Selected Topics of Information Theory and Mathematical Statistics

Marat V. Burnashev (Moscow)

1 Testing of hypotheses and Identification

Identification became a quite popular topic in Information Theory. It is a very good example of interaction between Mathematical Statistics (Testing of Hypotheses) and Information Theory. There are still many interesting open problems in it. I am going to devote a good part of these lectures to that topic. In order to come smoothly to identification problems we will need first some knowledge on testing of hypotheses and related notions.

1.1 Preliminaries. Testing of hypotheses and L_1 - distance.

Since later in identification we will deal only with finite alphabets all necessary preliminaries will be given here also only for finite alphabets.

L_1 - distance. Let $Y = (y_1, y_2, \ldots, y_K)$ be some finite set (alphabet). *Probability measure* P on Y is defined by any collection of *nonnegative values* $\{P(y_1), P(y_2), \ldots, P(y_K)\}$ satisfying the condition:

$$\sum_{i=1}^{K} P(y_i) = 1$$

L_1 - *distance* between a pair of probability measures P and Q on Y is defined as

$$\|P - Q\| = \sum_{i=1}^{K} |P(y_i) - Q(y_i)|$$

(sometimes it is called a *variational distance* and sometimes the factor $1/2$ in front of the sum is used).

Support of a measure P is defined as

$$\text{supp } P = \{y \in Y : P(y) > 0\}$$

(i.e. it consists of all points y of the set Y where the measure P is strongly positive).

Examples. If $P = Q$ (i.e. $P(y) = Q(y)$ for any $y \in Y$) then obviously $\|P-Q\| = 0$. If supports of measures P and Q do not intersect (i.e. measures P and Q are *orthogonal* to each other) then $\|P - Q\| = 2$.

L_1 - distance between a probability measure P and some collection of probability measures $\{Q_\alpha, \alpha \in \mathcal{A}\}$ (where $\mathcal{A}$ is an arbitrary index set) is defined as

$$\|P - \{Q_\alpha, \alpha \in \mathcal{A}\}\| = \inf_{\alpha \in \mathcal{A}} \|P - Q_\alpha\|$$

We will need one more operation with probability measures.

Let $\{Q_\alpha, \alpha \in \mathcal{A}\}$ (where $\mathcal{A}$ is an arbitrary index set) be some collection of probability measures. *Convex hull* of the family $\{Q_\alpha, \alpha \in \mathcal{A}\}$ (denoted by conv $\{Q_\alpha, \alpha \in \mathcal{A}\}$ or simply conv $\{Q_\alpha\}$) is the set of all possible finite convex linear combinations of measures from $\{Q_\alpha\}$. In other words, conv $\{Q_\alpha\}$ consists of all measures F that can be represented in the form

$$F = \sum_{i \in \mathcal{A}} c_i Q_i$$

with

$$c_i \geq 0 \text{ for all } i \ , \text{ and } \sum_{i \in \mathcal{A}} c_i = 1 \ ,$$

and only finite number of c_i's are not equal to zero.

We denote by $\mathcal{P}(Y)$ the set of *all* probability measures on Y. The set $\mathcal{P}(Y)$ is a *complete* set with respect to L_1 - distance (metric). In other words, if $\{Q_n, \ n = 1, 2, \ldots, \}$ is a *fundamental* sequence (i.e. $\|Q_k - Q_m\| \leq \epsilon$ if $k, m \geq n(\epsilon)$ for any $\epsilon > 0$) then it has a *limit* $Q_0 \in \mathcal{P}(Y)$ (i.e. it is also a *probability* measure).

Example. Let on the set $Y = (y_1, y_2, \ldots, y_K)$ be given the collection of probability measures $\{Q_i \ , \ i = 1, \ldots, K\}$ such that $Q_i(y_i) = 1$ for any $i = 1, \ldots, K$ (it is clear that then $Q_i(y_j) = 0$ for any $i \neq j$). Then $\text{conv}\{Q_i\} = \mathcal{P}(Y)$.

Testing of simple hypotheses. Let on a finite space Y there are given two probability measures: P and Q. We get *one* observation $y \in Y$ according to some distribution F. We know that an unknown distribution F satisfies one of the following two possibilities (hypotheses):

$$H_0 : \ F = P \ ,$$

or

$$H_1 : \ F = Q$$

In other words, due to hypothesis H_0 the distribution F coincides with P and due to hypothesis H_1 the distribution F is equal to Q.

Remark. Any hypothesis H is called *simple* if the observation distribution F is completely determined provided that H is a true hypothesis (in contrast to *composite* hypothesis when provided H is true we know only that F belongs to some *class* of distributions). Therefore we consider here the testing of *two simple hypotheses* problem.

Now we want to decide in some optimal way which of those hypotheses has really taken place (i.e. to *test* those hypotheses). Any *nonrandomized* way of decision making is determined by some set $\mathcal{D} \subseteq Y$ such that

$$y \in \mathcal{D} \Rightarrow H_0 \ ,$$

$$y \in \mathcal{D}^c = Y \setminus \mathcal{D} \Rightarrow H_1$$

(so if our observation belongs to $\mathcal{D}$ we make a decision in favor of hypothesis H_0 and if it belongs to $\mathcal{D}^c$ we make a decision in favor of hypothesis H_1).

Since we test here *two* hypotheses there are possible *two* kinds of errors:

the *1-st kind error* when we make a decision in favor of hypothesis H_1 while hypothesis H_0 is true, with probability

$$\alpha(\mathcal{D}) = P(\mathcal{D}^c) \ ,$$

and the *2-nd kind error* when we make a decision in favor of hypothesis H_0 while hypothesis H_1 is true, with probability

$$\beta(\mathcal{D}) = Q(\mathcal{D})$$

Of course, we would like to minimize both error probabilities, but usually it is impossible (except for orthogonal measures P and Q) and therefore we should find some compromise between them. For that reason, in order to find the *optimal way* of testing those hypotheses we need some *quality criteria* (i.e. some characteristics allowing us to compare different ways of decision making).

There are widely known several such quality criteria, but we will be interested in one particular quality criteria: we want to have the test that minimizes the *sum of both error probabilities.* Probably, this criteria is not very popular in classic mathematical statistics, but it fits the best our needs in Identification. In particular, we will be interested in the value:

$$\delta(P,Q) = \min_{\mathcal{D}}\{\alpha(\mathcal{D}) + \beta(\mathcal{D})\}$$

Next result gives a simple answer to that question.

Lemma 1. *When testing two simple hypotheses P and Q for minimal possible sum $\alpha + \beta$ the following equality holds true:*

$$\delta(P,Q) = \min_{\mathcal{D}} \{\alpha(\mathcal{D}) + \beta(\mathcal{D})\} = 1 - \frac{1}{2}\|P - Q\| \ , \tag{1}$$

where $\|P - Q\|$ is the L_1 - distance between measures P and Q.

Proof. For any set $\mathcal{D} \subseteq Y$ we have

$$\alpha(\mathcal{D}) + \beta(\mathcal{D}) = P(Y \setminus \mathcal{D}) + Q(\mathcal{D}) = 1 - [P(\mathcal{D}) - Q(\mathcal{D})] \geq$$

$$\geq 1 - \sup_{\mathcal{D}} \left\{ \sum_{y \in \mathcal{D}} (P(y) - Q(y)) \right\} = 1 - \sum_{\{y: P(y) > Q(y)\}} (P(y) - Q(y)) =$$

$$= 1 - \frac{1}{2}\|P - Q\| \ ,$$

where we used easy-to-check equalities

$$\|P - Q\| = \sum_{\{y: P(y) > Q(y)\}} (P(y) - Q(y)) + \sum_{\{y: Q(y) > P(y)\}} (Q(y) - P(y))$$

and

$$\sum_{\{y: P(y) > Q(y)\}} (P(y) - Q(y)) = \sum_{\{y: Q(y) > P(y)\}} (Q(y) - P(y)) \ .$$

¿From that proof it is easy to see also that the optimal set $\mathcal{D}_{\rm opt}$ to make a decision in favor of hypothesis H_0 has the form

$$\mathcal{D}_{\rm opt} = \{y : P(y) > Q(y)\}$$

Moreover, we can add to this set any points y with $P(y) = Q(y)$. $\triangle$

Testing of simple hypothesis against a composite one. Now we come to the main part of our introduction. Let on a finite space Y there are given some probability measure P and some *collection* Ω of probability measures Q. The collection Ω may be finite or infinite, but contains more then one element. We assume that an unknown distribution F satisfies one of the following two hypotheses:

$$H_0 : \ F = P \ ,$$

or

$$H_1 : \ F \in \Omega = \{Q\} \ .$$

In other words, due to hypothesis H_0 the distribution F coincides with P and due to hypothesis H_1 the distribution F belongs to the class Ω (i.e. coincides with some distribution Q from Ω).

When dealing with discrete sets (like our finite set Y) and composite hypothesis we should enlarge our ways of decision making (i.e. allowing to make a *randomized decisions*).

When testing two hypotheses on the set Y any *randomized decision (test)* is determined by choosing some *decision function* $0 \leq \varphi(y) \leq 1$, $y \in Y$. Each value $\varphi(y)$ of that function is simply the probability to make a decision in favor of hypothesis H_0 when y is observed.

For nonrandomized tests the function $\varphi(y)$ takes only values 0 and 1. Clear that in our case of testing two simple hypotheses we had
$\varphi(y) = 1$, $y \in \mathcal{D}$, and $\varphi(y) = 0$, $y \in \mathcal{D}^c$.

