

On Logarithmically Asymptotically Optimal Testing of Hypotheses and Identification



We introduce a new aspect of the influence of the information-theoretical methods on the statistical theory. The procedures of the probability distributions identification for $K(\geq 1)$ random objects each having one from the known set of $M(\geq 2)$ distributions are studied. N -sequences of discrete independent RV's represent results of N observations for each of K objects. On the base of such samples decisions must be made concerning probability distributions of the objects. For $N \rightarrow \infty$ the exponential decrease of the test's error probabilities is considered. The reliability matrices of logarithmically asymptotically optimal procedures are investigated for some models and formulations of the identification problems. The optimal subsets of reliabilities which values may be given beforehand and conditions guaranteeing positiveness of all the reliabilities are investigated.

“In statistical literature such a problem is referred to as one of classification or discrimination, but identification seems to be more appropriate”

Radhakrishna Rao [27].

1 Problem Statement

Let $\mathbf{X}_k = (X_{k,n}, n \in [N])$, $k \in [K]$, be $K(\geq 1)$ sequences of N discrete i.i.d. RV's representing possible results of N observations, respectively, for each of K randomly functioning objects.

For $k \in [K]$, $n \in [N]$, $X_{k,n}$ assumes values $x_{k,n}$ in the finite set \mathcal{X} of cardinality $|\mathcal{X}|$. Let $\mathcal{P}(\mathcal{X})$ be the space of all possible distributions on \mathcal{X} . There are $M(\geq 2)$ probability distributions G_1, \dots, G_M from $\mathcal{P}(\mathcal{X})$ in inspection, some of which are assigned to the vectors $\mathbf{X}_1, \dots, \mathbf{X}_K$. This assignment is unknown and must be determined on the base of N -samples (results of N independent observations) $\mathbf{x}_k = (x_{k,1}, \dots, x_{k,N})$, where $x_{k,n}$ is a result of the n th observation of the k th object.

When $M = K$ and all objects are different (any two objects cannot have the same distribution), there are $K!$ possible decisions. When objects are independent, there are M^K possible combinations.

Bechhofer, Kiefer, and Sobel presented investigations on sequential multiple-decision procedures in [7]. This book is concerned principally with a particular class of problems referred to as ranking problems.

Chapter “Models with Prior Knowledge at the Sender” of the book by Ahlswede and Wegener [5] is devoted to statistical identification and ranking problems.

We study models considered in [7] and [5] and variations of these models inspired by the pioneering papers by Ahlswede and Dueck [4] (see chapter “Identification via Channels” in Part I) and by Ahlswede [1], applying the concept of optimality developed in [9, 16, 22–24, 28] for the models with $K = 1$.

Consider the following family of error probabilities of a test

$$\alpha_{m_1, m_2, \dots, m_K | l_1, l_2, \dots, l_K}^{(N)}, \quad (m_1, m_2, \dots, m_K) \neq (l_1, l_2, \dots, l_K), \quad m_k, l_k \in [M], \quad k \in [K],$$

which are the probabilities of decisions l_1, l_2, \dots, l_K when actual indices of the distributions of the objects were, respectively, m_1, m_2, \dots, m_K .

The probabilities to reject all K hypotheses when they are true are the following

$$\alpha_{m_1, m_2, \dots, m_K | m_1, m_2, \dots, m_K}^{(N)} = \sum_{(l_1, l_2, \dots, l_K) \neq (m_1, m_2, \dots, m_K)} \alpha_{m_1, m_2, \dots, m_K | l_1, l_2, \dots, l_K}^{(N)}.$$

We study exponential decrease of the error probabilities when $N \rightarrow \infty$ and define (using logarithms and exponents to the base e)

$$\limsup_{N \rightarrow \infty} -\frac{1}{N} \log \alpha_{m_1, m_2, \dots, m_K | l_1, l_2, \dots, l_K}^{(N)} = E_{m_1, m_2, \dots, m_K | l_1, l_2, \dots, l_K} \geq 0. \quad (1)$$

These are exponents of error probabilities which we call reliabilities (in association with Shannon’s reliability function [15]). We shall examine the matrix $\mathbf{E} = \{E_{m_1, m_2, \dots, m_K | l_1, l_2, \dots, l_K}\}$ and call it the reliability matrix.

Our criterion of optimality is: given M, K and values of a part of reliabilities to obtain the best (the largest) values for others. In addition it is necessary to describe the conditions under which all these reliabilities are positive. The procedure that realizes such testing is identification, which following Birgé [9], we call “logarithmically asymptotically optimal” (LAO).

Let $N(x|\mathbf{x})$ be the number of repetitions of the element $x \in \mathcal{X}$ in the vector $\mathbf{x} \in \mathcal{X}^N$, and let

$$Q = \{Q(x) = N(x|\mathbf{x})/N, \quad x \in \mathcal{X}\}$$

is the distribution, called “the empirical distribution” (ED) of the sample \mathbf{x} in statistics, in information theory called “the type” [14, 15] and in algebraic literature “the composition”.

Denote the space of all empirical distributions for given N by $\mathcal{P}^{(N)}(\mathcal{X})$ and by $\mathcal{T}_Q^{(N)}$ the set of all vectors of the ED $Q \in \mathcal{P}^{(N)}(\mathcal{X})$.

Consider for $k \in [K]$, $m \in [M]$, relative entropies

$$D(Q_k || G_m) = \sum_{x \in \mathcal{X}} Q_k(x) \log \frac{Q_k(x)}{G_m(x)},$$

and entropies

$$H(Q_k) = - \sum_{x \in \mathcal{X}} Q_k(x) \log Q_k(x).$$

We shall use the following relations for the probability of the vector \mathbf{x} when G_m is the distribution of the object:

$$G_m^{(N)}(\mathbf{x}) = \prod_{n=1}^N G_m(x_n) = \exp\{-N[D(Q || G_m) + H(Q)]\}.$$

For $m_k \in [M]$, $k \in [K]$, when the objects are independent and G_{m_k} is the distribution of the k th object:

$$P_{m_1, m_2, \dots, m_K}^{(N)}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K) = \exp\{-N[\sum_{k=1}^K D(Q_k || G_{m_k}) + H(Q_k)]\}. \quad (2)$$

The equalities follow from the independence of N observations of K objects and from the definitions of relative entropies and entropies. It should be noted that the equality (2) is valid even when its left part is equal to 0, in that case for one of \mathbf{x}_k the distribution Q_k is not absolutely continuous relative to G_{m_k} and $D(Q_k || G_{m_k}) = \infty$.

Our arguments will be based on the following fact: the “maximal likelihood” test accepts as the solution values m_1, m_2, \dots, m_K , which maximize the probability $P_{m_1, m_2, \dots, m_K}^{(N)}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K)$, but from (2) we see that the same solution can be obtained by minimization of the sum $\sum_{k=1}^K [D(Q_k || G_{m_k}) + H(Q_k)]$, that is the comparison with the help of relative entropies of the ED’s of observed vectors with their hypothetical distributions may be helpful.

In this lecture we consider the following models.

1. K objects are different, they have different distributions among $M \geq K$ possibilities. For simplicity we restrict ourselves to the case $K = 2, M = 2$. It is the identification problem in formulations of the books [7] and [5].
2. K objects are independent, that is some of them may have the same distributions. We consider an example for $K, M = 2$. It is surprising, but this model has not been considered earlier in the literature.

