Large Deviations Techniques for Error Exponents to Multiple Hypothesis LAO Testing

Leader Navaei
Yerevan State University

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm
Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
DOI: 10.22237/jmasm/1193890320
Available at: http://digitalcommons.wayne.edu/jmasm/vol6/iss2/13

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.
Large Deviations Techniques for Error Exponents to Multiple Hypothesis LAO Testing

Cover Page Footnote
I am grateful for Professor E. A. Haroutunian for his very helpful comments which substantially improved the presentation of the paper.
Large Deviations Techniques for Error Exponents to Multiple Hypotheses LAO Testing

Leader Navaei
Yerevan State University

In this article the problem of multiple hypotheses testing using a theory of large deviations is studied. The reliability matrix of Logarithmically Asymptotically Optimal (LAO) tests is introduced and described, and the conditions for the positive of all its elements are indicated.

Key words: hypotheses testing, empirical distributions, the method of types, reliability matrix, Sanov's theorem.

Introduction

Many studies have been devoted to the study of exponential decrease, as the sample size $N$ goes to infinity, of the error probabilities $\alpha_1^N = \alpha_1$. For example Stain’s lemma determines the exponential rate of convergence to zero of the error probability of the second kind $\alpha_2^N$ as $N$ goes to infinity. Perez (1984) considered independent identically distributed observations and different asymptotical aspects of two hypotheses, as the interdependence of exponents. Csiszar and Shields (2004) considered independent identically distributed observations different asymptotical aspects of the two hypotheses testing via the theory of large deviations. This article is based on Haroutunian (1990), and provides a proof based on Sanov’s theorem.

Preliminaries

Let $\chi = \{1, 2, \ldots, K\}$ be the finite set of size $K$. The set of all probability distributions by $(PD's)$ on $\chi$ is denoted by $P(\chi)$. For $PD's$, $P$ and $Q$, $H(P)$ denotes entropy and $D(P \parallel Q)$ denotes information divergence (or the Kullback-Leibler distance).

$$H(P) = -\sum_{x \in \chi} P(x) \log P(x),$$

$$D(P \parallel Q) = \sum_{x \in \chi} P(x) \log \frac{P(x)}{Q(x)}.$$

In this article, exps and logs are used at base 2. Also considered are the standard conventions that $0 \log 0 = 0$, $0 \log \frac{0}{0} = 0$, $P \log \frac{P}{0} = \infty$ if $P > 0$.

The type of a vector $X = (x_1, x_2, \ldots, x_N) \in \chi^N$ is the empirical distribution given by $Q(x) = \frac{N(x | X)}{N}$ for all $x \in \chi$, where $N(x | X)$ denotes the number of occurrences of $x$ in $X$ or

$$Q(x) = \left( \frac{N_1}{N}, \frac{N_2}{N}, \ldots, \frac{N_N}{N} \right) \in \chi^N$$

that $N_i \equiv$ number of times out of $N$ trials that the
random variables \( x_1, x_2, \ldots, x_N \) occurrences in \( \mathcal{X} \).

The subset of \( P(\mathcal{X}) \) consisting of the possible types of sequences \( X \in \mathcal{X}^N \) is denoted by \( P_N(\mathcal{X}) \). For \( Q \in P_N(\mathcal{X}) \), the set of sequences of type class \( Q \) will be denoted by \( T_Q^N \).

The probability that \( N \) independent drawings from a \( PD, P \in P(\mathcal{X}) \) give \( X \in \mathcal{X}^N \), is denoted by \( P_N(X) \). If \( X \in T_Q^N \) then:

\[
P_N(X) = \exp\{-N(H(Q) + D(Q \parallel P))\}.
\]

Lemma

The number of types for sequences of length \( N \) grows at most polynomially with \( N \):

\[
|P_N(\mathcal{X})| \leq (N+1)^{|\mathcal{X}|},
\]

For any type \( Q \in P_N(\mathcal{X}) \):

\[
(N+1)^{-|\mathcal{X}|} \exp\{NH(Q)\} \leq |T_Q^N| \leq \exp\{NH(Q)\},
\]

