Far East J. Math. Sci. (FJMS) **25**(2) (2007), 335-344
This paper is available online at http://www.pphmj.com

ON MANY HYPOTHESES LOGARITHMICALLY ASYMPTOTICALLY OPTIMAL TESTING VIA THE THEORY OF LARGE DEVIATIONS

LEADER NAVAEI

(Received July 10, 2006)

Abstract

In this paper with the usage of theory of large deviations the problem proposed by R. Dobrushin for $M > 2$ hypotheses testing is solved. We notice Sanov's theorem and its applications in hypotheses testing and show that this method of investigation for the problem is easier and gives identical results as in [5].

1. Introduction

We shall present illustrations of usefulness of application of theory of large deviations to investigation of the logarithmically asymptotically optimal (LAO) testing of statistical hypotheses. Many papers were devoted to the study of exponential decrease, as the sample size N goes to infinity, of the error probabilities $\alpha_1^{(N)}$ of the first kind and $\alpha_2^{(N)}$ of the second kind of the optimal tests for two simple statistical hypotheses.

In case of the fixed probability $\alpha_1^{(N)} = \alpha_1$ 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. In the book of Csiszár and Shields

2000 Mathematics Subject Classification: 62P30.

Keywords and phrases: hypotheses testing, error probability, reliability matrix, logarithmically asymptotically optimal (LAO).

[3] for independent identically distributed observations different asymptotical aspects of two hypotheses testing are considered via theory of large deviations.

In [5] Haroutunian solve the problem proposed by R. Dobrushin for the case $M > 2$ hypotheses. In the present paper we introduce the proof of this theorem by using Sanov's theorem.

In the next section we shall express notations, basic concepts and theorem of Sanov and also in Section 3 we shall present the result and its proof.

2. Preliminaries

At first we review the technical background of the method of types. Assume that $\mathcal{X}$ is a finite set of the size $|\mathcal{X}|$. The set of all probability distributions (*PDs*) on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. For *PD's*, P and Q, $H(P)$ denotes entropy and $D(P \| Q)$ denotes information divergence (or the Kullback-Leibler distance)

$$H(P) \triangleq -\sum_{x \in \mathcal{X}} P(x) \log P(x), \quad D(P \| Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

In this paper we use exps and logs at base 2. We also consider 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 $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathcal{X}^N$ is the empirical distribution given by $Q(x) \triangleq N^{-1} \cdot N(x \mid \mathbf{x})$ for all $x \in \mathcal{X}$, where $N(x \mid \mathbf{x})$ denotes the number of occurrences of x in $\mathbf{x}$.

The subset of $\mathcal{P}(\mathcal{X})$ consisting of the possible types of sequences $\mathbf{x} \in \mathcal{X}^N$ is denoted by $\mathcal{P}_N(\mathcal{X})$. For $Q \in \mathcal{P}_N(\mathcal{X})$ the set of sequences of type class Q is denoted by $\mathcal{T}_Q^N(\mathbf{X})$.

The probability that N independent drawings from $P \in \mathcal{P}(\mathcal{X})$ give $\mathbf{x} \in \mathcal{X}^N$, is denoted by $P^N(\mathbf{x})$. If $\mathbf{x} \in \mathcal{T}_Q^N(\mathbf{x})$, then

$$P^N(\mathbf{x}) \triangleq \prod_{x \in \mathcal{X}} P(x)^{NQ(x)} = \exp\{-N[H(Q) + D(Q \| P)]\}.$$

Lemma ([3], [4])**.** (a) *The number of types of length N for sequences grows at most polynomially with N*:

$$|\mathcal{P}_N(\mathcal{X})| < (N+1)^{|\mathcal{X}|}.$$

(b) *For any type* $Q \in \mathcal{P}_N(\mathcal{X})$ *we have*:

$$(N+1)^{-|\mathcal{X}|} \exp\{NH(Q)\} \le |\mathcal{T}_Q^N(\mathbf{X})| \le \exp\{NH(Q)\}.$$

(c) *For any PD* $P \in \mathcal{P}(\mathcal{X})$ *we have*:

$$\frac{P^N(\mathbf{x})}{Q^N(\mathbf{x})} = \exp\{-ND(Q \| P)\}, \ \ \textit{if} \ \ \mathbf{x} \in \mathcal{T}_Q^N(\mathbf{X}),$$

and

$$(N+1)^{-|\mathcal{X}|} \exp\{-ND(Q \| P)\} \le P^N(\mathcal{T}_Q^N(\mathbf{X})) \le \exp\{-ND(Q \| P)\}.$$