3. We investigate one object, $K = 1$, and M possible probability distributions. The question is whether the m th distribution occurred or not. This is the problem of identification of distributions in the spirit of chapter “Identification via Channels”.
4. Ranking, or ordering problem [1]. We have one vector of observations $\mathbf{X} = (X_1, X_2, \dots, X_N)$ and M hypothetical distributions. The receiver wants to know whether the index of the true distribution of the object is in $\{1, 2, \dots, r\}$ or in $\{r + 1, \dots, M\}$.
5. r -identification of distribution [1]. Again $K = 1$. One wants to identify the observed object as a member either of the subset \mathcal{S} of $[M]$, or of its complement, with r being the number of elements in \mathcal{S} .

Section 2 of this lecture presents necessary notions and results on hypothesis testing. The models of identification for independent objects are considered in 3 and for different objects in 4. Section 5 is devoted to the problem of identification of an object distribution and 6 to the problems of r -identification and ranking. Some results are illustrated by numerical examples and graphs. Many directions of further research are indicated in the course of the text and in the 7.

2 Background

The study of interdependence of exponential rates of decrease, as the sample size N goes to the infinity, of the error probabilities $\alpha_{1|2}^{(N)}$ of the “first kind” and $\alpha_{2|1}^{(N)}$ of the “second kind” was started by the works of Hoeffding [23], Csiszár and Longo [16], Tusnády [28], Longo and Sgarro [24], Birgé [9], and for multiple hypotheses by Haroutunian [22]. Similar problems for Markov dependence of experiments were investigated by Natarajan [26], Haroutunian [21], Gutman [18] and others. As it was remarked by Blahut in his book [11], it is unfortunately confusing that the errors are denoted type I and type II, while the hypotheses are subscripted 0 and 1. The word “type” is also used in another sense to refer to the type of a measurement or the type of a vector. For this reason we do not use the names “0” and “1” for hypotheses and the name “type” for errors. Note that in [10, 11, 17] an application of the methods of hypothesis testing to the proper problems of information theory is developed.

It will be very interesting to combine investigation of described models with the approach initiated by the paper of Ahlswede and Csiszár [3] and developed by many authors, particularly, for the exponentially decreasing error probabilities by Han and Kobayashi [20].

In [8] Berger formulated the problem of remote statistical inference. Zhang and Berger [29] studied a model of an estimation system with compressed information. Similar problems were examined by Ahlswede and Burnashev [2] and by Han and Amari, S. [19]. In the paper of Ahlswede, Yang and Zhang [6] identification in channels via compressed data was considered. Fu and Shen [17] studied hypothesis testing for an arbitrarily varying source.

Our further considerations will be based on the results from [22] on multiple hypotheses testing, so now we expose briefly corresponding formulations and proofs. In our terms it is the case of one object ($K = 1$) and M possible distributions (hypotheses) G_1, \dots, G_M . A test $\varphi(\mathbf{x})$ on the base of N -sample $\mathbf{x} = (x_1, \dots, x_N)$ determines the distribution. Since experiments are independent the probability of the sample \mathbf{x} if the distribution is G_m will be

$$G_m^{(N)}(\mathbf{x}) = \prod_{n=1}^N G_m(x_n), \quad m \in [M].$$

We study error probabilities $\alpha_{m|l}^{(N)}$ for $m, l \in [M]$. Here $\alpha_{m|l}^{(N)}$ is the probability that the distribution G_l was accepted instead of true distribution G_m . For $m = l$ the probability to reject G_m when it is true, is denoted by $\alpha_{m|m}^{(N)}$ thus:

$$\alpha_{m|m}^{(N)} = \sum_{l:l \neq m} \alpha_{m|l}^{(N)}.$$

This probability is called [12] the test's "error probability of the kind m ". The matrix $\{\alpha_{m|l}^{(N)}\}$ is sometimes called the "power of the test" [12].

In this lecture we suppose that the list of possible hypotheses is complete. Remark that, as it was noted by Rao [27], the case, when the objects may have also some distributions different from G_1, \dots, G_M , is interesting too.

Let us analyze the reliability matrix

$$\mathbf{E} = \begin{pmatrix} E_{1|1} & \dots & E_{1|l} & \dots & E_{1|M} \\ \dots & \dots & \dots & \dots & \dots \\ E_{m|1} & \dots & E_{m|l} & \dots & E_{m|M} \\ \dots & \dots & \dots & \dots & \dots \\ E_{M|1} & \dots & E_{M|l} & \dots & E_{M|M} \end{pmatrix}$$

with components

$$E_{m|l} = \limsup_{N \rightarrow \infty} -\frac{1}{N} \log \alpha_{m|l}^{(N)}, \quad m, l \in [M].$$

According to this definition and the definition of $\alpha_{m|l}^{(N)}$ we can derive that

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

Really,

$$\begin{aligned}
 E_{m|m} &= \limsup_{N \rightarrow \infty} -\frac{1}{N} \log \sum_{l:m \neq l} \alpha_{m|l}^{(N)} \\
 &= \limsup_{N \rightarrow \infty} -\frac{1}{N} \log \max_{l:m \neq l} \alpha_{m|l}^{(N)} + \limsup_{N \rightarrow \infty} -\frac{1}{N} \log \left[\left(\sum_{l:m \neq l} \alpha_{m|l}^{(N)} \right) / \max_{l:m \neq l} \alpha_{m|l}^{(N)} \right] \\
 &= \min_{l:m \neq l} E_{m|l}.
 \end{aligned}$$

The last equality is a consequence of the fact that for all m and N

$$1 \leq \left(\sum_{l:m \neq l} \alpha_{m|l}^{(N)} \right) / \max_{l:m \neq l} \alpha_{m|l}^{(N)} \leq M - 1.$$

In the case $M = 2$, the reliability matrix is

$$\mathbf{E} = \begin{pmatrix} E_{1|1} & E_{1|2} \\ E_{2|1} & E_{2|2} \end{pmatrix} \quad (4)$$

and it follows from (3) that there are only two different values of elements, namely

$$E_{1|1} = E_{1|2} \text{ and } E_{2|1} = E_{2|2}, \quad (5)$$

so in this case the problem is to find the maximal possible value of one of them, given the value of the other.

In the case of M hypotheses for given positive and finite $E_{1|1}, E_{2|2}, \dots, E_{M-1, M-1}$ let us consider the regions of distributions

$$\mathcal{R}_l = \{Q : D(Q||G_l) \leq E_{l|l}\}, \quad l \in [M - 1], \quad (6)$$

$$\mathcal{R}_M = \{Q : D(Q||G_l) > E_{l|l}, l \in [M - 1]\} = \mathcal{P}(\mathcal{X}) - \bigcup_{l=1}^{M-1} \mathcal{R}_l, \quad (7)$$

$$\mathcal{R}_l^{(N)} = \mathcal{R}_l \cap \mathcal{P}^{(N)}, \quad l \in [M]. \quad (8)$$

Let

$$E_{l|l}^* = E_{l|l}^*(E_{l|l}) = E_{l|l}, \quad l \in [M - 1], \quad (9)$$

$$E_{m|l}^* = E_{m|l}^*(E_{l|l}) = \inf_{Q \in \mathcal{R}_l} D(Q||G_m), \quad m \in [M], \quad m \neq l, \quad l \in [M - 1], \quad (10)$$

$$E_{m|M}^* = E_{m|M}^*(E_{1|1}, \dots, E_{M-1, M-1}) = \inf_{Q \in \mathcal{R}_M} D(Q||G_m), \quad m \in [M-1], \quad (11)$$

$$E_{M|M}^* = E_{M|M}^*(E_{1|1}, \dots, E_{M-1, M-1}) = \min_{l \in [M-1]} E_{M|l}^*. \quad (12)$$

If some distribution G_m is not absolutely continuous relative to G_l the reliability $E_{m|l}^*$ will be equal to the infinity, this means that corresponding $\alpha_{m|l}^{(N)} = 0$ for some large N .