For any \( PD, P \in P(\mathcal{X}) \):

\[
\frac{P_N(X)}{Q_N(X)} = \exp\{-ND(Q \parallel P)\}, \text{ if } X \in T_Q^N,
\]

and

\[
(N+1)^{-|\mathcal{X}|} \exp\{-ND(Q \parallel P)\} \leq |T_Q^N| \leq \exp\{NH(Q)\},
\]

Theorem 1 (Sanov's theorem (Csiszar & Shields, 2004, Dembo & Zeitoni, 1993))

Let \( A \) be a set of distributions from \( P(\mathcal{X}) \) such that its closure is equal to the closure of its interior, then for the empirical distribution \( Q_X \) of a vector \( X \) from a strictly positive distribution \( P \) on \( \mathcal{X} \):

\[
\lim_{N \to \infty} \left( \frac{1}{N} \log P_N(X : Q_X \in A) \right) = \inf_{Q_X \in A} \left( D(Q_X \parallel P) \right).
\]

Problem Statement and Formulation of Results

The problem of multiple hypotheses testing is the following. Let \( \mathcal{X} = \{1, 2, \ldots, K\} \) be the finite set such that \( M \) incompatible hypotheses \( H_1, H_2, \ldots, H_M \) consist in that the random variable \( X \) taking values on \( \mathcal{X} \) has one of \( M \) distributions \( P_1, P_2, \ldots, P_M \). For decision making \( N \) independent experiences are carried out. When \( H_m \) is true, the sample \( X = \{x_1, x_2, \ldots, x_N\} \) of the experiments results has the probability

\[
P_m(X) = \prod_{i=1}^{N} P_m(x_i), \ m = 1, M.
\]

By means of non-randomized test \( \varphi(X) \) on the basis of a sample of length \( N \) one of the hypotheses must be accepted. For this aim one can divide the sample space \( \mathcal{X}^N \) on \( M \) disjoint subsets,

\[
\omega^N_m \equiv \{X : \varphi(X) = m\}, \ m = 1, M.
\]

The probability of the erroneous acceptance of hypothesis \( H_l \) provided that hypothesis \( H_m \) is true, for \( m \neq l \) is denoted:

\[
\alpha^N_{ml}(\varphi_X) \equiv \sum_{X \in \omega^N_l} P^N_m(X).
\]

For \( m = l \) denote by \( \alpha^N_{mm}(\varphi_X) \) the probability to reject \( H_m \) when it is true and this is:

\[
\alpha^N_{mm}(\varphi_X) \equiv \sum_{m \neq l} \alpha^N_{ml}(\varphi_X). \quad (1)
\]

The matrix \( \omega(\varphi_X) \equiv \{ \alpha^N_{ml}(\varphi_X) \} \) is called power of the test. Take into consideration the
rates of exponential decrease of the error probabilities and call them reliabilities:

\[ E_{mj}(\varphi) \equiv \lim_{N \to \infty} \left( -\frac{1}{N} \log \alpha_{mj}(\varphi_N) \right) \quad (2) \]

According to (1) and (2)

\[ E_{m/l} = \min_{m \neq l} E_{ml}(\varphi) \quad (3) \]

can be derived because

\[ E_{ml} = \lim_{N \to \infty} \frac{-1}{N} \log \sum_{\varphi_{m/l}} \alpha_{mj}(\varphi) \]
\[ = \lim_{N \to \infty} \frac{-1}{N} \log\left( \max_{\varphi_{m/l}} \left\{ \sum_{j} \alpha_{mj}(\varphi) \right\} \right) + 1 \]
\[ = \lim_{N \to \infty} \frac{-1}{N} \log\left( \max_{\varphi_{m/l}} \alpha_{mj}(\varphi) \right) + 0 \]
\[ = \min_{m \neq l} \left( -\frac{1}{N} \log(\alpha_{mj}(\varphi)) \right) \]
\[ = \min_{m \neq l} \left( E_{ml}(\varphi) \right) \]

The matrix \( E(\varphi) = \{ E_{mj}(\varphi) \} \) is called the reliability matrix of the tests sequences \( \varphi \).

The problem is to find the matrix \( E(\varphi) \) with largest elements, which can be achieved by tests when a part of elements of the matrix \( E(\varphi) \) is fixed.

