

INFORMATION THEORY

MEASURES SEPARATED IN L_1 METRICS AND ID-CODES

M. V. Burnashev and S. Verdú

UDC 621.391.15

A geometrical approach to ID-codes, based on their equivalence to some natural notions from mathematical statistics, is described. This not only enlarges the available analytical apparatus, but also enables us to strengthen some known results.

1. Introduction

Let A and B be 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 $Q = PW^{(n)}$ we denote the generated distribution on the channel output B^n .

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

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

It should be noted that it is allowed here for regions \mathcal{D}_i to intersect each other (for $\delta > 0$).

The notion of ID-codes was first introduced in [1], where it was shown that the maximal cardinality $M(n, \delta)$ of an ID-code of length n satisfies the inequality

$$\lim_{n \rightarrow \infty} \frac{\ln \ln M(n, \delta)}{n} \geq C, \quad 0 < \delta \leq 1. \quad (1)$$

where C is the channel capacity in natural units. In order to construct an ID-code of such cardinality it is sufficient to choose any usual code over A^n with cardinality of the order of ϵ^{nC} (and with small error probability) and to consider all equiprobable distributions on all of its subsets with some fixed cardinality of the order of $\delta \epsilon^{nC}$.

The converse of inequality (1)

$$\lim_{n \rightarrow \infty} \frac{\ln \ln M(n, \delta)}{n} \leq C, \quad 0 \leq \delta \leq \delta_0, \quad (2)$$

where δ_0 is some positive constant, was obtained in [3]. It should be mentioned that earlier in [1] a weaker form of inequality (2) was proved for $M(n, \delta_n)$, where $\delta_n \leq \epsilon^{-\epsilon n}$ and ϵ is any positive constant. In fact, the inequality (2) in [3] was obtained as a corollary of a stronger result (Lemma 1) showing that any output distribution can be rather accurately approximated by using some input distribution $P(x)$, $x \in A^n$, whose masses take values on a lattice with a span of the order of ϵ^{-nC} .

It follows from (1), (2) that

$$\lim_{n \rightarrow \infty} \frac{\ln \ln M(n, \delta)}{n} = C, \quad 0 < \delta \leq \delta_0. \quad (3)$$

Translated from *Problemy Peredachi Informatsii*, Vol. 30, No. 3, pp. 3-14, July-September, 1994. Original article submitted November 9, 1993.

Such a double exponential growth rate of the cardinality of an ID-code (in contrast to an ordinary exponential one for the usual codes) created a certain theoretical and practical interest in these codes [1–5]. ID-codes are also close to some cryptography problems [6].

In this paper a more general geometrical approach to ID-codes is proposed. This 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 the testing of composite hypotheses in mathematical statistics) not only enlarges the research analytical apparatus but also enables us to strengthen some results from [1, 3]. In particular, it will be shown (Theorem 1) that any distribution on the channel output B^n can be rather accurately approximated by using only the order of e^{n^C} of input blocks $x \in A^n$. From this result, Lemma 1 of [3] and inequality (2) follow, moreover, for any $0 \leq \delta < 1/2$. This gives a certain completeness to the converse inequality (2), since for $\delta \geq 1/2$ the number $M(n, \delta)$ becomes infinite provided that randomized decision rules are adopted for use (see example in [1], p. 16).

We consider first some properties of families of “almost orthogonal” measures on an arbitrary finite alphabet Y (i.e., for noiseless communication channels) and after that carry them over to noisy channels.

Definition 2. A collection $(P_i, \mathcal{D}_i, i = 1, \dots, M)$ of probability measures P_i on Y and regions $\mathcal{D}_i \subseteq Y$ is called an (M, δ) ID-code if the following conditions are satisfied:

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

It is clear that Definitions 1 and 2 coincide if $Y = A^n = B^n$ and $P = Q$ (i.e., the channel is noiseless).