The principal result of [22] is:

Theorem 272 *If all the distributions G_m are different and all elements of the matrix $\{D(G_l||G_m)\}$, $l, m \in [M]$, are positive, but finite, two statements hold:*

(i) *when the positive numbers $E_{1|1}, E_{2|2}, \dots, E_{M-1, M-1}$ satisfy conditions*

$$\begin{aligned} E_{1|1} &< \min_{l \in [2, M]} D(G_l||G_1), \\ &\vdots \end{aligned} \quad (13)$$

$$E_{m|m} < \min_{l \in [m-1]} E_{m|l}^*(E_{l|l}), \quad \min_{l \in [m+1, M]} D(G_l||G_m), \quad m \in [2, M-1],$$

then there exists a LAO sequence of tests, the reliability matrix of which $E^ = \{E_{m|l}^*\}$ is defined in (9)–(12) and all elements of it are positive;*

(ii) *even if one of conditions (13) is violated, then the reliability matrix of any such test has at least one element equal to zero (that is the corresponding error probability does not tend to zero exponentially).*

The essence of the proof of Theorem 272 consists in construction of the following optimal tests sequence. Let the decision l will be taken when \mathbf{x} gets into the set

$$\mathcal{B}_l^{(N)} = \bigcup_{Q \in \mathcal{R}_l^{(N)}} \mathcal{T}_Q^{(N)}, \quad l \in [M], \quad N = 1, 2, \dots \quad (14)$$

The non-coincidence of the distributions G_m and the conditions (13) guarantee that the sets from (14) are not empty, they meet conditions

$$\mathcal{B}_l^{(N)} \cap \mathcal{B}_m^{(N)} = \emptyset, \quad l \neq m,$$

and

$$\bigcup_{l=1}^M \mathcal{B}_l^{(N)} = \mathcal{X}^N,$$

and so they define a sequence of tests, which proves to be LAO.

For the simplest particular case $M = 2$ elements of the reliability matrix (4) satisfy equalities (5) and for given $E_{1|1}$ from (5) and (7) we obtain the value of $E_{2|1}^* = E_{2|2}^*$:

$$E_{2|1}^*(E_{1|1}) = \inf_{Q: D(Q||G_1) \leq E_{1|1}} D(Q||G_2). \tag{15}$$

Here, according to (13), we can take $E_{1|1}$ from $(0, D(G_2||G_1))$ and $E_{2|1}^*(E_{1|1})$ will range between $D(G_1||G_2)$ and 0.

3 Identification Problem for Model with Independent Objects

We begin with study of the second model. To illustrate possibly arising developments and essential features we consider a particular case $K = 2, M = 2$. It is clear that the case with $M = 1$ is trivial. The reliability matrix is (see (1))

$$\mathbf{E} = \begin{pmatrix} E_{1,1|1,1} & E_{1,1|1,2} & E_{1,1|2,1} & E_{1,1|2,2} \\ E_{1,2|1,1} & E_{1,2|1,2} & E_{1,2|2,1} & E_{1,2|2,2} \\ E_{2,1|1,1} & E_{2,1|1,2} & E_{2,1|2,1} & E_{2,1|2,2} \\ E_{2,2|1,1} & E_{2,2|1,2} & E_{2,2|2,1} & E_{2,2|2,2} \end{pmatrix}.$$

Let us denote by $\alpha_{m_1|l_1}^{(1)}, \alpha_{m_2|l_2}^{(2)}$ and $E_{m_1|l_1}^{(1)}, E_{m_2|l_2}^{(2)}$ the error probabilities and the reliabilities as in (4) for, respectively, the first and the second objects.

Lemma 273 *If $0 < E_{1|1}^{(i)} < D(G_2||G_1), i = 1, 2$, then the following equalities hold true:*

$$E_{m_1, m_2|l_1, l_2} = E_{m_1|l_1}^{(1)} + E_{m_2|l_2}^{(2)}, \quad \text{if } m_1 \neq l_1, m_2 \neq l_2, \tag{16}$$

$$E_{m_1, m_2|l_1, l_2} = E_{m_i|l_i}^{(i)}, \quad \text{if } m_{3-i} = l_{3-i}, m_i \neq l_i, i = 1, 2, \tag{17}$$

Proof From the independence of the objects it follows that

$$\alpha_{m_1, m_2|l_1, l_2}^{(N)} = \alpha_{m_1|l_1}^{(N,1)} \alpha_{m_2|l_2}^{(N,2)}, \quad \text{if } m_1 \neq l_1, m_2 \neq l_2, \tag{18}$$

$$\alpha_{m_1, m_2|l_1, l_2}^{(N)} = \alpha_{m_i|l_i}^{(N,i)} (1 - \alpha_{m_{3-i}|l_{3-i}}^{(N,3-i)}), \quad \text{if } m_{3-i} = l_{3-i}, m_i \neq l_i, i = 1, 2, \tag{19}$$

According to (1), from (18) we obtain (16), from (19) and the conditions of positiveness of $E_{1|1}^{(i)}$ and $E_{2|2}^{(i)}, i = 1, 2$, (17) follows. □

Theorem 274 *If the distributions G_1 and G_2 are different, the strictly positive elements $E_{1,1|1,2}$, $E_{1,1|2,1}$ of the reliability matrix \mathbf{E} are given and bounded above:*

$$E_{1,1|1,2} < D(G_2||G_1), \text{ and } E_{1,1|2,1} < D(G_2||G_1), \quad (20)$$

then the other elements of the matrix \mathbf{E} are defined as follows:

$$\begin{aligned} E_{2,1|2,2} &= E_{1,1|1,2}, & E_{1,2|2,2} &= E_{1,1|2,1}, \\ E_{2,2|1,1} &= E_{1,2|1,1} + E_{2,1|1,1}, & E_{2,1|1,2} &= E_{2,1|1,1} + E_{1,2|2,2}, \\ E_{1,2|2,1} &= E_{1,2|1,1} + E_{1,2|2,2}, & E_{1,1|2,2} &= E_{1,1|1,2} + E_{1,1|2,1}, \\ \\ E_{1,2|1,1} &= E_{2,2|2,1} = \inf_{Q:D(Q||G_1) \leq E_{1,1|1,2}} D(Q||G_2), \\ E_{2,1|1,1} &= E_{2,2|1,2} = \inf_{Q:D(Q||G_1) \leq E_{1,1|2,1}} D(Q||G_2), & (21) \\ E_{m_1, m_2 | m_1, m_2} &= \min_{(l_1, l_2) \neq (m_1, m_2)} E_{m_1, m_2 | l_1, l_2}, \quad m_1, m_2 = 1, 2. \end{aligned}$$

If one of the inequalities (20) is violated, then at least one element of the matrix \mathbf{E} is equal to 0.

