Journal of Modern Applied Statistical Methods

Volume 6 | Issue 2 Article 13

11-1-2007

Large Deviations Techniques for Error Exponents to Multiple Hypothesis LAO Testing

Leader Navaei
Yerevan State University

Recommended Citation
Navaei, Leader (2007) "Large Deviations Techniques for Error Exponents to Multiple Hypothesis LAO Testing," *Journal of Modern Applied Statistical Methods*: Vol. 6 : Iss. 2 , Article 13.
DOI: 10.22237/jmasm/1193890320
Available at: http://digitalcommons.wayne.edu/jmasm/vol6/iss2/13

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.

Journal of Modern Applied Statistical Methods
November, 2007, Vol. 6, No. 2, 487-491

1538 – 9472/07/$95.00

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 α_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.

Leader Navaei is Assistant Professor in the Faculty of Mathematics. His research interests are in Markov chains, large deviations techniques, and applied information theory in multiple hypotheses testing. Address correspondence to Yerevan State University, St. Alex Manoogian 1,Yerevan 375049, Republic of Armenia. Email: ashkan_l1380@yahoo.com, or l_navaei@ysu.am

Preliminaries

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

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

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

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 $X = (x_1, x_2, ..., x_N) \in \chi^N$ is the empirical distribution given by $Q(x) = \frac{N(x \mid X)}{N}$ for all $x \in \chi$, where $N(x \mid X)$ denotes the number of occurrences of x in X or $Q(x) = (\frac{N_1}{N}, \frac{N_2}{N}, ..., \frac{N_N}{N}) \in \chi^N$ that $N_i \equiv$ number of times out N trials that the random variables $x_1, x_2, ..., x_N$ occurrences in χ.

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

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

$$P^N(X) = \exp\{-N(H(Q) + D(Q \| P)\}.$$

Lemma

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

$$| P_N(\chi) | < (N+1)^{|\chi|},$$

For any type $Q \in P_N(\chi)$:

$$(N+1)^{-|\chi|} \exp\{NH(Q)\} \\ \leq | T_Q^N | \leq \exp\{NH(Q)\},$$

For any $PD, P \in P(\chi)$:

$$\frac{P^N(X)}{Q^N(X)} = \exp\{-ND(Q \| P)\}, \text{ If } X \in T_Q^N,$$

and

$$(N+1)^{-|\chi|} \exp\{-ND(Q \| P)\} \leq \\ \leq | T_Q^N | \leq \exp\{NH(Q)\},$$

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

Let A be a set of distributions from $P(\chi)$ 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 χ:

$$\lim_{N \to \infty} (-\frac{1}{N} \log P^N(X : Q_X \in A)) \\ = \inf_{Q_X \in A} (D(Q_X \| P)).$$

Problem Statement and Formulation of Results

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

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

By means of non-randomized test $\varphi_N(X)$ 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 χ^N on M disjoint subsets,

$$\omega_m^N \equiv \{X : \varphi_N(X) = m\}, \ m = \overline{1, M}.$$

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

$$\alpha_{m|l}^N(\varphi_N) \equiv P_m^N(\omega_l^N) = \sum_{X \in \omega_l^N} P_m^N(X).$$

For $m = l$ denote by $\alpha_{m|m}^N(\varphi_N)$ the probability to reject H_m when it is true and this is:

$$\alpha_{m|m}^N(\varphi_N) \equiv \sum_{m \neq l} \alpha_{m|l}^N(\varphi_N). \ (1)$$

The matrix $\omega(\varphi_N) \equiv \{ \alpha_{m|l}^N(\varphi_N) \}$ is called power of the test. Take into consideration the rates of exponential decrease of the error probabilities and call them reliabilities:

$$E_{m|l}(\varphi) \equiv \overline{\lim_{N\to\infty}}(-\frac{1}{N}\log \alpha_{m|l}(\varphi_N)) \quad (2)$$