Definition 3. A collection $(P_i, i = 1, \dots, M)$ of probability measures P_i on Y is called an (M, δ) pairwise separated collection (family) if their minimum distance satisfies the condition

$$\min_{i \neq j} \|P_i - P_j\| = \min_{i \neq j} \sum_Y |P_i(y) - P_j(y)| \geq 2(1 - \delta). \quad (4)$$

Definition 4. A collection $(P_i, i = 1, \dots, M)$ of probability measures P_i on Y is called an (M, δ) completely separated collection (family) if for any i the following condition is satisfied:

$$\|P_i - \text{conv}\{P_j, j \neq i\}\| = \min_{\mathbf{c}} |P_i - \sum_{j \neq i} c_j P_j| \geq 2(1 - \delta), \quad (5)$$

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

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

In Sec. 2 a certain equivalence between ID-codes and completely separated families of measures is established, and some estimates of the maximum possible cardinalities for all three families of measures are obtained. In Sec. 3 these results are carried over to noisy channels. In Sec. 4 some relation between pairwise and completely separated families is analyzed.

Remark. In Definitions 1 and 2 it would be more correct to admit also randomized decision rules when for every $y \in Y$ and $i = 1, \dots, M$ the conditional probability $P(i|y)$ of making a decision in favor of the i th message is chosen. All of the results of the paper (in particular, the upper bounds for $M(n, \delta)$) also remain valid for such an extended notion of ID-codes. We did not introduce the possibility of such randomization in Definitions 1 and 2 in order to avoid complications.

2. Noiseless channels

We first cite one simple and useful result from mathematical statistics.

Let there be given a measure P and a family of measures $\{Q_i, i \in \mathcal{A}\}$ on a set Y . The minimax problem of testing by one observation of the simple hypothesis P against the family of alternatives $\{Q_i, i \in \mathcal{A}\}$ is

considered. In this connection, if some set $D \subseteq Y$ of making a decision in favor of hypothesis P is chosen, then the error probabilities α and β of the first and the second kind are defined as follows:

$$\alpha(D) = P(Y \setminus D), \quad \beta(D) = \sup_i Q_i(D). \quad (6)$$

Lemma 1. (e.g., [6], p. 112) *For the minimal possible sum $\alpha + \beta$ the following equality holds:*

$$\min_D \{\alpha(D) + \beta(D)\} = 1 - \frac{1}{2} \|P - \text{conv}\{Q_i, i \in \mathcal{A}\}\|. \quad (7)$$

Proposition 1. *For any (M, δ) completely separated collection $\{P_i\}$ it is possible to define sets $\{\mathcal{D}_i\}$ such that $\{P_i, \mathcal{D}_i\}$ is an (M, δ) ID-code. Conversely, any (M, δ) ID-code is an $(M, 2\delta)$ completely separated collection as well.*

PROOF. If some (M, δ) completely separated collection $\{P_i\}$ is given, then we choose as the region \mathcal{D}_i the set D giving the minimum in (7) (when testing hypotheses P_i against all remaining alternatives). Then, owing to (6), Definition 2 will hold. Conversely, if some (M, δ) ID-code is given, then, owing to (6), (7), it is an $(M, 2\delta)$ completely separated collection as well. \triangle

We denote now the cardinality of the set Y by N , and let $M(N, \delta)$, $M_c(N, \delta)$, and $M_p(N, \delta)$ be the maximal possible cardinalities of the ID-code, completely separated collection, and pairwise separated collection on the set Y , respectively. By virtue of Proposition 1 we have

$$M(N, \delta) \leq M_c(N, 2\delta) \leq M_p(N, 2\delta), \quad 0 < \delta < 1/2. \quad (8)$$

Let us estimate now the lower bound of $M(N, \delta)$ and the upper bound of $M_p(N, 2\delta)$.

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

$$M(N, \delta) \geq \exp\{N\delta^2/(2\epsilon^2)\}, \quad (9)$$

$$M_p(N, \delta) \leq (2/(1 - \delta))^{N-1}. \quad (10)$$

PROOF. An analog of inequality (9) in a weaker form was proved in [1] (Statement 1) by an "exhaustive" method. A similar inequality in a more complicated situation was obtained in [5] (Theorem 1) also by an "exhaustive" method. For a change we use here some "random choice" arguments. We put as every measure P_i , $i = 1, \dots, M$, an equiprobable distribution on some subset $\mathcal{D}_i \subset Y$ of cardinality ϵN (the parameter $\epsilon < \delta$ will be chosen later). If 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 clearly the collection $(P_i, \mathcal{D}_i, i = 1, \dots, M)$ is an (M, δ) ID-code. 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 obvious:

