



Soviet-era science, translated into English

R. L. Dobrushin

1963

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196301.63982>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

R. L. Dobrushin

Uniform Methods of Information Transmission for Discrete Memoryless Channels and Messages with Independent Components

(Presented by Academician A. N. Kolmogorov, 31 VIII 1962)

In the well-known Shannon scheme of information transmission, the statistical parameters of the channel are assumed to be exactly known, and this knowledge is essentially used in choosing the optimal encoding and decoding. However, in many real situations one may regard as known only some constraint on the totality of statistical parameters, and then one must seek uniform methods of transmission for the entire class of channels whose parameters satisfy the prescribed constraints. The study of this question was begun in the works of Blackwell, Breiman, and Thomasian (¹), Wolfowitz (²) (see also (³, ⁴)) and the author (⁵), with the first two works treating, with detailed proofs, the case of homogeneous memoryless channels. A somewhat more general result was formulated in the author's paper (⁶). Here we wish to formulate results applicable to a rather general case.

Let X, Y be finite sets. A **channel segment of length n** is specified by the spaces

$$X^n = \{(x_1, \dots, x_n), x_i \in X\}$$

and

$$Y^n = \{(y_1, \dots, y_n), y_i \in Y\}$$

of the values of the signal at the input and output of the channel and by a set

$$P = \{p(b/a), b \in Y^n, a \in X^n\}$$

of transition probabilities such that

$$0 \leq p(b/a) \leq 1, \quad \sum_{b \in Y^n} p(b/a) = 1, \quad a \in X^n.$$

A channel segment is called a **memoryless channel segment** if, for $b = (y_1, \dots, y_n)$, $a = (x_1, \dots, x_n)$,

$$p(b/a) = p_1(y_1/x_1)p_2(y_2/x_2) \cdots p_n(y_n/x_n), \quad (1)$$

where $\{p_i(y/x)\}$ are the transition probabilities of a channel segment of length 1.

We shall call a **subsegment** of a memoryless channel segment with transition probabilities (1) the segment of length $k-j+1$, specified for some $1 \leq j \leq k \leq n$ by the transition probabilities

$$p(b/a) = p_j(y_1/x_1)p_{j+1}(y_2/x_2) \cdots p_k(y_{k-j+1}/x_{k-j+1}). \quad (2)$$

For fixed X and Y , for each n we shall consider an arbitrary collection T^n of memoryless channel segments of length n , and speak of a **system of channels** $T = \{T^n, n = 1, 2, \dots\}$. Introduce the following **consistency condition**: if a channel segment $P \in T^n$, then any of its subsegments having length $m < n$ belongs to T^m .

A **code** \mathfrak{A} of size N and length n will mean a collection of probability distributions $\{q_j(a), a \in X^n, j = 1, \dots, N\}$ and a function $\varphi(b), b \in Y^n$, with values $j = 1, \dots, N$. A code is called **nonrandomized** if, for each j , the distribution $\{q_j(a)\}$ is concentrated at one point. The **error probability** for a code \mathfrak{A} and a channel segment P will mean

$$r(\mathfrak{A}, P) = \frac{1}{N} \sum_{j=1}^N \sum_{a \in X^n} \sum_{b: \{\varphi(b) \neq j\}} q_j(a) p(b/a). \quad (3)$$

The principal quantity of interest to us is defined as

$$r(n, N) = \inf_{\mathfrak{A}} \sup_{P \in T^n} r(\mathfrak{A}, P), \quad (4)$$

where \mathfrak{A} ranges over all corrections of volume N and length n . We shall call the **capacity** $C(T)$ of the system of channels T the least upper bound of the numbers R such that $r(n, [2^{nR}]) \rightarrow 0$ as $n \rightarrow \infty$. Our aim is to compute $C(T)$. An essential feature of this formulation of the problem is that, unlike other works by authors on this topic, we allow randomization in encoding. It is easy to give examples showing that if one restricts oneself to nonrandomized corrections, then Theorem 1 will be false (for example, in the setting of Remark 3). The intuitive meaning of the randomization used here is that, for a given value of the message j , the transmitted signal is chosen with distribution $\{q_j(a)\}$. From the point of view of the final result of transmission, the random “errors” introduced by the random choice of transmitted signals are indistinguishable from the “random errors” arising in transmission over the channel and present also for nonrandomized corrections. This is the essential difference between our formulation of the problem and the formulation introduced in [7], where correlated randomization at the input and output of the channel is allowed. To implement it, an additional communication channel between the input and output of the main channel is needed. Under such an approach, the value of $C(T)$ turns out, generally speaking, to be larger. On the other hand, expression (6) would remain valid for $C(T)$ if we allowed randomization in the choice of the decoding function $\varphi(b)$.