Coming back to our problem of testing a simple hypothesis H_0 against a composite one H_1, suppose that we chose some decision function $\varphi(y)$, $y \in Y$. Then we can also make two kinds of errors: *1-st kind* and *2-nd kind*. We define their probabilities as :

probability of the *1-st kind error* when we make a decision in favor of hypothesis H_1 while hypothesis H_0 is true

$$\alpha(\varphi) = E_P \left(1 - \varphi(Y)\right) = \sum_{y \in Y} \left(1 - \varphi(y)\right) P(y) \ ;$$

and probability of the *2-nd kind error* when we make a decision in favor of hypothesis H_0 while hypothesis H_1 is true

$$\beta(\varphi) = \sup_{Q \in \Omega} E_Q \varphi(Y) = \sup_{Q \in \Omega} \sum_{y \in Y} \varphi(y) Q(y)$$

As can be seen in defining the 2-nd kind of error probability we consider the *worst* possible case in error probability.

Now we want to have the test (i.e. decision function $\varphi(y)$) that minimizes the sum of both error probabilities (it is some kind of the *minimax* problem of hypotheses testing). In particular, we will be interested in the value:

$$\delta(P, \Omega) = \min_{\varphi} \delta(P, \Omega, \varphi) = \min_{\varphi} \{\alpha(\varphi) + \beta(\varphi)\}$$

Next result gives the answer to that question.

Theorem 1. *When testing a simple hypothesis P against a composite hypothesis Ω the minimal possible sum $\alpha + \beta$ satisfies the following relation:*

$$\delta(P, \Omega) = \min_{\varphi} \{\alpha(\varphi) + \beta(\varphi)\} = 1 - \frac{1}{2} \| P - \operatorname{conv} \Omega \| \ , \tag{2}$$

where $\operatorname{conv} \Omega$ *denotes the convex hull of all probability measures Q from the collection Ω.*

Let us give first some statistical meaning to the relation (2). Let us take some finite convex linear combination $\sum \pi_i Q_i$ of some elements Q_i from Ω (of course, we get then some element of conv Ω). We may think that we put some *prior distribution* on the set Ω and replace the composite hypothesis Ω by the simple hypothesis $\sum \pi_i Q_i$. Then clearly $\delta(P,\Omega) \geq \delta(P, \sum \pi_i Q_i)$ and moreover due to Lemma 1 we have

$$\delta(P,\Omega) \geq \delta(P, \sum \pi_i Q_i) = 1 - \frac{1}{2}\|P - \sum \pi_i Q_i\|$$

Since the last relation holds true for any finite linear combination $\sum \pi_i Q_i$ it then follows that

$$\delta(P,\Omega) \geq 1 - \frac{1}{2}\|P - \operatorname{conv}\Omega\|$$

Therefore, in order to establish the relation (2) it remains to show the sign $\leq$ in it.

Remark. When we put some *prior distribution* $\pi(Q)$, $Q \in \Omega$, and define the 2-nd kind of error probability as

$$\beta(\varphi) = E_\pi \, E_Q \varphi(Y) = \sum_{Q \in \Omega} \pi(Q) \sum_{y \in Y} \varphi(y) Q(y) \ ,$$

we get a *bayesian* statement of the testing problem. From that point of view Theorem 1 is some variation of one of the most crucial principles in mathematical statistics that can be stated here as follows:

minimax statement of the problem is equivalent to the *bayesian* statement with the *worst (least-favorable) prior distribution* on $\overline{\operatorname{conv}\Omega}$, where $\overline{\operatorname{conv}\Omega}$ denotes the closure of the set $\operatorname{conv}\Omega$ in metric L_1 (remind that $\operatorname{conv}\Omega \subseteq \mathcal{P}(Y)$ and the set of all probability measures $\mathcal{P}(Y)$ on finite Y is complete in metric L_1).

Unfortunately, Theorem 1 (like any other that kind of result in mathematical statistics) is an *existance* result since it shows only the existance of the least-favorable distribution and says nothing about how to find it (it is usually a difficult problem).

Sketch of the proof of sign $\leq$ in Theorem 1. We have

$$\delta(P,\Omega) = \inf_{\varphi} \sup_{Q \in \Omega} \{\alpha(\varphi) + \beta(\varphi, Q) \ \leq \inf_{\varphi} \sup_{\pi} \{\alpha(\varphi) + \beta(\varphi,\pi)\} \ .$$

Asume now that we can change the order of *inf* and *sup* in the last relation (without changing the value of the resulting function). Then we can continue it as

$$\delta(P,\Omega) \leq \sup_{\pi} \inf_{\varphi} \{\alpha(\varphi) + \beta(\varphi,\pi)\} = \sup_{\pi} \left\{ 1 - \frac{1}{2}\|P - \sum \pi_i Q_i\| \right\} =$$

$$= 1 - \frac{1}{2}\|P - \operatorname{conv}\Omega\| \ ,$$

that completes the proof of Theorem 2.

It remains to understand when we can change that order. But it is the well-known *minimax* theorem in dynamic programming, convex analysis, etc. It is easy to check that in our case the spaces of φ and π are separable and complete, and moreover the function has necessary properties.

Historical notes. Wald's book [1] is a classic book on decision functions (including relations between minimax and bayesian approaches). Book [2] is another good source for testing of hypotheses. All essential for us results can be found also in [3]. Relations (1) – (2) appeared first in the paper [4] and both easily follow from [1] or [3].

1.2 Measures separated in L_1 – metrics.

Let Y still be a finite alphabet. Probability measures P and Q on Y are *orthogonal* if $\|P - Q\| = 2$. We consider first some properties of families of *"almost orthogonal"* probability measures on Y.

Definition 1. A collection $(P_i\,, i = 1, \dots, M)$ of probability measures P_i on Y is called (M, δ) – *pairwise separated* collection (family) if for any $i \neq j$ the following condition is satisfied:

$$\|P_i - P_j\| = \sum_Y |P_i(y) - P_j(y)| \geq 2(1 - \delta)\ . \tag{3}$$

Definition 2. A collection $(P_i\,, i = 1, \dots, M)$ of probability measures P_i on Y is called (M, δ) – *completely separated* collection (family) if for any i the following condition is satisfied:

$$\|P_i - \mathrm{conv}\{P_j\,, j \neq i\}\| = \min_{\mathbf{c}} |P_i - \sum_{j \neq i} c_j P_j| \geq 2(1 - \delta)\ , \tag{4}$$

where $\mathrm{conv}\{\mathcal{A}\}$ means the convex hull of measures from the family $\mathcal{A}$ and minimum is taken over all probability vectors $\mathbf{c} = (c_1, \dots, c_M)$.

It is clear that any δ–completely separated collection of measures is a δ–pairwise separated collection as well.

It will turn out later that a completely separated collection of measures is essentially equivalent to some identification–code that is of our main interest in this topic. It is natural then to investigate first the maximum possible cardinality of a completely separated family of measures.

Denote the cardinality of the set Y by N and let $M_c(N, \delta)$ and $M_p(N, \delta)$ be the maximal possible cardinalities of completely separated and pairwise separated collections on the set Y, respectively.

It is obvious that

$$M_c(N, \delta)\ \leq\ M_p(N, \delta), \quad 0 \leq \delta \leq 1\ . \tag{5}$$

Since as usual in that kind of problems we are not able to find the values $M_c(N, \delta)$ and $M_p(N, \delta)$, we will get some estimates for them through lowerbounding the value $M_c(N, \delta)$ and upperbounding the value $M_p(N, \delta)$. All bounds below are oriented to the case when N is large, δ is small and moreover $N\delta^2$ is also large.

Proposition 1. *For any $N \geq 2$ and $0 < \delta < 1$ the following bounds are valid :*

$$M_c(N, \delta)\ \geq\ \exp\{N\delta^2/(2e^2)\}\ , \tag{6}$$

$$M_p(N, \delta)\ \leq\ (2/(1 - \delta))^{N-1}\ . \tag{7}$$

Remark. Upperbound (7) will be essentially improved further but in that proof we will use upperbound (7)).

Proof. An analog of the inequality (6) (with a smaller constant instead of $1/(2e^2)$ was proved in [5] (Statement 1) by an *"exhaustive"* method. Similar inequality in a more complicated situation was obtained in [9] (Theorem 1) also by the "exhaustive" method. In order to prove the inequality (6) here for the sake of variety we use some *"random choice"* arguments.

We put as every measure $P_i\,, i = 1, \ldots, M$, an equiprobable distribution on some subset $\mathcal{D}_i \subset Y$ of cardinality ϵN (parameter $\epsilon < \delta$ will be chosen later). Assume that the cardinality of the intersection of any pair of regions $\mathcal{D}_i$ and $\mathcal{D}_j\,,\, i \neq j$, does not exceed $\epsilon\delta N$. Then we get for any i

$$\|P_i - \mathrm{conv}\{P_j\,, j \neq i\}\| \geq 2\,[P_i\,(\mathcal{D}_i) - \mathrm{conv}\{P_j\,, j \neq i\}\,(\mathcal{D}_i)] =$$

$$= 2 - 2\sup_{j \neq i} P_j\,(\mathcal{D}_i) \geq 2(1-\delta)\ .$$

Therefore such collection $(P_i\,, i = 1, \ldots, M)$ would be a (M, δ) – completely separated collection. Now, we choose every region $\mathcal{D}_i$ randomly from equiprobable elements of the set Y. Moreover all regions are chosen independently from each other. Then for the probability P that there will be a pair $\mathcal{D}_i$ and $\mathcal{D}_j\,,\, i \neq j$, with $|\mathcal{D}_i \cap \mathcal{D}_j| > \epsilon\delta N$ the following estimate is valid

$$P \leq M(M-1) \sum_{i > \epsilon\delta N} \binom{\epsilon N}{i} \binom{N - \epsilon N}{\epsilon N - i} \left[2 \binom{N}{\epsilon N}\right]^{-1}$$

