Mathematical Problems of Computer Science 46, 103–106, 2016.

On Neyman-Pearson Testing for Pair of Independent Objects

Evgueni A. Haroutunian and Parandzem M. Hakobyan

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

Abstract

The Neyman-Pearson principle for a model consisting of two independent objects is considered. It is supposed that two probability distributions are known and each object follows one of them independently. Two approaches of Neyman-Pearson testing are considered for this model. The aim is to compare error probabilities to the corresponding two cases.

Keywords: Independent objects, Statistical hypotheses, Neyman-Pearson testing, Error probability.

1. Introduction

In this paper we discuss Neyman-Pearson hypotheses testing problem for a model consisting of two independent objects. This model was proposed by Ahlswede and Haroutunian [1]. The characteristics of the objects are X_1 and X_2 independent random variables (RVs) taking values in the same finite set $\mathcal{X}$. So, that considered model is described by the random vector (X_1, X_2), which assumes values $(x^1, x^2) \in \mathcal{X} \times \mathcal{X}$.

The Neyman-Pearson lemma is the foundation of the mathematical theory of statistical hypothesis testing. This principle plays a central role in the theory and practice of statistics. There exist many works where the Neyman-Pearson lemma is expounded for the case of two hypotheses [2]–[10]. This principle to the case of multiple hypotheses is solved in [11, 12].

2. Problem Statement and Result Formulation

Let X_1 and X_2 be independent random variables taking values in the same finite set $\mathcal{X}$, each of them with one of two hypothetical probability distributions $G_m = \{G_m(x),\ x \in \mathcal{X}$, $m = 1, 2$.

Let $(\mathbf{x_1}, \mathbf{x_2}) = ((x_1^1, x_1^2), ..., (x_n^1, x_n^2), ..., (x_N^1, x_N^2)),\ x_n^i \in \mathcal{X},\ i = \overline{1, 2},\ n = \overline{1, N}$, be a sequence of results of N independent observations of the vector (X_1, X_2). It is necessary to define unknown PDs of the objects on the base of the observed data.

The decision for each object must be made from the same set of hypotheses: $H_m : G = G_m,\ m = \overline{1, 2}$. There are four hypothetical probability distributions for random vector (X_1, X_2):

$G_1 \circ G_1(x^1, x^2) = \{G_1(x^1)\dot{G}_1(x^2),\ \ (x^1, x^2) \in \mathcal{X} \times \mathcal{X}\}$,
$G_1 \circ G_2(x^1, x^2) = \{G_1(x^1)\dot{G}_2(x^2),\ \ (x^1, x^2) \in \mathcal{X} \times \mathcal{X}\}$,
$G_2 \circ G_1(x^1, x^2) = \{G_2(x^1)\dot{G}_1(x^2),\ \ (x^1, x^2) \in \mathcal{X} \times \mathcal{X}\}$
and
$G_2 \circ G_2(x^1, x^2) = \{G_2(x^1)\dot{G}_1(x^2),\ \ (x^1, x^2) \in \mathcal{X} \times \mathcal{X}\}$.

We call the procedure of making decision on the base of N observations the test. It can be defined by division of the sample space $\mathcal{X}^N \times \mathcal{X}^N$ on 4 disjoint subsets $\mathcal{B}^N_{i,j}$, $i, j = \overline{1,2}$. The set $\mathcal{B}^N_{i,j}$ consists of all vectors $(\mathbf{x_1}, \mathbf{x_2})$ for which the hypothesis $G_i \circ G_j$ is adopted.

Let $\alpha_{l_1,l_2|m_1,m_2} = G_{m_1} \circ G^N_{m_2}(\mathcal{B}^N_{l_1,l_2})$ be the probability of the erroneous acceptance $G_{l_1} \circ G_{l_2}$ by the test provided that $G_{m_1} \circ G_{m_2}$ is true, where $(l_1, l_2) \neq (m_1, m_2)$, $l_i, m_i = 1, 2$, $i = 1, 2$. The probability to reject a true distribution $G_{m_1} \circ G_{m_2}$ is as follows:

$$\alpha_{m_1,m_2|m_1,m_2} = \sum_{(l_1,l_2)\neq(m_1,m_2)} \alpha_{l_1,l_2|m_1,m_2}.$$

