necessary to consider its one-point compactification $\bar{E} = E \cup \{\infty\}$ and the companion argument about coffin state and lifetime of a process. This sub-Markov case is reduced to the preceding Markov theory by extending the given sub-Markov transition function in E to a Markov transition function on $\bar{E}$. Indeed, set $\bar{p}(t, x, B) = p(t, x, B)$ for any Borel set B in E, $\bar{p}(t, x, \{\infty\}) = 1 - p(t, x, E)$, $\bar{p}(t, \infty, \{\infty\}) = 1$ and $\bar{p}(t, \infty, E) = 0$. As mentioned early, this extra point is often called the coffin state. The extra coffin state ∞ does not belong to E, it is the point at "infinity" when E is non compact and it is an isolated point when E is compact.

Hence, the sub-Markov case is reduced to a Markov case on a compact space $\bar{E}$, which has a particular isolated point. For the Markov process X corresponding to the Markov transition function $\bar{p}$ on the one-point compactification $\bar{E}$, the *lifetime functional*

$$\varsigma = \inf \big\{ t \geq 0 : X(t) = \infty \big\} \tag{5.93}$$

acting on the canonical space plays a fundamental role. Again, observe that the canonical cad-lag space still being $D([0, \infty), E)$ and not $D([0, \infty), \bar{E})$, because $\bar{X}$ with values in $\bar{E}$ is not "fully" cad-lag, the limit as $t \to \varsigma$ does not necessarily exist. Hence, a Markov process with values in the compact space $\bar{E}$ with lifetime $\varsigma = \infty$ (almost surely for every P_μ) is actually a realization of a Markov transition on the initial locally compact space E. A sub-Markov transition function produces either a E-valued cad-lag *sub-Markov process* with a lifetime $\varsigma < \infty$ with positive probability or equivalently a $\bar{E}$-valued cad-lag (up to its lifetime) Markov process.

Most of the interest is on state spaces E, which are actually Polish spaces (complete separable metric spaces) so that the canonical cad-lag sample space $D([0,\infty),E)$ is also a Polish space. The locally compact character is used when dealing with sub-Markov transition functions.

The canonical process $X_t(\omega) = X(t,\omega) = \omega(t)$ and the shift map $\vartheta_t(\omega) = \vartheta(t,\omega) = \omega(\cdot + t)$, which are defined on the canonical space $D([0,\infty),E)$, may be considered as E-valued and D-valued cad-lag processes, respectively (note that $X = \{X(t) : t \geq 0\}$ can be regarded as a D-valued random variable). If we work on the one-point compactification state space $\bar{E}$ and the canonical filtration $\mathbb{F}$ is used (i.e., generated by the canonical process) then the lifetime functional (5.93) can be interpreted as a stopping time. There are other type of functionals or processes that we may consider

Definition 5.28 (functional)**.** Let A and L be measurable functions from the canonical cad-lag space $D([0,\infty),E)$ into $D([0,\infty),\mathbb{R})$. Then A is called an *additive* functional if $A_0(\omega) = 0$ and $A_t(\omega) - A_s(\omega) = A_{t-s}(\omega(\cdot + s))$, for every $t \geq s \geq 0$ and ω in $D([0,\infty),E)$. Also L is called a *multiplicative* functional if $L_0(\omega) = 1$ and $L_t(\omega)L_s(\omega) = L_{t-s}(\omega(\cdot + s))$, for every $t \geq s \geq 0$ and ω in $D([0,\infty),E)$. Similarly, an additive functional A is called *increasing* (or nondecreasing) if $A_t - A_s \geq 0$, for every $t > s \geq 0$, and a multiplicative functional L is called *positive* (or nonnegative) if $L_t \geq 0$, for every $t > 0$. If A or L is considered on the one-point compactification state space $\bar{E}$ then we also require A or L to be constant on the stochastic interval $[\![\varsigma,\infty[\![$, i.e., it has been extended to $[0,\infty)$ with values in $\bar{E}$. □

It is possible to generalize and suppose that A_t has bounded variation trajectories, still having the possibility to integrate with respect to A_t in the Stieltjes sense. Typical examples are

$$A_t = \int_0^t c(X_s)\mathrm{d}s \quad \text{and} \quad L_t = \exp\Big\{\int_0^t c(X_s)\mathrm{d}s\Big\},$$

where c is a positive (measurable) function defined on E, and in the case of $\bar{E}$ we suppose $c(\infty) = 0$.

If one is working on an abstract measurable space $(\Omega,\mathcal{F})$ then the shift operator ϑ is used to re-write the conditions as $A_t - A_s = A_{t-s}\circ\vartheta_s$ and $L_tL_s = L_{t-s}\circ\vartheta_s$, for every $t > s \geq 0$. In this case, functionals are regarded as either $\mathbb{R}$-valued cad-lag processes or $D([0,\infty),\mathbb{R})$-valued random variables.

As usually, the difficulty appears as soon as a probability (or a family of probabilities, as in the case of Markov processes) is assigned on the canonical space $D([0,\infty),E)$. The functionals are almost surely defined and we are interested in having "good" versions (or modifications) of them.

5.12.2 Classification of Markov processes

Concrete examples of Markov processes like diffusion processes, jump processes and Lévy processes have many properties in common besides the Markov property. They have all the Feller property, i.e., the semigroup associated to the

transition functions is a (strongly) continuous on $C_0(E)$, in particular, it maps the function space $C_0(E)$ in itself or equivalent the Feller property is satisfied. Here the state space E is a locally compact Hausdorff space (usually a Polish space), and $C_0(E)$ is the space of continuous functions null at infinity. Note that the Feller processes are defined just starting from the transition functions.

It can be proved that given a Feller transition function $p(t, x, A)$ in E, there exist strong Markov processes (with cad-lag quasi left-continuous trajectories) having p as its transition function, e.g. see Rogers and Williams [112, Chapters III and VI]. The class of processes defined by these properties is the class of Hunt processes, which contains the Feller processes. A Hunt process can be constructed from any regular symmetric Dirichlet form (see Fukushima et al. [51]). However, to extend this result to regular non-symmetric Dirichlet forms (see Ma and Röckner [91]), it is necessary to consider a light generalization of Hunt process, that of *standard process*, introduced by Blumenthal and Getoor [17].

Roughly speaking, we can schematize

$$\text{Feller processes} \subset \begin{matrix}\text{Hunt processes} \subset \text{standard processes} \subset \text{right processes} \\ \text{Ray processes}\end{matrix}$$

where formal definitions are given below.

Definition 5.29 (Hunt and standard)**.** Let us be a time-homogeneous cad-lag Markov process (as in Definition 5.27) with values in a Hausdorff space E (i.e., obtained form a Markov transition function on E where its lifetime functional $\varsigma = \infty$ by definition) or with values in the one-point compactification $\bar{E}$ of a locally compact Hausdorff space E (i.e., obtained form a sub-Markov transition function on E where ς denotes its lifetime functional).

(a) It is called a *Hunt process* if the strong Markov property (5.92) is satisfied and the paths are quasi left-continuous on $[0, \infty)$, with respect to any probability P^μ.

(b) It is called a *standard process* if the strong Markov property (5.92) is satisfied and the paths are quasi left-continuous on $[0, \varsigma)$, with respect to any probability P^μ.

(c) Finally, it is *special standard* if also the random variable X_τ is $\bigvee_n \mathcal{F}_{\tau_n}$-measurable, for any increasing sequence $\{\tau_n\}$ of stopping times converging to τ, with $\tau_n < \tau$ almost surely. □

Here $\bigvee_n \mathcal{F}_{\tau_n}$ is the smaller σ-algebra which contains $\mathcal{F}_{\tau_n}$, for every $n \geq 1$. Note the small difference with Hunt and Standard processes on the quasi left-continuous property that is valid almost surely on $\{\tau < \zeta\}$, instead of $\{\tau < \infty\}$. Thus a standard process is a Hunt process if its lifetime is infinite or is it is quasi left-continuous at ς, i.e., a Hunt process is realized in the canonical space $D([0, \infty), E)$ or $D([0, \infty], \bar{E})$. Clearly, it can be proved that any Hunt process is a special standard process.

The following result holds:

Theorem 5.30. *A Hunt process admits a Lévy system, i.e., there exists a continuous additive functional H and a family of kernel $N(x, \mathrm{d}y)$ on E, such*

that $t \mapsto H_t$ is a continuous, $N(x,\{x\}) = 0$ for any $x \in E$, and

$$\mathbb{E}^\mu\Big(\sum_{0<s\le t} f(X_{s-}, X_s)\mathbb{1}_{J_X}(s,\omega)\Big) = \\ = \mathbb{E}^\mu\Big(\int_0^t \mathrm{d}H_s \int_E f(X_s, y)N(X_s, \mathrm{d}y)\Big) \quad (5.94)$$

for any Borel positive f, defined on $E \times E$, and any initial distribution μ on E, where $J_X = \{(t,\omega), X(t_-,\omega) \neq X(t,\omega)\}$. □

Recall that a kernel $N(x, \mathrm{d}y)$ on E means (a) for each x in E the map $B \mapsto N(x,B)$ is a (σ-finite) measure on $\mathcal{B}(E)$ and (b) for each B in $\mathcal{B}(E)$ the map $x \mapsto N(x,B)$ is measurable. For instance, a proof can be found in Benveniste and Jacod [10] or Sharpe [119, Section 73, pp. 342–350].

Note that if we consider any Borel positive f, defined on $E\times E$ with $f(x,x) = 0$ then (5.94) reads

$$\mathbb{E}^\mu\Big(\sum_{0<s\le t} f(X_{s-}, X_s)\Big) = \mathbb{E}^\mu\Big(\int_0^t \mathrm{d}H_s \int_E f(X_s, y)N(X_s, \mathrm{d}y)\Big),$$

and for $f = 1$ out of diagonal we get that $\mathbb{E}^\mu\{\mathsf{number\ of\ jumps\ in}\ (0,t]\} = \mathbb{E}^\mu\{H_t\}$.

Following the discussion in Rogers and Williams [112, Chapter III], the class of Ray processes (defined in a axiomatic way) is the most convenient, among the various classes of Markov processes, to cope all the (time-continuous) Markov chains. They are in a sense "equivalent" to the class of right processes, introduced by Meyer (*la classe droite*), see also Dellacherie and Meyer [36, Chapters XI–XVI] or Sharpe [119, Chapter I].