$$\begin{aligned} P &\leq M^2 \sum_{i > \epsilon\delta N} \binom{\epsilon N}{i} \binom{N - \epsilon N}{\epsilon N - i} \left[2 \binom{N}{\epsilon N} \right]^{-1} \\ &\leq M^2 \delta \binom{\epsilon N}{\epsilon\delta N} \binom{N - \epsilon N}{\epsilon N - \epsilon\delta N} \left[2(\delta - \epsilon) \binom{N}{\epsilon N} \right]^{-1}. \end{aligned} \quad (11)$$

We can loosen this bound with the easily verifiable inequality

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

Then, (11) becomes

$$P \leq \frac{\delta M^2}{2(\delta - \epsilon)} \exp\{N\delta\epsilon \ln \frac{\epsilon\epsilon}{\delta}\}. \quad (12)$$

We now put $\varepsilon = \lfloor \varepsilon_0 N \rfloor / N$, $\varepsilon_0 = \delta e^{-2}$. Taking into account that $0 \leq (\varepsilon_0 - \varepsilon)N \leq 1$, we get from (12), after some simple calculations, assuming $N\delta \geq 15$,

$$P \leq M^2 \exp\{-N\delta^2 \varepsilon^{-2}\}. \quad (13)$$

Now if the right side of (13) is less than 1, then there exists a collection of M regions having the required properties, from which the lower bound (9) 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 (9) is valid for any $N \geq 2$ and $0 \leq \delta \leq 1$.

We now prove the upper bound in (10). Let $P = (x_1, \dots, x_N)$ 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 :

$$v_N = \int \dots \int_{A(N-1,1)} dx_1 \dots dx_{N-1} = \frac{1}{(N-1)!}, \quad (14)$$

where the following notation was used

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

Indeed,

$$\begin{aligned} \int \dots \int_{A(N-1,a)} dx_1 \dots dx_{N-1} &= \int_0^a \int \dots \int_{A(N-2, a-x_{N-1})} dx_1 \dots dx_{N-2} dx_{N-1} \\ &= \int_0^a v_{N-1} (a - x_{N-1})^{N-2} dx_{N-1} = v_{N-1} a^{N-1} / (N-1) = v_N a^{N-1}. \end{aligned}$$

Therefore, $v_N = v_{N-1} / (N-1)$, $v_2 = 1$, whence formula (14) follows.

We now fix any measure $P_0 = (z_1, \dots, z_N)$ 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 the Appendix) that $D(P_0)$ has the minimal "volume" $w_N(\delta)$ when $z_1 = 1$, $z_2 = \dots = z_N = 0$ (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 (14) we get ($b = A(N-1, 1) \cap \{x : x_1 \geq (1 + \delta)/2\}$)

$$\begin{aligned} w_N(\delta) &= \int \dots \int_B dx_1 \dots dx_{N-1} = v_{N-1} \int_{(1+\delta)/2}^1 (1 - x_1)^{N-2} dx_1 \\ &= v_N (2/(1 - \delta))^{N-1}. \end{aligned} \quad (15)$$

Since $M_p(N, \delta) \leq v_N / w_N(\delta)$, we now get the upper bound (10) from (14), (15). Δ

Corollary 1. *The maximal cardinality $M(N, \delta)$ of an ID-code for any $N \geq 2$ and $0 < \delta < 1/2$ satisfies the following inequalities:*

$$\exp\{N\delta^2/(2e^2)\} \leq M(N, \delta) \leq (2/(1 - 2\delta))^{N-1}. \quad (16)$$

Remark. If $N = |A|^n$ (which holds when a communication channel is used), then the left and right sides of (16) grow (when $n \rightarrow \infty$) with approximately the same double exponential rate. It should also be mentioned that upper bound (16) yields to the simple bound $M(n, \delta) \leq 2^N$, $\delta < 1$ (see [1], p. 17), but the bound (16) also remains valid for ID-codes with randomized decision rules, in contrast to the mentioned bound from [1].