In order to simplify the last sum we use a standard way. Notice that for $i > \epsilon\delta N$ we have

$$\binom{\epsilon N}{i+1} \binom{N - \epsilon N}{\epsilon N - i - 1} \left[\binom{\epsilon N}{i} \binom{N - \epsilon N}{\epsilon N - i}\right]^{-1} \leq \frac{\epsilon(1-\delta)^2}{\delta(1 - 2\epsilon + \delta\epsilon)} \leq \frac{\epsilon}{\delta}\ .$$

Therefore replacing that sum by the geometrical progression we get

$$P \leq M(M-1)\delta \binom{\epsilon N}{\epsilon\delta N} \binom{N - \epsilon N}{\epsilon N - \epsilon\delta N} \left[2(\delta - \epsilon) \binom{N}{\epsilon N}\right]^{-1}\ . \tag{8}$$

We can loosen this bound with the easy verifiable inequality

$$\binom{K-a}{a-i} \Big/ \binom{K}{a} \leq \left(\frac{a}{K-a}\right)^i \left(\frac{K-a}{K}\right)^{a-i}.$$

Indeed

$$\binom{K-a}{a-i} \Big/ \binom{K}{a} =$$

$$= \frac{(a-i+1)\ldots a}{(K-a+1)\ldots(K-a+i)} \frac{(K-2a+i+1)\ldots(K-a)}{(K-a+i+1)\ldots K} \leq$$

$$\leq \left(\frac{a}{K-a}\right)^i \left(\max_{1\leq j\leq a-i} \frac{K-2a+i+j}{K-a+i+j}\right)^{a-i} \leq \left(\frac{a}{K-a}\right)^i \left(\frac{K-a}{K}\right)^{a-i}$$

Then, (8) becomes

$$P \leq \frac{\delta M(M-1)}{2(\delta-\epsilon)} \exp\left\{N\delta\epsilon \ln \frac{e\epsilon}{\delta}\right\} . \tag{9}$$

We put now $\epsilon = [\epsilon_0 N]/N$, $\epsilon_0 = \delta e^{-2}$. Since $0 \leq (\epsilon_0 - \epsilon)N \leq 1$, we get from (9) after some simple calculations, assuming $N\delta \geq 15$

$$P \leq M(M-1)\exp\{-N\delta^2 e^{-2}\} . \tag{10}$$

Now if the right side of (10) is less than 1, then there exists a collection of M regions having the required properties, from where the lower bound (6) follows provided $N\delta \geq 15$. Since there are always N orthogonal measures on Y (and using that fact when $N\delta < 15$), we get that lower bound (6) is valid for any $N \geq 2$ and $0 \leq \delta \leq 1$.

We now prove the upper bound (7). Let $P = (x_1, \ldots, x_N)$ be any probability measure on Y and let $\mathcal{P}_N$ be the set of all probability measures on Y:

$$\mathcal{P}_N = \left\{P : x_i \geq 0,\ \sum_{i=1}^N x_i = 1\right\} .$$

The set $\mathcal{P}_N$ has a "volume" v_N :

$$v_N = \int \cdots \int_{A(N-1,1)} dx_1 \ldots dx_{N-1} = \frac{1}{(N-1)!}\ , \tag{11}$$

where the following notation was used

$$A(K,a) = \left\{x \in R^K : x_i \geq 0\ ,\ i = 1, \ldots, K\ ;\ \sum_{i=1}^K x_i \leq a\right\} .$$

Indeed

$$v_N = \int_0^a \int \cdots \int_{A(N-2,a-x_{N-1})} dx_1 \ldots dx_{N-2} dx_{N-1} =$$

$$= \int_0^a v_{N-1}(a - x_{N-1})^{N-2} dx_{N-1} = v_{N-1} a^{N-1}/(N-1) = v_N a^{N-1} .$$

Therefore $v_N = v_{N-1}/(N-1)$, $v_2 = 1$, from where formula (11) follows.

We fix now any measure $P_0 = (z_1, \ldots, z_N)$ and consider the set $D(P_0)$ of all measures P such that $\|P - P_0\| \leq 1 - \delta$. It is possible to show (see Appendix) that $D(P_0)$ has the minimal "volume" $w_N(\delta)$ when $z_1 = 1$, $z_2 = \ldots = z_N = 0$ (i.e., when P_0 is an extreme point of the set $\mathcal{P}_N$). Then,

$\|P - P_0\| = 2 - 2x_1$ and with the help of (11) we get $(b = A(N-1,1) \cap \{x : x_1 \geq (1+\delta)/2\})$

$$w_N(\delta) = \int \ldots \int_B dx_1 \ldots dx_{N-1} =$$

$$= v_{N-1} \int_{(1+\delta)/2}^{1} (1-x_1)^{N-2} dx_1 = v_N (2/(1-\delta))^{N-1} \, . \tag{12}$$

Since $M_p(N, \delta) \leq v_N / w_N(\delta)$, we now get the upper bound (7) from (11) – (12). $\triangle$

The following result improves the upper bound (7).

Proposition 2. *For any $0 < \delta < 1$ the following inequality holds true:*

$$M_p(N,\delta) \leq N + \frac{1}{\delta^2} + \frac{1}{2\delta^2} \exp\left\{ \frac{\delta^2 N}{(1-\sqrt{\delta})^3} \ln \frac{2e}{\delta^2} \right\} \, . \tag{13}$$

To prove Proposition 2 we will need the lemma.

Lemma 2. *Let μ_i, $i = 1, \ldots, M$, be the collection of δ–pairwise separated probability measures on the set Y of cardinality N.*
a) If $\max_{y,i} \mu_i(y) \leq \mu$ then

$$M \leq \frac{(1-\delta)\mu N}{(1-\delta\mu N)_+} \, . \tag{14}$$

b) If $\mu_i(y) \geq \mu \geq \delta/N$ for all $y \in Y_i = \{y : \mu_i(y) > 0\}$, $i = 1, \ldots, M$, then

$$M \leq \frac{(1-\delta)\mu N}{2\delta} \exp\left\{ \frac{\delta}{\mu} \ln \frac{2e\mu N}{\delta(1-\delta)} \right\} \, . \tag{15}$$

Proof. We start with inequality (14). It is clear that

$$\sum_{i=1}^{M} \sum_{j=1}^{M} ||\mu_i - \mu_j|| \geq 2(1-\delta)M(M-1) \, . \tag{16}$$

¿From the other hand, using the representation

$$||\mu - \nu|| = 2\left(1 - \sum_{y \in Y} \min\{\mu(y), \nu(y)\}\right)$$

(which follow from simple formula $|a-b| = a + b - 2\min\{a,b\}$) and inequality $\min\{\mu_i(y), \mu_j(y)\} \geq \mu_i(y)\mu_j(y)/\mu$, we get applying Cauchy - Buniakovsky inequality:

$$\sum_{i=1}^{M} \sum_{j=1}^{M} \|\mu_i - \mu_j\| \leq 2\left[M^2 - \frac{1}{\mu} \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{Y} \mu_i(y)\mu_j(y)\right] =$$

$$= 2\left[M^2 - \frac{1}{\mu}\sum_{Y}\left(\sum_{i=1}^{M}\mu_i(y)\right)^2\right] \le \tag{17}$$

$$\le 2\left[M^2 - \frac{1}{\mu N}\left(\sum_{Y}\sum_{i=1}^{M}\mu_i(y)\right)^2\right] = 2M^2\left(1 - \frac{1}{\mu N}\right) \quad .$$

Now from (16) - (17) the inequality (14) follows.

We will prove now the inequality (15). Assume first that $\mu \ge \delta$. Then obviously $Y_i \bigcap Y_j = \emptyset$, $i \ne j$ and therefore $M \le N$. In order to show that the right side of (15) is greater than N it is sufficient to show that

$$0 \le \frac{\delta}{\mu}\ln\frac{2e\mu N}{\delta(1-\delta)} + \ln\frac{(1-\delta)\mu}{2\delta} =$$

$$= \left(1 - \frac{\delta}{\mu}\right)\ln(1-\delta) + \left(1 + \frac{\delta}{\mu}\right)\ln(\frac{\mu}{\delta}) + \frac{\delta}{\mu}\ln(2eN) - \ln 2 \quad .$$

Using simple inequality $\ln(1/x) \ge 1 - x$ we get that the last expression is greater or equal than

$$1 - \ln 2 - \frac{\delta^2}{\mu^2} + \frac{\delta}{\mu}\ln(2N) \ge 0\ ,\ N \ge 2\ ,\ \delta \le \mu \le 1\ ,$$

from where validity of (15) for $\delta \le \mu$ follows.

Therefore we assume now that $\mu < \delta$. Since

$$\mu|Y_i \cap Y_j| \le \sum_{Y}\min\{\mu_i(y), \mu_j(y)\} \le \delta\,,$$

then $|Y_i \cap Y_j| \le [\delta/\mu] = T$, $i \ne j$; $T \ge 1$. Therefore the number of measures μ_i with $|Y_i| > T$ does not exceed $\binom{N}{T+1}$ and the number of measures μ_i with $|Y_i| \le T$ obviously does not exceed $\binom{N}{T}M_p(T,\delta)$. Therefore

$$M \le \binom{N}{T+1} + \binom{N}{T}M_p(T,\delta) \le \frac{N}{T}\binom{N}{T}M_p(T,\delta) \ . \tag{18}$$

Using inequality (7) to upperbound the value $M_p(T,\delta)$ and simple inequality $\binom{n}{k} \le (en/k)^k$ we get from (18)

$$M \le \frac{N(1-\delta)}{2}\exp\left\{T\ln\frac{2eN}{T(1-\delta)} - \ln T\right\} \quad .$$

The exponent function in the right side of last inequality is $\cap$–convex on T and moreover its derivative at $T = \delta/\mu$ is positive for $N \ge 2$. Therefore we can upperbound it replacing T by δ/μ, that gives the inequality (15). $\triangle$

Proof of Proposition 2. Let $(\mu_i\,;\, i = 1, \ldots, M)$ be a set of δ–pairwise separated measures on the set Y of cardinality N. It is clear that it contains not more than N measures μ_i with $\max_y \mu_i(y) > \delta$. Therefore below we consider only measures μ_i with $\max_y \mu_i(y) \le \delta$.

