Mathematical Problems of Computer Science 46, 37–43, 2016.

On Optimality of Tests of M-ary Hypotheses for Fixed Number of Independent Observations

Evgueni A. Haroutunian

Institute for Informatics and Automation Problems of NAS RA
e-mail: eghishe@sci.am

Abstract

The paper is devoted to the design of optimal approaches of testing of multiple simple hypotheses with samples of a fixed number of independent observations.
Keywords: Statistical hypothesis testing, Error probability, Compairing of tests, Optimal test, Neyman - Pearson approach.

1. Introduction

We address the classical detection problem. Let $M \geq 2$ simple hypotheses, which we note H_1, ..., H_M, correspond to distributions $G_1, ..., G_M$, given on the finite space $\mathcal{X}$. It is not known which hypothesis is true. The observed sample is a vector $\mathbf{x} = (x_1, x_2, , ..., x_N)$ in N-dimentional space $\mathcal{X}^N$ of N independent observations identically distributed according to one of hypotheses. The statistician using the obtained sample must guess which hypothesis is true. It is necessary to choose an optimal criterion for making such a decision. The procedure of decision is called a test.

Defining a test can be formulated as partitioning of the space $\mathcal{X}^N$ into M disjoint sets $\mathcal{X}_1^N$,..., $\mathcal{X}_M^N$ on each of which one of the hypotheses H_1, ..., H_M will be, respectively, accepted.

The quality of a test φ may be characterized by the matrix of error probabilities

$$\alpha_{l|m} = \alpha_{l|m}(\varphi) = G_m(\mathbf{x} \in \mathcal{X}_l^N), \quad l, m = \overline{1, M}, \quad l \neq m,$$

$$\alpha_m = \alpha_{m|m}(\varphi) = G_m(\mathbf{x} \notin \mathcal{X}_m^N), \quad m = \overline{1, M},$$

where $\alpha_{l|m}$ is the probability to accept H_l, when H_m is true, and $\alpha_{m|m}(\varphi)$ is the probability of rejection of hypothesis H_m when it is true. Evidently

$$\alpha_m = \sum_{l \neq m} \alpha_{l|m}, \quad m = \overline{1, M}.$$

Rich literature is dedicated to the problem of hypothesis testing, we quote [1-3]. The case of multiple hypotheses testing is considered in detail in [4] and [5].

To compare tests there exist different approaches. A Bayes approach is based on assumption that the observations are governed by given probabilities called the a priori probabilities. In this case the set of tests is ordered by the values of the average probabilities of errors of tests [4].

Neyman - Pearson approach [6] primarily was formulated for two hypotheses. In case of M hypotheses the error probabilities are fixed for $m = \overline{1, M-1}$. The test is found by maximizing the probability $1 - \alpha_M$ of correct detection of the hypothesis H_M.

Another direction of investigations of quality and optimality of tests is the asymptotic approach when the length of sample is not limited and when the volume n of the samples goes to the infinity, error probabilities $\alpha_{l|m}$ decrease exponentially as $2^{-nE_{l|m}}$ and the interdependence of these exponential coefficients $E_{l|m}$, called reliabilities, characterize the test. This direction is founded by Hoeffding [7] and the corresponding sequence of tests is called logarithmically asymptotically optimal (LAO), or exponentional rate optimal (ERO) [7,8]. The problems of LAO tests for multiple hypotheses for diverse models were studied in [9-14].

In the minimax approach the comparision of maximal values of error probabilities of a test also allows to order the set of all tests.

The interchange of ideas and methods of Statistics and Information theory is presented in [15,16].

Our purpose in this paper is comparing different tests and introducing an order on the set of all tests for $M \geq 2$ hypotheses when both, list of hypothetical distributions and the volume N of samples, are fixed. We proceed analogically as in [4], where the Bayesian tests are studied.

2. Formulation of Results

We begin by consideing the case of M hypothetical distributions when a priori probabilities are absent. It is clear that the test is better if its error probabilites are smaller.