Theorem 1 (Sanov's theorem [3], [4])**.** *Let* $\mathcal{A}$ *be a set of distributions from* $\mathcal{P}$ *such that its closure is equal to the closure of its interior. Then for the empirical distribution* $Q_{\mathbf{x}}$ *of a vector* $\mathbf{x}$ *from a strictly positive distribution P on* $\mathcal{X}$:

$$\lim_{N \to \infty} \left(-\frac{1}{N} \log P(Q_{\mathbf{x}} \in \mathcal{A}) \right) = \inf_{Q_{\mathbf{x}} \in \mathcal{A}} D(Q_{\mathbf{x}} \| P).$$

Proof. Let $\mathcal{A}_N \triangleq \mathcal{A} \cap \mathcal{P}_N$. Then from above Lemma, upper and lower bounds for all finite N are deduced. By the upper bound of Lemma we can write:

$$\begin{aligned} P(Q_{\mathbf{x}} \in \mathcal{A}_N) &= P^N\left(\bigcup_{Q_{\mathbf{x}} \in \mathcal{A}_N} \mathcal{T}_{Q_{\mathbf{x}}}^N(X) \right) \\ &= \sum_{Q_{\mathbf{x}} \in \mathcal{A}_N} P^N(\mathcal{T}_{Q_{\mathbf{x}}}^N(X)) \\ &\le \sum_{Q_{\mathbf{x}} \in \mathcal{A}_N} \exp\{-ND(Q_{\mathbf{x}} \| P)\} \\ &\le (N+1)^{|\mathcal{X}|} \exp\{-N \inf_{Q_{\mathbf{x}} \in \mathcal{A}_N} D(Q_{\mathbf{x}} \| P)\} \end{aligned}$$

and also lower bound is:

$$P(Q_{\mathbf{x}} \in \mathcal{A}_N) = \sum_{Q_{\mathbf{x}} \in \mathcal{A}_N} P^N(\mathcal{T}^N_{Q_{\mathbf{x}}}(X))$$

$$\geq \sum_{Q_{\mathbf{x}} \in \mathcal{A}_N} (N+1)^{-|\mathcal{X}|} \exp\{-ND(Q_{\mathbf{x}} \,\|\, P)\}$$

$$\geq (N+1)^{-|\mathcal{X}|} \exp\{-N \inf_{Q_{\mathbf{x}} \in \mathcal{A}_N} D(Q_{\mathbf{x}} \,\|\, P)\}.$$

Since $\lim\limits_{N\to\infty} N^{-1} \log(N+1)^{|\mathcal{X}|} = 0$ and $D(Q_{\mathbf{x}} \,\|\, P)$ is continuous in Q, the hypotheses on $\mathcal{A}$ implies that $D_{Q_{\mathbf{x}} \in \mathcal{A}_N}(Q_{\mathbf{x}} \,\|\, P)$ is arbitrarily close to $D_{Q_{\mathbf{x}} \in \mathcal{A}}(Q_{\mathbf{x}} \,\|\, P)$ if N is large. The proof is complete.

3. Problem Statement and Formulation of Results

The problem of many 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 correspondingly one of the 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 $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ of the experiments results has the probability

$$P_m^N(\mathbf{x}) \triangleq \prod_{i=1}^{N} P_m(x_i), \quad m = \overline{1, M}.$$

By means of non-randomized test $\varphi_N(\mathbf{x})$ and on the basis of a sample $\mathbf{x}$ of length N, we must accept one of the hypotheses. For this aim we can divide the sample space $\mathcal{X}^N$ on M disjoint subsets

$$\mathcal{A}_m^N \triangleq \{\mathbf{x} : \varphi_N(\mathbf{x}) = m\}, \quad m = \overline{1, M}.$$

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

$$\alpha_{m|l}^N(\varphi_N) \triangleq P_m^N(\mathcal{A}_l^N) = \sum_{\mathbf{x} \in \mathcal{A}_l^N} P_m^N(\mathbf{x}).$$

For $m = l$ we denote by $\alpha^N_{m|m}(\varphi_N)$ the probability to reject H_m when it is true:

$$\alpha^N_{m|m}(\varphi_N) \triangleq \sum_{l \neq m} \alpha^N_{m|l}(\varphi_N). \tag{1}$$

The matrix $\mathcal{A}(\varphi_N) \triangleq \{\alpha^N_{m|l}(\varphi_N)\}$ is called *power* of the test.

We consider the rates of exponential decrease of the error probabilities and call them reliabilities

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

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

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

The problem is to find the matrix E with largest elements, which can be achieved by tests when a part of elements of the matrix E is fixed. According to (1) and (2) we can derive that:

$$E_{m|m} = \min_{l \neq m} E_{m|l}. \tag{3}$$

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

Our aim is to define conditions on $E_{1|1}, \ldots, E_{M-1|M-1}$ under which there exists LAO sequence of tests φ^* and show how other elements $E_{m|l}(\varphi^*)$ for the matrix $E(\varphi^*)$ can be found from them.