Fix now some μ and q such that $0 < \mu \le \delta < q < 1$ and put $Y_i(\mu) = \{y \in Y : \mu_i(y) \ge \mu\}$. Consider first all measures μ_i such that $\mu_i\left(Y_i^c(\mu)\right) \ge 1-q$. Denoting $M_1(\mu, q)$ the total number of such measures we introduce on their basis new probability measures ν_i with supports $Y_i^c(\mu)$:

$$\nu_i(y) = \frac{\mu_i(y)}{\mu_i(Y_i^c(\mu))} \ , \ y \in Y_i^c(\mu) \ ; \ i = 1, \dots, M_1 \quad .$$

For these measures for any $i\,,\,j = 1, \dots, M_1 \ ; \ i \ne j$, the following relations hold true:

$$\|\nu_i - \nu_j\| \ge 2\left\{1 - \frac{\delta}{(1-q)}\right\} \quad , \quad \max_{i,y} \nu_i(y) < \frac{\mu}{\mu_i(Y_i^c(\mu))} \le \frac{\mu}{(1-q)} \quad .$$

Therefore from Lemma 2a) we get the upperbound for the value $M_1(\mu, q)$:

$$M_1(\mu, q) \le \frac{\mu N}{((1-q)^2 - \delta\mu N)_+} \quad . \tag{19}$$

For remaining $M - M_1(\mu, q) = M_2(\mu, q)$ measures $\mu_i\,,\, i = 1, \dots, M_2$, we have $\mu_i(Y_i(\mu)) \ge q$. Since all measures values do not exceed δ then in every set $Y_i(\mu)$ there exists a subset $Y_i'(\mu)$ such that $q \le \mu_i(Y_i'(\mu)) \le \delta + q$. Introduce the new probability measures σ_i with supports $Y_i'(\mu)$:

$$\sigma_i(y) = \frac{\mu_i(y)}{\mu_i(Y_i'(\mu))} \ , \ a \in Y_i'(\mu) \ ; \ i = 1, \dots, M_2 \quad .$$

For these measures for any $i\,,\,j = 1, \dots, M_2 \ ; \ i \ne j$, and $y \in Y_i'(\mu)$ we have

$$\|\sigma_i - \sigma_j\| \ge 2\left(1 - \frac{\delta}{q}\right) \quad , \quad \sigma_i(y) \ge \frac{\mu}{\delta + q} \quad .$$

Therefore from Lemma 2b) we get the upperbound for the value $M_2(\mu, q)$:

$$M_2(\mu, q) \le \frac{N\mu}{2\delta} \exp\left\{\frac{\delta(\delta+q)}{\mu q} \ln \frac{2e\mu N q}{\delta(q-\delta)}\right\} \quad , \tag{20}$$

provided

$$\mu/(\delta+q) \ge \delta/Nq \quad . \tag{21}$$

It follows from (19) - (20) and the remark made at the Proposition's proof beginning that

$$M_p(N, \delta) \le N + \frac{\mu N}{((1-q)^2 - \delta\mu N)_+} + \tag{22}$$

$$+\frac{N\mu}{2\delta} \exp\left\{\frac{\delta(\delta+q)}{\mu q} \ln \frac{2e\mu N q}{\delta(q-\delta)}\right\} \quad .$$

Choosing now $\mu\delta N = (1-\sqrt{\delta})^3(1+\sqrt{\delta})$, $q = \sqrt{\delta}$, we get from (21) - (22) (and from (14) for $\delta^2 N \leq 1-\delta$) that

$$M_p(N,\delta) \leq N + \frac{1}{\delta^2} + \frac{1}{2\delta^2}\exp\left\{\frac{\delta^2 N}{(1-\sqrt{\delta})^3}\ln\frac{2e}{\delta^2}\right\} \, ,$$

for $\delta \leq 1/4$. If $1/4 < \delta < 1$ then this upperbound is weaker than upperbound (7) that concludes the proof of Proposition 2. $\triangle$

Statements of Propositions 1 – 2 are combained in the following theorem.

Theorem 2. *For any* $N \geq 2$ *and* $0 < \delta < 1$ *the following bounds are valid* :

$$\exp\{N\delta^2/(2e^2)\} \leq M_c(N,\delta) \leq$$

$$\leq M_p(N,\delta) \leq N + \frac{1}{\delta^2} + \frac{1}{2\delta^2}\exp\left\{\frac{\delta^2 N}{(1-\sqrt{\delta})^3}\ln\frac{2e}{\delta^2}\right\} \, . \tag{23}$$

Remark. Lower and upper bounds (23) are essentially oriented to the case when $\delta \to 0$ and $N\delta^2 \to \infty$. In that case, as can be seen from (23), the upperbound contains an additional factor of order $\ln(1/\delta)$. From that factor point of view it is not known even which of bounds (23) (lower or upper ?) can be improved. And the reason for that is quite serious. In order to understand it, assume that each probability measure μ_i is the equiprobable distribution on set Y_i and moreover, $|Y_i| = \delta N$, $i = 1,\ldots,M$. Then $\|\mu_i - \mu_j\| = 2(1-|Y_i \cap Y_j|/(\delta N) \geq 2(1-\delta)$ if $|Y_i \cap Y_j| \leq \delta^2 N$ for any $i \neq j$. Now what is the largest possible number M of such measures ?

This question is equivalent to the following question from coding theorem. We consider a *constant weight* $w = \delta N$ binary code of length N with minimal distance $\geq (1-\delta)w$. What is the maximal cardinality $M(N,\delta)$ of such code ? The best known (in coding theory) upperbound for the number $M(N,\delta)$ (it is the so called *Johnson's bound* or some its small improvements) contains already that factor $\ln(1/\delta)$! (see also [9]).

Historical notes. Notions of pairwise and completely separated families were introduced in [10]. They reflect some geometrical approach to identification problem considered in [5, 6]. Proposition 1 was proved in [10]. Proposition 2 was proved in [11].

1.3 Identification codes or How large is the set of all output measures for noisy channel ?

Identification codes. Let A and B be a finite input and output alphabets of a stationary memoryless channel with conditional transition probabilities $W(b|a), a \in A, b \in B$. If P is some probability distribution (measure) on the channel input A^n then by $Q = W^{(n)}P$ we denote the generated probability distribution on the channel output B^n.

Definition 3. *A collection $(P_i, \mathcal{D}_i, i = 1, \ldots, M)$ of probability measures P_i on A^n and regions $\mathcal{D}_i \subseteq B^n$ is called an (M, n, δ, W) – ID–code if the following conditions are satisfied*:

$$Q_i(\mathcal{D}_i) \geq 1 - \delta \ \text{ and } \ Q_i(\mathcal{D}_j) \leq \delta \ \text{ for any } \ i \neq j \ .$$

Remark. As can be seen it is allowed here for regions $\mathcal{D}_i$ to intersect between each other (for $\delta > 0$). We can generalize also ID–codes allowing to use randomized decisions (such possibility will be used once below). We didn't include such randomization in the definition of ID–codes in order to avoid some overcomplification (nevertheless, all upperbounds for the maximal cardinalities below remain valid for such generalized ID–codes as well).

The notion of ID–codes was first introduced in [5], where it was shown that the maximal cardinality $M(n, \delta, W)$ of ID–code of length n satisfies the following lowerbound.

Proposition 3.

$$\lim_{n \to \infty} \frac{\ln \ln M(n, \delta, W)}{n} \geq C(W) \ , \qquad 0 < \delta \leq 1 \ , \tag{24}$$

where $C(W)$ is the capacity of the channel W in natural units.

The converse of the inequality (24)

$$\lim_{n \to \infty} \frac{\ln \ln M(n, \delta, W)}{n} \leq C(W), \ \ 0 \leq \delta \leq \delta_0 \ , \tag{25}$$

where δ_0 is some unspecified positive constant, was obtained in [7]. It should be mentioned that earlier in [5] some weaker form of the inequality (25) was proved for $M(n, \delta_n, W)$, where $\delta_n \leq e^{-\epsilon n}$ and ϵ is any positive constant.

It follows from (24)–(25) that

$$\lim_{n \to \infty} \frac{\ln \ln M(n, \delta, W)}{n} = C(W), \ \ 0 < \delta \leq \delta_0 \ . \tag{26}$$

Such a double exponential growth rate of the cardinality of ID–code (in contrast to an ordinary exponential one for usual codes) created a certain theoretical and practical interest to these codes [5 – 11]. ID–codes are also close to some cryptography problems [9, 11].

Our more general geometrical approach to ID–codes was introduced in [10]. That approach is based on a certain equivalence between ID–codes and some families of "almost orthogonal" measures. Such an approach (and its natural connection with testing of composite hypotheses in mathematical statistics) not only enlarges the research analytical apparatus but also enables us to strengthen some results from [5, 7]. In particular, we will show that the inequality (25) remains valid for any $0 \le \delta < 1/2$.

As a result we will be able to prove the following

Theorem 3.

$$\lim_{n\to\infty} \frac{\ln \ln M(n,\delta,W)}{n} = \begin{cases} 0 & , \quad \delta = 0 \quad ; \\ C(W) & , \quad 0 < \delta < 1/2 \quad ; \\ \infty & , \quad 1/2 \le \delta \le 1 \quad . \end{cases} \tag{27}$$

Before going to the most interesting case of $0 < \delta < 1/2$ let us consider the remaining cases.