Proof The last equalities in (21) follow (as (3)) from the definition of

$$\alpha_{m_1, m_2 | m_1, m_2}^{(N)} = \sum_{(l_1, l_2) \neq (m_1, m_2)} \alpha_{m_1, m_2 | l_1, l_2}^{(N)}, \quad m_1, m_2 = 1, 2.$$

Let us consider the reliability matrices of each of the objects X_1 and X_2

$$\mathbf{E}^{(1)} = \begin{pmatrix} E_{1|1}^{(1)} & E_{1|2}^{(1)} \\ E_{2|1}^{(1)} & E_{2|2}^{(1)} \end{pmatrix} \quad \text{and} \quad \mathbf{E}^{(2)} = \begin{pmatrix} E_{1|1}^{(2)} & E_{1|2}^{(2)} \\ E_{2|1}^{(2)} & E_{2|2}^{(2)} \end{pmatrix}.$$

From (5) we know that $E_{1|1}^{(i)} = E_{1|2}^{(i)}$ and $E_{2|1}^{(i)} = E_{2|2}^{(i)}$, $i = 1, 2$. From (20) it follows that $0 < E_{1|1}^{(1)} < D(G_2||G_1)$, $0 < E_{1|1}^{(2)} < D(G_2||G_1)$. Really, if $0 < E_{1,1|1,2} < D(G_2||G_1)$, but $E_{1|1}^{(2)} \geq D(G_2||G_1)$, then from (19) and (1) we arrive to

$$\limsup_{N \rightarrow \infty} -\frac{1}{N} \log(1 - \alpha_{1|2}^{(N,2)}) < 0,$$

therefore index N_0 exists, such that for sub-sequence of $N > N_0$ we will have $1 - \alpha_{1|2}^{(N,2)} > 1$. But this is impossible because $\alpha_{1|2}^{(N,2)}$ is the probability and must be positive.

Using Lemma 273 we can deduce that the reliability matrix \mathbf{E} can be obtained from matrices $\mathbf{E}^{(1)}$ and $\mathbf{E}^{(2)}$ as follows:

$$\mathbf{E} = \begin{pmatrix} \min(E_{1|2}^{(1)}, E_{1|2}^{(2)}) & E_{1|2}^{(2)} & E_{1|2}^{(1)} & E_{1|2}^{(1)} + E_{1|2}^{(2)} \\ E_{2|1}^{(2)} & \min(E_{1|2}^{(1)}, E_{2|1}^{(2)}) & E_{1|2}^{(1)} + E_{2|1}^{(2)} & E_{1|2}^{(1)} \\ E_{2|1}^{(1)} & E_{2|1}^{(1)} + E_{1|2}^{(2)} & \min(E_{2|1}^{(1)}, E_{1|2}^{(2)}) & E_{1|2}^{(2)} \\ E_{2|1}^{(1)} + E_{2|1}^{(2)} & E_{2|1}^{(1)} & E_{2|1}^{(2)} & \min(E_{2|1}^{(1)}, E_{2|1}^{(2)}) \end{pmatrix},$$

in other words, providing, that conditions (20) are fulfilled, we find that

$$\begin{aligned} E_{1,1|1,2} &= E_{1|2}^{(2)} = E_{1|1}^{(2)}, & E_{1,1|2,1} &= E_{1|2}^{(1)} = E_{1|1}^{(1)}, \\ E_{2,1|2,2} &= E_{1,1|1,2} = E_{1|2}^{(2)}, & E_{1,2|2,2} &= E_{1,1|2,1} = E_{1|2}^{(1)}, \\ E_{1,2|1,1} &= E_{2,2|2,1} = E_{2|1}^{(2)}, & E_{2,1|1,1} &= E_{2,2|1,2} = E_{2|1}^{(1)}, \\ E_{2,2|1,1} &= E_{2|1}^{(1)} + E_{2|1}^{(2)}, & E_{2,1|1,2} &= E_{2|1}^{(1)} + E_{1|2}^{(2)}, \\ E_{1,2|2,1} &= E_{1|2}^{(1)} + E_{2|1}^{(2)}, & E_{1,1|2,2} &= E_{1|2}^{(1)} + E_{1|2}^{(2)}, \end{aligned} \tag{22}$$

and

$$E_{m_1, m_2 | m_1, m_2} = \min\{E_{m_1 | m_1}^{(1)}, E_{m_2 | m_2}^{(2)}\}, \quad m_1, m_2 = 1, 2,$$

From Theorem 272 we know that if $E_{1|1}^{(i)} \in (0, D(G_2 || G_1))$, $i = 1, 2$, then the tests of both objects are LAO and the elements $E_{2|1}^{(i)}$, $i = 1, 2$, can be calculated (see (15)) by

$$E_{2|1}^{(i)} = \inf_{Q: D(Q || G_1) \leq E_{1|1}^{(i)}} D(Q || G_2), \quad i = 1, 2, \tag{23}$$

and if $E_{1|1}^{(i)} \geq D(G_2 || G_1)$, then $E_{2|1}^{(i)} = 0$.

According to (22) and (23), we obtain, that when (20) takes place, the elements of the matrix \mathbf{E} are determined by relations (21). When one of the inequalities (20) is violated, then from (23) and the first and the third lines of (22) we see, that some elements in the matrix \mathbf{E} must be equal to 0 (namely, either $E_{1,2|1,1}$, or $E_{2,1|1,1}$ and others).

Now let us show that the compound test for two objects is LAO, that is it is optimal. Suppose that for given $E_{1,1|1,2}$ and $E_{1,1|2,1}$ there exists a test with matrix \mathbf{E}' , such that it has at least one element exceeding the respective element of the matrix \mathbf{E} . Comparing elements of matrices \mathbf{E} and \mathbf{E}' different from $E_{1,1|1,2}$ and $E_{1,1|2,1}$, from (22) we obtain that either $E_{1,2|1,1} < E'_{1,2|1,1}$, or $E_{2,1|1,1} < E'_{2,1|1,1}$,

i.e. either $E_{2|1}^{(2)} < E_{2|1}^{(2)'}$, or $E_{2|1}^{(1)} < E_{2|1}^{(1)'}$. It is contradiction to the fact, that LAO tests have been used for the objects X_1 and X_2 .

When it is demanded to take the same values for the reliabilities of the first and the second objects $E_{1|2}^{(1)} = E_{1|2}^{(2)} = a_1$ and, consequently, $E_{2|1}^{(1)} = E_{2|1}^{(2)} = a_2$, then the matrix \mathbf{E} will take the following form

$$\mathbf{E} = \begin{pmatrix} a_1 & a_1 & a_1 & 2a_1 \\ a_2 & \min(a_1, a_2) & a_1 + a_2 & a_1 \\ a_2 & a_1 + a_2 & \min(a_1, a_2) & a_1 \\ 2a_2 & a_2 & a_2 & a_2 \end{pmatrix}. \quad \square$$

4 Identification Problem for Models with Different Objects

The K objects are not independent, they have different distributions, and so the number M of the distributions is not less than K . This is the model studied in [7]. For brevity we consider the case $K = 2, M = 2$. The matrix of reliabilities will be the following:

$$\mathbf{E} = \begin{pmatrix} E_{1,2|1,2} & E_{1,2|2,1} \\ E_{2,1|1,2} & E_{2,1|2,1} \end{pmatrix}. \quad (24)$$

Since the objects are strictly dependent this matrix coincides with the reliability matrix of the first object (see (4))

$$\mathbf{E}^{(1)} = \begin{pmatrix} E_{1|1}^{(1)} & E_{1|2}^{(1)} \\ E_{2|1}^{(1)} & E_{2|2}^{(1)} \end{pmatrix},$$

because the distribution of the second object is uniquely defined by the distribution of the first one.