According to (1) and (2)

$$E_{m|m} = \min_{m\neq l} E_{m|l} \quad (3)$$

can be derived because

$$\begin{aligned}
E_{m|m} &= \lim_{N\to\infty}\frac{-1}{N}\log \alpha_{m|m}(\varphi) \\
&= \lim_{N\to\infty}\frac{-1}{N}\log \sum_{m\neq l}\alpha_{m|l}(\varphi) \\
&= \lim_{N\to\infty}\frac{-1}{N}[\log(Max\alpha_{m|l}\{\frac{\sum_{m\neq l}\alpha_{m|l}}{Max\alpha_{m|l}}+1\})] \\
&= \lim_{N\to\infty}\frac{-1}{N}\log(Max_{\,m\neq l}\,\alpha_{m|l})+0 \\
&= \min_{m\neq l}\lim_{N\to\infty}\frac{-1}{N}\log(\alpha_{m|l}) \\
&= \min_{m\neq l} E_{m|l}(\varphi)
\end{aligned}$$

The matrix $E(\varphi)=\{E_{m|l}(\varphi)\}$ is called the reliability matrix of the tests sequences φ.

$$E(\varphi) = \begin{bmatrix} E_{1|1}.....E_{1|l}.....E_{1|M} \\ \vdots \\ E_{m|1}.....E_{m|l}.....E_{m|M} \\ \vdots \\ E_{M|1}.....E_{M|l}.....E_{M|M} \end{bmatrix}.$$

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

Consider for a given positive and finite $E_{1|1}, E_{2|2}, ..., E_{M-1|M-1}$ the following family of regions:

$$\Re_l \equiv \{Q : D(Q \| P_l) \le E_{l|l}\}, \quad (4.a)$$
$$l = \overline{1, M-1}$$
$$\Re_M \equiv \{Q : D(Q \| P_l) > E_{l|l},\ l = \overline{1, M-1}\} \quad (4.b)$$
$$\Re_l^N = \Re_l \cap P_N(\chi), \quad (4.c)$$
$$l = \overline{1, M}$$

and introduce the functions:

$$E_{l|l}^* = E_{l|l}^*(E_{l|l}) \equiv E_{l|l},\ l = \overline{1, M-1}, \quad (5.d)$$

$$E_{m|l}^* = E_{m|l}^*(E_{l|l}) \equiv \inf_{Q\in\Re_l}(D(Q \| P_m)), \quad (5.b)$$
$$l = \overline{1, M-1}\ , m = \overline{1, M}\ , m \neq l.$$

$$E_{m|M}^* = E_{m|l}^*(E_{1|1}, E_{2|2}, ..., E_{M-1|M-1}) \equiv \quad (5.c)$$
$$\inf_{Q\in\Re_M}(D(Q \| P_m)),\ m = \overline{1, M-1},$$

$$E_{M|M}^* = E_{M|M}^*(E_{1|1}, E_{2|2}, ..., E_{M-1|M-1}) \equiv \quad (5.d)$$
$$\min_{l=\overline{1,M-1}} E_{M|l}.$$

With the assumption $A = \Re_l$, $P = P_m$ in Sanov's theorem for conditions (4), (5) there is :

$$\begin{aligned}
&\lim_{N\to\infty}(-\frac{1}{N}\log \alpha_{m|l}^N(\varphi_N^*)) \\
&= \lim_{N\to\infty}(-\frac{1}{N}\log P_m^N(\Re_l)) \qquad (6) \\
&= \inf_{Q_X\in\Re_l}(D(Q_X \| P_m))
\end{aligned}$$

The notation $y_1^N \approx y_2^N$ 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)

$$E_{m|l}(\varphi^*) = \inf_{Q_X \in \Re_l} (D(Q \parallel P_m)). \quad (7)$$

Therefore the value of $\alpha_{m|l}^N(\varphi_N^*)$ is equal to

$$\alpha_{m|l}^N(\varphi_N^*) \approx \exp(-N \inf_{Q_X \in \Re_l} (D(Q \parallel P_m)) \quad (8)$$
$$\approx \exp(-NE_{m|l}(\varphi_N{}^*)).$$

In fact, the error probability $\alpha_{m|l}^N(\varphi_N^*)$ goes to zero with exponential rate $\inf_{Q_X \in \Re_l} (D(Q \parallel P_m))$ for P_m not in the set of $\Re_l$.