If $\delta = 0$ then obviously there are not more than $|A|^n$ orthogonal measures on the output B^n of the channel and therefore $M(n, \delta = 0) \le |A|^n$ from where the first line of (27) follow.

If $\delta \ge 1/2$ then we can use any number M of input measures $\{P_i\,,\, i = 1,\ldots,M\}$ (e.g., all of them may even coincide) and make the following stupid decision at the output of channel: independently of the output we make a decision in favor of measure P_i with probability $1/2$ and against it with the same probability $1/2$. Such a strange ID–code with randomized decion rules will provide all error probabilities equal to $1/2$ from where the last line of (27) follows (this is the example from [5], p. 16).

Let us consider now the most interesting case $0 < \delta < 1/2$.

How large is the set of all output measures for noisy channel ?

For a noisy channel we are able to choose only any *input* distribution P on A^n and so we are able to get *output* distributions only of the form $Q = W^{(n)}P$. Therefore the set $\mathcal{P}_{\text{out}}(B^n)$ of all possible output distributions on B^n is $\mathcal{P}_{\text{out}}(B^n) = W^{(n)}\mathcal{P}(A^n)$.

In order to stay in our geometrical framework we need to introduce a natural definitions of pairwise and completely separated families of probability measures for noisy channel W.

Definition 1'. A collection $(P_i\,, i = 1,\ldots,M)$ of probability measures P_i on A^n (or collection $(Q_i\,, i = 1,\ldots,M)$ of probability measures $Q_i = W^{(n)}P_i$ on B^n) is called (M,n,δ,W) – *pairwise separated* collection if for any $i \ne j$ the following condition is satisfied:

$$\left\| W^{(n)}P_i - W^{(n)}P_j \right\| = \|Q_i - Q_j\| \ge 2(1-\delta)\ . \tag{28}$$

Definition 2'. A collection $(P_i\,, i = 1,\ldots,M)$ of probability measures P_i on A^n (or collection $(Q_i\,, i = 1,\ldots,M)$ of probability measures $Q_i =$

$W^{(n)}P_i$ on B^n) is called (M, n, δ, W) – *completely separated* collection if for any i the following condition is satisfied:

$$\left\| W^{(n)}P_i - \text{conv}\left\{W^{(n)}P_j \, , \, j \neq i\right\}\right\| =$$

$$= \|Q_i - \text{conv}\,\{Q_j \, , \, j \neq i\}\| \geq 2(1-\delta) \; . \tag{29}$$

Proposition 4. For any (M, n, δ, W) – completely separated collection $\{P_i\}$ (or $\{Q_i\}$) it is possible to define regions $\{\mathcal{D}_i \subseteq B^n\}$ such that $\{P_i \, , \mathcal{D}_i\}$ will be (M, n, δ, W) – ID–code. Conversely, any (M, n, δ, W) – ID–code is $(M, n, 2\delta, W)$ – completely separated collection.

Proof. Let $\{P_i\}$ (and therefore $\{Q_i\}$ as well) be (M, n, δ, W) – completely separated collection. Then as the region $\mathcal{D}_i$ we choose the set D giving the minimum to the sum of error probabilities when testing Q_i against all remaining measures $\{Q_j \, , \, j \neq i\}$. Due to (2) that sum will not exceed δ and therefore each error probability also will not exceed δ. It means that such collection $\{Q_i \, , \mathcal{D}_i\}$ will be (M, n, δ, W) – ID–code. Converse part of proposition is obvious. $\triangle$

Remark. The factor 2 that comes out when from (M, n, δ, W) – ID–code we get $(M, n, 2\delta, W)$ – completely separated collection is the reason why the value $\delta = 1/2$ becomes the critical value for the maximal cardinality $M(n, \delta, W)$ of ID–code (cf. Theorem 3).

We denote also by $M_p(n, \delta, W)$ and $M_c(n, \delta, W)$ maximal possible cardinalities of pairwise and completely δ–separated families of probability measures for noisy channel W , respectively. Then from Proposition 4 similar to (5) we have for $0 < \delta < 1/2$:

$$M_c(n, \delta, W) \leq M(n, \delta, W) \leq M_c(n, 2\delta, W) \leq M_p(n, 2\delta, W) \; . \tag{30}$$

Due to (30) instead of investigation the maximal cardinality $M(n, \delta, W)$ of ID–codes (and so dealing with error probabilities and regions $\mathcal{D}_i$) it is sufficient to investigate the maximal cardinality $M_c(n, 2\delta, W)$ of completely separated family of measures on the output set $\mathcal{P}_{\text{out}}(B^n) = W^{(n)}\mathcal{P}(A^n)$.

We start that investigation first with getting the lower bound (24). Its validity follows from (30) and the folowing

Proposition 5. *For any $\delta > 0$ and $R < C$ there exists $n_0(R, \delta)$ such that for any $n > n_0(R, \delta)$ the following inequality holds true :*

$$M_c(n, \delta, W) \geq \exp\left\{\frac{\delta^2 e^{nR}}{20}\right\} \; ; \; R < C \; , \; \delta > 0 \; , \; n > n_0(R, \delta) \; .$$

Proof. Due to the direct part of the well–known *coding theorem* [13, 14] for any $R < C, \;\; \lambda > 0,$ there exists some $n_0(R, \lambda)$ such that for any $n > n_0(R,\lambda)$ there exist $M > e^{nR}$ input blocks $i \in A^n$ generating measures $\{Q_i\}$ from $\mathcal{P}_{\rm out}(B^n) = W^{(n)}\mathcal{P}(A^n)$ and regions $\mathcal{D}_i \subseteq B^n$, $i = 1,\ldots,M$, such that for any $i,j = 1,\ldots,M$; $i \neq j$, the following conditions are satisfyed

$$\mathcal{D}_i \cap \mathcal{D}_j = \emptyset \ , \ \ Q_i(\mathcal{D}_i) \geq 1-\lambda \ .$$

Now we are almost in the situation of Section 1.2 where we constructed δ–completely separated family of measures on the alphabet Y (i.e. for the noiseless channel). Instead of the alphabet Y we have now the family of measures $\{Q_i\}$ and their "almost" supports $\{\mathcal{D}_i\}$. We can make that analogy complete if we replace (for a while) every measure Q_i by another probability measure Q_i' in the following way:

$$Q_i'(y) = \frac{Q_i(y)}{Q_i(\mathcal{D}_i)} \ \text{ for } \ y \in \mathcal{D}_i \ ; \ \ Q_i'(y) = 0 \ \text{ for } \ y \notin \mathcal{D}_i \ ,$$

(another words, we bound the measure Q_i to the set $\mathcal{D}_i$ and normalize it to have the probability measure). Now we are exactly in the situation of Section 1.2 with the collection of sets $\{\mathcal{D}_i\}$ (or the collection of measures $\{Q_i'\}$) instead of the alphabet Y. It is easy to understand now that any δ–completely separated collection build on the "alphabet" $\{Q_i'\}$ generates a similar $(1-(1-\delta)(1-\lambda) = \delta+\lambda-\delta\lambda)$– completely separated collection build on the "alphabet" $\{Q_i\}$. Therefore denoting $\epsilon = \delta+\lambda-\delta\lambda$, we get from Proposition 1

$$M_c(n,\epsilon,W) \geq \exp\left\{\frac{(\epsilon-\lambda)^2 e^{nR}}{2(1-\lambda)^2 e^2}\right\} \ ; \ R < C \ , \ \epsilon > 0 \ , \ n > n_0(R,\lambda) \ .$$

Putting finally $\lambda = \epsilon/10$, we get the assertion of Proposition 5. $\triangle$

Remark. It is not difficult to upperbound the function $n_0(R,\lambda)$. For example, from [13] (Problem 5.23) follows that

$$n_0(R,\lambda) \leq \frac{(1+4\ln^2 B)}{(C-R)_+^2} \ln \frac{4}{\lambda} \ .$$

We switch now to the proof of converse of (24) for any $0 < \delta < 1/2$.

It is a good example of a problem that, *if formulated in a right form*, will be reduced to an essentially technical (in mathematical sense) task. In order to ask ourselves such a *right question*, let $f_i(y)$ be the generated probability measure on the channel output $Y = B^n$ when input block $i \in A^n$ is used. Assume that we put some prior distribution $\{\pi_i\}$ on the input set A^n and therefore get the output distribution $Q_\pi(y) = \sum \pi_i f_i(y)$ on B^n. Generally, dimension of the input distribution $\{\pi_i\}$ is equal to A^n. But do we really need to use *all blocks* from A^n in order to approximate (in L_1–metrics) the generated distribution $Q_\pi(y) = \sum \pi_i f_i(y)$ on B^n ? And the answer is *no*.

It turns out that for that purpose it is sufficient to use only the order of e^{nC} input blocks, where $C = C(W)$ is the channel capacity.

Essentially, this kind of result was first obtained in [7] (Lemma 1) where authors showed that any output distribution can be rather accurately approximated by using some input distribution $\{\pi_i\}$ on A^n whose masses take values on a lattice with a span of the order of e^{-nC}.

We shall give here another proof of an equivalent (but a more geometrical) result from [10]. It can be formulated in the following way.

Proposition 6. *For any $\delta > 0$ there exists $n_0(\delta)$ such that for any output measure Q on B^n, $n \geq n_0(\delta)$, there exist $B^{n\delta}e^{nC}$ input blocks $i \in A^n$ such that for their generated measures $\{Q_i\}$ the following inequality holds true :*

$$\left\| Q - \mathrm{conv}\left\{ Q_i, i = 1, \ldots, B^{n\delta}e^{nC} \right\} \right\| \leq \delta \, . \tag{31}$$

In other words, any output distribution Q can be arbitrary closely approximated by using only the order of e^{nC} of input blocks. We should emphasize here that blocks used for that purpose, generally speaking, *depend on measure* Q, but the number $n_0(\delta)$ *does not depend on* Q,

Putting aside for a while the proof of that Proposition, we shall show now how to get necessary corrolaries from it.