Definition 1: *An average error probability $\alpha(\varphi)$ for test φ is*

$$\alpha(\varphi) = \frac{1}{2M} \sum_{l,m} \alpha_{l|m} = \frac{1}{M} \sum_{1}^{M} \alpha_m.$$

Definition 2: *The test φ^0 we call optimal if $\alpha(\varphi^0) = \min_{\varphi} \alpha(\varphi)$.*

Now it is possible to arrange the set of tests φ according to the values of the avarage error probabilities $\alpha(\varphi)$. In the following theorem we state the construction of optimal test.

Let probability distributions G_m, $m = \overline{1, M}$, have densities $f_m(x)$ with respect to some measure μ, then the likelihood function is $f_m(\mathbf{x}) = \prod_{n=1}^{N} f_m(x_n), \quad m = \overline{1, M}.$

Theorem 1:

1. The average probability of error $\alpha(\varphi)$ of any test φ satisfies the inequality

$$\alpha(\varphi) \geq 1 - \frac{1}{M} \int \max_m f_m(\mathbf{x}) \mu^n(d\mathbf{x}). \tag{1}$$

2. In order the test φ^0 be optimal it is necessary and sufficient that for almost all with respect to measure G ($G(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^{M} G_m(x)$) values of $\mathbf{x}$ the test satisfies the relations

$$\varphi^0(\mathbf{x}) = H_m, \quad \text{if} \quad f_m(\mathbf{x}) = \max_l f_l(\mathbf{x}). \tag{2}$$

For such φ^0 relation (1) becomes an equality.

It is necessary to note that when two or more values $f_m(\mathbf{x})$ are maximal it does not matter which of these hypotheses to choose and we can make a random decision, this will not change the value of $\alpha(\varphi^0)$.

It is not difficult to apply this criterion to two particular extremal cases. If all densities are equal $f_1(\mathbf{x}) = ... = f_M(\mathbf{x})$ it is not possible to construct a test with $\alpha(\varphi^0)$ less than $\frac{M-1}{M}$. In the other case, when $\mathcal{X}^N = \bigcup\limits_{m=1}^{M} \mathcal{X}_m^N$ and $G_m(\mathcal{X}_m^N) = 1$, $G_m(\mathcal{X}_l^N) = 0$, $m = \overline{1, M}$, $l \neq m$, the best test has $\alpha(\varphi^0) = 0$.

Proof: 1. By Definition 1 for each test φ for M hypotheses the avarage error probability of the test φ is

$$\alpha(\varphi) = \frac{1}{M} \sum_{m=1}^{M} \alpha_m,$$

where

$$\alpha_m = G_m(\varphi(\mathbf{x}) \neq m) = \int\limits_{\mathbf{x}:\varphi(\mathbf{x})\neq m} f_m(\mathbf{x})\mu^n(d\mathbf{x}).$$

Then

$$\alpha(\varphi) = \frac{1}{M} \sum_{m=1}^{M} \int\limits_{\mathbf{x}:\varphi(\mathbf{x})\neq m} f_m(\mathbf{x})\mu^n(d\mathbf{x}) =$$

$$= 1 - \frac{1}{M} \sum_{m=1}^{M} \int\limits_{\mathbf{x}:\varphi(\mathbf{x})= m} f_m(\mathbf{x})\mu^n(d\mathbf{x}) \geq$$

$$\geq 1 - \frac{1}{M} \int \max_m f_m(\mathbf{x})\mu^n(d\mathbf{x}).$$

2. Optimal test φ^0 defined in (2) reaches the lower bound in (1), that is condition (2) is sufficient. To prove the necessity of (2) suppose that the optimal test φ is such that $\varphi(\mathbf{x}) = H_m$ with $f_m(\mathbf{x}) < \max_l f_l(\mathbf{x})$ for $\mathbf{x} \in \mathcal{A}$ with a set $\mathcal{A}$ of positive probability $G(\mathcal{A}) > 0$. Such test φ can be improved on $\mathcal{A}$ by giving for $\mathbf{x} \in \mathcal{A}$, $\varphi^0(\mathbf{x}) = H_{m_1}$, with $f_{m_1}(\mathbf{x}) = \max_l f_l(\mathbf{x})$.