We can conclude that among 4 elements of the reliability matrix of two dependent objects only 2 elements are distinct, the second of which is defined by given $E_{1|1}^{(1)} = E_{1,2|1,2}$.

From symmetry it follows that the reliability matrix of the second object also may determine the matrix (24).

5 Identification of the Probability Distribution of an Object

Let we have one object, $K = 1$, and there are known $M \geq 2$ possible distributions. The question is whether r th distribution occurred, or not. There are two error probabilities for each $r \in [M]$ the probability $\alpha_{m=r|l \neq r}^{(N)}$ to accept l different from r ,

when r is in reality, and the probability $\alpha_{m \neq r | l=r}^{(N)}$ that r is accepted, when it is not correct.

The probability $\alpha_{m=r | l \neq r}^{(N)}$ is already known, it coincides with the probability $\alpha_{r|r}^{(N)}$ which is equal to $\sum_{l:l \neq r} \alpha_{r|l}^{(N)}$. The corresponding reliability $E_{m=r | l \neq r}$ is equal to $E_{r|r}$ which satisfies the equality (3).

We have to determine the dependence of $E_{m \neq r | l=r}$ upon given $E_{m=r | l \neq r} = E_{r|r}$, which can be assigned values satisfying conditions (13), this time we will have the conditions:

$$0 < E_{r|r} < \min_{l:l \neq r} D(G_l \| G_r), \quad r \in [M].$$

We need the probabilities of different hypotheses. Let us suppose that the hypotheses G_1, \dots, G_M have, say, probabilities $\Pr(r)$, $r \in [M]$. The only supposition we shall use is that $\Pr(r) > 0$, $r \in [M]$. We will see, that the result formulated in the following theorem does not depend on values of $\Pr(r)$, $r \in [M]$, if they all are strictly positive.

Now we can make the following reasoning for each $r \in [M]$:

$$\alpha_{m \neq r | l=r}^{(N)} = \frac{\Pr^{(N)}(m \neq r, l=r)}{\Pr(m \neq r)} = \frac{1}{\sum_{m:m \neq r} \Pr(m)} \sum_{m:m \neq r} \Pr^{(N)}(m, r).$$

From here we see that for $r \in [M]$

$$\begin{aligned} E_{m \neq r | l=r} &= \limsup_{N \rightarrow \infty} \left(-\frac{1}{N} \log \alpha_{m \neq r | l=r}^{(N)} \right) \\ &= \limsup_{N \rightarrow \infty} \frac{1}{N} \left(\log \sum_{m:m \neq r} \Pr(m) - \log \sum_{m:m \neq r} \alpha_{m|r}^{(N)} \Pr(m) \right) \\ &= \min_{m:m \neq r} E_{m|r}^*. \end{aligned} \tag{25}$$

Using (25) by analogy with the formula (15) we conclude (with \mathcal{R}_r defined as in (6) for each r including $r = M$ by the values of $E_{r|r}$ from $(0, \min_{l:l \neq r} D(G_l \| G_r))$) that

$$\begin{aligned} E_{m \neq r | l=r}(E_{r|r}) &= \min_{m:m \neq r} \inf_{Q \in \mathcal{R}_r} D(Q \| G_m) \\ &= \min_{m:m \neq r} \inf_{Q: D(Q \| G_r) \leq E_{r|r}} D(Q \| G_m), \quad r \in [M]. \end{aligned} \tag{26}$$

We can summarize this result in

Theorem 275 *For the model with different distributions, for the given sample \mathbf{x} define its ED Q , and when $Q \in \mathcal{R}_r^{(N)}$ we accept the hypothesis r . Under condition that the probabilities of all M hypotheses are positive the reliability of such test $E_{m \neq r | l=r}$ for given $E_{m=r | l \neq r} = E_{r | r}$ is defined by (26).*

For presentation of examples let us consider the set $\mathcal{X} = \{0, 1\}$ with only 2 elements. Let 5 probability distributions are given on \mathcal{X} :

$$G_1 = \{0.1, 0.9\}$$

$$G_2 = \{0.65, 0.35\}$$

$$G_3 = \{0.45, 0.55\}$$

$$G_4 = \{0.85, 0.15\}$$

$$G_5 = \{0.23, 0.77\}$$

On Fig. 1, the results of calculations of $E_{m \neq r | l=r}$ as function of $E_{m=r | l \neq r}$ are presented.

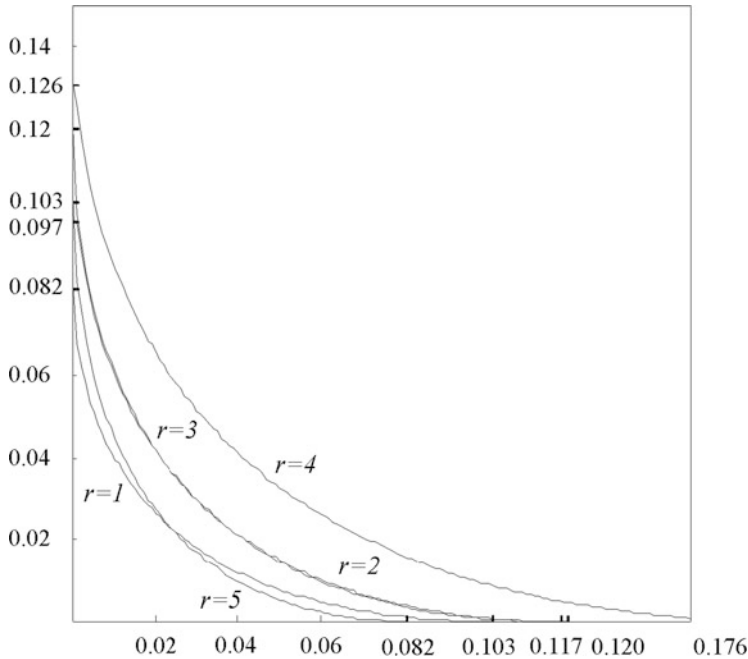


Fig. 1 $E_{m \neq r | l=r}$ as function of $E_{m=r | l \neq r}$ for $r = 1, 2, 3, 4, 5$