We shall call the **information** corresponding to some joint probability distribution $\pi(e, f)$, $e \in E$, $f \in F$, where E and F are finite sets, the number

$$I(\{\pi(e, f)\}) = \sum_{\substack{e \in E \\ f \in F}} \pi(e, f) \log \frac{\pi(e, f)}{\sum_{e \in E} \pi(e, f) \sum_{f \in F} \pi(e, f)}.$$

For any set G of channel segments of length n , we shall denote by

$$c(G) = \sup_{\{g(a)\}} \inf_{P \in G} I(\{g(a)p(b/a)\}), \quad (5)$$

where $\{g(a)\}$ are all possible probability distributions on X^n . For each fixed $a_0 \in X^n$, let Q_{a_0} denote the minimal convex closed system of probability distributions on Y^n containing all distributions $\{p(b/a_0)\}$, where $P \in G$. The **row-wise convex closure** $[G]$ of the set G is defined to be the set of all segments $P = \{p(b/a)\}$ of length n such that, for any a_0 , the distribution $\{p(b/a_0)\} \in Q_{a_0}$. The visual difference between the row-wise convex closure and the ordinary convex closure of a system of matrices is that for each row the linear combination is taken with its own coefficients.

Theorem 1. For any system of memoryless channels with the consistency condition,

$$C(T) = \lim_{n \rightarrow \infty} \frac{1}{n} c([T^n]). \quad (6)$$

Remark 1. Suppose a certain transition probability matrix $\Pi = \{\pi(y/x), y \in Y, x \in X\}$ is given (a channel segment of length 1) and T^n consists of a single channel $P = \{p(b/a)\}$ with $b = (y_1, \dots, y_n)$, $a = (x_1, \dots, x_n)$,

$$p(b/a) = \pi(y_1/x_1) \cdots \pi(y_n/x_n). \quad (7)$$

Then $c([T^n]) = nc(\Pi)$, and Theorem 1 reduces to the well-known assertion of Feinstein's lemma for homogeneous memoryless channels.

Remark 2. Suppose that a certain set Γ of matrices Π is given and that the collection T^n consists of all channel segments of the form (7), where $\Pi = \{\pi(y/x)\} \in \Gamma$ (i.e., the system T^n is a system of homogeneous channels without memory). Then $c([T^n]) \sim nc(\Gamma)$ as $n \rightarrow \infty$, so that $C(T) = c(\Gamma)$. Theorem 1 is thus reduced to a result proved in ^(1,2).

Remark 3. Suppose that the system T^n consists of all segments $P = \{p(b/a)\}$ such that, for $b = (y_1, \dots, y_n)$, $a = (x_1, \dots, x_n)$,

$$p(b/a) = p_1(y_1/x_1)p_2(y_2/x_2) \cdots p_n(y_n/x_n),$$

where $\{p_i(y/x)\}$ belongs to Γ for all $i = 1, \dots, n$ (i.e., the system T^n consists of all nonhomogeneous channels without memory with prescribed transition probabilities at one instant). Then $c([T^n]) = nc([\Gamma])$ and $C(T) = c([\Gamma])$. This result is formulated in (6).

Problems analogous to those considered above also arise as applied to messages (see (6)). Let U and V be finite sets, and let $\rho(u, v)$, $u \in U$, $v \in V$, be a nonnegative function, called the **loss function**. Let $U^n = \{(u_1, \dots, u_n), u_i \in U\}$ and $V^n = \{(v_1, \dots, v_n), v_i \in V\}$ be the spaces of values of the message at the input and at the output, and let, for $d = (u_1, \dots, u_n)$, $e = (v_1, \dots, v_n)$,

$$\rho_n(d, e) = \sum_{i=1}^n \rho(u_i, v_i). \quad (8)$$

Intuitively, $\rho_n(d, e)$ specifies the “loss” arising if, as a result of transmission, the input message d is transformed into the output message e . A segment of an input message Q of length n is specified as a probability distribution $Q = \{q(d), d \in U^n\}$. A message segment is called a **message segment with independent components** if

$$q(d) = q_1(u_1)q_2(u_2) \cdots q_n(u_n), \quad (9)$$

where $q_i(u)$, $u \in U$, are certain probability distributions. We shall call a **subsegment of a message with independent components**, specified by the probabilities (8), a segment of length $k - j + 1$, specified, for some $1 \leq j \leq k \leq n$, by the probabilities

$$q(d) = q_j(u_1)q_{j+1}(u_2) \cdots q_k(u_{k-j+1}). \quad (10)$$

For fixed U and V , for each n we shall consider an arbitrary collection S^n of input message segments with independent components of length n , and speak of a **system of messages** $S = \{S^n, n = 1, 2, \dots\}$. We shall assume that the following **consistency condition** is satisfied: if a message segment $Q \in S^n$, then any of its subsegments having length $m < n$ belongs to S^m .