As already pointed out, to give a Markov transition function $p(t,x,A)$ on E (which is jointly measurable in t and x) is equivalent to give the Laplace transform, called in this context a *resolvent kernel*

$$R_\lambda(x,A) = \int_0^\infty \mathrm{e}^{-\lambda t}\, p(t,x,A)\ \mathrm{d}t$$

For any bounded Borel function f on E we consider also the resolvent operator

$$R_\lambda f(x) = \int_E f(y)\ R_\lambda(x, \mathrm{d}y)$$

that verifies the resolvent identity $R_\lambda - R_\mu + (\lambda - \mu)R_\lambda R_\mu = 0$ and $\lambda R_\lambda 1 = 1$. If it happens that the resolvent operator maps $C_0(E)$ into itself and $\lambda R_\lambda \to I$ in the uniform convergence of $C_0(E)$, then the Markov process associated is Feller process.

Ray weakened the condition of strong continuity in zero, introducing the so called Ray resolvent. To this end, let us introduce the concept of *α-supermedian function*, i.e. a positive, continuous and bounded function on E such

that

$$\lambda R_{\lambda+\alpha} f \leq f, \qquad \forall \lambda > 0,$$

and let us denote the *cone* of α-super-median functions by CSM^α.

Now, $\{R_\lambda\}$ is called a *Ray resolvent* on E if each R_λ maps $C_b(E) \to C_b(E)$ and

$$\bigcup_{\alpha \geq 0} \mathsf{CSM}^\alpha \text{ separates points of } E.$$

In other words, given two points x, x' in E, there exist a α and a α-super-median function f such that $f(x) \neq f(x')$. The canonical Markov process associated to a Ray resolvent admits a cad-lag modification that has the strong Markov property.

• *Remark* 5.31. In potential theory the resolvent operator R_λ is called the λ-*potential* operator. There is a concept similar to α-super-median functions in term of the semigroup P_t associated to the Markov transition function $p(t, x, A)$, namely, a positive Borel function f is called α-*super-mean-valued* (also called α-*super-averaging*, see Chung [27, p. 45]) if $\mathrm{e}^{-\alpha t} P_t f \leq f$ for all $t \geq 0$. A α-super-mean-valued function is also a α-super-median function, but, in general, the converse is not true, i.e., a α-super-median function is not necessarily a α-super-mean-valued function. In potential theory it is important also the concept of α-*excessive* function f, i.e., a α-super-mean-valued function f such that $\mathrm{e}^{-\alpha t} P_t f \uparrow f$ as $t \downarrow 0$. This class of functions are useful generalization of *super-harmonic* functions in classical potential theory. □

Definition 5.32 (right and Ray processes)**.** The class of *right processes* with state space E is the class of time-homogeneous cad-lag strong Markov processes, such that the process $f(X_t)$ is almost surely right continuous for any $\alpha > 0$ and any α-super-mean-valued function f. On the other hand, a *Ray process* is one with Ray resolvent, i.e., the super-median functions $\cup_{\alpha \geq 0} \mathsf{CSM}^\alpha$ separates points of E. □

A useful result is the fact that any standard process is a right process, which is not so simple to prove. As mentioned early, Ray processes are (time-homogeneous) cad-lag strong Markov processes.

5.12.3 Some Examples and Remarks

First we give a couple of examples.

Example 5.1. Consider jump Markov processes, for instance, see the books Bremaud [21], Feller [48, II, Section X.3, pp. 316–320], Ethier-Kurtz [47, Section 8.1, pp. 376–382], Sharpe [119, Section 72, pp. 339–342]. Let $\{X_t\}$ be a cad-lag piecewise constant process. In order that $\{X_t\}$ be a Markov process, we introduce a kernel $Q(x, B)$ on E, which describes the probability distribution of the points where the process jumps away from x, and a function $\lambda(x) \geq 0$, that

describes the rate of jumps, i.e. $P^x(\tau > t) = \mathrm{e}^{-\lambda(x)t}$, where $\tau = \inf\{t : X_t \neq X_0\}$. We set $\tau_0 = 0$ and define recursively $\tau_n = \tau_{n-1} + \tau(\vartheta_{\tau_{n-1}})$ the subsequent jump times. Alternatively we describe the process by means of infinitesimal generator

$$Af(x) = \lambda(x) \int_E \big[f(y) - f(x)\big]\, Q(x, \mathrm{d}y).$$

In this case, if λ is bounded then then Lévy system can be taken as

$$H_t = \int_0^t \lambda(X_s)\mathrm{d}s, \qquad N(x, B) = Q(x, B).$$

If λ is not bounded we can take

$$H_t = t, \qquad N(x, B) = \lambda(x)\mathbb{1}_{\lambda(x)<\infty} Q(x, B).$$

In general the Lévy system for a Markov process is not uniquely determined. □

Example 5.2. In the case of a Lévy process in $\mathbb{R}^n$, with Lévy measure ν, then we choice for a Lévy system can be

$$H_t = t, \qquad N(x, B) = \nu(B - x).$$

For example, consider the symmetric Cauchy process, i.e. a purely jump Lévy process with Lévy measure

$$\nu(\mathrm{d}y) = \frac{1}{\pi}\frac{1}{y^2}\mathrm{d}y$$

on $\mathbb{R} \setminus \{0\}$. □

Now, let us show some application of additive functionals. Given a Hunt process $\{X_t\}$ with Markov transition function $p(t, x, A)$, consider a continuous positive additive functional A_t and the multiplicative functional

$$L_t = \mathrm{e}^{-A_t}$$

as in Definition 5.28. We will consider two other Hunt processes, starting from X_t.

The first is X_t^A, the *canonical subprocess* of X_t with respect to the multiplicative functional L_t. It is constructed starting from the Markov transition function $p^A(t, x, B)$, given by the semigroup

$$P_t^A f(x) = \mathbb{E}^x\{L_t f(X_t))\}, \quad x \in E,\ f \text{ Borel and positive.}$$

It is possible also to construct X_t^A by introducing a new "lifetime"

$$\varsigma^A = \inf\{t < \zeta : A_t \geq \zeta\},$$

where ζ is a random variable exponentially distributed with mean value 1 and independent of the Hunt process X_t with respect to P^x, for every $x \in E$. Then define

$$X_t^A = X_t, \quad \text{for } t < \varsigma^A$$

and $X_t^A = \infty$ otherwise. For details see Fukushima et al. [51, pag. 326].

The second process is given by

$$\check{X}_t = X_{\tau_t}, \qquad \tau_t(\omega) = \inf\{s > 0 : A_s(\omega) > t\}.$$

with its natural filtration $\check{\mathcal{F}}_t = \mathcal{F}_{\tau_t}$, for $t \geq 0$. It can proved that the process $\{\check{X}_t\}$ is a strong Markov process, with respect to the filtration $\{\check{\mathcal{F}}_t : t \geq 0\}$. Moreover, if A_t is a strictly positive continuous additive functional, then $\{\check{X}_t\}$ is a Hunt process.

There are other functionals (besides additive or multiplicative) that can be used to transform Markov processes, e.g., the first exit time from a region (usually a smooth open or closed subset of $\mathbb{R}^d$). For instance, if $\mathcal{O}$ is an open connected subset of $\mathbb{R}^d$ (which is the interior of its closure) then the procedure of stopping the process at the first exit time from the closure of $\mathcal{O}$, namely,

$$X_t^{\mathcal{O}} = X_{t\wedge\tau}, \qquad \tau(\omega) = \inf\{s \geq 0 : X_s(\omega) \in \mathbb{R}^d \setminus \mathcal{O}\},$$

produces a Hunt Markov process if the initial we were so. However, $X^{\mathcal{O}}$ may not be a Feller process when X is Feller. Note that the Hunt character of the process has to do mainly with time-regularity of the paths, which is clearly preserved by the above functional (even if $\mathcal{O}$ is only a Borel set), while the Feller character involves the study of $\tau = \tau_x$ as a functional depending on the initial condition $X_0 = x$. The reader may consult, among other sources, the papers Stroock and Varadhan [127, 128] for a complete study on the (degenerate) diffusion processes with boundary conditions.

Perhaps, the prototype of Hunt processes are the diffusions process (with jumps) in infinite dimension or with boundary conditions. For instance, a Levy or Wiener process in infinite dimension does not produce a strongly continuous semigroup in C_0, even if the Feller property (i.e., mapping C_0 into itself) is satisfied. If we stop a degenerate diffusion in $\mathbb{R}^d$ at the first exit time from a smooth domain then we produce a simple example of a Markov process without the Feller property, which is stochastically continuous and produces a Hunt process. Moreover, piecewise deterministic processes (see Davis [34] and references therein) may not be Feller processes (this is mainly due to the nature of the boundary conditions or jump-mechanism, they may have predictable jumps, so they are not necessarily quasi left-continuous, see Section 5.9), but they are cad-lag strong Markov processes, actually, Ray and right processes. If a Markov transition function yields a Hunt process then we may expect that a sub-Markov transition functions (with the same degree of regularities) should yield a special standard process. On the other hand, most examples of sub-Markov transition functions are obtained form Markov processes and multiplicative functional, so

that really they produce Hunt processes. More representative are stochastic differential equations with unbounded coefficients (e.g., see Stroock and Varadhan [129, Chapter 10, pp. 248–260], which may yield explosions or solutions with a finite lifetime, i.e., special standard processes which are not Hunt processes.

5.13 Final Comment on Markov Processes

In general, a Markov Process (MP) has a state space X and a time space T, and depending on where T is typically $\mathbb{N}_0 = \{0, 1, \ldots\}$ or $\mathbb{R}_0^+ = [0, \infty[$, or a subset set of $\mathbb{N}_0$ or of $\mathbb{R}_0^+$ (moreover, sometimes the notation is $\mathbb{N}$ for $\mathbb{N}_0$ and $\mathbb{R}^+$ for $\mathbb{R}_0^+$); or adding $\{\infty\}$ as in $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$, or using only a subset of them (e.g., a finite set $\{1, 2, \ldots, N\}$ or an interval $[a, b]$, $]a, b[$, etc., or even $\mathbb{Z} = \{0, \pm 1, \pm, 2, \ldots\}$ or $\mathbb{R}$. More general situations can also be considered, but always T is a 'linearly ordered set' (i.e., an order relation is given and for any two elements t_1 and t_2 it is known whether $t_1 < t_2$, $t_1 > t_2$ or $t_1 = t_2$, this means that 'partial order' are in other category!) within one of the following types: (1) discrete (= countable set of isolated points, e.g., $\mathbb{N}_0$), or (2) countable (with possible non-isolated points and under the trivial topology, i.e. every subset is open, e.g. the rational numbers $\mathbb{Q}$), or (3) trivial-uncountable (i.e., with the trivial topology, e.g., the first uncountable ordinal Ω_0), or (4) continuous (or continuous-type = non of the previous and a Borel space with a linear order, recall Borel = measurable subset of a complete separable metric space).