We notice first that any channel W (or $W^{(n)}$) acts like a "compressing" operator in the following sense.

Lemma 3. *For any channel W, any pair of input distributions P_1, P_2 and corresponding pair of output distributions Q_1, Q_2 the following inequality holds true :*

$$\|Q_1 - Q_2\| \leq \|P_1 - P_2\| \, . \tag{32}$$

Indeed

$$\|Q_1 - Q_2\| = \sum_Y \left| \sum_X W(y|x)\left(p_1(x) - p_2(x)\right) \right| \leq$$

$$\leq \sum_X |p_1(x) - p_2(x)| \sum_Y W(y|x) = \sum_X |p_1(x) - p_2(x)| = \|P_1 - P_2\| \, . \; \triangle$$

Now we can prove the converse result. It will follow from (30) and the following

Proposition 7. *For any $0 < \delta < 1/2$ the following inequality holds true:*

$$\lim_{n\to\infty} \frac{\ln\ln M_p(n, 2\delta, W)}{n} \leq C(W) \, , \quad 0 < \delta < 1/2 \, . \tag{33}$$

Proof. Let $\{Q_i, i = 1, \ldots, M\}$ be some 2δ – pairwise separated collection of output measures. We fix some small $\epsilon > 0$ such that $2(\delta + \epsilon) < 1$. By virtue of Proposition 6 any measure Q_i can be ϵ–approximated by another output measure Q'_i where the measure Q'_i is generated by some $B^{n\epsilon}e^{nC}$ input blocks from A^n. Clear that the collection $\{Q'_i, i = 1,\ldots,M\}$ is $2(\delta+\epsilon)$ – pairwise separated. Since the channel $W^{(n)}$ is a "compressing" operator (Lemma 3), the maximal number of $2(\delta+\epsilon)$ – pairwise separated measures, generated by every collection of $N = B^{n\epsilon}e^{nC}$ input blocks, is upperbounded by formula (7). The total number of collections with the cardinality N on the alphabet A^n does not exceed A^{nN}. Therefore, for the maximal possible cardinality $M_p(n, 2\delta, W)$ of pairwise separated collection we get ($N = B^{n\epsilon}e^{nC}$, $2(\delta+\epsilon) < 1$)

$$M_p(n, 2\delta, W) < (2/(1-2\delta-2\epsilon))^N A^{nN} \leq$$

$$\leq \exp\left\{\left(\ln\frac{2}{(1-2\delta-2\epsilon)} + n\ln A\right)e^{n(C+\epsilon\ln B)}\right\} \quad , \quad n > n_0(\epsilon) ,$$

from where relation (33) and (25) follow. $\triangle$

Remark. It was sufficient for us to use here the upperbpund (7) instead of a much tighter upperbound (13). The reason is that *log log* kind of asymptotics in *Identification* (e.g. (33)) is very insensitive to their difference. Absolutely different situation is in a closely related *Authentication* [11].

Now for the purpose of completeness we shall show also how Lemma 1 from [7] follows from Proposition 6.

Corollary 1. *Any output distribution Q on B^n can be arbitrary closely approximated by using some input distribution $\{p_i, i = 1,\ldots,N\}, N \sim e^{nC}$, on A^n, taking values only of the form $p_i = j/N$, j – integer.*

Proof. For any $\epsilon > 0$ and any distribution $\{p_i, i = 1,\ldots,N\}$ there exists some distribution $\{\hat{p}_i\}$, taking values only of the form $\hat{p}_i = j\epsilon/N$ with $|p_i - \hat{p}_i| \leq \epsilon/N$, and moreover,

$$\|p - \hat{p}\| = \sum_{i=1}^{N} |p_i - \hat{p}_i| \leq \epsilon . \tag{34}$$

Indeed, we choose $\hat{p}_1 = j\epsilon/N$ with the minimal possible $|p_1 - \hat{p}_1|$. Then $|p_1 - \hat{p}_1| \leq \epsilon/(2N)$. Now we choose $\hat{p}_2$ of the same form with the minimal possible $|(p_1+p_2)-(\hat{p}_1+\hat{p}_2)|$. Then $|(p_1+p_2)-(\hat{p}_1+\hat{p}_2)| \leq \epsilon/(2N)$ and $|p_2 - \hat{p}_2| \leq \epsilon/(2N)$. Repeating this process, we get as a result the collection $\{\hat{p}_i\}$, having desired properties. Since the channel is a compressing operator, for measures Q and $\hat{Q}$, generated by distributions $\{p_i\}$ and $\{\hat{p}_i\}$, respectively, due to (34), we have $\|Q - \hat{Q}\| \leq \epsilon$. From this result and Proposition 6 the assertion of Corollary 1 follows. $\triangle$

It remained us only to establish the validity of Proposition 6. The proof below is essentially technical and is based on "random choice" of approximating input distribution.

Proof of Proposition 6. We shall need a few simple pure statistical lemmas.

Let Y be a finite alphabet and there are given K probability distributions $f_i(y),\ i=1,\dots,K$, on it with prior probabilities $\{p_i\}$. Consider also the "averaged" distribution

$$p(y)=\sum_{i=1}^{K} p_i f_i(y)\ .$$

We choose now randomly and independently s distributions from $\{f_i\}$ with respective probabilities $\{p_i\}$ (with returns) and put

$$\hat{p}(y)=\frac{1}{s}\sum_{i=1}^{K}\nu_i f_i(y)\ ,$$

where $\nu_i=$ number of measures f_i among s chosen distributions.

Lemma 4. *The following estimate on the variance of $\hat{p}(y)$ holds :*

$$E|p(y)-\hat{p}(y)|^2\le\frac{1}{s}\sum_{j=1}^{K}p_j(1-p_j)f_j^2(y)\ ,\ y\in Y\ . \tag{35}$$

Proof. Let μ_l be the index of measure chosen from $\{f_i\}$ on the l–th step. Then $\mu_1,\dots,\mu_s$ are i.i.d.r.v.'s, and moreover

$$Ef_{\mu_l}=p(y)\quad ,\quad \sum_{j=1}^{K}\nu_j f_j(y)=\sum_{l=1}^{s}f_{\mu_l}(y)\ .$$

Therefore,

$$E|\hat{p}(y)-p(y)|^2=E\left|\frac{1}{s}\sum_{l=1}^{s}\left(f_{\mu_l}(y)-p(y)\right)\right|^2=$$

$$=\frac{1}{s}E\left(f_{\mu_1}-p(y)\right)^2=\frac{1}{s}\left(Ef_{\mu_1}^2(y)-p^2(y)\right)=$$

$$=\frac{1}{s}\left\{\sum_{j=1}^{K}p_jf_j^2(y)-\left[\sum_{j=1}^{K}p_jf_j(y)\right]^2\right\}\le\frac{1}{s}\sum_{j=1}^{K}p_j(1-p_j)f_j^2(y)\ .\ \triangle$$

Now for $\epsilon>0$ and $i=1,\dots,K$, we choose some sets $Y_i(\epsilon)$ such that $P_i\{Y_i(\epsilon)\}\ge 1-\epsilon\ ,\ i=1,\dots,K$, and put

$$K(y)=\{i:y\in Y_i(\epsilon)\}\ ,\ \ Y(\epsilon)=\bigcup_{i=1}^{K}Y_i(\epsilon)\ \ .$$

Lemma 5. *For any $\epsilon > 0$ the following estimate is valid* :

$$\sum_Y E|\hat{p}(y) - p(y)| \le 2\epsilon + \left(\frac{|Y(\epsilon)|}{s} \max_{\substack{j=1,\dots,K \\ y \in Y_j(\epsilon)}} f_j(y) \right)^{1/2} \tag{36}$$

($|A|$ means the cardinality of the set A).

Proof. We have

$$\sum_Y E|\hat{p}(y) - p(y)| \le \sum_{Y(\epsilon)} E \left| \sum_{j \in K(y)} \left(\frac{\nu_j}{s} - p_j \right) f_j(y) \right| +$$

$$+ \sum_Y E \left| \sum_{j \notin K(y)} \left(\frac{\nu_j}{s} - p_j \right) f_j(y) \right| = \Sigma_1 + \Sigma_2 \ .$$

Using (35) and Cauchy–Buniakovsky inequality we get for Σ_1

$$\Sigma_1 \le \sum_{Y(\epsilon)} \left[E \left| \sum_{j \in K(y)} \left(\frac{\nu_j}{s} - p_j \right) f_j(y) \right|^2 \right]^{1/2} \le$$

$$\le \sum_{Y(\epsilon)} \left(\frac{1}{s} \sum_{j \in K(y)} p_j f_j^2(y) \right)^{1/2} \le \left(\frac{|Y(\epsilon)|}{s} \sum_{Y(\epsilon)} \sum_{j \in K(y)} p_j f_j^2(y) \right)^{1/2} \le$$

$$\le \left(\frac{|Y(\epsilon)|}{s} \sum_{j=1}^{K} p_j \sum_{Y_j} f_j^2(y) \right)^{1/2} \le \left(\frac{|Y(\epsilon)|}{s} \max_{\substack{j=1,\dots,K \\ y \in Y_j(\epsilon)}} f_j(y) \right)^{1/2} .$$

For Σ_2 we have obviously

$$\Sigma_2 \le \sum_Y \sum_{j \notin K(y)} E \left| \frac{\nu_j}{s} - p_j \right| f_j(y) \le 2 \sum_Y \sum_{j \notin K(y)} p_j f_j(y) =$$

$$= 2 \sum_{j=1}^{K} p_j \sum_{Y \setminus Y_j(\epsilon)} f_j(y) \le 2\epsilon \ .$$

Combaning these two inequalities we get (25) . $\triangle$

In the setting of a noisy channel measure $f_i(y)$ represents the conditional distribution of the channel output $y \in B^n$ given the "input" $i \in A^n$. As sets $Y_i(\epsilon) \subseteq B^n$ some "supports" of measures $f_i(y)$ will be chosen. In that meaning Lemma 5 is supposed to be used for the same type codeblocks i. The next result generalizes it for codeblocks of different types.