**Definition**

The test sequence \( \varphi^* = (\varphi_1, \varphi_2, \ldots) \) is called LAO if for given values of the elements \( E_{lj}, E_{2j}, \ldots, E_{M-1j} \) it provides maximal values for all other elements of \( E(\varphi^*) \).

Consider for a given positive and finite \( E_{lj}, E_{2j}, \ldots, E_{M-1j} \) the following family of regions:

\[ \mathcal{R}_l \equiv \{ Q : D(Q \parallel P_l) \leq E_{lj} \}, \quad (4.a) \]
\[ l = 1, 2, \ldots, M-1 \]
\[ \mathcal{R}_M \equiv \{ Q : D(Q \parallel P_l) > E_{lj}, \quad l = 1, 2, \ldots, M \} \quad (4.b) \]
\[ \mathcal{R}_l^N = \mathcal{R}_l \cap P_N(\mathcal{X}), \quad (4.c) \]
\[ l = 1, 2, \ldots, M \]

and introduce the functions:

\[ E_{lj}^* = E_{lj}^*(E_{lj}) \equiv E_{lj}, \quad l = 1, 2, \ldots, M-1, \quad (5.d) \]
\[ E_{m/l}^* = E_{m/l}^*(E_{m/l}) \equiv \inf_{Q_{\mathcal{R}_m}} (D(Q \parallel P_m)), \quad l = 1, 2, \ldots, M-1, \quad m \neq l \]
\[ \inf_{Q_{\mathcal{R}_M}} (D(Q \parallel P_M)), \quad m = 1, 2, \ldots, M-1 \]
\[ E_{ll}^* = E_{ll}^*(E_{ll}) \equiv \inf_{Q_{\mathcal{R}_l}} (D(Q \parallel P_l)), \quad l = 1, 2, \ldots, M \]

With the assumption \( A = \mathcal{R}_1, \quad P = P_m \) in Sanov’s theorem for conditions (4), (5) there is:

\[ \lim_{N \to \infty} \left( -\frac{1}{N} \log \alpha_{mj}^N(\varphi_N^*) \right) \]
\[ = \lim_{N \to \infty} \left( -\frac{1}{N} \log P_m^N(\mathcal{R}_l) \right) \]
\[ = \inf_{Q_{\mathcal{R}_m} \in P_m} (D(Q_{\mathcal{X}} \parallel P_m)) \]

The matrix \( E(\varphi) = \{ E_{mj}(\varphi) \} \) is called the reliability matrix of the tests sequences \( \varphi \).
The notation \( y_1^N = y_2^N \) can be used when \( g(y_1^N) = g(y_2^N) + \varepsilon_N \), where \( \varepsilon_N \to 0 \), for \( N \to \infty \). Using (6)

\[
E_{ml}(\phi^*) = \inf_{Q_i \in \mathcal{R}_i} (D(Q \parallel P_m)). \tag{7}
\]

Therefore the value of \( \alpha^{m*}_{ml} (\phi^*_N) \) is equal to

\[
\alpha^{m*}_{ml} (\phi^*_N) = \exp(-N \inf_{Q_i \in \mathcal{R}_i} (D(Q \parallel P_m))) \tag{8}
\]

\[
= \exp(-NE_{ml}(\phi^*_N)).
\]

In fact, the error probability \( \alpha^{m*}_{ml} (\phi^*_N) \) goes to zero with exponential rate \( \inf_{Q_i \in \mathcal{R}_i} (D(Q \parallel P_m)) \) for \( P_m \) not in the set of \( \mathcal{R}_i \).

Theorem 2

For fixed on finite set \( \mathcal{X} \) family of distributions \( P_1, P_2, \ldots, P_M \) the following two statements hold: If the positive finite numbers \( E_{11}, E_{22}, \ldots, E_{M-1,M-1} \) satisfy conditions:

\[
E_{ij} \leq \min_{i=2, M} D(P_j \parallel P_i), \tag{9}
\]

\[
E_{M,M} = \min \left[ \min_{i=1, m=1, M} E^*_{ml}(E_{ij}), \min_{i=1, m=1, M} D(P_i \parallel P_m) \right],
\]

Hence:

a) There exists a LAO sequence of tests \( \phi^*_N \), the reliability matrix of which \( E^* = \{E^*_{ml}(\phi^*)\} \) is defined in (5), and all elements of it are positive.

b) Even if one of conditions (9) is violated, then the reliability matrix of an arbitrary test necessarily has an element equal to zero, the corresponding error probability does not tend exponentially to zero.