Perhaps, the classification of T is more important (or should go first) than the classification of the state space X. Actually, it seems that only the order structure on T and a topology on X are necessary for the so-called Markov property, which is the central point of MP. The state space X could be classified as (1) discrete (finite or infinite), or (2) countable, or (3) trivial-uncountable, or (4) continuous (= Borel space), or better (5) Polish space, etc. Usually, for measure-type properties a Borel space X gives a good setting, and this is used for a discrete times space T (like $T = \mathbb{N}$), while for a Polish space X is preferred for a continuous times space (like $T = \mathbb{R}^+$). Thus, typically we say 'discrete-time MP' when $T = \mathbb{N}_0$ and 'continuous-time MP' when $T = [0, \infty[$.

The name MP becomes Markov chain (MC) when the state space X is not continuous, and also when only the time space T is not continuous. This means that a MC could be in discrete-time or continuous-time as in MP. However, typically a MC has a finite state space $X = \{1, 2, \ldots, N\}$ and $T = \mathbb{N}_0$. The theory specify that a MP has: (i) a transition probability function (or kernel) and this TPF or TPK is, in general, hard to give explicitly; (ii) an infinitesimal generator and this IG, in general, is easier to manipulate than the TPK; actually, properties on the TPK or IG (weak or strong) are used to fully determine the 'specific of MP', so that a construction or 'realization' of the MP can be given in a suitable probability space, where the Markov property holds, mostly usable in its 'strong' form. Actually, IGs are related to semigroups (SGs) and this gives a natural connection with 'semigroup theory' in Functional Analysis, where certainly, not every semigroup or IG corresponds to a MP. Moreover, adding

the t, we get the 'theory of evolution operators', etc., etc., ...

It is usual to study only homogeneous MP, since any non-homogeneous MP can be considered as a homogeneous MP with an enlarged state space, namely $T \times X$. Note that technically, a homogeneous MP is necessarily a particular case of a non-homogeneous MP. There is also a large class called 'semi-Markov processes' (SMP), which in short, they are MPs set in any of the four possibilities for $T \subset \mathbb{R}^+$, where the state space has the form $\mathbb{R}^+ \times X$, and if a SMP is regarded in the state space X then it is not necessarily a non-homogeneous MP with $T = \mathbb{R}^+$. Markov Renewal Processes (MRPs) is a well know class of SMPs on Borel spaces X (= state), which are MPs is the discrete time $\mathbb{N}_0$ with state in $\mathbb{R}^+ \times X$.

Usually, MCs are homogeneous in discrete time $T = \mathbb{N}_0$ or continuous time $T = \mathbb{R}^+$, and on Borel state spaces X, i.e., a (either homogeneous or non-homogeneous) continuous time (or discrete time) MC could be considered as a MRP (or a SMP with $T = \mathbb{N}_0$). A (homogeneous) Poison processes can be regarded as a (homogeneous) MC with either $T = \mathbb{R}_0^+$ and $X = \mathbb{N}_0$ or $T = \mathbb{N}_0$ and $X = \mathbb{R}_0^+ \times \mathbb{N}_0$, while a (homogeneous) compound Poison processes can be regarded as a (homogeneous) MC with $T = \mathbb{R}_0^+$ and X a Borel space. Indeed, the IG of a homogeneous MC with discrete time $T = \mathbb{N}_0$ and a Borel state space X has the form

$$\varphi \mapsto A\varphi(x) = \int_X [\varphi(y) - \varphi(x)] m(\mathrm{d}y),$$

where $m(\cdot)$ is a probability on X, and if the MC is set in continuous time $T = \mathbb{R}_0^+$ then $m(\cdot)$ is finite measure on X, and $m(X)$ determines the exponential law of transitions in time. In particular, this expression of A becomes a sum when X is countable, i.e., the known matrices in the finite case. A homogeneous Poisson measure $p(t, B)$, $t > 0$, $B \in \mathcal{B}(\mathbb{R}_*^d)$ with Levy measure $m(B) = \mathbb{E}\{p(t, B)\}/t$ of order 1 defines a MP with a possible unbounded IG

$$A\varphi(x) = \int_{\mathbb{R}_*^+} [\varphi(y) - \varphi(x)] m(\mathrm{d}y),$$

where $m(\cdot)$ is a measure on $\mathbb{R}_*^d = \mathbb{R}^d \setminus \{0\}$ integrating the function $x \mapsto |x|$, i.e., the above expression of A makes sense when φ is C^1. Adding a variable t in the measures $m(B)$ as $m(t, B)$, we could get a non-homogeneous MC or a non-homogeneous Poisson measure (or a point process), and even when $m(]a, b], B)$ we get a SMP with $T = \mathbb{R}_0^+$. Clearly, the second case is even more general than MRPs, and usually called Markov or semi-Markov point processes. Certainly, stochastic differential equations (SDEs) driven by Lévy processes (or a combination of a Wiener process and a Poisson measure) are an excellent class of more complicated examples.

Appendix A

Exercises - Chapter (1) Elementary Probability

All exercises are re-listed here, but now, most of them have a (possible) solution. **Certainly, this is not for the first reading**. This part is meant to be read after having struggled (a little) with the exercises. Sometimes, there are many ways of solving problems, and depending of what was developed "in the theory", solving the exercises could have alternative ways. In any way, some exercises are trivial while other are not simple. It is clear that what we may call "Exercises" in one textbook could be called "Propositions" in others.

(1.1)Preliminary Examples

(1.1.1) Discrete Probabilities

(1.1.2) Other Probabilities

(1.1.3) Independent Random Variables

Exercise 1.1. Let $\{F_i : i \geq 1\}$ be a sequence of distributions in $\mathbb{R}$, i.e., each F_i is a cad-lag non-decreasing function such that $\lim_{r\to-\infty} F_i(r) = 0$ and $\lim_{r\to\infty} F_i(r) = 1$. Show that there exits a sequence $\{X_i : i \geq 1\}$ of independent real-valued random variables defined on the universal probability space $(\Omega, \mathcal{F}, P)$, $\Omega = [0, 1)$, $\mathcal{F}$ the Borel σ-algebra and P the Lebesgue measure, such that $P(X_i \leq r) = F_i(r)$, i.e., each X_i has distribution F_i. Hint: First, complete the above arguments so that it is clear the construction of a sequence $\{\xi_i : i \geq 1\}$ of independent random variable uniformly distributed. Next, define the inverse of each F_i as $F_i^{-1}(s) = \inf\{r \in \mathbb{R} : s \leq F_i(r)\}$, for every s in $[0, 1)$ and verify that the sequence $\{X_i = F_i^{-1}(\xi_i) : i \geq 1\}$ has the required properties. □

(1.2) Laws of Large Numbers

(1.3) Convergence of Probabilities

(1.3.1) Tightness

(1.3.2) Approximation

(1.3.3) Various Types of Convergence

Exercise 1.2. Consider the Hilbert cube $\mathcal{H} = [0,1]^\infty$, i.e., h belongs to $\mathcal{H}$ if and only if $h\colon \{1,2,\ldots\} \to [0,1]$ endowed with the product norm $\mathrm{d}_{\mathcal{H}}(g,h) = \sum_i 2^{-i}|g(i) - h(i)|$. Verify that $\mathrm{d}_{\mathcal{H}}(h_n,h) \to 0$ if and only if $h_n(i) \to h(i)$ for every i. Let (X,d) be a metric space with a countable dense subset $\{e_i : i \geq 1\}$ and define the map $\Phi\colon X \to \mathcal{H}$ by the formula $h = \Phi(x)$, $h(i) = \min\{\mathrm{d}(x,e_i),1\}$. Prove $\mathrm{d}(x_n,x) \to 0$ if and only if $\mathrm{d}_{\mathcal{H}}(\Phi(x_n),\Phi(x)) \to 0$. Deduce Urysohn's Theorem, namely, any separable metric space is homeomorphic to a subset of $\mathcal{H}$, i.e., Φ is injective continuous and open. The same map Φ can be used to convert the Borel measures on X to the Borel measure on $\mathcal{H}$. □

Exercises - Chapter (2) Basic Probability

(2.1) Characteristic Functions

Exercise 2.1. Beside the computation of the characteristic function of the normal distribution in the real line, we should be able to verify the following calculations:

(1) if x is a Normal random variable with mean a and variance b, i.e., with (Lebesgue) density $(2\pi)^{-1/2}\exp\big(-(x-a)^2/(2b^2)\big)$ then $\mathbb{E}\{e^{itx}\} = \exp\big(iat - b^2t^2/2\big)$,

(2) if x is a Poisson random variable, i.e., $P\{x = k\} = e^{-\lambda}\lambda^k/k!$, for any $k = 0, 1, \ldots$, then $\mathbb{E}\{e^{itx}\} = \exp\big(\lambda(e^{it} - 1)\big)$,

(3) if x is a random variable with a uniform distribution on (a, b), i.e., with (Lebesgue) density $\mathbb{1}_{(a,b)}/(b-a)$ then $\mathbb{E}\{e^{itx}\} = \big(e^{ibt} - e^{iat}\big)/\big(it(b-a)\big)$,

(4) if x is a random variable with a triangular distribution on $(-1/a, 1/a)$, i.e., with (Lebesgue) density $\mathbb{1}_{(-1/a,1/a)}(a - a^2|x|)$ then $\mathbb{E}\{e^{itx}\} = 2(1-\cos at)/(a^2t^2)$,

(5) if x is a random variable with an exponential distribution, i.e., with (Lebesgue) density $\mathbb{1}_{(0,\infty)}e^{-\lambda x}\lambda$ then $\mathbb{E}\{e^{itx}\} = \lambda/(\lambda - it)$,

(6) if x is a random variable with a bilateral distribution, i.e., with (Lebesgue) density $e^{-\lambda|x|}\lambda/2$ then $\mathbb{E}\{e^{itx}\} = \lambda/(\lambda^2 + t^2)$.

(7) if x is a Polya random variable, i.e., with (Lebesgue) density $(1-\cos x)(\pi x^2)$ then $\mathbb{E}\{e^{itx}\} = (1 - |t|)^+$.

(8) if x is a Cauchy random variable, i.e., with (Lebesgue) density $1/\big(\pi(1+x^2)\big)$ then $\mathbb{E}\{e^{itx}\} = e^{-|x|}$.

Essentially, we should calculate (3) to deduce the following expressions by using linearity and convolution, e.g., see Durrett [42, Section 2.3, pp. 91-98]. □

(2.2) Central Limit Theorem

Exercise 2.2. **(1)** Consider the dyadic numbers $R_n = \{i2^{-n} : i = 1, \ldots, 4^n\}$, $R = \bigcup_n R_n$ and prove that $\sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \le r} = r2^n$, for every r in R.