3. Noisy channels

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

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

We now choose 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 ν_i is the number of measures f_i among k chosen distributions.

Proposition 3. *The following estimate on the variance of $\hat{p}(y)$ holds:*

$$\mathbf{E}|p(y) - \hat{p}(y)|^2 \leq \frac{1}{s} \sum_{j=1}^K p_j(1 - p_j) f_j^2(y). \quad (17)$$

PROOF. Let μ_l be the index of the measure at the l th step, chosen from $\{f_i\}$. Then μ_1, \dots, μ_s are i.i.d.r.v.'s, and, moreover,

$$\mathbf{E}f_{\mu_l} = p(y) \quad \text{and} \quad \sum_{j=1}^K \nu_j f_j(y) = \sum_{l=1}^s f_{\mu_l}(y).$$

Therefore,

$$\begin{aligned} \mathbf{E}|\hat{p}(y) - p(y)|^2 &= \mathbf{E}\left|\frac{1}{s} \sum_{l=1}^s (f_{\mu_l}(y) - p(y))\right|^2 \\ &= \frac{1}{s} \mathbf{E}(f_{\mu_1} - p(y))^2 = \frac{1}{s} (\mathbf{E}f_{\mu_1}^2(y) - p^2(y)) \\ &= \frac{1}{s} \left\{ \sum_{j=1}^K p_j f_j^2(y) - \left[\sum_{j=1}^K p_j f_j(y) \right]^2 \right\} \leq \frac{1}{s} \sum_{j=1}^K p_j(1 - p_j) f_j^2(y). \quad \Delta \end{aligned}$$

Now for $\varepsilon > 0$ and $i = 1, \dots, K$ we choose some arbitrary sets $Y(i, \varepsilon)$ such that $P_i\{Y(i, \varepsilon)\} \geq 1 - \varepsilon$ and put

$$K(y) = \{i : y \in Y(i, \varepsilon)\}, \quad Y_\varepsilon = \bigcup_{i=1}^K Y(i, \varepsilon).$$

Lemma 2. *For any $\varepsilon > 0$ the following estimate is valid:*

$$\sum_Y \mathbf{E}|\hat{p}(y) - p(y)| \leq \varepsilon + \sqrt{|Y_\varepsilon| s^{-1} \max_{j=1, \dots, K, y \in Y(j, \varepsilon)} f_j(y)} \quad (18)$$

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

PROOF. We have

$$\begin{aligned} \sum_Y \mathbf{E}|\hat{p}(y) - p(y)| &= \sum_{Y_\varepsilon} \mathbf{E} \left| \sum_{j \in K(y)} \left(\frac{\nu_j}{s} - p_j \right) f_j(y) \right| \\ &\quad + \sum_Y \mathbf{E} \left| \sum_{j \notin K(y)} \left(\frac{\nu_j}{s} - p_j \right) f_j(y) \right| = \Sigma_1 + \Sigma_2; \end{aligned}$$

$$\begin{aligned}
\Sigma_2 &\leq \sum_Y \sum_{j \notin K(y)} p_j f_j(y) = \sum_{j=1}^K p_j \sum_{Y \setminus Y(j)} f_j(y) \leq \varepsilon, \\
\Sigma_1 &\leq \sum_{Y_\varepsilon} \left[\mathbf{E} \left| \sum_{j \in K(y)} \left(\frac{\nu_j}{s} - p_j \right) f_j(y) \right|^2 \right]^{1/2} \\
&\leq \sum_{Y_\varepsilon} \left(s^{-1} \sum_{j \in K(y)} p_j f_j^2(y) \right)^{1/2} \leq \left(|Y_\varepsilon| s^{-1} \sum_{Y_\varepsilon} \sum_{j \in K(y)} p_j f_j^2(y) \right)^{1/2} \\
&\leq \left(|Y_\varepsilon| s^{-1} \sum_{j=1}^K p_j \sum_{Y(j)} f_j^2(y) \right)^{1/2} \leq \left(|Y_\varepsilon| s^{-1} \max_{j=1, \dots, K, y \in Y(j, \varepsilon)} f_j(y) \right)^{1/2} \cdot \Delta
\end{aligned}$$

In the setting of a noisy channel, the measure $f_i(y)$ represents the conditional distribution of the channel output B^n given the “input” $i \in A^n$. Some “supports” of measures $f_i(y)$ will be chosen as the sets $Y(i, \varepsilon)$. In this respect, Lemma 2 is reserved for the same type codeblocks i . The next result generalizes it for codeblocks of different types.