Proof: At first it is remarked that \( D(P_i \parallel P_m) > 0 \), for \( m \neq l \), because all measures \( P_i, l = 1, M \), are distinct. Now for the proof of the sufficiency of the conditions (9). Consider the following sequence of tests \( \phi^* \) given by the sets

\[
B^N_l = \bigcup_{Q_i \in \mathcal{R}_i} T^N_l, \quad l = 1, M. \tag{10}
\]

The sets \( B^N_l, \quad l = 1, M \), satisfies conditions to give test, by means:

\[
B^N_l \cap B^N_m = \phi, \quad l \neq m,
\]

and

\[
\bigcup_{l=1}^M B^N_l = \mathcal{X}^N.
\]

The following shows that exponent \( E_{m,m} (\phi^*) \) for sequence of tests \( \phi^* \) defined in (10) is not less than \( E_{m,m} \). The following is known from lemma,

\[
| T^N_l | = \exp\{NH(Q)\}
\]

and

\[
P^N(T^N_l) = \exp\{-N(D(Q \parallel P)\}, \quad m = 1, M\]

and also by using the result of theorem 1 there is:

\[
\alpha^{m*}_{ml} (\phi^*) = \exp\{-NE_{ml}\}
\]

\[
\alpha^{m*}_{ml} (\phi^*) = \exp\{-NE_{ml}^* (E_{ij})\}, \quad l = 1, M-1,
\]

\[
\alpha^{m*}_{ml} (\phi^*) = \exp\{-NE_{ml}^* (E_{ij}, E_{22}, \ldots, E_{M-1,M-1})\}, \quad m = 1, M.
\]

Using (9) and (4 - 5), all \( E^*_{ml} \) are strictly positive. The proof of part (a) will be finished if one demonstrate that the sequence of the test \( \phi^* \) is LAO, that is, at given finite
$E_{1|1}, E_{2|2}, \ldots, E_{M-1|M-1}$ for any other sequence of tests $\phi^{**}$

$$E_{m|l}^{**}(\phi^{**}) \leq E_{m|l}^{**}(\phi^{*}), \ m, l = 1, M.$$  

For this purpose it is sufficient to see that the sequence of tests asymptotically dose not became better if the sets $B^N_m$ will not be union of some number of whole types $T^N_Q$, in other words, if a test $\phi^{**}$ is defined, for example, by sets $G^N_1, G^N_2, \ldots, G^N_M$ and, in addition, $Q$ is such that $0 < |G^N_l \cap T^N_Q| = |T^N_Q|$, 

The test $\phi^{**}$ will not became worse if instead of the set $G^N_l$, one takes $G^N_l \supset T^N_Q$, it $G^N_l$ nonempty intersection with $T^N_Q$. At last is is able to prove the necessity of the condition (9).

If the sequence of the tests is LAO, then it can be given by sets of (10) form. But, the non-fulfillment of the conditions (9) is equivalent either to violation of (3), or to equality to zero some of $E_{m|l}^{*}$ given in (9), and this again contradicts with (3) because $E_{m|m}, \ m = 1, M-1,$ must be positive.

Remark 1

From definition (5) and (9) it follows that:

$$E_{m|m}^{*} = E_{m|M}^{*}, \ m = 1, M-1.$$  

Remark 2

After the change of hypotheses enumeration the theorem remains valid with corresponding changes in conditions (9).

Remark 3

The maximal likelihood test accepts the hypotheses maximising the probability of sample $X$. In fact

$$r^* = \arg \max_r P^N_r(X).$$  

But it follows from equality 

$$P^N_r(X) = \exp \{-N[H(Q) + D(Q \parallel P)]\}$$  

that at the same time $r^* = \arg \min_r D(Q \parallel P)$. In fact the principle of maximum of likelihood is equivalent to the principle of minimum ok Kullback-Leibler distance.

Acknowledgement

I am grateful to Professor E. A. Haroutunian for his very helpful comments which substantially improved the presentation of the paper.

References


