Mathematics

R. L. Dobrushin

1960

SovietRxiv

Abstract

Full Text

Mathematics

R. L. Dobrushin

ASYMPTOTICS OF ERROR PROBABILITIES IN THE TRANSMISSION OF INFORMATION OVER A MEMORYLESS CHANNEL WITH A SYMMETRIC MATRIX OF TRANSITION PROBABILITIES

(Presented by Academician A. N. Kolmogorov, 4 III 1960)

A stationary discrete memoryless channel is specified by the set $\mathcal{E} = (E_1, \dots, E_M)$ of input states, the set $\overline{\mathcal{E}} = (\overline{E}_1, \dots, \overline{E}_N)$ of output states, and the matrix of transition probabilities $P = \{p_{ij},\ i = 1, \dots, M,\ j = 1, \dots, N\}$. We shall call the matrix P **symmetric** if each of its rows is obtained by some permutation of the elements of any other row and each column is obtained by some permutation of the elements of any other column, and we shall consider only channels with a symmetric matrix P. The most important special case of such channels is the symmetric binary channel, where $M = N = 2$, $p_{11} = p_{22} = p$, $p_{12} = p_{21} = q$, $p + q = 1$, for which the asymptotics of the error probability was studied by Elias ([9,10]). By the space of input (output) signals of length n for a memoryless channel we shall mean the space $\mathcal{E}^{(n)}$ ($\overline{\mathcal{E}}^{(n)}$) of all sequences $(E_{i_1}, \dots, E_{i_n})$, $i_k = 1, \dots, M$ $(\overline{E}_{j_1}, \dots, \overline{E}_{j_n})$, $j_k = 1, \dots, N$. The transition probabilities $p(\overline{e}/e)$, $e = (E_{i_1}, \dots, E_{i_n}) \in \mathcal{E}^{(n)}$, $\overline{e} = (\overline{E}_{j_1}, \dots, \overline{E}_{j_n}) \in \overline{\mathcal{E}}^{(n)}$, are defined as

$$p(\overline{e}/e) = p_{i_1 j_1} p_{i_2 j_2} \cdots p_{i_n j_n}. \tag{1}$$

By a **method of transmitting** K messages we shall mean the totality of a code, which is a set $\mathfrak{A} = (e_1, \dots, e_K)$, $e_l \in \mathcal{E}^{(n)}$, and a decoding method, specified by a system of functions $r_l(\overline{e})$, $l = 1, \dots, K$, $\overline{e} \in \overline{\mathcal{E}}^{(n)}$, such that $r_l(\overline{e}) \geq 0$,

$$\sum_{l=1}^{K} r_l(\overline{e}) = 1, \qquad \overline{e} \in \overline{\mathcal{E}}^{(n)}.$$

The **error probability** for the code $\mathfrak{A}$ is defined as

$$p(\mathfrak{A}) = \inf \frac{1}{K} \sum_{l=1}^{K} \sum_{\overline{e} \in \overline{\mathcal{E}}^{(n)}} p(\overline{e}/e_l)\,[1 - r_l(\overline{e})], \tag{2}$$

where the lower bound is taken over all decoding methods. The **optimal error probability** is defined as

$$p_n(K) = \inf_{\mathfrak{A}} p(\mathfrak{A}), \tag{3}$$

where the lower bound is taken over all codes $\mathfrak{A}$.

Let ξ_l, $l = 1, \dots, K$, be independent uniformly distributed random variables taking values in $\mathcal{E}^{(n)}$. The set $\widetilde{\mathfrak{A}} = \{\xi_1, \dots, \xi_K\}$ will be called a **random code**. The **mean error probability**

call the mathematical expectation

$$\overline{p}_n(K) = M\{p(\mathfrak{A})\}. \tag{4}$$

Consider the set of all possible values of the sum

$$x = \log p_{1j_1} + \log p_{1j_2} + \cdots + \log p_{1j_n}, \tag{5}$$

where $j_l = 1, \dots, N$, and number them so that $x_1 > x_2 > \cdots$. Let u_i be the number of sets $(j_1, j_2, \dots, j_n)$ such that the sum (5) is equal to x_i; define s and t from the condition

$$\frac{1}{N^n}\left[\sum_{i=1}^{s-1} u_i + t - 1\right] < \frac{1}{K} \le \frac{1}{N^n}\left[\sum_{i=1}^{s-1} u_i + t\right]. \tag{6}$$

We shall call the number

$$\hat{p}_n(K) = 1 - \sum_{i=1}^{s-1} u_i e^{x_i} - t e^{x_s}. \tag{7}$$

the **Hamming lower bound**. Then

$$\hat{p}_n(K) \le p_n(K) \le \overline{p}_n(K). \tag{8}$$

Denote by C the capacity of our channel. Then

$$C = \log N + \sum_{j=1}^{N} p_{1j} \log p_{1j}. \tag{9}$$

We shall write that $a_n \asymp b_n$ if $\overline{\lim\limits_{n\to\infty}} \dfrac{a_n}{b_n} < \infty$, $\underline{\lim\limits_{n\to\infty}} \dfrac{a_n}{b_n} > 0$.