Let L classes of measures $f_{li}(y)$, $l = 1, \dots, L$, $i = 1, \dots, K(l)$, with prior probabilities $\{p_{li}\}$ be 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 also choose some set $Y_{li}(\varepsilon)$ such that $P_{li}\{Y_{li}\} \geq 1 - \varepsilon$ and put (ν_{li} is the number of measures f_{li} among s_l chosen distributions)

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

Lemma 3. *For any $\varepsilon > 0$ the following estimate is valid:*

$$\sum_Y \mathbf{E} |\hat{p}(y) - p(y)| \leq \varepsilon + \left(\sum_{l=1}^L |Y_l| s_l^{-1} \max_{i=1, \dots, K(l); y \in Y_{li}} f_{li}(y) \right)^{1/2}. \quad (19)$$

The proof follows from the following chain of inequalities:

$$\begin{aligned}
\sum_Y \mathbf{E} |\hat{p}(y) - p(y)| &\leq \varepsilon + \sum_{l=1}^L \left(|Y_l| s_l^{-1} \max_{i=1, \dots, K(l); y \in Y_{li}} f_{li}(y) \sum_{j=1}^{K(l)} p_{lj} \right)^{1/2} \\
&\leq \varepsilon + \left(\sum_{l=1}^L |Y_l| s_l^{-1} \max_{i=1, \dots, K(l); y \in Y_{li}} f_{li}(y) \right)^{1/2} \cdot \Delta
\end{aligned}$$

Now we return to the channel $W^{(n)}$ with input A^n and output B^n alphabets. The application of Lemma 3 to this case is rather standard and is based on the consideration of codeblocks from A^n of the same type (composition) (see [8], Ch. 1.2 and Ch. 2.1). For this reason we shall omit some technical details that are easily restored using [8].

As usual, we partition all codeblocks from A^n into 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 the set $Y_{li} \subseteq B^n$ we choose a set of probability “almost 1” and consisting of approximately equiprobable points. More precisely, we fix some arbitrarily small $\varkappa > 0$ and let P_l denote the type of codeblock (l, i) [8] and Q_{li} denote the output distribution generated by that codeblock. Then, for any $n \geq n_0(\varkappa)$, it is possible to choose sets $Y_{li} \subseteq B^n$ with $Q_{li}\{Y_{li}\} \geq 1 - \varepsilon$ such that simultaneously for all l, i the following conditions will be satisfied:

$$\begin{aligned} |Y_{li}| &< \exp\{n(1 + \varkappa)H(W|P_l)\}, \\ \max_{i=1, \dots, K(l); y \in Y_{li}} f_{li}(y) &< \exp\{-n(1 - \varkappa)H(W|P_l)\}, \\ |Y_l| &< \exp\{n(1 + \varkappa)H(WP_l)\}. \end{aligned} \quad (20)$$

Taking into account that $H(WP_l) - H(W|P_l) = I(P_l, W)$ and the total number of classes $L \leq (n + 1)^{|A|}$, we put

$$s_l = \varkappa^{-1} \exp\{n[I(P_l, W) + 2\varkappa \ln B]\}(n + 1)^{|A|}. \quad (21)$$

Then

$$s = \sum_{l=1}^L s_l < \varkappa^{-1} \exp\{n[C + 2\varkappa \ln B]\}(n + 1)^{2|A|}. \quad (22)$$

Now from (19)–(22) follows

$$\sum_{B^n} \mathbf{E}|\hat{p}(y) - p(y)| \leq 4\varepsilon. \quad (23)$$

We can formulate the result just obtained in (23) in the following way.