By a **quantization** \mathfrak{B} of size M and length n we shall mean a collection of M elements e_1, \dots, e_M of the space V^n and a function $\psi(d)$, $d \in U^n$, with values e_1, \dots, e_M . The intuitive meaning of this definition (cf. (8, 9)) is that whenever the input message has taken a value d such that $\psi(d) = e_i$, the “ i -th signal” is transmitted over the channel, which at the output is decoded as the value of the output message e_i . The **mean loss for the quantization** \mathfrak{B} and the message segment Q is called

$$\varepsilon(\mathfrak{B}, Q) = \sum_{d \in U^n} q(d) \rho_n(d, \psi(d)). \quad (11)$$

The principal quantity of interest to us is defined as

$$\varepsilon(n, M) = \inf_{\mathfrak{B}} \sup_{Q \in S^n} \varepsilon(\mathfrak{B}, Q), \quad (12)$$

where \mathfrak{B} is all possible quantizations of volume M and length n . For any $\varepsilon \geq 0$ we shall call the ε -entropy $H(\varepsilon, S)$ of the message system S the greatest lower bound of the numbers R such that $\overline{\lim}_{n \rightarrow \infty} \varepsilon(n, [2^{nR}]) \leq \varepsilon$. Our aim is to compute $H(\varepsilon, S)$. We note that here we are considering the case of nonrandomized coding; however, Theorem 1 would remain true even if we allowed a random choice of $\psi(d)$ (cf. ⁽⁹⁾).

For any collection D of input message segments of length n and any $\varepsilon \geq 0$, we shall denote by

$$h(\varepsilon, D) = \sup_{Q \in D} \inf_{F_Q} I(\{q(d)f(e/d)\}), \quad (13)$$

where $F_Q = \{f(e/d), e \in V^n, d \in U^n\}$ is the collection of all possible sets of transition probabilities:

$$\sum_{e \in V^n} f(e/d) = 1, \quad d \in U^n, \quad 0 \leq f(e/d) \leq 1$$

such that

$$\sum_{d \in U^n} \sum_{e \in V^n} \rho_n(d, e) q(d) f(e/d) \leq \varepsilon. \quad (14)$$

By $[D]$ we shall denote the minimal closed convex system of probability distributions on U^n containing all distributions $Q \in D$.

Theorem 2. *For any message system with independent components $S = \{S^n\}$, for which the consistency conditions hold,*

$$H(\varepsilon, S) = \lim_{n \rightarrow \infty} \frac{1}{n} h(\varepsilon, [S^n]). \quad (15)$$

Remark 4. Let some probability distribution $\Pi = \{\pi(u), u \in U\}$ be given (an input message segment of length 1), and let S^n consist of a single message $Q = \{q(d)\}$, where for $d = (u_1, \dots, u_n)$

$$q(d) = \pi(u_1) \cdots \pi(u_n). \quad (16)$$

Then $h(\varepsilon, [S^n]) = nh(\varepsilon, \Pi)$, and Theorem 2 reduces to the result first proved by Shannon ⁽⁸⁾ and by the author ⁽⁹⁾.

Remark 5. Let some set Γ of distributions Π be given, and let the collection S^n consist of all messages of the form (16), where $\Pi \in \Gamma$ (i.e., the system S^n is a system of input messages with independent identically distributed components). Then $h(\varepsilon, [S^n]) \sim nh(\varepsilon, \Gamma)$, so that $H(\varepsilon, S) = h(\varepsilon, \Gamma)$. Theorem 1 then reduces to the result formulated by the author in ⁽⁶⁾.

Remark 6. Let the system S^n consist of all distributions $\{q(d)\}$ such that, for $d = (u_1, \dots, u_n)$,

$$q(d) = \pi_1(u_1)\pi_2(u_2) \cdots \pi_n(u_n), \quad (17)$$

where $\{\pi_i(u)\}$, for all $i = 1, \dots, n$, belong to Γ (i.e., the system S^n consists of all messages with independent components for which the component distributions belong to Γ). Then $h(\varepsilon, [S^n]) = nh(\varepsilon, [\Gamma])$ and $H(\varepsilon, S) = h(\varepsilon, [\Gamma])$. This result was also formulated in ⁽⁶⁾.

Moscow State University
named after M. V. Lomonosov

Received
4 VIII 1962

CITED LITERATURE

1. D. Blackwell, L. Breiman, A. J. Thomasian, *Ann. Math. Stat.*, **29**, 1223 (1959).
2. J. Wolfowitz, *Arch. Rat. Mech. and Anal.*, **4**, 371 (1960).
3. J. Wolfowitz, *IRE Trans., Circuit Theory*, **7**, 513 (1960).
4. J. Wolfowitz, *Inform. and Decision Proc.*, N. Y.—Toronto—London, 1960, p. 178.
5. R. L. Dobrushin, *Radiotekhn. i elektronika*, **4**, 1951 (1959).
6. R. L. Dobrushin, *Proc. Fourth Berkl. Symp. on Math. Stat. and Prob.*, **1**, 1961, p. 211.
7. D. Blackwell, L. Breiman, A. J. Thomasian, *Ann. Math. Stat.*, **31**, 558 (1960).
8. C. E. Shannon, *IRE Conv. Rec.*, **7**, 142 (1959).
9. R. L. Dobrushin, *UMN*, **14**, No. 6, 3 (1959).

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

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.