Let L classes of measures $\{f_{li}(y)\ ,\ i=1,\ldots,K(l)\}$, $l=1,\ldots,L$, with prior probabilities $\{p_{li}\}$ are given and

$$p(y)=\sum_{l=1}^{L}\sum_{i=1}^{K(l)} p_{li}\,f_{li}(y)\ .$$

In order to approximate $p(y)$ we choose in every class l randomly and independently s_l measures f_{li} proportionally to the distribution p_{li}. Moreover, for every pair l,i we choose also some set $Y_{li}(\epsilon)$ such that $P_{li}\{Y_{li}\}\geq 1-\epsilon$ and put (ν_{li} = number of measures f_{li} among s_l chosen distributions)

$$\hat{p}(y)=\sum_{l=1}^{L}\frac{p_l}{s_l}\sum_{i=1}^{K(l)}\nu_{li}f_{li}(y)\ ,\quad p_l=\sum_{i=1}^{K(l)}p_{li}\ ,\quad Y_l(\epsilon)=\bigcup_{i=1}^{K(l)}Y_{li}(\epsilon)\ .$$

Lemma 6. *For any $\epsilon>0$ the following estimate is valid* :

$$\sum_{Y} E|\hat{p}(y)-p(y)|\leq 2\epsilon+\left(\sum_{l=1}^{L}\frac{p_l|Y_l(\epsilon)|}{s_l}\max_{\substack{i=1,\ldots,K(l)\\ y\in Y_{li}(\epsilon)}} f_{li}(y)\right)^{1/2}\ . \qquad (37)$$

Proof. We have from Lemma 5:

$$\sum_{Y} E|\hat{p}(y)-p(y)|\leq 2\epsilon+\sum_{l=1}^{L}p_l\left(\frac{|Y_l(\epsilon)|}{s_l}\max_{\substack{i=1,\ldots,K(l)\\ y\in Y_{li}(\epsilon)}} f_{li}(y)\right)^{1/2}\leq$$

$$\leq 2\epsilon+\left(\sum_{l=1}^{L}\frac{p_l|Y_l(\epsilon)|}{s_l}\max_{\substack{i=1,\ldots,K(l)\\ y\in Y_{li}(\epsilon)}} f_{li}(y)\right)^{1/2}\ .\ \triangle$$

Now, we return to the channel $W^{(n)}$ with input A^n and output B^n alphabets. Application of Lemma 6 to that case is rather standard and it is based on considering codeblocks from A^n of the same type (composition) (see [14], Ch. 1.2 and 2.1 or [5]). Necessary formulas (38) below can be also found in [14] or [5].

As usual, we partition all codeblocks from A^n on classes l consisting of codeblocks of the same composition (type). In every pair l,i the parameter i will denote the block's index inside the class l. An input block (l,i) generates the distribution f_{li} on the channel output B^n. As a set $Y_{li} \subseteq B^n$ we choose a set of "almost 1" probability and consisting of approximately equiprobable points. More precisely, we fix some arbitrary small $\epsilon > 0$ and let P_l denotes the type of codeblock (l,i) [13] and Q_{li} denotes the output distribution generated by that codeblock. Then for any $n \geq n_0(\epsilon)$ it is possible to choose some sets $Y_{li}(\epsilon) \subseteq B^n$ with $Q_{li}\{Y_{li}(\epsilon)\} \geq 1-\epsilon$ such that simultaneously for all l,i the following conditions will be satisfied:

$$|Y_{li}(\epsilon)| < \exp\{n(1+\epsilon)H(W|P_l)\} \, ,$$

$$\max_{\substack{i=1,\dots,K(l)\\ y \in Y_{li}(\epsilon)}} f_{li}(y) < \exp\{-n(1-\epsilon)H(W|P_l)\} \, , \tag{38}$$

$$|Y_l(\epsilon)| = \left| \bigcup_i Y_{li}(\epsilon) \right| < \exp\left\{n(1+\epsilon)H(WP_l)\right\} \, .$$

Now we put

$$s_l = \epsilon^{-2} \exp\left\{n\left[I(P_l, W) + \epsilon H(WP_l) + \epsilon H(W|P_l)\right]\right\} \, . \tag{39}$$

Then taking into account that $H(WP_l) - H(W|P_l) = I(P_l, W) \leq C$, $H(WP_l) + H(W|P_l) \leq 2\ln B$ and that the total number of classes $L \leq (n+1)^{|A|}$, we have from (38)–(39) for $n > n_0(\epsilon)$

$$s = \sum_{l=1}^{L} s_l \leq \epsilon^{-2} L \exp\{n[C + 2\epsilon \ln B]\} \leq B^{3\epsilon n} e^{nC} \, . \tag{40}$$

Similarly we have from (37)–(39)

$$\sum_{B^n} E|\hat{p}(y) - p(y)| \leq 3\epsilon \, . \tag{41}$$

Now from (40) - (41) Proposition 6 follows. $\triangle$

On a "basis" for the set of all output measures. In connection with Proposition 6 some natural question arises: is it possible to choose some *universal collection* of $N \sim e^{nC}$ input blocks (like a basis) such that using them it is possible arbitrary closely to approximate *any* output distribution ? The answer, generally speaking, is *no.* To see this consider the example of the binary symmetrical channel with crossover probability $p < 1/2$.

Let x_0 be the all-zero input block of the length n and $x_1, \dots, x_N$ be all possible input blocks of the Hamming weight $w(x_i) \geq d$ (d will be chosen later). Let also $\{Q_i, i = 0,1,\dots,N\}$ be the generated measures on the channel output B^n, respectively. We shall evaluate now how large should be d such that it will be impossible to approximate sufficiently accurately the measure Q_0 with the help of all remaining measures $\{Q_i, i=1,\dots,N\}$. It is convenient to use here some statistical interpretation and the formula (1). Let us consider the problem of testing hypothesis Q_0 against $\{Q_i, i = 1,\dots,N\}$. Notice first that if the vector x_i, $i = 0,1,\dots,N$, is transmitted then for the Hamming weight $w(y)$ of output vector y we have

$$E\left(w(y)\,|x_i\right) = w(x_i) + p\left(n - 2w(x_i)\right) \ ,$$

$$E\Big\{[w(y) - E(w(y)|x_i)]^2|x_i\Big\} = p(1-p)n \ .$$

Now we choose as the decision set $\mathcal{D}_0$ of acceptance hypothesis Q_0 the following set:

$$\mathcal{D}_0 = \Big\{y : w(y) \le pn + [2znp(1-p)]^{1/2}\Big\} \ ,$$

where the parameter $z = z(\delta) > 0$ will be chosen later. Then, due to the central limit theorem, we have

$$Q_0(Y^n \setminus \mathcal{D}_0) = \Phi\left(-\sqrt{2z}\right) + o(1) \ , \ \ n \to \infty \ , \tag{42}$$

$$Q_i(\mathcal{D}_0) = \Phi\left[-\frac{w(x_i)(1-2p)}{\sqrt{np(1-p)}} + \sqrt{2z}\right] + o(1) \ , \ \ i = 1,\dots,N \ .$$

Now, if $w^2(x_i)(1-2p)^2 \ge 8znp(1-p)$, $i = 1,...,N$, then due to (42) we have from formula (1)

$$\begin{aligned} &\|Q_0 - \mathrm{conv}\,\{Q_i,\, i=1,\dots,N\}\| \ge \\ &\ge 2\left[1 - Q_0(Y^n \setminus \mathcal{D}_0) - \mathrm{conv}\,\{Q_i,\, i=1,\dots,N\}(\mathcal{D}_0)\right] \ge \\ &\ge 2\left[1 - 2\Phi\left(-\sqrt{2z}\right)\right] + o(1) \ , \ \ n \to \infty \ . \end{aligned} \tag{43}$$

We use now simple upperbound $2\Phi(-x) \le e^{-x^2/2}$, $x > 0$, and put $z = \ln(1/\delta)$. Then we get from (43)

$$\|Q_0 - \mathrm{conv}\,\{Q_i,\, i=1,\dots,N\}\| \ge 2(1-\delta) \ . \tag{44}$$

Therefore, if the minimum distance d of arbitrary code $\{x_1, ..., x_N\}$ satisfies the condition $d^2(1-2p)^2 \ge 8np(1-p)\ln(1/\delta)$, then the collection $\{Q_i, i = 1,\dots,N\}$ is (N,δ) – completely separated. It is easy to understand that the maximal cardinality N of such a collection has the order of $\exp\{n\ln 2 - a(\delta,p)n^{1/2}\ln n\}$, i.e. , it differs negligibly from the cardinality 2^n of the whole space of input blocks.

Completely and pairwise separated collections of measures. As we know already, ID–codes and completely separated collections of measures are essentially equivalent. But it is rather difficult to check the complete separability of measures (e.g. condition (4)). It is much easier to check the pairwise separability of measures. Then the following question naturally arises: if we are given some collection of δ – pairwise separated measures, is it possible then to select some its subcollection such that it will be ϵ – completely separated (with ϵ close to δ) and will have "practically" the same cardinality ?

Or, equivalently, under which conditions a collection of δ – pairwise separated measures is simultaneously a collection of ϵ – completely separated measures (with ϵ slightly worse than δ) ?