(2) Let $\{X_{i,n} : i = 1, \dots, 4^n,\ n \geq 1\}$ be a countable family of identically distributed random variables with $\mathbb{E}\{X_{i,n}\} = 0$ and $\mathbb{E}\{|X_{i,n}|^2\} = 1$, and such that for every index $n \geq 1$ fixed, $\{X_{i,n} : i = 1, \dots, 4^n\}$ is a set of independent random variables. Define

$$W_{n,r} = 2^{-n/2} \sum_{i=1}^{4^n} X_{i,n} \mathbb{1}_{i2^{-n} \leq r}, \quad \forall n \geq 1,\ r \in R,$$

and revise the arguments in Theorem 2.2 to show that the distribution of the sequence $\{W_{n,r} : n \geq 1\}$ converges to the normal distribution $N(0, r)$.

(3) If, besides the condition on (2), we assume that $\{X_{i,n} : i = 1, \dots, 4^n,\ n \geq 1\}$ is a set of independent random variables then, use the technique of Theorem 2.2 to show that the distribution of series

$$W_r = \sum_n 2^{-n} \sum_{i=1}^{4^n} X_{i,n} \mathbb{1}_{i2^{-n} \leq r}, \quad \forall r \in R,$$

converges to the normal distribution $N(0, r)$. □

(2.3) Conditional Expectation

Exercise 2.3. If $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{A}$ is a sub σ-algebra of $\mathcal{F}$ then denote by $L^2_0(\mathcal{A})$ the closed subspace of $L^2(\Omega, \mathcal{F}, P)$ containing all $\mathcal{A}$-measurable functions with zero mean, i.e.,

$$L^2_0(\mathcal{A}) = \big\{ f \in L^2(\Omega, \mathcal{F}, P) : f \text{ is } \mathcal{A}\text{-measurable and } \mathbb{E}\{f\} = 0 \big\}.$$

Show that two sub σ-algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ of $\mathcal{F}$ are independent if and only if $L^2_0(\mathcal{A}_1)$ is orthogonal to $L^2_0(\mathcal{A}_2)$, i.e.,

$$\mathbb{E}\{fg\} = 0, \quad \forall f \in L^2_0(\mathcal{A}_1),\ g \in L^2_0(\mathcal{A}_2).$$

Prove or disprove an analogue result for a family of σ-algebras $\{\mathcal{A}_i : i \in I\}$. □

Exercise 2.4. Let $(\Omega, \mathcal{F})$ be a measurable space. Recall that a π-systems $\mathcal{F}_0$ is a subset of $\mathcal{F}$ which is stable under finite intersections, i.e., if A and B belongs to $\mathcal{F}_0$ then $A \cap B$ also belongs to $\mathcal{F}_0$. Also, we denote by $\sigma(\mathcal{F}_0)$ the minimal sub σ-algebra of $\mathcal{F}$ containing all the elements of $\mathcal{F}_0$, i.e. the σ-algebra generated by $\mathcal{F}_0$. Prove that if $\mathcal{H}$ and $\mathcal{G}$ are two sub σ-algebras which are generated by the π-systems $\mathcal{H}_0$ and $\mathcal{G}_0$, then $\mathcal{H}$ and $\mathcal{G}$ are independent if and only if $\mathcal{H}_0$ and $\mathcal{G}_0$ are independent, i.e., if and only if $P(H \cap G) = P(H)P(G)$ for any H in $\mathcal{H}_0$ and G in $\mathcal{G}_0$ (e.g., see the book by Bauer [6, Section 5.1, pp. 149–154]). □

Exercise 2.5. Establish the existence for the conditional expectation on a given probability space $(\Omega, \mathcal{F}, P)$ for an integrable random variable X with respect to a sub σ-algebra $\mathcal{G}$ by two ways. Firstly (a) by means of the Radon-Nikodym theorem, i.e., on the measurable space $(\Omega, \mathcal{G})$ consider the probability measures

$\nu(G) = \mathbb{E}\{X\mathbb{1}_G\}$ and $\mu(G) = \mathbb{E}\{\mathbb{1}_G\}$ satisfying $\nu \ll \mu$. Secondly (b) by means of the orthogonal projection π from the Lebesgue space $L^2(\Omega, \mathcal{F}, P)$ into the closed subspace $L^2(\Omega, \mathcal{G}, P)$, i.e., π satisfies $(X - \pi(X), Y) = 0$, for any Y in $L^2(\Omega, \mathcal{G}, P)$, where $(\cdot, \cdot)$ denotes the scalar product. □

Exercise 2.6. Let $G_1, \ldots, G_n$ be a measurable disjoint sets in probability space $(\Omega, \mathcal{F}, P)$ with $P(G_i) > 0$. If $\mathcal{G}$ is the σ-algebra generated by $\{A_1, \ldots, A_n\}$ then show that $\mathbb{E}\{X \mid \mathcal{G}\} = \sum_{i=1}^n p_i(X)\mathbb{1}_{G_i}$, where $p_i(X) = \mathbb{E}\{X\mathbb{1}_{A_i}\}/P(A_i)$. Finally, discuss the validity of the expression $\mathbb{E}\{X \mid \mathcal{G}\} = \sum_{i=1}^\infty (X, g_i)g_i$, where $(\cdot, \cdot)$ denotes the scalar product in $L^2(\Omega, \mathcal{F}, P)$, and now $\mathcal{G}$ is the σ-algebra generated by a sequence of random variables $\{g_i,\, i \geq 1\}$, which is assumed to be an orthonormal system. Perhaps, one should consider first the case when g_i assumes only a finite number of values or even $f_i = \mathbb{1}_{A_i}/\sqrt{P(A_i)}$ or when $\{g_i\}$ have disjoint supports. □

Exercise 2.7. Let X, Y be real random variables on a complete probability space. If Z is a random variable with values in some Polish space E then prove that the relation $X = \mathbb{E}\{Y \mid Z\}$ is characterized by the condition $\mathbb{E}\{Y\varphi(Z)\} = \mathbb{E}\{X\varphi(Z)\}$, for all $\varphi : E \longrightarrow \mathbb{R}$ which is bounded and continuous. Moreover, if E is locally compact, then the class of continuous function with compact support is sufficient to characterized the conditional expectation. Furthermore, any class of Borel functions that approximate any continuous and bounded function in the pointwise and bounded topology is sufficient. In particular simple functions, i.e., $\mathbb{E}\{Y\mathbb{1}_{a<Z\leq b}\} = \mathbb{E}\{X\mathbb{1}_{a<Z\leq b}\}$, for every $b > a$ in $\mathbb{R}$. □

Exercise 2.8. On a probability space $(\Omega, \mathcal{F}, P)$, let X be a real-valued random variable independent of a sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$, and f be a bounded Borel measurable function in $\mathbb{R}^2$. Define $f_1(y) = \mathbb{E}\{f(X, y)\}$. Prove that f_1 is Borel measurable and $f_1(Y) = \mathbb{E}\{f(X, Y) \mid \mathcal{G}\}$ almost surely. □

Exercise 2.9. Prove that two σ-algebras $\mathcal{G}$ and $\mathcal{H}$ are independent in a probability space $(\Omega, \mathcal{F}, P)$ if and only if the subspace $L^2(\mathcal{G})$ and $L^2(\mathcal{H})$ are orthogonal on the constant functions, i.e., X in $L^2(\mathcal{G})$, Y in $L^2(\mathcal{H})$, and $\mathbb{E}\{X\} = \mathbb{E}\{Y\} = 0$ imply $\mathbb{E}\{XY\} = 0$, this is a rewording of Exercise 2.3. Next, deduce that $\mathcal{G}$ and $\mathcal{H}$ are independent if and only if $\mathbb{E}\{XY\} = \mathbb{E}\{X\}\mathbb{E}\{Y\}$ for every X in $L^2(\mathcal{G})$ and Y in $L^2(\mathcal{H})$. □

Exercise 2.10. Show that a family of σ-algebras $\{\mathcal{G}_i : i \in I\}$ is independent (sometimes called mutually independent) if and only if for any finite subset J of indexes I, and for any random variables X_i in $L^\infty(\mathcal{G}_i)$ we have $\mathbb{E}\{\prod_{i\in J} X_i\} = \prod_{i\in J} \mathbb{E}\{X_i\}$, e.g., see Malliavin [92, Section IV.3, pp. 190–198]. □

(2.4) Regular Conditional Probability

Exercise 2.11. Let $\mathcal{G}$ be a finitely-generated σ-algebra, i.e., $\mathcal{G} = \sigma(F_1, \ldots, F_n)$. First, show that $\mathcal{G}$ can be expressed as $\sigma(G_1, \ldots, G_m)$, where the sets $G_1, \ldots, G_m$ are disjoint and minimal in the sense that any proper subset of $\{G_1, \ldots, G_m\}$

does not generate $\mathcal{G}$. Actually, $\{G_1, \dots, G_m\}$ is a partition and the set G_i are called *atoms* of $\mathcal{G}$, which has exactly 2^m elements. Second, gives an explicit expression of $P\{A \,|\, \mathcal{G})\}(\omega)$ in term of the family of sets $G_1, \dots, G_m$. Third, if X is a simple random variable (i.e., having a finite number of values, say $x_1, \dots, x_m$ with $P\{X = x_i\} > 0$ and $\sum_i P\{X = x_i\} = 1$) then show that $\sigma(X)$ (i.e., the minimal σ-algebra for which X is measurable) is finitely-generated, calculate $P\{A \mid X = x_i\}$, for $i = 1, \dots, m$ and consider the function $x \mapsto P(x, A)$ defined as $P(x, A) = P\{A \mid X = x_i\}$ if $x = x_i$ for some $i = 1, \dots, m$, and $P(x, A) = P(A)$ otherwise. Fourth, show that the expression $P(X, A)$ is a regular conditional probability of A given X, i.e., for any A measurable set we have $P\{A \mid X\} = P(X, A)$ almost surely. □

*** 2.11

Exercises - Chapter (3) Canonical Sample Spaces

(3.1) Continuous and cad-lag Functions

(3.2) Modulus of Continuity

(3.3) Skorokhod Topology

(3.4) Skorokhod Topology for BV functions

(3.5) Integer Measures

(3.6) Sequences of Probability

Exercises - Chapter (4) Probability Spaces

This part is not intended as real exercises, but as guide and a complement to the previous sections, helping to clarify and specify some statements given previously. The reader may take a look at the book Chaumont and Yor [24].