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

$$\mathcal{R}_l \triangleq \{Q : D(Q \| P_l) \leq E_{l|l}\}, \quad l = \overline{1, M-1}, \tag{4}$$

$$\mathcal{R}_M \triangleq \{Q : D(Q \| P_l) > E_{l|l}, \; l = \overline{1, M-1}\}, \tag{5}$$

$$\mathcal{R}_l^N \triangleq \mathcal{R}_l \cap \mathcal{P}(\mathcal{X}), \quad l = \overline{1, M} \tag{6}$$

and the following numbers:

$$E_{l|l}^* \triangleq E_{l|l}^*(E_{l|l}) \triangleq E_{l|l}, \quad l = \overline{1, M-1}, \tag{7}$$

$$E_{m|l}^* \triangleq E_{m|l}^*(E_{l|l}) \triangleq \inf_{Q \in \mathcal{R}_l}(D(Q \| P_m)), \;\; m = \overline{1, M}, \;\; m \neq l, \;\; l = \overline{1, M-1}, \tag{8}$$

$$E_{m|M}^* \triangleq E_{m|M}^*(E_{l|l}, \ldots, E_{M-1|M-1}) \triangleq \inf_{Q \in \mathcal{R}_l}(D(Q \| P_m)), \quad m = \overline{1, M-1}, \tag{9}$$

$$E_{M|M}^* \triangleq E_{M|M}^*(E_{l|l}, \ldots, E_{M-1|M-1}) \triangleq \min_{l=\overline{1, M-1}} E_{M|l}. \tag{10}$$

$E_{m|l}$ can be equal to ∞, this can happen when some measures are not absolutely continuous one relative to the others.

Now we explain application of Sanov's theorem in hypotheses testing.

We know that with assumption $\mathcal{A} = \mathcal{R}_l$, $P = P_m$ in Sanov's theorem for conditions (4)-(6), (7)-(10) we have (see Figure 1)

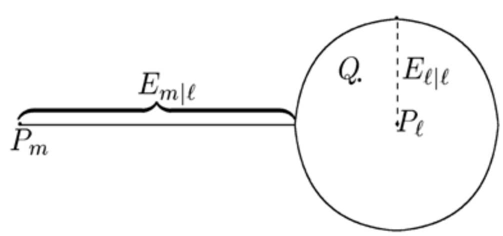

Figure 1. Interpretation of the construction of the test.

$$\lim_{N \to \infty} -\frac{1}{N} \log P_m(Q \in \mathcal{R}_l) = \inf_{Q \in \mathcal{R}_l} D(Q \| P_m) \tag{11}$$

and also

$$\lim_{N\to\infty} -\frac{1}{N}\log P_m(Q \in \mathcal{R}_l) = \lim_{N\to\infty} -\frac{1}{N}\log \alpha_{m|l}(\varphi^*). \tag{12}$$

We introduce also $y_1^N \approx y_2^N$, when $g(y_1^N) = g(y_2^N) + \varepsilon_N$, where $\varepsilon_N \to 0$ for $N \to \infty$. Now using (11) and (12) we can write

$$E_{m|l}(\varphi^*) \approx \inf_{Q\in\mathcal{R}_l} D(Q \,\|\, P_m). \tag{13}$$

Therefore the value of:

$$\alpha_{m|l}(\varphi_N^*) \approx \exp(-N \inf_{Q\in\mathcal{R}_l} D(Q\,\|\,P_m)) \approx \exp(-NE_{m|l}(\varphi_N^*)). \tag{14}$$

In fact the error $\alpha_{m|l}(\varphi_N)$ still goes to zero with exponential rate $\inf\limits_{Q\in\mathcal{R}_l} D(Q\,\|\,P_m)$ for P_m not in the set of $\mathcal{R}_l$.

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

$$E_{1|1} < \min_{l=\overline{2,\, M}} D(P_l \,\|\, P_1),$$

$$\vdots$$

$$E_{M|M} < \min\left[\min_{l=\overline{1,\, m-1}} E^*_{m|l}(E_{l|l}),\ \min_{l=\overline{m+1,\, M}} D(P_l\,\|\,P_m)\right], \quad m = \overline{2,\, M-1} \tag{15}$$

hence:

(a) *There exists a LAO sequence of tests* φ_N^*, *the reliability matrix of which* $E^* = \{E^*_{m|l}(\varphi^*)\}$ *is defined in* (7)-(10), *and all elements* $E^*_{m|m}$ *of it are positive.*

(b) *Even if one of conditions* (15) *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 we remark that $D(P_l\,\|\,P_m) > 0$, for $l \neq m$. That is all measures P_l, $l = \overline{1,\, M}$ are distinct. Now we prove the sufficiency of the conditions (15). Consider the following sequence of tests φ^* given by the sets

$$\mathcal{B}_l^N = \bigcup_{Q \in R_l^N} \mathcal{T}_Q^N, \quad l = \overline{1, M}. \tag{16}$$