Really $\alpha(\varphi^0) < \alpha(\varphi)$ because

$$\alpha(\varphi) - \alpha(\varphi^0) = \frac{1}{M} \int\limits_{\mathcal{A}} [f_{m_1}(\mathbf{x}) - f_m(\mathbf{x})]\mu^n(d\mathbf{x}) > 0.$$

Now we shall consider the minimax approach. We will compare tests with maximal values of error probabilities.

Definition 3: *Denote maximal error probability in the matrix for each test φ by*

$$\overline{\alpha}(\varphi) = \max_m \alpha_{m|m}(\varphi).$$

It is clear that it is possible to put in order tests by values of $\overline{\alpha}(\varphi)$. We name minimax the test $\overline{\varphi}$ which has the minimal value of $\overline{\alpha}(\overline{\varphi})$

$$\overline{\alpha}(\overline{\varphi}) = \min_\varphi \overline{\alpha}(\varphi).$$

Theorem 2: *The optimal test $\overline{\varphi}$ such that*

$$\alpha_1(\overline{\varphi}) = \alpha_2(\overline{\varphi}) = ... = \alpha_M(\overline{\varphi}) \tag{3}$$

will be a minimax test.

Proof: For every test φ using (3) we have

$$\overline{\alpha}(\varphi) \geq \sum_{m=1}^{M} \frac{1}{M}\alpha_m(\varphi) \geq \sum_{m=1}^{M} \frac{1}{M}\alpha_m(\overline{\varphi}) = \overline{\alpha}(\overline{\varphi}).$$

Now we take into consideration the notion of the most powerful test in a class of tests. In order to compare tests it is possible to introduce the classes having given values of error probabilities α_1, ..., α_{M-1}

$$K_{\alpha_1,...,\alpha_{M-1}} = \{\varphi : \alpha_m(\varphi) = \alpha_m, \ \ m = \overline{1, M-1}\}$$

and then order tests by the values of $\alpha_M(\varphi)$, naturally the smaller $\alpha_m(\varphi)$ corresponds to the better test.

Definition 4: *A test $\widetilde{\varphi} \in K_{\alpha_1,...,\alpha_{M-1}}$ is the most powerful test in the class $K_{\alpha_1,...,\alpha_{M-1}}$ if for any φ from the same class*

$$\alpha_m(\widetilde{\varphi}) \leq \alpha_m(\varphi).$$

If $M = 2$ the optimal test $\widetilde{\varphi}$ is given by the Fundamental Lemma of Neyman-Pearson [2,3,6,16]. For more than two hypotheses $M \geq 2$ the solution can be found by generalization of the Neyman-Pearson Lemma [17]. This important result deserves to be recured.

As it was noted in [15] the case $N = 1$ contains the general one and there is no need to restrict attention to multiple independent drawings.

For given preassingned values $0 < \alpha^*_{1|1}, \alpha^*_{2|2}, ..., \alpha^*_{M-1|M-1} < 1$ we choose numbers $T_1, T_2, ..., T_{M-1}$ and sets $\mathcal{A}^*_m$, $m = \overline{1, M}$, such that

$$\mathcal{A}^*_1 = \left\{\mathbf{x} : \min\left(\frac{G_1(\mathbf{x})}{G_2(\mathbf{x})}, ..., \frac{G_1(\mathbf{x})}{G_M(\mathbf{x})}\right) > T_1\right\}, \quad 1 - G_1(\mathcal{A}^*_1) = \alpha^*_{1|1},$$

$$\mathcal{A}^*_2 = \overline{\mathcal{A}^*_1} \cap \left\{\mathbf{x} : \min\left(\frac{G_2(\mathbf{x})}{G_3(\mathbf{x})}, ..., \frac{G_2(\mathbf{x})}{G_M(\mathbf{x})}\right) > T_2\right\}, \quad 1 - G_2(\mathcal{A}^*_2) = \alpha^*_{2|2},$$

....