Theorem 1. Let $0 < H < C$. Suppose* that among the p_{ij} there are two elements $p_{ij} \neq p_{kl} \neq 0$, and let

$$R(h) = \frac{1}{M} \sum_{i=1}^{M} (p_{1i})^h, \qquad m(h) = \frac{d \log R(h)}{dh}, \qquad \sigma^2(h) = \frac{dm(h)}{dh}. \tag{10}$$

Then there exists a unique h_0 (depending on H) such that

$$\log R(h_0) - h_0 m(h_0) = -H. \tag{11}$$

Let

$$H_{\text{crit}} = \frac{1}{2} m(1/2) - \log R(1/2). \tag{12}$$

Then for all H and $n \to \infty$,

$$\hat{p}_n([e^{nH}]) \sim \underline{I}_n n^{-\frac{1}{2h_0}} e^{n[\log R(h_0) + (1-h_0)m(h_0) + \log N]}. \tag{13}$$

For $H > H_{\text{crit}}$ and $n \to \infty$,

$$\overline{p}_n([e^{nH}]) \sim \overline{I}_n n^{-\frac{1}{2h_0}} e^{n[\log R(h_0) + (1-h_0)m(h_0) + \log N]}. \tag{14}$$

For $H < H_{\text{crit}}$ and $n \to \infty$,

$$\overline{p}_n([e^{nH}]) \sim \overline{I}_n n^{-1/2} e^{n[H + 2\log R(1/2) + \log N]}. \tag{15}$$

Here, if d is the greatest common divisor of the system of numbers $\log p_{1i} - \log p_{1j}$, $i = 1, \dots, N$, $j = 1, \dots, N$, and $d = 0$ when such divisors

---

* The special case in which all $p_{ij} \neq 0$ coincide can be investigated in an analogous manner.

No, for $d = 0$

$$\underline{I}_n = \frac{\left(\sqrt{2\pi}\, h_0\sigma(h_0)\right)^{1/h_0-1}}{\sqrt{2\pi}(1-h_0)\sigma(h_0)};$$

$$\overline{I}_n = \frac{\left(\sqrt{2\pi}\, h_0\sigma(h_0)\right)^{1/h_0-3}\Gamma(2-1/h_0)}{\sqrt{2\pi}(1-h_0)\sigma(h_0)} \quad \text{for } H > H_{\text{crit}}; \tag{16}$$

$$\overline{I}_n = \frac{2}{\sqrt{\pi}\,\sigma(1/2)} \quad \text{for } H < H_{\text{crit}},$$

and for $d \neq 0$

$$\underline{I}_n = \left[\frac{\sqrt{2\pi}\sigma(h_0)(1-e^{-h_0 d})}{d}\right]^{1/h_0-1} \frac{d(1+\theta_n^1)}{\sqrt{2\pi}\sigma(h_0)(1-e^{-(1-h_0)d})}; \tag{16'}$$

$$\overline{I}_n = \frac{\left[\sqrt{2\pi}d(1-e^{-h_0 d})\sigma(h_0)\right]^{1/h_0-3}\Gamma(2-1/h_0)}{\sqrt{2\pi}d(1-e^{-(1-h_0)d})\sigma(h_0)}(1+\theta_n^2) \quad \text{for } H > H_{\text{crit}};$$

$$\overline{I}_n = \frac{d(1+\theta_n^3)}{\sqrt{\pi}\sigma(1/2)(1-e^{-d/2})} \quad \text{for } H < H_{\text{crit}},$$

where $|\theta_n^1| \leq 1-e^{-h_0 d}$, $0 \geq \theta_n^2 \geq -(1-e^{-2(1-h_0)d})$, $0 \geq \theta_n^3 \geq -(1-e^{-d/2})$.

Corollary 1. For $H > H_{\text{crit}}$

$$\hat{p}_n([e^{nH}]) \asymp p_n([e^{nH}]) \asymp \overline{p}_n([e^{nH}]) \asymp$$

$$\asymp n^{-1/2h_0} e^{-n[\log R(h_0)+(1-h_0)m(h_0)+\log N]}. \tag{17}$$

If Theorem 1 is applied to the binary symmetric channel, then we obtain Elias' s result, with, however, the change that in Elias' s result $n^{-1/2h_0}$ is replaced by $n^{-1/2}$, which is explained by an error in Elias' s reasoning.

Suppose that $M = p^k$, where p is a prime number and k is an integer. We identify $\mathcal{E}$ with the direct product of k copies of the cyclic group of order p, and $\mathcal{E}^{(n)}$ with the direct product of n copies of the group $\mathcal{E}$. Addition in the commutative group $\mathcal{E}^{(n)}$ will be denoted by the sign $+$. We shall call a code $\mathfrak{A}$ **group** if, for $e_1 \in \mathfrak{A}$, $e_2 \in \mathfrak{A}$, also $e_1+e_2 \in \mathfrak{A}$. The **optimal error probability** of a group code will be called

$$q_n(K) = \inf_{\mathfrak{A}} p(\mathfrak{A}),$$

where the lower bound is taken over all group codes.