Theorem 1. *For any $\delta > 0$ it is possible to find $\varepsilon > 0$ and n_0 such that for any output measure Q on B^n , $n \geq n_0$, there exist $\exp\{n(C + \varepsilon)\}$ input blocks $i \in A^n$ such that for their generated measures $\{Q_i\}$ the following inequality holds:*

$$\|Q - \text{conv}\{Q_i, i = 1, \dots, e^{n(N+\varepsilon)}\}\| \leq \delta. \quad (24)$$

In other words, any output distribution Q can be arbitrarily closely approximated by using only the order of e^{nC} of input blocks. We should mention here that blocks used for this purpose, generally speaking, depend on measure Q .

For channel $W^{(n)}$ we can, analogously to Sec. 1, introduce the notions of (M, n, δ) pairwise separated collection and (M, n, δ) completely separated collection. We also denote by $M_p(n, \delta)$ and $M_c(n, \delta)$ their maximal possible cardinalities, respectively. Then, analogously to (8), we have

$$M(n, \delta) \leq M_c(n, 2\delta) \leq M_p(n, 2\delta), \quad 0 < \delta < 1/2. \quad (25)$$

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

Lemma 4. *For any channel W , any pair of input distributions P_1, P_2 and the corresponding pair of output distributions Q_1, Q_2 the following inequality holds:*

$$\|Q_1 - Q_2\| \leq \|P_1 - P_2\|. \quad (26)$$

Indeed,

$$\begin{aligned} \|Q_1 - Q_2\| &= \sum_Y \left| \sum_X W(y|x)(p_1(x) - p_2(x)) \right| \\ &\leq \sum_X |p_1(x) - p_2(x)| \sum_Y W(y|x) = \sum_X |p_1(x) - p_2(x)| = \|P_1 - P_2\|. \quad \Delta \end{aligned}$$

Proposition 4. *The limiting relation (3) holds for any $0 < \delta < 1/2$.*

PROOF. It is sufficient to establish formulas similar to (1) and (2). In order to obtain an analog of (1), on the input alphabet A^n we choose a usual code with the cardinality of the order of e^{nC} and a small error

probability. For the alphabet consisting of this code the estimate (9) will be valid (with e^{nC} instead of N), from which the analog of (1) follows (see also [1], pp. 18-19).

We now proceed to the proof of the analog of (2). By virtue of Theorem 1 any measure on the channel output B^n can be arbitrarily closely approximated by using the order of e^{nC} input blocks from A^n . Since the channel $W^{(n)}$ is a "compressing" operator (Lemma 4), the maximal number of 2δ -pairwise separated measures, generated by every collection of $N = e^{nC}$ input blocks, is estimated from above by formula (10). The total number of collections with cardinality N on the alphabet A^n does not exceed A^{nN} . Therefore, for the maximal possible cardinality $M_p(n, 2\delta)$ of a pairwise separated collection we get

$$M_p(n, 2\delta) < (2/(1-2\delta))^{e^{nC}} A^{n e^{nC}}. \quad (27)$$

Now from (25) and (27), the inequality (2) for any $0 < \delta < 1/2$ and, hence, the validity of Proposition 4 follows. \triangle

Corollary 2. *Any output distribution can be arbitrarily closely approximated by using some input distribution $\{p_i, i = 1, \dots, N\}$, $N \sim e^{nC}$, taking values only of the form $p_i = j/N$, where j is an integer (this result is similar to Lemma 1 from [3]).*

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

$$\|p - \hat{p}\| = \sum_{i=1}^N |p_i - \hat{p}_i| \leq \varepsilon. \quad (28)$$

Indeed, we choose $\hat{p}_1 = j\varepsilon/N$ with the minimal possible $|p_1 - \hat{p}_1|$. Then $|p_1 - \hat{p}_1| \leq \varepsilon/(2N)$. 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 $|(p_1 + p_2) - (\hat{p}_1 + \hat{p}_2)| \leq \varepsilon/(2N)$ and $|p_2 - \hat{p}_2| \leq \varepsilon/(2N)$. Repeating this process, we get as a result the collection $\{\hat{p}_i\}$, having the desired properties. Since the channel is a compressing operator, for measures Q and \hat{Q} generated by these distributions, owing to (28), we have $\|Q - \hat{Q}\| \leq \varepsilon$. From this result and Theorem 1, the assertion of Corollary 2 follows. \triangle

In connection with Theorem 1, a natural question arises: Is it possible to choose and fix some collection of $N \sim e^{nC}$ input blocks (like a basis) such that using them it is possible to approximate arbitrarily closely any output distribution? The answer, generally speaking, is negative. To see this, consider the example of a binary symmetric channel with crossover probability $p < 1/2$.