(4.1) Random Variables

Exercise 4.1. If $\mathcal{F}$ is a collection (or class) of subsets of Ω then $\mathcal{F}$ is called a π-system if for any A and B in $\mathcal{F}$ we have $A \cap B$ in $\mathcal{F}$ and a λ-system if Ω belongs to $\mathcal{F}$ and satisfies (a) for any A and B in $\mathcal{F}$ with $A \subset B$ we have $B \setminus A$ belongs to $\mathcal{F}$ and (b) for any monotone increasing sequence of sets $A_i \subset A_{i+1}$ we have $A = \bigcup_i A_i$ in $\mathcal{F}$. Prove the following versions of monotone class theorem: (1) If $\mathcal{G} \subset \mathcal{F}$ and $\mathcal{G}$ is a π-system and $\mathcal{F}$ is a λ-system then the σ-field or σ-algebra $\sigma(\mathcal{G})$ generated by $\mathcal{G}$ is contained in $\mathcal{F}$. (2) If $\mathcal{F}$ is a π-system and $\mathcal{H}$ a linear space of functions from Ω into $\mathbb{R}$ such that $\mathbb{1}_\Omega$, $\mathbb{1}_A$ and $\varphi = \lim_i \varphi$ belong to $\mathcal{H}$, for every A in $\mathcal{F}$ and for any sequence φ_i in $\mathcal{H}$ such that $0 \leq \varphi_i \leq \varphi_{i+1}$, $\varphi_i(\omega) \to \varphi(\omega)$ and $\varphi(\omega)$ is finite for any ω, then $\mathcal{H}$ contains all $\sigma(\mathcal{F})$-measurable functions. □

Exercise 4.2. Let $\vee_n \mathcal{F}_n$ be the σ-algebra generated by sequence $\{\mathcal{F}_n : n \geq 1\}$ of sub σ-algebra in a probability space $(\Omega, \mathcal{F}, P)$. Use an argument of monotone class to show that for any set A in $\vee_n \mathcal{F}_n$ there exists a sequence $\{A_k : k \geq 1\}$ of sets in $\bigcup_n \mathcal{F}_n$ such that $P(A \setminus A_k)$ and $P(A_k \setminus A)$ converge to 0, e.g., see Kallenberg [71, Lemma 3.16, p. 54]. □

Exercise 4.3. Let x be a function from a measurable space $(\Omega, \mathcal{F})$ into a Polish space E (complete metric space), so X is a random variable with values in E. Denote by $\mathcal{F}_x$ the σ-algebra generated by x. Use an argument of monotone class to show that any real valued random variable φ which is $\mathcal{F}_x$-measurable has the form $\varphi(\omega) = f(x(\omega))$, for some Borel function f from E into $\mathbb{R}$, e.g., see He et al. [59, Theorem 1.5, p. 5] or Kallenberg [71, Lemma 1.13, p. 7]. □

Exercise 4.4. Let $X(t, \omega)$ be a function from $T \times \Omega$ into $\mathbb{R}^d$, where T is a countable and $(\Omega, \mathcal{F})$ is a measurable space. Prove that the following statements

are equivalent:

(a) The function $(t, \omega) \mapsto X(t, \omega)$ from is Borel measurable with respect to the product σ-algebra $\mathcal{B}(T) \times \mathcal{F}$.

(b) The function $\omega \mapsto X(\cdot, \omega)$ from Ω into $\mathcal{L}^0(T, \mathbb{R}^d)$.

Discuss possible extensions to the case where T is a Borel subset of $\mathbb{R}$, e.g. see Doob [39, Theorem 2.1.13 in p. 408]. □

(4.2) Distributions and Independence

Exercise 4.5. Calculate the mean and the covariance of random variables with a Gaussian, Poisson and exponential distributions. Moreover, show that if x is a Gaussian variable with variance r then the even moments can be calculate by recurrence, i.e., $\mathbb{E}\{|x|^{2n+2}\} = r(2n-1)\mathbb{E}\{|x|^{2n}\}$, for any integer $n \geq 1$. □

Exercise 4.6. Prove that a linear combination of Gaussian random variables is also a Gaussian random variable. Calculate its mean and covariance, and check that all moments are finite. □

(4.3) Filtrations and Optional Times

Exercise 4.7. Prove properties (a) to (i) for optional or stopping times. □

Exercise 4.8. Let $(\Omega, \mathcal{F})$ be a measurable space. Recall that a π-systems $\mathcal{F}_0$ is a subset of $\mathcal{F}$ which is stable under finite intersections, i.e., if A and B belongs to $\mathcal{F}_0$ then $A \cap B$ also belongs to $\mathcal{F}_0$. Also, we denote by $\sigma(\mathcal{F}_0)$ the minimal sub σ-algebra of $\mathcal{F}$ containing all the elements of $\mathcal{F}_0$, i.e. the σ-algebra generated by $\mathcal{F}_0$. Prove that if $\mathcal{H}$ and $\mathcal{G}$ are two sub σ-algebras which are generated by the π-systems $\mathcal{H}_0$ and $\mathcal{G}_0$, then $\mathcal{H}$ and $\mathcal{G}$ are independent if and only if $\mathcal{H}_0$ and $\mathcal{G}_0$ are independent, i.e., if and only if $P(H \cap G) = P(H)P(G)$ for any H in $\mathcal{H}_0$ and G in $\mathcal{G}_0$ (e.g., see the book by Bauer [6, Section 5.1, pp. 149–154]). □

Exercise 4.9. Establish the existence for the conditional expectation on a given probability space $(\Omega, \mathcal{F}, P)$ for an integrable random variable x with respect to a sub σ-algebra $\mathcal{G}$ by two ways. Firstly (a) by means of the Radon-Nikodym theorem, i.e., on the measurable space $(\Omega, \mathcal{G})$ consider the probability measures $\nu(G) = \mathbb{E}\{x \mathbb{1}_G\}$ and $\mu(G) = \mathbb{E}\{\mathbb{1}_G\}$ satisfying $\nu \ll \mu$. Secondly (b) by means of the orthogonal projection π from the Lebesgue space $L^2(\Omega, \mathcal{F}, P)$ into the closed subspace $L^2(\Omega, \mathcal{G}, P)$, i.e., π satisfies $(x - \pi(x), y) = 0$, for any y in $L^2(\Omega, \mathcal{G}, P)$, where $(\cdot, \cdot)$ denotes the scalar product. □

Exercise 4.10. Let x, y be real random variables on a complete probability space. If z is a random variable with values in some Polish space E then prove that the relation $x = \mathbb{E}\{y \mid z\}$ is characterized by the condition $\mathbb{E}\{y\varphi(z)\} = \mathbb{E}\{x\varphi(z)\}$, for all $\varphi : E \longrightarrow \mathbb{R}$ which is bounded and continuous. Moreover, if E is locally compact, then the class of continuous function with compact support is sufficient to characterized the conditional expectation. □

Exercise 4.11. Prove properties (a) to (g) of the conditional expectation. □

Exercise 4.12. Prove properties (h) to (m) of the conditional expectation. □

(4.4) Versions of Processes

Exercise 4.13. Let $\mathcal{G}$ be the σ-algebra generated by a sequence $\{x_i : i = 1, 2, \dots\}$ of measurable functions from $(\Omega, \mathcal{F})$ into $(\mathbb{R}, \mathcal{B})$, and x be an integrable random variable in the complete probability space $(\Omega, \mathcal{F}, P)$ with values in $\mathbb{R}^n$. Use an argument of *monotone class* to show that $\mathbb{E}\{x \mid \mathcal{G}\} = 0$ if and only if $\mathbb{E}\{f(x_1, x_2, \dots, x_m)\, x\} = 0$ for any m and any bounded continuous function f from $\mathbb{R}^n$ into $\mathbb{R}$, e.g. see Yong and Zhou [137, Proposition 1.12 in p. 13]. □

Exercise 4.14. Prove that if x is in $L^1(\Omega, \mathcal{F}, P)$ then the family of elements in $L^1(\Omega, \mathcal{F}, P)$, defined by $\{y = \mathbb{E}\{x \mid \mathcal{G}\} : \mathcal{G}$, is a sub σ-algebra of $\mathcal{F}\}$ is uniformly integrable. Indeed use Jensen's inequality to establish that $kP\{|y| > k\} \le \mathbb{E}\{|x|\}$, for any k, and in view of

$$\int_{|y|>k} |y(\omega)| P(\mathrm{d}\omega) \le \int_{|y|>k} |x(\omega)| P(\mathrm{d}\omega),$$

the desired result follows. □

Exercise 4.15. On a probability space $(\Omega, \mathcal{F}, P)$, let x be a real random variable independent of a sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$, and f be a bounded Borel measurable function in $\mathbb{R}^2$. Define $f_1(\eta) = \mathbb{E}\{f(x, \eta)\}$. Prove that f_1 is Borel measurable and $f_1(y) = \mathbb{E}\{f(x, y) \mid \mathcal{G}\}$ almost surely. □

Exercise 4.16. Let $\mathcal{G}$ be a finitely-generated σ-algebra, i.e., $\mathcal{G} = \sigma[F_1, \dots, F_n]$. First, prove that also $\mathcal{G}$ can be expressed as $\sigma[G_1, \dots, G_m]$, where the sets $G_1, \dots, G_m$ are pairwise disjoint and minimal in the sense that any proper subset of $\{G_1, \dots, G_m\}$ does not generate $\mathcal{G}$. Actually, $\{G_1, \dots, G_m\}$ is a partition and the set G_i are called *atoms* of $\mathcal{G}$, which has exactly 2^m elements. Second, gives an explicit expression of $P\{A \mid \mathcal{G})\}(\omega)$ in term of the family of sets $G_1, \dots, G_m$. Third, if X is a simple random variable (i.e., having a finite number of values, say $x_1, \dots, x_m$ with $P\{X = x_i\} > 0$ and $\sum_i P\{X = x_i\} = 1$) then show that $\sigma(X)$ (i.e., the minimal σ-algebra for which X is measurable) is finitely-generated, calculate $P\{A \mid X = x_i\}$, for $i = 1, \dots, m$ and consider the function $x \mapsto P(x, A)$ defined as $P(x, A) = P\{A | X = x_i\}$ if $x = x_i$ for some $i = 1, \dots, m$, and $P(x, A) = P(A)$ otherwise. Fourth, show that the expression $P(X, A)$ is a regular conditional probability of A given X, i.e., for any A measurable set we have $P\{A \mid X\} = P(X, A)$ almost surely, see Remark 2.13. □

Exercise 4.17. Let $\{X(t) : t \ge 0\}$ be a (separable) stochastic process on the probability space $(\Omega, \mathcal{F}, P)$ with valued into $\mathbb{R}$. Prove that if X is either right or left continuous in probability then any dense set Q on $[0, \infty)$ is separant. □

Exercise 4.18. On a probability space $(\Omega, \mathcal{F}, P)$, let $\{X_n(t) : t \geq 0\}$, $n \geq 1$ be a sequence of families of random variables and $\mathcal{F}_n(t)$ be the σ-algebra generated by the random variables $\{X_n(s) : 0 \leq s \leq t\}$ and all sets of measure zero in $\mathcal{F}$. Assume that $X_n(t)$ converges in probability to $X(t)$, for every $t \geq 0$. Prove that

$$\lim_{n\to\infty} \mathbb{E}\{y \,|\, \mathcal{F}_n(t)\} = \mathbb{E}\{y \,|\, \mathcal{F}(t)\}, \quad \forall t \geq 0,$$

for every integrable random variable y such that the above limit exists in probability. □

(4.5) Continuous Markov Chains

Exercise 4.19. Let $\{\rho(t,i,j) : i,j = 1,\dots,n\}$ be a family of continuous functions from $[0,\infty)$ into $\mathbb{R}$ satisfying

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

Consider the n-dimensional system of ordinary differential equations

$$\dot{p}_s(t,i,j) = \sum_k p_s(t,i,k)\,\rho(t,k,j), \quad \forall t > s,\ i,j.$$

where the dot means derivative in t and p_s is the fundamental solution, i.e. it satisfies $p_s(s,i,j) = \delta_{i,j}$.