Theorem 2

For fixed on finite set χ family of distributions $P_1, P_2, \ldots\ldots, P_M$ the following two statements hold: If the positive finite numbers $E_{1|1}, E_{2|2}, \ldots\ldots, E_{M-1|M-1}$ satisfy conditions:

$$E_{l|l} \le \min_{l=2,M} D(P_l \parallel P_1), \quad (9)$$
$$E_{M|M} < \min\left[\min_{l=1,m-1} E_{m|l}^*(E_{l|l}), \min_{l=m+1,M} D(P_l \parallel P_m)\right],$$

Hence:

a) There exists a LAO sequence of tests φ_N^*, the reliability matrix of which $E^* = \{E_{m|l}^*(\varphi^*)\}$ 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 $D(P_l \parallel P_m) > 0$, for $m \neq l$, because all measures P_l, $l = \overline{1, M}$, are distinct. Now for the proof of the sufficiency of the conditions (9). Consider the following sequence of tests φ^* given by the sets

$$B_l^N = \bigcup_{Q \in \Re_l} T_Q^N, \; l = \overline{1, M}. \quad (10)$$

The sets B_l^N, $l = \overline{1, M}$, satisfies conditions to give test, by means:

$$B_l^N \cap B_m^N = \phi, \; l \neq m,$$

and

$$\bigcup_{l=1}^{M} B_l^N = \chi^N.$$

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

$$|T_Q^N| \approx \exp\{NH(Q)\}$$

and

$$P^N(T_Q^N) \approx \exp\{-N(D(Q \parallel P)\}, \; m = \overline{1, M}$$

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

$$\alpha_{m|m}^N(\varphi^*) \approx \exp\{-NE_{m|m}\},$$

and

$$\alpha_{m|l}^N(\varphi^*) \approx \exp\{-NE_{m|l}^*(E_{l|l})\}, \; l = \overline{1, M-1},$$
$$m = \overline{1, M}, \; m \neq l,$$

$$\alpha_{m|l}^N(\varphi^*) \approx \exp\{-NE_{m|M}^*(E_{1|1}, E_{2|2}, \ldots\ldots, E_{M-1|M-1})\},$$
$$m = \overline{1, M}.$$

Using (9) and (4 - 5), all $E_{m|l}^*$ are strictly positive. The proof of part (a) will be finished if one demonstrate that the sequence of the test φ^* is LAO, that is, at given finite

$E_{1|1}, E_{2|2}, ..., E_{M-1|M-1}$ for any other sequence of tests φ^{**}

$$E^*_{m|l}(\varphi^{**}) \leq E^*_{m|l}(\varphi^*), \ m,l = \overline{1,M}.$$

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 φ^{**} is defined, for example, by sets $G^N_1, G^N_2, ..., G^N_M$ and, in addition, Q is such that $0 < | G^N_l \bigcap T^N_Q | \approx | T^N_Q |$,

The test φ^{**} 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 $E_{m|m}$, $m = \overline{1, M-1}$, must be positive.

Remark 1

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

$$E^*_{m|m} = E^*_{m|M}, \ m = \overline{1, M-1}$$

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

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

But it follows from equality $P^N(X) = \exp\{-N[H(Q) + D(Q \,||\, P)]\}$ that at the same time $r^* = \arg\min_r D(Q \,||\, P)$. 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

Csiszar I., & Korner J. (1981). *Information theory: coding theorem for discrete memoryless systems.* NY: Academic Press.

Csiszar I., & Longo G. (1971). *On the error exponent for source coding and for testing simple statistical hypotheses, Studio sc. Mathem. Hungarica*, *6*, 181-191

Csiszar I., & Shields P. (2004). *Information theory and statistics*: *Fundamentals and trends in communications and information theory*. Hanover: MA.

Dembo A., &Zeitoni O. (1993). *Large deviations techniques and applications*, London: Jons and Bartlet .

Haroutunian, E. A. (1990). Logarithmically asymptotically optimal testing of multiple statistical hypotheses, *Problems of Control and Information Theory*, *19*, 413-421.

Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Annals of Mathematical Statistics*, *36*, 369-401.

Tusnady, G. (1977). On asymptotically optimal tests, *Annals of Statistics*, *5*(2), 385-393.

Longo, G., & Sgarro A. (1980). The error exponent for the testing of simple statistical hypotheses, A combinatorial. approach. *Journal of Combinotorics, Information. and System Sciences*, *5*(1), 58-67.

Perez, A. (1984). *Second-type-error exponent given the first-type-error exponent in the testing statistical hypotheses by unfitted procedures.* Abstract of papers of the Sixth International Symposium on Information Theory. (Part 1). Tashkent, 277-279.