The matrix of error probabilities is as follows:

$$\begin{pmatrix} \alpha_{1,1|1,1} & \alpha_{1,2|1,1} & \alpha_{2,1|1,1} & \alpha_{2,2|1,1} \\ \alpha_{1,1|1,2} & \alpha_{1,2|1,2} & \alpha_{2,1|1,2} & \alpha_{2,2|1,2} \\ \alpha_{1,1|2,1} & \alpha_{1,2|2,1} & \alpha_{2,1|2,1} & \alpha_{2,2|2,1} \\ \alpha_{1,1|2,2} & \alpha_{1,2|2,2} & \alpha_{2,1|2,2} & \alpha_{2,2|2,2} \end{pmatrix}$$

Now we will consider Neyman-Pearson testing for this model with two approaches: a) direct method and b) renumbering method of hypotheses pairs.

Our aim is to compare the corresponding error probabilities of those two approaches.

Let us describe those cases.

a) Direct approach

Let $\alpha^I_{1|1}$ and $\alpha^I_{1|1}$ be error probabilities of the first and the second objects, respectively. According to fundamental Neyman-Pearson lemma, we can choose T^I and T^{II} positive numbers and we can divide the sample spaces of the first and the second objects as follows:

1) For the first object:

$$\mathcal{A}^I_1 = \left\{\mathbf{x_1} : \frac{G^N_1(\mathbf{x_1})}{G^N_2(\mathbf{x_1})} > T^I\right\},\ \ 1 - G^N_1(\mathcal{A}^I_1) = \alpha^I_{1|1},$$

and

$$\mathcal{A}^I_2 = \mathcal{X}^N / \mathcal{A}^I_1.$$

2) For the second object:

$$\mathcal{A}^{II}_1 = \left\{\mathbf{x_2} : \frac{G^N_1(\mathbf{x_2})}{G^N_2(\mathbf{x_2})} > T^{II}\right\},\ \ 1 - G^N_1(\mathcal{A}^{II}_1) = \alpha^{II}_{1|1},$$

and

$$\mathcal{A}^{II}_2 = \mathcal{X}^N / \mathcal{A}^{II}_1.$$

Finally, the Neyman-Pearson division of the sample space $\mathcal{X}^N \times \mathcal{X}^N$ is the following:

$$A_{1,1} = \mathcal{A}^I_1 \times \mathcal{A}^{II}_1,\ \ A_{1,2} = \mathcal{A}^I_1 \times \mathcal{A}^{II}_2,\ \ A_{2,1} = \mathcal{A}^I_2 \times \mathcal{A}^{II}_1,\ \ A_{2,2} = \mathcal{A}^I_2 \times \mathcal{A}^{II}_2.$$

According to the definition of error probabilities and independents of objects we can see that the error probabilities of this test are formulated as follows:

$$\alpha_{l_1,l_2|m_1,m_2} = G^N_{m_1} \circ G^N_{m_2}(\mathcal{A}^N_{l_1,l_2}) = \sum_{\mathbf{x_1}\in\mathcal{A}^N_{l_1}} G_{m_1}(\mathbf{x_1}) \times \sum_{\mathbf{x_2}\in\mathcal{A}^N_{l_2}} G_{m_2}(\mathbf{x_2}).$$

b) Renumbering approach

Here we have to renumber the each pair of hypotheses one at a time and we have to apply Neyman-Pearson lemma for multiple hypotheses, which is investigated in [12].

In that case for the given positive values $\alpha^*_{1,1|1,1}$, $\alpha^*_{1,2|1,2}$ and $\alpha^*_{2,1|2,1}$ and chosen numbers T_1, T_2 and T_3 sets $\mathcal{A}_{i,j}$, $i,j=\overline{1,2}$, are the following:

$$\mathcal{A}_{1,1} = \left\{(\mathbf{x_1},\mathbf{x_2}) : \min\left(\frac{G^N_1(\mathbf{x_1})}{G^N_2(\mathbf{x_2})}, \frac{G^N_1(\mathbf{x_2})}{G^N_2(\mathbf{x_2})}, \frac{G^N_1(\mathbf{x_1})\cdot G^N_1(\mathbf{x_2})}{G^N_2(\mathbf{x_1})\cdot G^N_2(\mathbf{x_2})}\right) > T_1\right\}, \ 1-G^N_1\circ G^N_1(\mathcal{A}_{1,1}) = \alpha^*_{1,1|1,1},$$