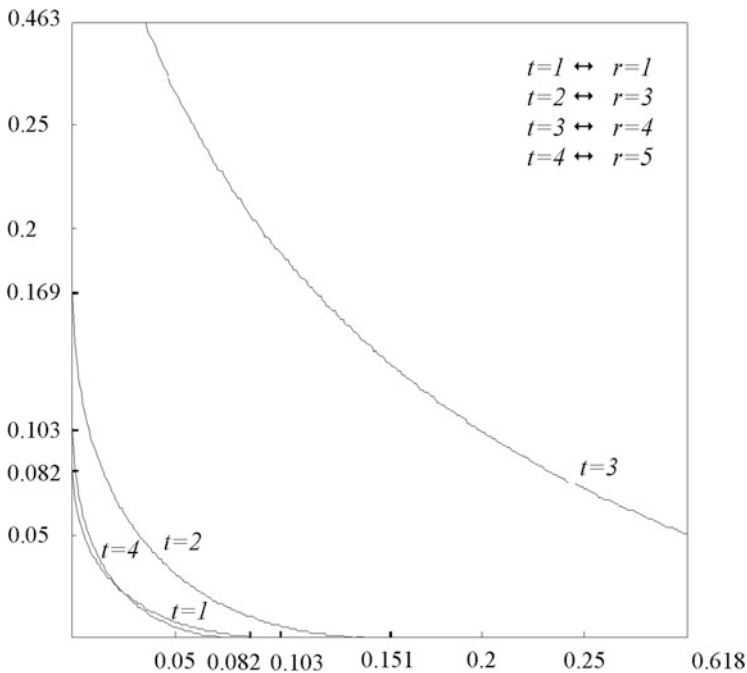


Fig. 2 $E_{m \neq t | l=t}$ as function of $[E_{t|t}]$ for $t = 1, 2, 3, 4$

The elements of the matrix of relative entropies of all pairs of distributions are used for calculation of conditions (13) for this example.

$$\{D(G_m \| G_l)\}_{m \in [5]}^{l \in [5]} = \begin{pmatrix} 0 & 0.956 & 0.422 & 2.018 & 0.082 \\ 1.278 & 0 & 0.117 & 0.176 & 0.576 \\ 0.586 & 0.120 & 0 & 0.618 & 0.169 \\ 2.237 & 0.146 & 0.499 & 0 & 1.249 \\ 0.103 & 0.531 & 0.151 & 1.383 & 0 \end{pmatrix}.$$

In Figs. 2 and 3 the results of calculations of the same dependence are presented for 4 distributions taken from previous 5.

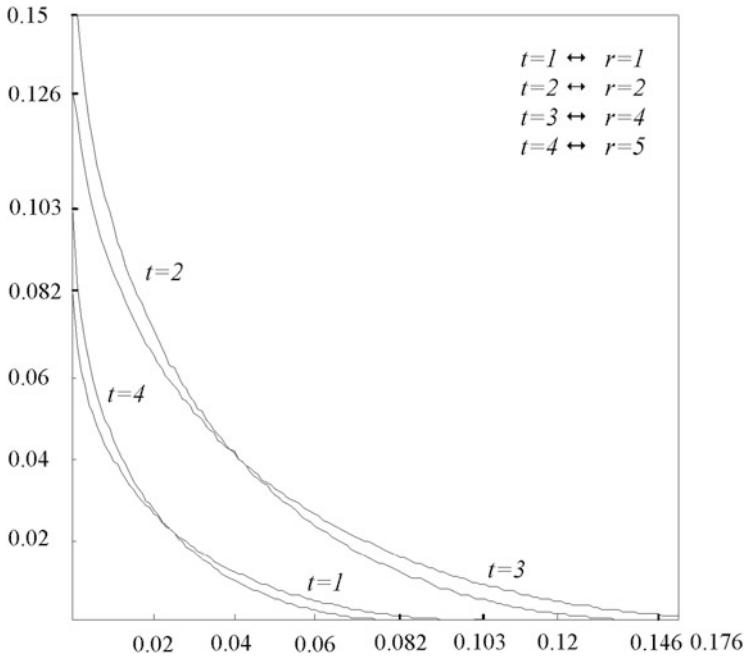


Fig. 3 $E_{m \neq l | l=t}$ as function of $[E_{t|t}]$ for $t = 1, 2, 3, 4$

6 r-Identification and Ranking Problems

The model was introduced in [1] and named K -identification. Since in this lecture the letter K is already used we speak of r -identification. Given N -sample \mathbf{x} of measurements of the object the problem is to answer to the question: is the distribution of the object in the part S of M possible distributions or in its complement, here r is the number of elements of the set S .

Again we can make decision on the base of the ED Q of the sample \mathbf{x} and suppose that before experiments all hypotheses have some positive probabilities

$$\Pr(1), \dots, \Pr(M).$$

Using (6)–(8) with some $E_{1,1}, \dots, E_{M-1,M-1}$ meeting the conditions (13) when $Q \in \bigcup_{l \in S} \mathcal{R}_l^{(N)}$ decision “ l is in S ” follows.

The model of ranking is the particular case of the model of r -identification with $\mathcal{S} = \{1, 2, \dots, r\}$. But conversely the r -identification problem without loss of generality may be considered as the ranking problem, to this end we can renumber the hypotheses placing the hypotheses of \mathcal{S} in the r first places. Because these two models are mathematically equivalent we shall speak below only of the ranking model.

It is enough to consider the cases $r \leq \lceil M/2 \rceil$, because in the cases of larger r we can replace \mathcal{S} with its complement. Remark that the case $r = 1$ was considered in 5.

We study two error probabilities of a test: the probability $\alpha_{m \leq r | l > r}^{(N)}$ to make incorrect decision when m is not greater than r and the probability $\alpha_{m > r | l \leq r}^{(N)}$ to make error when m is greater than r . The corresponding reliabilities are

$$E_1(r) = E_{m \leq r | l > r} \text{ and } E_2(r) = E_{m > r | l \leq r}, \quad 1 \leq r \leq \lceil M/2 \rceil. \quad (27)$$

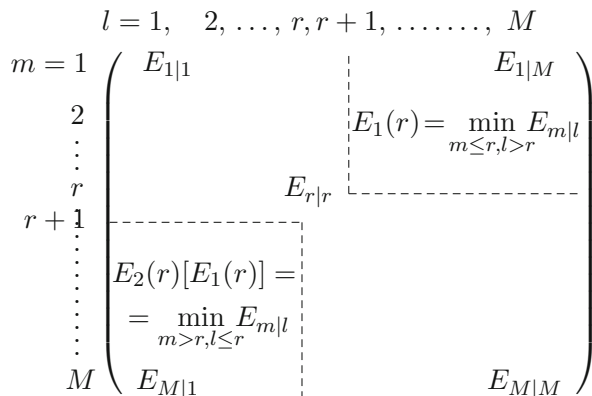
With supposition (6) we have

$$\begin{aligned} \alpha_{m \leq r | l > r}^{(N)} &= \frac{\Pr^{(N)}(m \leq r, l > r)}{\Pr(m \leq r)} \\ &= \frac{1}{\sum_{m \leq r} \Pr(m)} \sum_{m \leq r} \sum_{l > r} \Pr^{(N)}(m, l) \\ &= \frac{1}{\sum_{m \leq r} \Pr(m)} \sum_{m \leq r} \sum_{l > r} \alpha_{m | l}^{(N)} \Pr(m). \end{aligned} \quad (28)$$

The definition (27) of $E_1(r)$ and the equality (28) give

$$\begin{aligned} E_1(r) &= \limsup_{N \rightarrow \infty} -\frac{1}{N} \log \alpha_{m \leq r | l > r}^{(N)} \\ &= \limsup_{N \rightarrow \infty} -\frac{1}{N} \left[\log \sum_{m \leq r} \sum_{l > r} \Pr(m) \alpha_{m | l}^{(N)} - \log \sum_{m \leq r} \Pr(m) \right] \\ &= \min_{m \leq r, l > r} E_{m | l}. \end{aligned} \quad (29)$$