¿From statistics point of view similar question can be formulated in the following way. Let some set $\{Q_i, i = 0, 1, \ldots, N\}$ of probability measures be given. Assume that when we test hypothesis Q_0 against any simple hypothesis Q_i we can achieve sum of error probabilities $\delta(Q_0, Q_i) \leq \delta$. Then due to formula (1) we have $\|Q_0 - Q_i\| \geq 2(1-\delta)$ for any $i = 1, \ldots, N,$. What can we say about the minimal sum of error probabilities when we test the hypothesis Q_0 against *all* hypothesis Q_i, $i = 1, \ldots, N$, ? Or, due to formula (2), what can we say about $\|Q_0 - \text{conv}\,\{Q_i, i = 1, \ldots, N\}\|$? Clear that this new sum $\delta(Q_0, \text{conv}\{Q_i\})$ will be $\geq \delta$, but when we can expect that it will be rather close to δ ?

Before giving a partial answer to this question we bring two examples, showing what kind of results can be expected here.

Example 1. Let there be given the equiprobable distribution P_0 and measures P_i, $i = 1, \ldots, |X|$, on the alphabet X , such that $P_i\{x_i\} = 1$. Then $\|P_0 - P_i\| = 2 - 2|X|^{-1} \geq 2 - 2\delta$, if $|X| \geq 1/\delta$. On the other hand, it is clear that $\|P_0 - \text{conv}\,\{P_i,\ i = 1, \ldots, |X|\}\| = 0$, i.e., this collection of measures is not completely separable for any $\delta < 1$. From ID–codes point of view for any choice of a nonempty region $\mathcal{D}_0$ we shall have $P_i(\mathcal{D}_0) = 1$ for some $i > 0$. But if we remove the measure P_0 from that collection, the remaining orthogonal measures will represent completely separated collection.

It becomes clear from the next (a more meaningful example) that it is hardly possible to get some easy–to check necessary and sufficient conditions in this problem. The infinite alphabet used in that example is not crucial. It is always possible to approximate it arbitrary precisely by a sufficiently large finite alphabet.

Example 2. Let there be given gaussian measures P_i, $i = 1, \ldots, N$, in Euclidean space R^n with the identity covariance matrix I_n and the mean vector $A^{1/2}\mathbf{e}_i$, where $\mathbf{e}_i$ is the n – dimensional vector with the i–th coordinate be 1 and all the remaining coordinates be zeros. That model corresponds to the observation of one of N orthogonal signals with energy A in white gaussian noise. Then with the help of the formula (7) it is simple to get

$$\|P_i - P_j\| = 2\left[1 - 2P\left\{\xi > (A/2)^{1/2}\right\}\right] = \tag{45}$$
$$= 2\left[2\Phi\left((A/2)^{1/2}\right) - 1\right],\ i \neq j\ ,$$

where ξ is the gaussian random variable with parameters $(0,1)$ and $\Phi(x)$ is the standard distribution function of this random variable.

It is hardly possible to calculate explicitly the distance between the measure P_0 and the convex combination of all remaining measures. But it is possible to upperbound it. Putting on the set $P_j,\ j \neq i$, the equiprobable prior distribution and using the simple inequality

$$\|P - Q\|^2 \leq E_P\left(\frac{dQ}{dP}\right)^2 - 1\ ,$$

we get after some simple calculations (see details in [4])

$$\|P_i - \operatorname{conv}\{P_j, j \neq i\}\|^2 \leq (N-1)^{-1}e^{2A}\ . \tag{46}$$

We can see from (45) that when A is sufficiently large, measures $\{P_i,\ i = 1,\ldots,N\}$ are pairwise almost orthogonal and hence are well pairwise separated. But from (46) it follows that if the number of measures N is much larger than $\exp\{2A\}$, then this collection of measures is not completely separated for small δ. Particularly, for such N it is impossible to discriminate hypotheses P_i and $\{P_j,\ j \neq i\}$ with small error probabilities. In fact, it is possible to show [15] that the true critical value here is $\exp\{A\}$.

We present now a sufficient condition for a complete separability of measures which is rather easy to check in some cases.

Proposition 8. *Let there be given probability measures $P_i,\ i = 1,\ldots,N$, on a finite alphabet X. For any subset $A \subseteq X$ we denote*

$$\epsilon_{ij}(A) = \sum_A P_i(x)P_j(x)\ .$$

Then for any i the following estimate holds true:

$$\|P_i - \operatorname{conv}\{P_j, j \neq i\}\| \geq 2\max_A\left[P_i(A) - \left(|A|\max_{j\neq i}\epsilon_{ij}(A)\right)^{1/2}\right] \tag{47}$$

and $\{P_i,\ i = 1,\ldots,N\}$ is an (N,δ) – completely separated collection with

$$\delta = \max_i \min_A\left[P_i(X \setminus A) + \left(|A|\max_{j\neq i}\epsilon_{ij}(A)\right)^{1/2}\right]\ . \tag{48}$$

Proof. Let $Q = \sum c_i P_i$ be any convex combination of measures $P_i,\ i = 1,\ldots,N$, and $A \subseteq X$. Using representation $|p - q| = p + q - 2\min\{p,q,\}$

and the Cauchy-Bunyakowski inequality we carry out the following chain of calculations:

$$\|P_1 - Q\| = 2\left[1 - \sum_X \min\{P_1(x), Q(x)\}\right] \geq$$

$$\geq 2\left[P_1(A) - \sum_A \min\{P_1(x), Q(x)\}\right] \geq 2\left[P_1(A) - \sum_A (P_1(x)Q(x))^{1/2}\right] \geq$$

$$\geq 2\left[P_1(A) - \left(|A| \sum_A P_1(x)Q(x)\right)^{1/2}\right] \geq$$

$$\geq 2\left[P_1(A) - \left(|A| \max_{i \geq 2} \epsilon_{1i}(A)\right)^{1/2}\right] \,,$$

from where formulas (47)–(48) follow. $\triangle$

APPENDIX

1. We show that the set $D(P_0)$ in the proof of the inequality (8) has the minimum volume when P_0 is an extreme point of the set $\mathcal{P}_N$.

Let P_i , $i = 1, \dots, N$, denotes the N – dimensional vector whose i–th coordinate is 1 and all the remaining coordinates are 0. We fix arbitrary $a > 0$ and consider sets

$$V_i = \{x \in \mathcal{P}_N : \|P_i - x\| \leq a\} \ , \ i = 1, \dots, N \ .$$

Let $c = (c_1, \dots, c_N)$ be an arbitrary probability vector. Introduce sets

$$V(c) = \sum_{i=1}^{N} c_i V_i = \left\{ y \in \mathcal{P}_N : y = \sum_{i=1}^{N} c_i y_i \ ; y \in V_i \ , i = 1, \dots, N \right\} \ ,$$

$$D(c) = \{y \in \mathcal{P}_N : \|c - y\| \leq a\} \ .$$

Obviously, sets V_i , $V(c)$ and $D(c)$ are convex. It is simple also to check that $V(c) \subseteq D(c)$. Designating $v(A)$ the set's A volume and taking into account that volumes $v(V_i)$, $i = 1, \dots, N$ are equal, we get by virtue of Brunn-Minkowski theorem [12]

$$v(D(c)) \geq v(V(c)) \geq \left[\sum_{i=1}^{N} c_i v^{1/N}(V_i) \right]^N = v(V_1) \ . \qquad \triangle$$

References

[1] A. Wald. *Statistical Decision Functions.* Wiley, New York (1950).

[2] E. L. Leman. *Testing Statistical Hypotheses.* Chapman & Hall, New York (1959).

[3] P. J. Huber and V. Strassen, "Minimax Tests and The Neyman–Pearson Lemma for Capacities", *The Annals of Statistics*, **1**, No. 2, pp. 251–263 (1973).

[4] M. V. Burnashev, "Minimax Detection of Inaccurately Known Signal in the Background of White Gaussian Noise", *Theory of Prob. and Its Appl.*, **24**, No. 1, pp. 106–118 (1979).

[5] R. Ahlswede and G. Dueck, "Identification via Channels," *IEEE Trans. Inform. Theory*, **35**, No. 1, pp. 15–29 (1989).

[6] R. Ahlswede and G. Dueck, "Identification in the Presence of Feedback – A Discovery of New Capacity Formulas," *IEEE Trans. Inform. Theory*, **35**, No. 1, pp. 30–36 (1989).

[7] T. S. Han and S. Verdu, "New Results in the Theory of Identification via Channels," *IEEE Trans. Inform. Theory*, **38**, No. 1, pp. 14–25 (1992).

[8] T. S. Han and S. Verdu, "Approximation Theory of Output Statistics," *IEEE Trans. Inform. Theory*, **39**, No. 3, pp. 752–772 (1993).

[9] L. A. Bassalygo and M. V. Burnashev, "Estimate for the Maximal Number of Messages for a Given Probability of Succesful Deception," *Probl. Inform. Trans.*, **30** , No. 2, pp. 42–48 (1994).

[10] M. V. Burnashev and S. Verdu, "Measures Separated in L_1 metrics and ID–codes," *Probl. Inform. Trans.*, **30** , No. 3, pp. 3–14 (1994).

[11] L. A. Bassalygo and M. V. Burnashev, "Authentication, Identification and Pairwise Separated Measures," *Probl. Inform. Trans.*, **32** , No. 1, pp. 41–47 (1996).

[12] Ju. D. Burago and V. A. Zalgaller. *Geometrical Inequalities* [in Russian], Nauka, Leningrad (1980).

[13] R. G. Gallager. *Information Theory and Reliable Communication*, John Wiley & Sons, New York (1968).

[14] I. Csiszar and J. Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Akademiai Kiado, Budapest (1981).

[15] M. V. Burnashev and I. A. Begmatov, “On a Signal Detection Problem Leading to Stable Distributions”, *Theory of Prob. and Its Appl.*, **35**, No. 3, pp. 1169–1172 (1990).