$$\mathcal{A}^*_{M-1} = \overline{\mathcal{A}^*_1} \cap \overline{\mathcal{A}2^*} \cap .... \cap \overline{\mathcal{A}^*_{M-2}} \cap \left\{\mathbf{x} : \frac{G_{M-1}(\mathbf{x})}{G_M(\mathbf{x})}) > T_{M-1}\right\}, \quad 1 - G_{M-1}(\mathcal{A}^*_{M-1}) = \alpha^*_{M-1|M-1},$$

and

$$\mathcal{A}^*_M = \mathcal{X}^N - (\mathcal{A}^*_1 \cup \mathcal{A}^*_2 \cup ... \cup \mathcal{A}^*_{M-1}) = \overline{\mathcal{A}^*_1} \cap \overline{\mathcal{A}2^*} \cap .... \cap \overline{\mathcal{A}^*_{M-2}} \cap \overline{\mathcal{A}^*_{M-1}}.$$

The corresponding error probabilities are denoted by

$$\alpha^*_{l|m}(\varphi_N), \quad l, m = \overline{1, M}.$$

Theorem 3: (Generalization of Neyman-Pearson Lemma)

The test determined by sets $\mathcal{A}^*_1$, $\mathcal{A}^*_2$, ..., $\mathcal{A}^*_M$ is optimal in the sense that, for each other test defined by the sets $\mathcal{B}_1$, $\mathcal{B}_2$, ..., $\mathcal{B}_M$ with the corresponding error probabilities $\beta_{l|m}$, $l, m = \overline{1, M}$, if $\beta_{m|m} \leq \alpha^*_{m|m}$, for some $m \in [1, M-1]$, then there exists at least one index j, $j \in [m+1, M]$ such that $\beta_{m|j} \geq \alpha^*_{m|j}$.

Proof: Let $\Phi_{\mathcal{A}_m^*}$ and $\Phi_{\mathcal{B}_m}$ be the indicator functions of the decision regions $\mathcal{A}_m^*$ and $\mathcal{B}_m, m = \overline{1, M}$. For all $\mathbf{x} = (x_1, x_2, ..., x_N) \in \mathcal{X}^N$, the following inequality is correct

$$(\Phi_{\mathcal{A}_m^*}(\mathbf{x}) - \Phi_{\mathcal{B}_m}(\mathbf{x}))(G_m(\mathbf{x}) - \max(T_m G_{m+1}(\mathbf{x}), ..., T_m G_m(\mathbf{x}))) \geq 0.$$

Multiplying and then summing over $\mathcal{X}^N$ we obtain

$$\sum_{\mathbf{x}:\mathbf{x}\in\mathcal{X}^N} \Big[\Phi_{\mathcal{A}_m^*}(\mathbf{x})G_m(\mathbf{x}) - \Phi_{\mathcal{A}_m^*}(\mathbf{x})\max(T_m G_{m+1}(\mathbf{x}), ..., T_m G_m(\mathbf{x})) -$$

$$-\Phi_{\mathcal{B}_m}(\mathbf{x})G_m(\mathbf{x}) + \Phi_{\mathcal{B}_m}(\mathbf{x})\max(T_m G_{m+1}(\mathbf{x}), ..., T_m G_m(\mathbf{x})) \geq 0\Big],$$

$$\sum_{\mathbf{x}:\mathbf{x}\in\mathcal{A}m^*} [G_m(\mathbf{x}) - T_m \max(G_{m+1}(\mathbf{x})), ..., G_M(\mathbf{x})] -$$

$$- \sum_{\mathbf{x}:\mathbf{x}\in\mathcal{B}m} [G_m(\mathbf{x}) - T_m \max(G_{m+1}(\mathbf{x})), ..., G_M(\mathbf{x})] \geq 0.$$

According to the definition of error probability we obtain the following inequality

$$1 - \alpha^*_{m|m} - T_m \max(\alpha^*_{m|m+1}, ..., \alpha^*_{m|M}) - (1 - \beta_{m|m}) + T_m \max(\beta_{m|m+1}, ..., \beta_{m|M}) \geq 0,$$

$$-\beta_{m|m} + \alpha^*_{m|m} \leq T_m \left[-\max(\alpha^*_{m|m+1}, ..., \alpha^*_{m|M}) + \max(\beta_{m|m+1}, ..., \beta_{m|M})\right].$$

We see now that from $\beta_{m|m} \leq \alpha^*_{m|m}$ it follows that

$$\max(\beta_{m|m+1}, ..., \beta_{m|M}) \geq \max(\alpha^*_{m|m+1}, ..., \alpha^*_{m|M}).$$

From this it follows that if the maximal is $\beta_{m|j}$, $j \in [m+1, M]$, then $\beta_{m|j} \geq \alpha^*_{m|j}$.