Fig. 4 Calculation of $E_2(r) [E_1(r)]$



Analogously, at the same time

$$\begin{aligned}
 E_2(r) &= \limsup_{N \rightarrow \infty} -\frac{1}{N} \log \alpha_{m>r|l \leq r}^{(N)} \\
 &= \limsup_{N \rightarrow \infty} -\frac{1}{N} \left[\log \sum_{m>r} \sum_{l \leq r} \alpha_{m|l}^{(N)} - \log \sum_{m>r} \Pr(m) \right] \\
 &= \min_{m>r, l \leq r} E_{m|l}. \tag{30}
 \end{aligned}$$

For any test the value of $E_1(r)$ must satisfy the condition (compare (3) and (29))

$$E_1(r) \geq \min_{m:m \leq r} E_{m|m}. \tag{31}$$

Thus for any test meeting all inequalities from (13) for $m \leq r$ and inequality (31) the reliability $E_2(r)$ may be calculated with the equality (30). For given value of $E_1(r)$ the best $E_2(r)$ will be obtained if we use liberty in selection of the biggest values for reliabilities $E_{m|m}$, $r < m \leq M - 1$, satisfying for those m -s conditions (13). These reasonings may be illuminated by Fig. 4 and resumed as follows:

Theorem 276 *When the probabilities of the hypotheses are positive, for given $E_1(r)$ for $m \leq r$ not exceeding the expressions on the right in (13), $E_2(r)$ may*

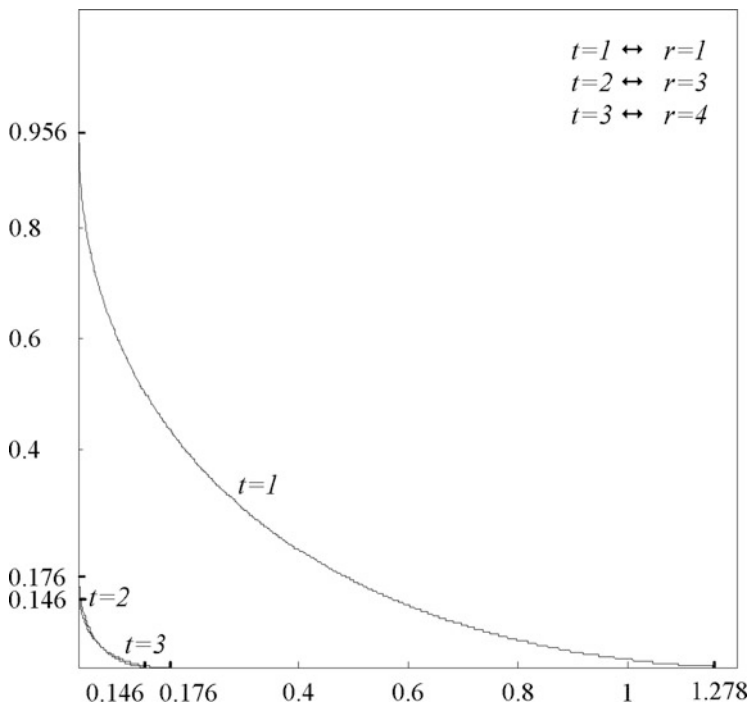


Fig. 5 $E_{m \neq t | l = t}$ as function of $[E_{t | t}]$ for $t = 1, 2, 3$

be calculated in the following way:

$$E_2(r) [E_1(r)] = \max_{\{E_{m|l}, m, l \in [M]\}: \min_{m \leq r, l > r} E_{m|l}^* = E_1(r)} \left[\min_{m > r, l \leq r} E_{m|l}^* \right] \tag{32}$$

with $E_{m|l}^*$ defined in (9)–(12).

Remark One can see from (32) that for $r = 1$ we arrive to (26) for $r = 1$.

In Figs. 5 and 6 for 2 subsets by 3 distributions taken from 5 defined for Fig. 1 the results of calculation of the dependence (26) and in Figs. 7 and 8 the corresponding results of the formula (33) are presented.