(a) First, assume that the data are constants i.e., $\rho(t,i,j) = \rho_(i,j)$ and denote by ρ the square matrix $(\rho_(i,j) : i,j = 1,\dots,n)$. Show that in this case, the fundamental solution $p_s(t,i,j) = \exp[-(t-s)\rho_d]\exp[(t-s)(\rho-\rho_d)]$, where ρ_d is the diagonal matrix with the coefficients $(\rho_{i,i} : i = 1,\dots,n)$. Conclude that all entries of $p_s(t,i,j)$ are non-negative and that each line adds (sum in j) to one.

(b) Extend the previous conclusion to the general case where the data ρ may depend on t.

(c) Assume the data are constant in t as in (a) and define $\lambda = -\inf_i \rho_i, i > 0$. Let Y_n be a Markov chain with transition probability $P(Y_n = j \,|\, Y_{n-1} = i) = \rho_{i,j}/\lambda$ if $i \neq j$ for $n = 1,2,\dots,$ and 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 $(Y_0, Y_1, \dots)$. Prove that $X_t = Y_n$ for t in the stochastic interval $[\![T_n, T_{n+1}[\![$, where $T_0 = 0$ and $T_n = \tau_1+\tau_2+\cdots+\tau_n$, gives a realization of the pure jumps Markov with the above infinitesimal rates (see Durrett [43, p. 250, Example 2.1]).

(d) Discuss the case of the double sequence $\{\rho(t,i,j) : i,j = 1,2,\dots\}$ of continuous functions. □

(4.6) Markov Processes

Exercise 4.20. Let $\mathcal{H}$ be a σ-algebra of a probability space $(\Omega, \mathcal{F}, P)$. Define $\mathcal{I}$ the collection of all sets in $\mathcal{F}$ independent of $\mathcal{H}$. Prove that $\mathcal{I}$ is closed under monotone union and intersection. Deduce that $\mathcal{I}$ is the σ-algebra $\mathcal{H}^{\perp}$ generated by all sets in $\mathcal{F}$ independent of $\mathcal{H}$. □

Exercise 4.21. Let $X = (X_t : t \geq 0)$ be a family of random variables with values in a complete separable metric (Polish) space E defined on a probability space $(\Omega, \mathcal{F}, P)$. Assume that X satisfies the Markov property 4.17 and suppose that X (regarded as a process) is right-continuous in probability, i.e., for every $\varepsilon > 0$ and $t \geq 0$ there exists $\delta > 0$ such that $P\{|X_s - X_t| \geq \varepsilon\} < \varepsilon$ for every s in $(t, t+\delta)$. Prove that X satisfies the Markov property as in Definition 4.17 with the natural filtration $(\mathcal{F}_t : t \geq 0)$, i.e., the minimal increasing family of σ-algebra satisfying the usual conditions such that X_t is $\mathcal{F}_t$-measurable for every $t \geq 0$. □

Exercise 4.22. Let X be a $\mathbb{R}^d$-valued adapted stochastic process in $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and suppose that for some transition function $p(s, x, t, \mathrm{d}y)$ we have

$$P\{h(X_t) \in B \mid \mathcal{F}_s\} = \int_{\mathbb{R}^d} h(y) p(s, X_s, t, \mathrm{d}y), \quad \forall t > s \geq 0,$$

almost surely, for every $t > s \geq 0$, and any real-valued continuous and bounded function h. Verify that X is indeed a Markov process. Which other classes of functions h could be used? How about processes taking values in some topological space E instead of $\mathbb{R}^d$? □

Exercise 4.23. Let $X = (X_t : t \geq 0)$ be a (strong) Markov process with values in a complete separable metric (Polish) space E defined on a probability space $(\Omega, \mathcal{F}, P)$, and with transition probability function $p(s, x, t, B)$, $t > s \geq 0$, x in E and B in $\mathcal{B}(E)$. If necessary, assume that for every B the mapping $(s, x, t) \mapsto p(s, x, t, B)$ is measurable. Define

$$\dot{p}((s, x), t, \mathrm{d}r \times \mathrm{d}e) := \delta_{s+t}(\mathrm{d}r)\, p(s, x, s+t, \mathrm{d}e), \quad \forall t > 0,$$

for all (s, x) in $\dot{E} := [0, \infty) \times E$, and where $\delta_s(\mathrm{d}r)$ is the Dirac unit mass in $[0, \infty)$ concentrated at $r = s$. Prove that $\dot{X} := (t, X_t)$ is a (strong) homogeneous Markov process with values in $\dot{E}$ and with transition probability function $\dot{p}((s, x), t, B)$, $t > 0$, (s, x) in $\dot{E}$ and B in $\mathcal{B}(\dot{E})$. □

(4.7) Construction of Processes

Exercise 4.24. Prove that if X is a d-dimensional stochastic process with independent and stationary increments (see properties (a) and (b) of Lévy processes) then

$$\mathbb{E}\{\varphi(X(s) - X(t))\psi(X(t))\} = \mathbb{E}\{\varphi(X(s) - X(t))\}\, \mathbb{E}\{\psi(X(t))\},$$
$$\mathbb{E}\{\varphi(X(s) - X(t))\} = \mathbb{E}\{\varphi(X(s-t))\},$$

for any continuous function from $\mathbb{R}^d$ into $\mathbb{R}$ with compact support. □

Exercise 4.25. Prove that any characteristic function possesses the properties (a) and (b) of continuity and positive define stated in the text. □

Exercise 4.26. Verify that the characteristic functions of a Lévy process have exponential form. Discuss to what this correspond on the finite-dimensional distributions viewpoint. □

Exercise 4.27. Verify that the consistency properties are satisfied for a family of finite-dimensional distributions constructed from a transition function $P(s, x, t, A)$, an initial time t_0 and probability P_0. □

Exercise 4.28. Prove that for a Lévy process (P_X, X) the continuity condition (4.4) reduces to

$$\mathbb{E}_X\{|X_h|^\alpha\} \leq Ch^{1+\beta} \qquad \forall h > 0,$$

for a some positive constants α, β and C. Similarly the cad-lag condition (4.5) can be expressed by

$$\mathbb{E}_X\{|X_h|^\alpha\} \leq Ch^{\frac{1}{2}+\beta} \qquad \forall h > 0,$$

for a some positive constants α, β and C. Extend this result to processes generated by a transition function i.e. to Markov processes. □

Exercise 4.29. By means of the finite-dimensional distributions proves that the Wiener process satisfies the continuity condition (4.4) so that its paths are continuous. □

Exercise 4.30. Show that the continuity condition (4.4) is not satisfied for the Poisson process but a direct calculation proves that it is continuous in probability, see property (c) of Definition 4.21. □

Exercise 4.31. Proved that the cad-lag condition (4.5) is satisfied for the Cauchy process i.e. there exist positive constants α, β and C such that

$$\int_E |x-y|^\alpha P(s, x, t, \mathrm{d}y) \leq C|t-s|^{\frac{1}{2}+\beta}, \quad \forall s, t \in [0, T],\ \forall x \in \mathbb{R}$$

for ant $T > 0$. □

(4.8) Examples of Markov processes

Exercise 4.32. Let $\mathcal{B}^T(\mathbb{R})$ be the product σ-algebra (i.e., generated by the cylinder sets), which may be smaller that $\mathcal{B}(\mathbb{R}^T)$ (the minimal σ-algebra containing all open sets in $\mathbb{R}^T$, endowed with the product topology). Prove that a typical set in $\mathcal{B}^T(\mathbb{R})$ has the form

$$\{\omega \in \mathbb{R}^T : (\omega(t_1), \omega(t_2), \dots) \in A\}$$

where $A \in \mathcal{B}(\mathbb{R}^{\{1,2,\dots\}})$ and $(t_1, t_2, \dots)$ is a sequence in $\mathbb{R}$. Verify that a singleton (i.e., a set of only one element) belongs to $\mathcal{B}(\mathbb{R}^T)$, but does not belong to $\mathcal{B}^T(\mathbb{R})$ if the index set T is uncountable. □

Exercise 4.33. Let $\mathcal{F}$ be the σ-algebra generated by the coordinate random variables $\omega \mapsto \omega(t)$ from $C([0,\infty), E)$ into E, where E is a complete separable metric space and t ranges over a dense set Q of $[0,\infty)$. Prove that $\mathcal{F} = \mathcal{B}$, where $\mathcal{B}$ is the (Borel) σ-algebra generated by open all sets in $C([0,\infty), E)$. Conclude 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 $[0,\infty)$. Prove the same result for the space $(D([0,\infty), E), \mathcal{B})$, where now $\mathcal{B}$ is the (Borel) σ-algebra generated by open all sets in $D([0,\infty), E)$. □

Exercise 4.34. For a function x from $[0,T]$ into $\mathbb{R}$ which have only discontinuities of first class we define two modulii of continuity $w(x,h)$ and $w'(x,h)$ one by

$$w(x,h,T) := \inf_{\{t_i\}} \sup_i \sup\{|x(t) - x(s)| : t_{i-1} \le s < t < t_i\}$$

where $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \cdots < t_n = T$, with $t_i - t_{i-1} \ge h$ and $n \ge 1$, and the other by

$$w'(x,h) := \sup\{\sqrt{|x(t+h) - x(s)|\,|x(s) - x(t)|} : \\ : -h < t < s < t + h < T + h\},$$

where we have extended $x(t) = x(0)$ for $t < 0$ and $x(t) = x(T)$ for $t > T$. Prove that for any x in $D([0,T], E)$ we have $w(x,h) \longrightarrow 0$ and $w'(x,h) \longrightarrow 0$ as $h \to 0$. By means of the above modulii of continuity, give a characterization of pre-compact sets in the the space $D([0,T], E)$ endowed with the Skorokhod topology. □