For $K = L^r$ (L an integer), we shall call a **random group code** $\overline{\mathfrak{A}}$ the collection of all elements of the form

$$k_1\xi_1 + k_2\xi_2 + \cdots + k_L\xi_L, \qquad k_i = 0, 1, \ldots, p-1,$$

where ξ_i are independent uniformly distributed random variables with values in $\mathcal{E}^{(n)}$. The **mean error probability** of a group code will be called

$$\overline{q}_n(K) = M\{p(\overline{\mathfrak{A}})\}. \tag{18}$$

Theorem 2. Suppose that the sets of signals at the input $\mathcal{E}$ and at the output $\overline{\mathcal{E}}$ coincide and that, for all $e \in \mathcal{E}$, $\overline{e} \in \mathcal{E}$, $\tilde{e} \in \mathcal{E}$, $p_{\overline{e}e} = p_{\overline{e}+e,\, e+\tilde{e}}$. Then the upper limit

$$\overline{\lim_{n\to\infty}} \frac{\overline{q}_n([e^{nH}])}{\overline{p}_n([e^{nH}])} < \infty. \tag{19}$$

Corollary 2. For $H > H_{\text{crit}}$

$$p_n([e^{nH}]) \asymp q_n([e^{nH}]). \tag{20}$$

For the symmetric binary channel this result was also obtained by Elias. Binary group codes were introduced in ([6,7]), and nonbinary ones were studied in ([1,2,11]), etc. Since it is essential that all elements of the group $\mathfrak{G}$ have the same order, and from the theorem on the structure of a commutative group ([5]) it follows that a commutative group has this property only if it is a power of a cyclic group of prime order, such a result cannot be obtained for other commutative groups.

We now examine the advantages afforded by feedback, i.e., by the possibility of using, when encoding the message at the input, information about the signal at the output at previous instants of time. We shall not give here the full corresponding definitions because of their bulkiness (they are given in our paper ([3]*)). We shall call the **optimal error probability for transmission with the use of feedback**

$$\pi_n(K) = \inf P\{\eta \neq \tilde{\eta}\},$$

where the lower bound is taken over all pairs $\eta, \tilde{\eta}$ such that the quantities η and $\tilde{\eta}$ each take K values, the quantity η has the uniform distribution, and (in the terminology of ([3])) the input message η is transformed into the output message $\tilde{\eta}$ as a result of transmission through a channel of length n.

Theorem 3. For arbitrary n and K

$$p_n(K) \leqslant \pi_n(K). \tag{21}$$

Corollary 3. If $H > H_{\text{crit}}$,

$$\pi_n([e^{nH}]) \sim p_n([e^{nH}]). \tag{22}$$

Concerning the proofs of the theorems, we note that in the proof of Theorem 1 use is made of asymptotic estimates of the probabilities of large deviations of sums of independent random variables, obtained by Cramér' s method ([4]). In the proof of Theorem 2 the main role is played by the fact that, if one introduces an auxiliary random code with elements of the form

$$k_1\xi_1 + k_2\xi_2 + \cdots + k_L\xi_L + \xi_{L+1}, \tag{23}$$

where ξ_{L+1} also has the uniform distribution and is independent of ξ_k, $k \leqslant L$, then all K quantities (23) have uniform distributions and are pairwise independent (but dependent in the aggregate!), which makes it possible to extend to the random code of the form (23) the estimation methods used for the random code $\mathfrak{A}$.

Moscow State University
named after M. V. Lomonosov

Received
4 III 1960

REFERENCES

1. L. F. Borodin, *Nauch.-tekhn. obshch. radiotekhn. i elektron. im. A. S. Popova*, Collected Papers, vol. 2, 1958, p. 110.

2. M. Golay, *IRE Trans., Inf. Theory*, IT-4, 103 (1958).

3. R. L. Dobrushin, *Teor. veroyatn. i ee primen.*, 3, 795 (1958).

4. H. Cramér, *Act. Sci. et Ind.*, No. 736, Paris, 1937; G. Cramér, *UMN*, No. 10, 166 (1944).

5. A. G. Kurosh, *Theory of Groups*, Moscow, 1955.

6. D. Slepian, *Bell Syst. Techn. J.*, 35, 203 (1956); D. Slepian, in: *Theory of Information Transmission*, Moscow, 1957, p. 82.

7. R. W. Hamming, *Bell Syst. Techn. J.*, 29, 147 (1950); R. Hamming, in: *Codes with Error Detection and Correction*, Moscow, 1956, p. 7.

8. K. Shannon, *IRE Trans., Inf. Theory*, IT-2, No. 3, 8 (1956).

9. P. Elias, *IRE Convent. Rec.*, Part 4, 37 (1959).

10. P. Elias, *Inform. Theory, Third Lond. Symp.*, Sept. 1955; P. Elias, in: *Theory of Information Transmission*, Moscow, 1957, p. 114.

11. W. Ulrich, *Bell Syst. Techn. J.*, 36, 1341 (1957).

* We take this opportunity to point out that already after the publication of ([3]) it became known to us that Shannon ([8]) had proved Theorem 2 of the present paper.

Note: Figure translations are in progress. See original paper for figures.