Let x_0 be the all-zero input block of 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 generated measures on the channel output B^n , respectively. We shall now evaluate how large d should be that it would be impossible to approximate sufficiently accurately the measure Q_0 with the help of the remaining measures $\{Q_i, i = 1, \dots, N\}$. It is convenient to use here some statistical interpretation and the formula (7). Let us consider the problem of testing two hypotheses Q_0 and $\{Q_i, i = 1, \dots, N\}$. We note that if d is sufficiently large, then, by the central limit theorem, we have for $i = 0, 1, \dots, N$

$$Q_i \left\{ y : |w(y) - pn - w(x_i)(1-2p)| > [2znp(1-p)]^{1/2} \right\} < e^{-z}. \quad (29)$$

We choose as the decision set \mathcal{D}_0 of accepting hypotheses Q_0 the following set:

$$\mathcal{D}_0 = \left\{ y : |w(y) - pn| \leq [2znp(1-p)]^{1/2} \right\}.$$

Now, if $(w(x_i)(1-2p))^2 \geq 2znp(1-p)$, $i = 1, \dots, N$, then owing to (29) we have

$$Q_0(Y^n \setminus \mathcal{D}_0) < e^{-z} \quad \text{and} \quad Q_i(\mathcal{D}_0) < e^{-z}, \quad i = 1, \dots, N.$$

We put now $z = \ln(2/\delta)$ and get from formula (7)

$$\|Q_0 - \text{conv} \{Q_i, i = 1, \dots, N\}\| \geq 2(1-\delta).$$

Therefore, if the minimum distance d of an arbitrary code x_1, \dots, x_N satisfies the condition $(d(1-2p))^2 \geq 2np(1-p)\ln(2/\delta)$, 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.

4. Completely and pairwise separated collections of measures

As was shown in the previous section, 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 (5)). It is much easier to check the pairwise separability of measures. Then the following question naturally arises: Is it possible to select from a given set of pairwise separated measures a subcollection of completely separated measures of “practically” the same cardinality. Or, equivalently, under which conditions is a collection of pairwise separated measures simultaneously a collection of completely separated measures (perhaps, with a slightly worse δ).

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

Example 1. Let there be given an equiprobable distribution P_0 and measures $P_i, i = 1, \dots, |X|$, 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 $\|P_0 - \text{conv}\{P_i, i = 1, \dots, |X|\}\| = 0$, i.e., this collection of measures is not completely separable for any $\delta < 1$. From the point of view of ID-codes 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 this collection, the remaining orthogonal measures will represent a completely separated collection.

It becomes clear from the next (more meaningful) example that it is hardly possible to obtain easily verified necessary and sufficient conditions in this problem. The infinite alphabet used in this example is not crucial. It is always possible to approximate it arbitrarily precisely by a sufficiently large finite alphabet.

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

$$\|P_i - P_j\| = 2 \left[1 - 2P \left\{ \xi > (A/2)^{1/2} \right\} \right] = 2 \left[2\Phi \left((A/2)^{1/2} \right) - 1 \right], \quad i \neq j, \quad (30)$$

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 estimate its upper bound. Putting on the set $P_j, j \neq i$, the equiprobable prior distribution and using the simple inequality

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

we get, after some simple calculations (see details in [6]),

$$\|P_i - \text{conv}\{P_j, j \neq i\}\|^2 \leq (N-1)^{-1}e^{2A}. \quad (31)$$

We can see from (30) that when A is sufficiently large, measures $\{P_i, i = 1, \dots, N\}$ are almost orthogonal and hence are well pairwise separated. But from (31) 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, following [8], it is not difficult to show that the critical value here is not $\exp\{2A\}$, but $\exp\{A\}$.