Disscussion: This paper deals with some central basic results of the Theory of testing statistical hypotheses ought to be included in textbooks. Different approaches of constructing optimal tests for finitely many simple hypotheses by application of samples of fixed length are presented.

Acknowledgment

The author thanks the reviewer for suggestions on improving the exposition of the material.

References

[1] R. R. Bahadur, "Stochastic comparison of tests", Ann. Math. Statist. vol. 31, no 3, pp. 276-295, 1960.
[2] E. L. Lehman and J.P. Romano, *Testing statistical hypotheses*, Third Edition. Springer, New York, 2005.
[3] A. K. Bera, "Hypothesis testing in the 20-th century with a special reference to testing with misspecified models", In: "Statistics for the 21-st century. Methodologies for Applications of the Future", Marcel Dekker, Inc., New York, Basel, pp. 33-92, 2000.
[4] A. A. Borovkov, *Mathematical Statistics*, (in Russian), Nauka, Novosibirsk, 1997.
[5] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Part 1. New York, Wiley, 2001.

[6] J. Neyman, E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses", *Phil. Trans. Roy. Soc. London, Ser. A*, vol. 231, pp. 289-337, 1933.

[7] W. Hoeffding, "Asymtotically optimal tests for multinominal distributions", Ann. Math. Statist., vol. 36, no. 2, pp. 369-401, 1965.

[8] G. Tusnady, "On asymptotically optimal tests", Ann. Statist., vol. 5, no. 2, pp. 385 - 393, 1977.

[9] E. A. Haroutunian, "On asymptotically optimal criteria for Markov chains", (in Russian), *First World Congress of Bernoulli Society,* section 2, vol. 2, no. 3, pp. 153-156, 1989.

[10] E. A. Haroutunian, "Logarithmically asymptotically optimal testing of multiple statistical hypotheses", *Problems of Control and Information Theory*, vol. 19, no. 5-6, pp. 413-421, 1990.

[11] E. Haroutunian, M. Haroutunian and A. Harutyunyan, "Reliability criteria in information theory and in statistical hypothesis testing", *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 2-3, 171 P, 2008.

[12] E. A. Haroutunian and P. M. Hakobyan,"On multiple hypotheses testing by informed statistician for arbitrarily varying object and application to source coding", *Transactions of the Institute for Informatics and Automation Problems NAS of RA, Mathematical Problems of Computer Science*, vol. 23, pp. 36-46, 2004.

[13] E. A. Haroutunian and P. M. Hakobyan, "On logarithmically asymptotically optimal testing of three distributions for pair of independent objects", *Transactions of the Institute for Informatics and Automation Problems NAS of RA, Mathematical Problems of Computer Science*, vol. 24, pp. 76-81, 2005.

[14] R. F. Ahlswede and E. A. Haroutunian, "On logarithmically asymptotically optimal testing of hypotheses and identification", in *Lecture Notes in Computer Science*, Vol. 4123, *General Theory of Information Transfer and Combinatorics,* pp. 462-478, Springer Verlag, 2006.

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

[16] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Second Edition, Wiley, New York, 2006.

[17] E.A. Haroutunian, "Neyman-Pearson Principle for more than two hypotheses", Abstracts of Annual session of the Armenian Mathematical Union, Yerevan, pp. 47-48, 2013.

Submitted 12.08.2016, accepted 10.11.2016.

Ֆիքսված թվով անկախ դիտարկումների M-ական վարկածների տեստերի օպտիմալության մասին

Ե. Հարությունյան

Ամփոփում

Հոդվածը նվիրված է ֆիքսված թվով անկախ դիտարկումների նմուշների հիման վրա բազմաթիվ պարզ վարկածների տեստավորման օպտիմալ մոտեցումների կառուցմանը:

Об оптимальности тестов M-арных гипотез при фиксированном числе независимых наблюдений

Е. Арутюнян

Аннотация

Статья посвящена построению оптимальных подходов к тестированию многих простых гипотез при фиксированном числе независимых наблюдений.