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

$$\mathcal{B}_l^N \bigcap \mathcal{B}_m^N = \varnothing, \quad l \neq m,$$

and

$$\bigcup_{l=1}^{M} \mathcal{B}_l^N = \mathcal{X}^N.$$

Now let us show that exponent $E_{m|m}(\varphi^*)$ for sequence of tests φ^* defined in (16) is not less than $E_{m|m}$. We know from Lemma that:

$$|\mathcal{T}_Q^N| \approx \exp\{NH(Q)\}$$

and

$$P^N(\mathcal{T}_Q^N) \approx \exp\{-N(D(Q \| P))\}, \quad m = \overline{1, M}$$

and also with (14) we have

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

and

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

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

And at last for $m = l = M$ we have:

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

With using (15) we know that all $E_{m|l}^*$ are strictly positive. The proof of part (a) will be finished if one demonstrates that the sequence of the test φ^* is LAO, that is at given finite $E_{1|1}, \ldots, E_{M-1|M-1}$ for any other sequence of tests φ^{**}

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

For this purpose it is sufficient to see that the sequence of tests asymptotically does not became better if the sets $\mathcal{B}_m^N$ will not be union of some number of whole types $\mathcal{T}_Q^N$, in other words, if a test φ^{**} is defined, for example, by sets $\mathcal{G}_1^N, \ldots, \mathcal{G}_M^N$ and, in addition, Q is such that

$$0 < \left| \mathcal{G}_l^N \bigcap \mathcal{T}_Q^N \right| \approx |\mathcal{T}_Q^N|.$$

The test φ^{**} will not became worse if instead of the set $\mathcal{G}_l^N$ one takes $\mathcal{G}_l^N \cap \mathcal{T}_Q^N$ correspondingly decreasing the sets, which had nonempty intersection with $\mathcal{T}_Q^N$, and at last we prove the necessity of the condition (15). It is just now shown that if the sequence of the tests is LAO, then it can be given by sets of (16) form. But the non-fulfillment of the conditions (15) is equivalent either to violation of (3) or to equality to zero some of $E^*_{m|l}$ given in (15), and this again contradicts with (3) because $E_{m|m}$, $m = \overline{1, M-1}$, must be positive.

4. Remarks

(1) After the change of hypotheses enumeration the theorem remains valid with corresponding changes in conditions (15).

(2) The maximal likelihood test accepts the hypotheses maximising the probability of sample $\mathbf{x}$ in fact

$$r^* = \arg\max_r P_r^N(\mathbf{x}).$$

But it follows from equation $P_r^N(\mathbf{x}) \triangleq \exp\{-N[H(Q) + D(Q \| P_r)]\}$ that at the same time $r^* = \arg\min_r D(Q \| P_r)$. In fact the principle of maximum of likelihood is equivalent to the principle of minimum of Kullback-Leibler distance.

(3) Instead of the diagonal elements of the reliability matrix it is

possible to give finite values to the one arbitrary element in each of the first $M-1$ columns. Assumption equalities in (8) as equations, and taking into account the monotone decrease and continuity, when they are finite, of function, defined in (8), it is possible to find a unique $E_{l|l}$ for given $E_{m|l}$, $l=\overline{1, M-1}$. It is not difficult to reestablish the compatibility conditions, in which matrix E has all positive elements. For further computation of the reliability matrix elements it will be necessary to replace the infinite values of the obtained $E_{l|l}$ by arbitrary large, but finite ones.

References

[1] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, Inc., New York, 1981.

[2] I. Csiszár and G. Longo, On the error exponent for source coding and for testing simple statistical hypotheses, Studia Sci. Math. Hungar. 6 (1971), 181-191.

[3] I. Csiszár and P. Shields, Information Theory and Statistics: A Tutorial, Fundamentals and Trends in Communications and Information Theory 1(4) (2004), 417-528.

[4] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, Jones and Bartlett Publishers, Boston, 1993.

[5] E. A. Haroutunian, Logarithmically asymptotically optimal testing of multiple statistical hypotheses, Problems Control Inform. Theory 19 (1990), 413-421.

[6] W. Hoeffding, Asymptotically optimal tests for multinomial distributions, Ann. Math. Statist. 36 (1965), 369-401.

[7] G. Longo and A. Sgarro, The error exponent for the testing of simple statistical hypotheses: a combinatorial approach, J. Combin. Inform. System Sci. 5(1) (1980), 58-67.

[8] G. Tusnády, On asymptotically optimal tests, Ann. Statist. 5(2) (1977), 385-393.

Yerevan State University
Faculty of Mathematics
St. Alex Manoogian 1, Yerevan 375049
Republic of Armenia
e-mail: l_navaei@ysu.am