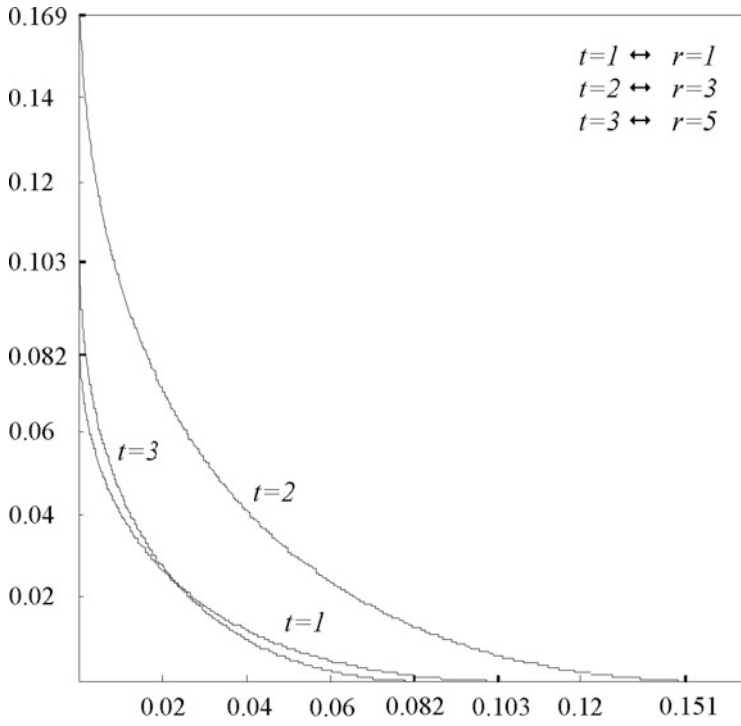


Fig. 6 $E_{m \neq t|l=t}$ as function of $[E_{t|t}]$ for $t = 1, 2, 3$

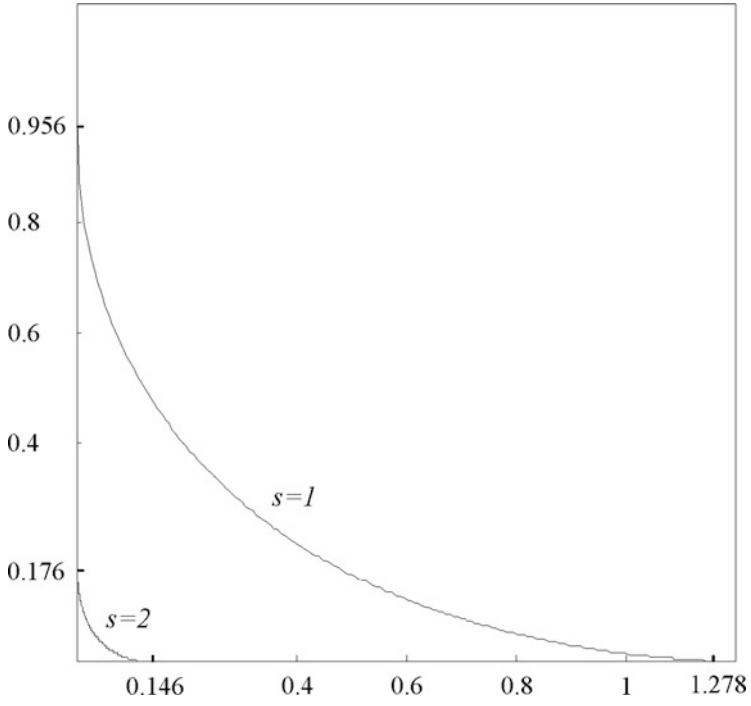


Fig. 7 $E_2(r), E_1(r)$

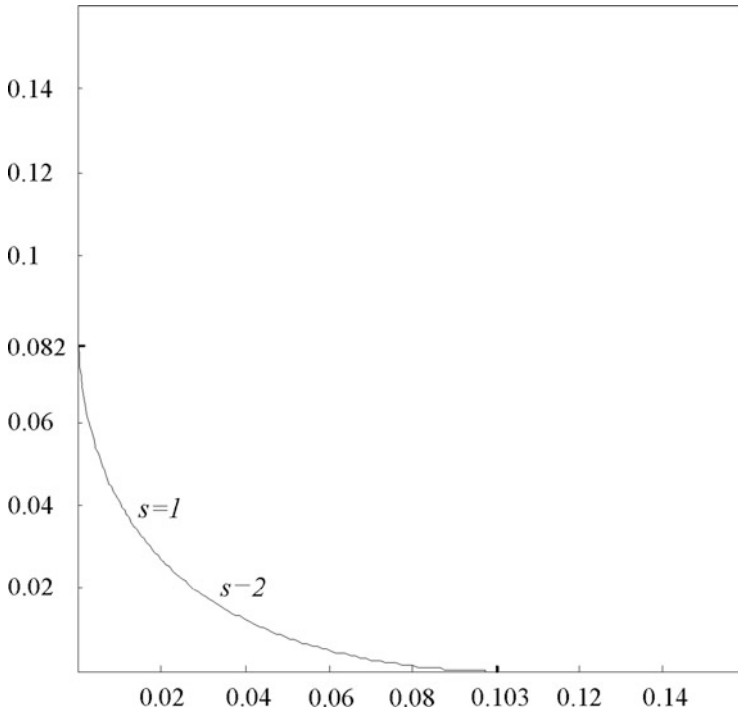


Fig. 8 $E_2(r), E_1(r)$

7 Conclusion and Extensions of Problems

The lecture is a contribution to influence of the information theory methods on statistical theory. We have shown by simple examples what questions arise in different models of statistical identification.

Problems and results of the lecture may be extended in several directions some of which have been already noted above.

It is necessary to examine models in which measurements are described by more general classes of RV's and processes [18, 19, 21, 26].

One of the directions is connected with the use of compressed data of measurements [2, 6, 8, 19, 29].

One may see perspectives in application of identification approach and methods to the authentication theory [25] and steganography [13].

References

1. R. Ahlswede, General theory of information transfer, Preprint 97–118, SFB 343 “Diskrete Strukturen in der Mathematik”, Universität Bielefeld, General Theory of Information Transfer and Combinatorics, Report on a Research Project at the ZIF (Center of interdisciplinary studies) in Bielefeld Oct. 1, 2002–August 31, 2003, edit R. Ahlswede with the assistance of L. Bäumer and N. Cai, also Special issue of Discrete Mathematics
2. R. Ahlswede, M. Burnashev, On minimax estimation in the presence of side information about remote data. *Ann. Stat.* **18**(1), 141–171 (1990)
3. R. Ahlswede, I. Csiszár, Hypotheses testing with communication constraints. *IEEE Trans. Inf. Theory* **32**(4), 533–542 (1986)
4. R. Ahlswede, G. Dueck, Identification via channels. *IEEE Trans. Inf. Theory* **35**(1), 15–29 (1989)
5. R. Ahlswede, I. Wegener, *Search Problems* (Wiley, New York, 1987)
6. R. Ahlswede, E. Yang, Z. Zhang, Identification via compressed data. *IEEE Trans. Inf. Theory* **43**(1), 48–70 (1997)
7. R.E. Bechhofer, J. Kiefer, M. Sobel, *Sequential Identification and Ranking Procedures*. University of Chicago Press, Chicago (1968)
8. T. Berger, Decentralized estimation and decision theory, in *Presented at IEEE Seven Springs Workshop on Information Theory, Mt. Kisco, NY* (1979)
9. L. Birgé, Vitesse maximales de décroissance des erreurs et tests optimaux associés. *Z. Wahrsch. verw. Gebiete* **55**, 261–273 (1981)
10. R.E. Blahut, Hypothesis testing and information theory. *IEEE Trans. Inform. Theory*, **IT-20**, 405–417 (1974)
11. R.E. Blahut, *Principles and Practice of Information Theory* (Addison-Wesley, Reading, 1991)
12. A.A. Borovkov, *Mathematical Statistics (in Russian)* (Nauka, Novosibirsk, 1997)
13. C. Cachin, An information-theoretic model for steganography, in *Proceedings of the 2nd Workshop on Information Hiding (David Ausmith, ed.)*. Lecture Notes in computer Science (Springer, Berlin, 1998)
14. I. Csiszár, Method of types. *IEEE Trans. Inf. Theory* **44**(6), 2505–2523 (1998)
15. I. Csiszár, J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Academic Press, New York, 1981)
16. I. Csiszár, G. Longo, On the error exponent for source coding and for testing simple statistical hypotheses. *Studia Sc. Math. Hungarica* **6**, 181–191 (1971)
17. F.W. Fu, S.Y. Shen, Hypothesis testing for arbitrarily varying source with exponents type constraint. *IEEE Trans. Inf. Theory* **44**(2), 892–895 (1998)
18. M. Gutman, Asymptotically optimal classification for multiple test with empirically observed statistics. *IEEE Trans. Inf. Theory* **35**(2), 401–408 (1989)
19. T.S. Han, S. Amari, Statistical inference under multiterminal data compression. *IEEE Trans. Inf. Theory* **44**(6), 2300–2324 (1998)
20. T.S. Han, K. Kobayashi, Exponential-type error probabilities for multiterminal hypothesis testing. *IEEE Trans. Inf. Theory* **35**(1), 2–13 (1989)
21. E.A. Haroutunian, On asymptotically optimal testing of hypotheses concerning Markov chain (in Russian). *Izvestia Acad. Nauk Armenian SSR. Seria Mathem.* **22**(1), 76–80 (1988)
22. E.A. Haroutunian, Logarithmically asymptotically optimal testing of multiple statistical hypotheses. *Probl. Control Inf. Theory* **19**(5–6), 413–421 (1990)
23. W. Hoeffding, Asymptotically optimal tests for multinomial distributions. *Ann. Math. Stat.* **36**, 369–401 (1965)
24. G. Longo, A. Sgarro, The error exponent for the testing of simple statistical hypotheses, a combinatorial approach. *J. Comb. Inf. Syst. Sci.* **5**(1), 58–67 (1980)
25. U.M. Maurer, Authentication theory and hypothesis testing. *IEEE Trans. Inf. Theory* **46**(4), 1350–1356 (2000)

26. S. Natarajan, Large deviations, hypotheses testing, and source coding for finite Markov chains. *IEEE Trans. Inf. Theory* **31**(3), 360–365 (1985)
27. R.C. Rao, *Linear Statistical Inference and its Applications* (Wiley, New York, 1965)
28. G. Tusnady, On asymptotically optimal tests. *Ann. Stat.* **5**(2), 385–393 (1977)
29. Z. Zhang, T. Berger, Estimation via compressed information. *IEEE Trans. Inf. Theory* **34**(2), 198–211 (1988)