Exercise 4.35. Show that the space $D([0,+\infty], E)$ complete with the locally uniform convergence [i.e., the topology in $C([0,+\infty], E)$], but is not separable. On the other hand, show that $C([0,+\infty], E)$ is a closed subspace of $D([0,+\infty], E)$. □

Exercise 4.36. Let X be a Borel measurable function form Ω into itself, where (Ω, d_Ω) be a separable and complete metric space. Suppose that $\{P_n : n = 1, 2, \dots\}$ is a sequence of probability measures on Ω which converges weakly to P. Define $\{Q_n : n = 1, 2, \dots\}$ and Q as the image measures through the mapping X of $\{P_n : n = 1, 2, \dots\}$ and P. Prove that if X is P-almost surely continuous then the sequence $\{Q_n : n = 1, 2, \dots\}$ converges weakly to the measure Q. □

Exercise 4.37. Let Ω be a complete metric space. Use a monotone class argument to show that the smallest class of functions M in $B(\Omega)$ satisfying:

(a) if $\{f_n\}$ is a sequence in M boundedly and pointwise convergent to f then f

belongs to $B(\Omega)$,

(b) if A is open in Ω then the characteristic (or indicator) function $\mathbb{1}_A$ belongs to M,

(c) if f and g are in M then $\alpha f + \beta g$ is in M for any constant α and β,

is actually $B(\Omega)$. □

========== NEED WORK ==========

Exercises - Chapter (5) Stochastic Processes

This part is not intended as real exercises, but as guide and a complement to the previous sections, helping to clarify and specify some statements given previously. The reader may take a look at the book Chaumont and Yor [24].

(5.1) Discrete Time

(5.2) Filtered Spaces

(5.3) Bounded Variation

(5.4) Martingales

(5.5) Semi-Martingales

(5.6) Strong Markov Processes

(5.7) Extended Generators

(5.8) Poisson Processes and Queues

(5.9) Piecewise Deterministic Processes

(5.10) Lévy Processes

(5.11) Transition Functions

(5.12) Hunt and Standard Processes

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 $\mathbb{F} = \{\mathcal{F}(t) : t \in T\}$, 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.

Polish space: is a separable and complete metric space (commonly used in Probability) and because the metric is rarely used one says 'metrizable' instead of metric, and clearly, an open subset of a Polish space is itself a Polish space. Also recall that a **Borel space** is a Borel subset of a complete and separable metric space, i.e., a Borel subset of a Polish space. Other common terminology is **LCCB** (locally compact with countable basis, also called locally-compact second-countable Hausdorff space), and any LCCB is a Polish space, but not the converse, since a Polish space could be not locally compact. e.g., $L^1(\mathbb{R})$.

$C([0,\infty),\mathbb{R}^d)$ 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. Note that occasional, the Euclidean space $\mathbb{R}^d$ could be replaced by a Borel space E, with the notation $C([0,\infty);E)$ or $D([0,\infty);E)$, and in general, these spaces are considered subspaces of the space of functions from $[0,\infty)$ into E, usually denoted by $E^{[0,\infty)}$.

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 $\mathbb{R}^d$, $1 \le p \le \infty$ 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. 207

[2] L. Arnold. *Stochastic Differential Equations*. Wiley, New York, 1974. 132

[3] R.B. Ash. *Probability and Measure Theory*. Harcourt/Academic Press, Burlington, MA, 2000. 18

[4] K.B. Athreya and S.N. Lahiri. *Measure theory and probability theory*. Springer, New York, 2006. 1

[5] R.F. Bass. *Stochastic Processes*. Cambridge University Press, Cambridge, 2011. 45, 72

[6] H. Bauer. *Probability Theory and Elements of Measure Theory*. Holt, Rinehart and Winston, New York, 1972. 26, 242, 248

[7] R. Bellman. *Dynamic programming*. Princeton University Press, Princeton, NJ, 2010. Reprint of the 1957 edition, With a new introduction by Stuart Dreyfus. xi

[8] A. Bensoussan. *Dynamic programming and inventory control*. IOS Press, Amsterdam, 2011. xi

[9] A. Bensoussan and J.L. Lions. *Impulse Control and Quasi-Variational Inequalities*. Gauthier-Villars, Paris, 1984. (pages are referred to the French edition). 129

[10] A. Benveniste and J. Jacod. Systems de Lévy des processus de Markov. *Inventiones Math.*, 21:183–198, 1973. 233

[11] J. Bertoin. *Lévy processes*. Cambridge University Press, Cambridge, 1996. 120, 121, 129, 206, 214, 216

[12] D.P. Bertsekas. *Dynamic programming and Stochastic Control*. Prentice Hall Inc., Englewood Cliffs, NJ, 1987. Deterministic and stochastic models. x

[13] D.P. Bertsekas and S.E. Shreve. *Stochastic optimal control*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1978. The discrete time case. xi

[14] K. Bichteler. *Stochastic Integration with jumps*. Cambridge University Press, Cambridge, 2002. 84, 108, 145, 147, 156

[15] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968. 15, 42, 45, 68, 73

[16] P. Billingsley. *Probability and Measure*. Wiley, New York, 2nd edition, 1986. 25, 102

[17] R.M. Blumenthal and R.K. Getoor. *Markov Processes and Potential Theory*. Academic Press, New York, 1968. 113, 117, 118, 123, 179, 214, 226, 232

[18] A. Bobrowski. *Functional analysis for probability and stochastic processes*. Cambridge University Press, Cambridge, 2005. 19

[19] K. Borovkov. *Elements of Stochastic Modelling*. World Scientific, Singapore, 2nd edition, 2014. 194

[20] L. Breiman. *Probability*. Society for Industrial and Applied Mathematics, Philadelphia, 1992. 96

[21] P. Bremaud. *Point Processes and Queues*. Springer-Verlag, New York, 1981. xi, 129, 134, 143, 144, 193, 194, 216, 234

[22] P. Bremaud. *Markov Chains*. Springer-Verlag, New York, 1998. xi, 110, 194

[23] M. Capinski and E. Kopp. *Measure, Integral and Probability*. Springer-Verlag London Ltd., London, 2nd edition, 2004. 1

[24] L. Chaumont and M. Yor. *Exercises in probability*. Cambridge University Press, Cambridge, 2nd edition, 2012. A guided tour from measure theory to random processes, via conditioning. 247, 255

[25] K.L. Chung. *Elementary Probability Theory with Stochasitic Processes*. Springer-Verlag, New York, 1975. 11

[26] K.L. Chung. *Markov Chains with Stationary Transition Probabilities*. Springer-Verlag, New York, 2nd edition, 1980. 110, 111, 131, 143

[27] K.L. Chung. *From Markov Processes to Brownian Motion*. Springer-Verlag, New York, 1982. 118, 126, 127, 234

[28] K.L. Chung and R.J. Williams. *Introduction to Stochastic Integration*. Birkhäuser, Boston, 2nd edition, 1990. 150

[29] E. Çınlar. *Probability and stochastics*. Springer, New York, 2011. 118

[30] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimension*. Cambridge University Press, Cambridge, 1992. 18, 80, 103, 204

[31] G. Da Prato and J. Zabczyk. *Ergodicity for Infinite Dimensional Systems*. Cambridge University Press, Cambridge, 1996. 73, 208

[32] D. J. Daley and D. Vere-Jones. *An introduction to the Theory of Point Processes. Vol. 2*. Springer-Verlag, New York, 2nd edition, 2008. General theory and structure. 194

[33] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Vol. 1*. Springer-Verlag, New York, 2nd edition, 2003. Elementary theory and methods. 194

[34] M.H.A. Davis. *Markov Models and Optimization*. Chapman and Hall, London, 1993. 144, 179, 182, 183, 184, 194, 200, 201, 203, 236

[35] B. De Finetti. *Theory of Probability*, volume 1 and 2. Wiley, London, 1975. 85

[36] 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. 85, 118, 143, 154, 161, 167, 172, 175, 177, 222, 226, 227, 233

[37] J. Dieudonné. *Foundation of Modern Analysis*. Academic Press, New York, 1969. 43

[38] J.L. Doob. *Stochastic Processes*. Wiley, New York, 1953. 74, 94, 102, 109, 125, 143

[39] J.L. Doob. *Classical Potential Theory and its Probabilistic Counterpart*. Springer-Verlag, New York, 1984. 31, 88, 100, 117, 119, 138, 145, 179, 248

[40] J.L. Doob. *Measure Theory*. Springer-Verlag, New York, 1984. 1, 143, 150, 157

[41] R.M. Dudley. *Real Analysis and Probability*. Cambridge University Press, Cambridge, 2002. 15, 94, 146

[42] R. Durrett. *Probability, Theory and Examples*. Duxbory Press, Belmont, 2nd edition, 1996. 22, 25, 241

[43] R. Durrett. *Stochastic Calculus. A practical introduction*. CRC Press, Boca Raton, 1996. 105, 169, 250

[44] E.B. Dynkin. *Theory of Markov Processes*. Prentice-Hall, Englewood Cliffs (NJ), 1961. 118

[45] E.B. Dynkin. *Markov Processes*, volume 1 and 2. Springer-Verlag, New York, 1965. 118, 123, 177, 179

[46] R.J. Elliott. *Stochastic Calculus and Applications*. Springer-Verlag, New York, 1982. 154, 216

[47] S.N. Ethier and T.G. Kurtz. *Markov Processes*. Wiley, New York, 1986. 45, 54, 68, 72, 73, 118, 123, 183, 226, 234

[48] W. Feller. *An Introduction to Probability Theory and its Applications*, volume Vols I and II. Wiley, New York, 1960 and 1966. 94, 111, 234

[49] G.B. Folland. *Real Analysis*. Wiley, New York, 2nd edition, 1999. 94

[50] A. Friedman. *Stochastic Differential Equations and Applications*, volume I and II. Academic Press, New York, 1975 and 1976. 208

[51] M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet Forms and Symmetric Markov Processes*. De Gruyter, Berlin, 1994. 232, 236

[52] 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. 85

[53] I. Gikhman and A. Skorokhod. *Introduction to the Theory of Random Process*. Dover Publications, New York, 1996. 45, 68, 102, 103, 104, 107

[54] B.V. Gnedenko. *The Theory of Probability*. Chelsea, New York, 1989. 75, 94

[55] R.A. Gordon. *The Integrals of Lebesgue, Denjoy, Perron, and Henstock*. Am. Math. Soc., Providence, Rhode Island, 1994. 148