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

Proposition 5. *Let there be given measures P_i , $i = 1, \dots, N$, on a finite alphabet X . For any subset $A \subseteq X$ we denote*

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

Then for any i the following estimate holds:

$$\|P_i - \text{conv} \{P_j, j \neq i\}\| \geq 2 \max_A \left[P_i(A) - \left(|A| \max_{j \neq i} \varepsilon_{ij}(A) \right)^{1/2} \right], \quad (32)$$

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

$$\delta = \max_i \min_A \left[P_i(X \setminus A) + \left(|A| \max_{j \neq i} \varepsilon_{ij}(A) \right)^{1/2} \right]. \quad (33)$$

PROOF. Let $Q = \sum c_i P_i$ be any convex combination of measures P_i , $i = 1, \dots, N$, and $A \subseteq X$. Using the representation $|p - q| = p + q - 2 \min\{p, q\}$ and the Cauchy–Bunyakowski inequality, we carry out the following chain of calculations:

$$\begin{aligned} \|P_1 - Q\| &= 2 \left[1 - \sum_X \min\{P_1(x), Q(x)\} \right] \\ &\geq 2 \left[P_1(A) - \sum_A \min\{P_1(x), Q(x)\} \right] \geq 2 \left[P_1(A) - \sum_A (P_1(x)Q(x))^{1/2} \right] \\ &\geq 2 \left[P_1(A) - \left(|A| \sum_A P_1(x)Q(x) \right)^{1/2} \right] \geq 2 \left[P_1(A) - \left(|A| \max_{i \geq 2} \varepsilon_{1i}(A) \right)^{1/2} \right]. \end{aligned}$$

from which formulas (32), (33) follow. \triangle

Using this proposition, it is not difficult to analyze the example considered at the end of Sec. 3.

The authors would like to thank R. Ahlswede, L. A. Bassalygo, and Te Sun Han for helpful discussions and constructive critical remarks.

Appendix

We show that the set $D(P_0)$ in the proof of inequality (10) has the minimum volume when P_0 is an extreme point of the set \mathcal{P}_N .

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

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

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

$$\begin{aligned} V(c) &= \sum_{i=1}^N c_i V_i = \left\{ y \in \mathcal{P}_N : y = \sum_{i=1}^N c_i y_i; y \in V_i, i = 1, \dots, N \right\}, \\ D(c) &= \{y \in \mathcal{P}_N : \|c - y\| \leq a\}. \end{aligned}$$

Obviously, sets V_i , $V(c)$, and $D(c)$ are convex. It is also easy to check that $V(c) \subseteq D(c)$. Designating by $v(A)$ the volume of the set A and taking into account that volumes $v(V_i)$, $i = 1, \dots, N$, are equal, we get, by virtue of the Brunn–Minkowski theorem [9],

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

REFERENCES

1. R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inf. Theory*, **35**, No. 1, 15–29 (1989).
2. R. Ahlswede and G. Dueck, "Identification in the presence of feedback - a discovery of new capacity formulas," *IEEE Trans. Inf. Theory*, **35**, No. 1, 30–36 (1989).
3. T. S. Han and S. Verdú, "New results in the theory of identification via channels," *IEEE Trans. Inf. Theory*, **38**, No. 1, 14–25 (1992).
4. T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, **39**, No. 3, 752–772 (1993).
5. S. Verdú and V. K. Wei, "Explicit construction of optimal constant-weight codes for identification via channels," *IEEE Trans. Inf. Theory*, **39**, No. 1, 30–36 (1993).
6. L. A. Bassalygo and M. V. Burnashev, "Estimate for the maximal number of messages for a given probability of successful deception," *Probl. Inf. Trans.*, **30**, No. 2, 42–48 (1994).
7. M. V. Burnashev, "Minimax detection of inaccurately known signal in the background of white Gaussian noise," *Theory Probab. Appl.*, **24**, No. 1, 106–118 (1979).
8. I. Csiszar and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Akademiai Kiado, Budapest (1981).
9. M. V. Burnashev and I. A. Begmatov, "On one signal detection problem leading to stable distributions," *Theory Probab. Appl.*, **35**, No. 3, 1169–1172 (1990).
10. Ju. D. Burago and V. A. Zalgaller, *Geometrical Inequalities* [in Russian], Nauka, Leningrad (1980).