$$\mathcal{A}_{1,2} = \overline{\mathcal{A}_{1,1}}\left\{(\mathbf{x_1},\mathbf{x_2}) : \min\left(\frac{G^N_1(\mathbf{x_1})}{G^N_2(\mathbf{x_2})}, \frac{G^N_1(\mathbf{x_1})\cdot G^N_2(\mathbf{x_2})}{G^N_2(\mathbf{x_1})\cdot G^N_1(\mathbf{x_2})}\right) > T_2\right\}, \ 1-G^N_1\circ G^N_2(\mathcal{A}_{1,2}) = \alpha^*_{1,2|1,2}$$

$$\mathcal{A}_{2,1} = \overline{\mathcal{A}_{1,1}}\cap\overline{\mathcal{A}_{1,2}}\left\{(\mathbf{x_1},\mathbf{x_2}) : \frac{G^N_1(\mathbf{x_2})}{G^N_2(\mathbf{x_2})} > T_3\right\}, \ 1-G^N_2\circ G^N_1(\mathcal{A}_{2,1}) = \alpha^*_{2,1|2,1}$$

and

$$\mathcal{A}^*_{2,2} = \overline{\mathcal{A}_{1,1}}\cap\overline{\mathcal{A}_{1,2}}\cap\overline{\mathcal{A}_{2,1}}.$$

The corresponding error probabilities are formulated as above.

Let us assume that we have found the error probabilities by direct approach. By considering the following $\alpha_{1,1|1,1}$, $\alpha_{1,2|1,2}$ and $\alpha_{2,1|2,1}$ error probabilities and applying the renumbering approach we will find all the other error probabilities. Because in [12] it is proved that these error probabilities are the smallest. Hence, we can insist that the renumbering approach of Neyman-Pearson testing is the best method.

References

[1] R. F. Ahlswede and E. A. Haroutunian, "Testing of hypotheses and identification", *Electronic Notes on Discrete Mathematics, vol. 21*, pp. 185–189, 2005.

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

[3] J. Neyman, *First Course in Probability and Statistics*, Holt, Rinehart and Winston, New York, 1950.

[4] E. L. Lehman and J.P. Romano, *Testing statistical hypotheses*, Third Edition. Springer, New York, 2005.

[5] A. A. Borovkov, *Mathematical Statistics*, in Russian, Nauka, Moscow, 1997.

[6] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Part 1. New York: Wiley, 1968.

[7] M. H. DeGroot, *Probability and Statistics*, 2nd ed., Reading, MA, Addison-Wesley, 1986.

[8] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 2, Inference and relationship, Third edition. Hafner publishing company, London, 1961.

[9] 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 Fitire”, Marcel Dekker, Inc., New York, Basel, pp. 33-92, 2000.

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

[11] E. A. Haroutunian, “Neyman-Pearson principle for more than two hypotheses”, *Abstracts of Armenian Mathematical Union Annual Session Dedicated to 90 Anniversary of Rafael Alexandrian*, Yerevan, pp.49–50, 2013.

[12] E. A. Haroutunian and P. Hakobyan, “Neyman-Pearson principle for more than two hypotheses”, *Mathematical Problems of Computer Science* vol. 40, pp. 34–38, 2013.

Submitted 24.08.2016, accepted 07.11.2016.

Երկու անկախ օբյեկտների նկատմամբ Նեյմանի-Պիսոնի տեստավորման մասին

Ե. Հարությունյան և Փ. Հակոբյան

Ամփոփում

Դիտրկվել է երկու անկախ օբյեկտներով բնութագրվող մոդելի վերաբերյալ Նեյմանի-Պիրսոնի սկզբունքը: Ենթադրվում է, որ հայտնի են երկու հավանականային բաշխումներ և օբյեկտներից յուրաքանչյուրը մյուսից անկախ բաշխված է դրանցից որևէ մեկով: Այս մոդելի նկատմամբ դիտարկվել է Նեյմանի-Պիսոնի տեստավորման երկու մոտեցում: Նպատակն է համեմատել այս երկու մոտեցումների համապատասխան սխալների հավանականությունները:

О тестировании Неймана-Пирсона для пары независимых объектов

Е. Арутюнян и П. Акопян

Аннотация

Рассматривается принцип Неймана-Пирсона для модели состоящей из двух независимых объектов. Предполагается, что два вероятностных распределения известны, и каждый объект следует одному из них независимо от другого. Рассматривается два подхода тестирования Неймана-Пирсона для данной модели. Цель состоит в том, чтобы сравнить вероятности ошибки в двух соответствующих случаях.