[56] A. Gut. *Probability: A Graduate Course*. Springer, New York, 2005. 19

[57] P.R. Halmos. *Measure Theory*. Springer-Verlag, New York, 1974. 94, 135

[58] B. Harlamov. *Continuous semi-Markov processes*. Applied Stochastic Methods Series. ISTE, London, 2008. 85

[59] S.W. He, J.G. Wang, and J.A. Yan. *Semimartingale theory and stochastic calculus*. Kexue Chubanshe (Science Press), Beijing, 1992. 34, 177, 247

[60] O. Hernández-Lerma and J.B. Lasserre. *Discrete-time Markov control processes*. Springer-Verlag, New York, 1996. Basic optimality criteria. xi

[61] H. Holden, B. Oksendal, J. Uboe, and T. Zhang. *Stochastic Partial Differential Equations*. Birkhäuser, Boston, 1996. 76

[62] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North Holland, Amsterdam, 2nd edition, 1989. 68, 169

[63] M. Iosifescu, N. Limnios, and G. Oprisan. *Introduction to Stochastic Models*. Wiley-ISTE, 2010. 85

[64] K. Itô. *Introduction to Probability Theory*. Cambridge University Press, Cambridge, 1978. 96

[65] K. Itô. *Stochastic Processes*. Springer-Verlag, New York, 2004. 206

[66] K. Itô and H.P. McKean. *Diffusion Processes and their Sample Paths*. Springer-Verlag, New York, 1965. 126

[67] 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. 75, 84

[68] J. Jacod and P. Protter. *Probability Essentials*. Springer-Verlag, New York, 2nd edition, 2004. 14, 101

[69] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York, 1987. 42, 45, 54, 68, 74, 129, 143, 153, 154, 167, 174, 228

[70] A. Jakubowski. A non-Skorohod topology on the Skorohod space. *Electronic J. Probab.*, 2:1–21, 1997. 54

[71] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, New York, 2nd edition, 2002. 27, 74, 97, 107, 119, 171, 172, 173, 247

[72] O. Kallenberg. *Random Measures, Theory and Applications*. Springer, New York, 2017. 62

[73] G. Kallianpur and R.L. Karandikar. *White Noise Theory of Prediction, Filtering and Smothing*. Gordon and Breach, London, 1988. 132

[74] 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. 83, 159

[75] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 2nd edition, 1991. 99, 103, 119, 167

[76] S. Karlin and H.M. Taylor. *A First Course in Stochastic Processes*. Academic Press, New York, 1975. 143

[77] S. Karlin and H.M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, New York, 2nd edition, 1981. 143

[78] A.F. Karr. *Probability*. Springer-Verlag, New York, 1993. 139

[79] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*. Springer-Verlag, New York, 1976. 110

[80] R.Z. Khasminskii. *Stochastic Stability of Differential Equations*. Sijthoff and Noordhoff, The Netherlands, 1980. 108

[81] L. Kleinrock. *Queueing Systems*, volume 1 and 2. Wiley, New York, 1975. 110

[82] F.B. Knight. *Essentials of Brownian Motion and Diffusions*. American Mathematical Society, Providence, R.I., 1981. 127

[83] N.V. Krylov. *Introduction to the Theory of Diffusion Processes*. American Mathematics Society, Providence, Rhode Island, 1995. 85, 126, 208

[84] N.V. Krylov. *Introduction to the Theory of Random Processes*. American Mathematics Society, Providence, Rhode Island, 1995. 205

[85] G.S. Ladde and V. Lakshmikantham. *Random Differential Inequalities*. Academic Press, New York, 1980. 198

[86] J. Lamperti. *Stochastic Processes. A survey of the mathematical theory*. Springer-Verlag, New York, 1977. 131

[87] G. Letta. *Martingales et intégration stochastique*. Quaderni della Scuola Normale Superiore, Pisa, 1984. 216

[88] R.S. Liptser and A.N. Shiryayev. *Theory of Martingales*. Kluwer Academic Publ., Dordrecht, 1986. 42, 72, 172, 174, 175, 177

[89] M. Loève. *Probability Theory*, volume 1 and 2. Springer-Verlag, New York, 4rd edition, 1977 and 1978. 85

[90] E. Lukacs. *Characteristic Functions*. Griffin Publ., London, 2nd edition, 1970. 24

[91] Z. Ma and M. Röckner. *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*. Springer-Verlag, Berlin, 1992. 232

[92] P. Malliavin. *Integration and Probability*. Springer-Verlag, New York, 1995. 34, 74, 243

[93] P. Malliavin. *Stochastic Analysis*. Springer-Verlag, New York, 1997. 119

[94] M.B. Marcus and J. Rosen. *Markov processes, Gaussian processes, and local times*. Cambridge University Press, Cambridge, 2006. 118, 124, 132, 207

[95] J.L. Menaldi. *Measure and Integration: Theory and Exercises.* Wayne State University / Internet, link: http://digitalcommons.wayne.edu/mathfrp/22, 2015. vii

[96] J.L. Menaldi. *Distributions and Function Spaces: Schwartz Theory of Distributions, Sobolev and Besov Spaces.* Wayne State University / Internet, link: http://digitalcommons.wayne.edu/mathfrp/23, 2016. vii

[97] J.L. Menaldi. *Stochastic Processes and Integrals.* Wayne State University / Internet, link: http://digitalcommons.wayne.edu/mathfrp/61, 2017. vii

[98] M. Métivier. *Semimartingales: A course on Stochastic Processes.* De Gruyter, Berlin, 1982. 159, 168

[99] P.A. Meyer. *Probability and potentials.* Blaisdell Publ. Co., Waltham, Mass., 1966. 86, 163

[100] H. Morimoto. *Stochastic Control and Mathematical Modeling.* Cambridge Univ Press, New York, 2010. 39, 116, 179

[101] R. Nelson. *Probability, Stochastic Processes and Queueing Theory.* Springer-Verlag, New York, 1995. 110, 143

[102] J. Neveu. *Bases Mathématiques du Calcul des Probabilités.* Masson, Paris, 1970. 102, 103, 136, 138

[103] J. Neveu. *Discrete-Parameters Martingales.* North Holland, Amsterdam, 1975. 143, 167

[104] R. Pallu de la Barrière. *Optimal Control Theory.* Dover, New York, 1967. 121, 132

[105] D. Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, New York, 1984. 42, 44, 68

[106] D. Pollard. *A User's Guide to Measure Theoretic Probability.* Cambridge University Press, Cambridge, 2002. 1, 13, 19, 84

[107] N. Prabhu. *Topics in Probability.* World Scientific Publishing, Singapur, 2011. 93

[108] P. Protter. *Stochastic Integration and Differential Equations.* Springer-Verlag, New York, 1990. 121, 128, 129, 144, 154, 168, 173, 175, 216

[109] M.M. Rao. *Stochastic Processes: General Theory.* Kluwer Academic, Dordrecht, 1995. 85

[110] D. Revuz. *Markov chains.* North-Holland Publishing Co., Amsterdam, 2nd edition, 1984. xi

[111] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer-Verlag, New York, 3rd edition, 1999. 101, 108, 119, 122, 133, 167, 173, 194

[112] 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. 118, 119, 134, 145, 146, 154, 161, 164, 171, 175, 176, 177, 179, 232, 233

[113] J.S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second edition, 2006. 1

[114] S. Ross. *Simulation*. Academic Press, San Diego, 1997. 98

[115] W. J. Runggaldier and L. Stettner. *Approximations of Discrete Time Partially Observed Control Problems*. Applied Mathematics Monographs, Comitato Nazionale per le Scienze Matematiche, Consiglio Nationale delle Ricerche (CNR). Giardini Editori e Stampatori in Pisa, 1994. x

[116] K.I. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Univ. Press, Cambridge, 1999. 91, 120, 128, 129, 148, 206, 211, 215, 216, 222, 226

[117] R.L. Schilling. *Measures, Integrals and Martingales*. Cambridge University Press, New York, 2005. 31, 35

[118] L. Schwartz. *Théorie des distributions*. Hermann, Paris, 1978. 131

[119] M. Sharpe. *General Theory of Markov Processes*. Academic Press, New York, 1988. 85, 233, 234

[120] P.S. Shields. *The Ergodic Theory of Discretes Sample Paths*. Am. Math. Soc., Providence, Rhode Island, 1996. 111

[121] A.N. Shiryayev. *Probability*. Springer-Verlag, New York, 1984. 15, 86, 120, 124, 136, 150

[122] A.V. Skorokhod. *Studies in the Theory of Random Processes*. Dover, New York, 1982. 18

[123] A.V. Skorokhod. *Asymptotic Methods in the Theory of Stochastic Differential Equations*. English transl., Amer. Math. Soc., Providence, Rhode Island, 1989. 108

[124] T. Soderstrom. *Discrete-time Stochastic Systems*. Springer, 2002. Estimation and Control. xi

[125] K.R. Stromberg. *Probability for Analysts*. Chapman and Hall, New York, 1994. 85, 94

[126] D.W. Stroock. *Probability Theory: An Analytic View*. Cambridge University Press, Cambridge, 1999. revised edition. 85, 94, 204

[127] D.W. Stroock and S.R. Varadhan. Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.*, 24:147–225, 1971. 236

[128] D.W. Stroock and S.R. Varadhan. On degenerate elliptic-parabolic operators of second order and their associagted diffusions. *Comm. Pure Appl. Math.*, 25:651–714, 1972. 236

[129] D.W. Stroock and S.R. Varadhan. *Multidimensional Diffusion Process*. Springer-Verlag, Berlin, 1979. 73, 84, 219, 237

[130] K. Taira. *Diffusion Processes and Partial Differential Equations*. Academic Press, Boston, 1988. 225

[131] K. Taira. *Semigroups, Boundary Value Problems and Markov Processes*. Springer-Verlag, Berlin, second edition, 2014. 40, 118

[132] J.C. Taylor. *An Introduction to Measure and Probability*. Springer-Verlag, New York, 1996. 1, 35, 85

[133] D. Williams. *Probabiliy with Martingales*. Cambridge Univ. Press, Cambridge, 1991. 101, 143

[134] E. Wong. *Stochastic Processes in Information and Dynamical Systems*. McGraw-Hill, New York, 1971. 104, 124

[135] G.G. Yin and Q. Zhang. *Discrete-Time Markov Chains*. Springer-Verlag, New York, 2005. Two-Time-Scale Methods and Applications, Stochastic Modelling and Applied Probability. xi

[136] G.G. Yin and Q. Zhang. *Continuous-time Markov chains and applications*. Springer-Verlag, New York, second edition, 2013. Two-time-scale Approach. 110

[137] J. Yong and X.Y. Zhou. *Stochastic Controls*. Springer-Verlag, New York, 1999. 249

Index