



Soviet-era science, translated into English

R. L. Dobrushin

1963

SovietRxiv

View the original and related papers at <https://sovietsrxiv.org/items/ru-196301.67919>

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

Abstract

Full Text

R. L. Dobrushin

Uniform Methods of Information Transmission. The General Case

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

In this note we study the same problems as in our paper ⁽¹⁾, but in a more general setting. We shall regard the definitions and notation from ⁽¹⁾ as known.

We investigate a message whose components are not independent. Let $1 \leq j \leq k \leq n$, and let, for $d = (u_1, \dots, u_n)$, the function

$$\gamma_{jk}(d) = (u_j, u_{j+1}, \dots, u_k) \in U^{k-j+1}.$$

Generalizing definition (10) from ⁽¹⁾, we shall call a segment of length $k - j + 1$, specified by the probabilities

$$\tilde{q}(\tilde{d}) = \sum_{\gamma_{jk}(d)=\tilde{d}} q(d), \quad \tilde{d} \in U^{k-j+1},$$

a **subsegment of a message segment** of length n , specified by the probabilities $\{q(d), d \in U^n\}$.

We shall call a **system of messages** a collection

$$S = \{S^n, n = 1, 2, \dots\},$$

where S^n is some collection of segments of length n , and we shall assume that the consistency condition formulated in ⁽¹⁾ is fulfilled.

For any $A \subset U^n$, put

$$q(A) = \sum_{d \in A} q(d).$$

We shall call the **σ -algebra of events** \mathfrak{B}_{jk} the system of sets A representable in the form

$$A = \{\gamma_{jk}(d) \in \tilde{A}\}, \quad \tilde{A} \subset U^{k-j+1}. \quad (1)$$

We shall call the m -th **coefficient of strong mixing** (see ⁽²⁾) ($0 < m \leq n$) the number

$$a_m(Q) = \sup |Q(A \cap B) - Q(A)Q(B)|,$$

where the supremum is taken over all s , $0 \leq s < s + m \leq n$, and all $A \in \mathfrak{B}_{0,s}$, $B \in \mathfrak{B}_{s+m,n}$.

Let

$$a_m(S) = \sup_{Q \in S^n, n \geq m} a_m(Q).$$

We introduce the following **condition of uniform strong mixing**: for some $\chi > 0$,

$$a_m(S) = O(e^{-\chi m}).$$

Theorem 1. For any system of messages $S = \{S^n\}$ for which the consistency and uniform strong mixing conditions hold,

$$H(\varepsilon, S) = \lim_{n \rightarrow \infty} \frac{1}{n} h(\varepsilon, [S^n]). \quad (2)$$

For special cases see (4, 5).

Let us consider channels with memory. We shall say that a channel segment

$$P = \{p(b/a)\}$$

is a **channel segment without anticipation** if, for $b = (y_1, \dots, y_n)$, $a = (x_1, \dots, x_n)$, and any specified $x_1, \dots, x_k, y_1, \dots, y_k$, the sum

$$\sum_{y_{k+1}, \dots, y_n \in Y} p(y_1, \dots, y_n, x_1, \dots, x_n)$$

does not depend on the choice of x_{k+1}, \dots, x_n .

We shall say that a **system of channels**

$$T = \{T^n, n = 1, 2, \dots\}$$

is given if T^n is some collection of channel segments without anticipation of length n .

We shall say that for the system the **condition of consistency and strong ergodicity** is satisfied if, for some $\chi > 0$ and $K < \infty$, and for any $0 \leq j \leq k \leq n$, to every probability distribution $g(x_1, \dots, x_j)$ on $(x_1, \dots, x_j) \in X^j$ one can assign a probability distribution $\bar{g}(x_1, \dots, x_k)$ on $(x_1, \dots, x_k) \in X^k$ such that:

a) for all x_1, \dots, x_j

$$g(x_1, \dots, x_j) = \sum_{x_{j+1}, \dots, x_k} \bar{g}(x_1, \dots, x_k);$$

b) for any $\bar{a} = (x_{k+1}, \dots, x_n) \in X^{n-k}$ and any segment $\{p(b/a)\} \in T^n$, the conditional probability distribution $\hat{p}(\bar{b}/\bar{a})$ on the set of sequences $\bar{b} = (y_{k+1}, \dots, y_n)$, defined by the formula

$$\hat{p}(\bar{b}/\bar{a}) = \sum_{x_1, \dots, x_k} \sum_{y_1, \dots, y_k} \bar{g}(x_1, \dots, x_k) p(y_1, \dots, y_n/x_1, \dots, x_n)$$

is such that there exists a channel segment $\{p'(\bar{b}/\bar{a})\} \in T^{n-k}$ such that

$$\sum_{\bar{b} \in Y^{n-k}} |\hat{p}(\bar{b}/\bar{a}) - p'(\bar{b}/\bar{a})| \leq K e^{-\chi(k-j)};$$

- c) for any $\bar{a} = (x_{k+1}, \dots, x_n) \in X^{n-k}$ and any segment $\{p(b/a)\} \in T^n$, and for the conditional probability distribution $\tilde{p}(b/\bar{a})$, defined by the formula

$$\tilde{p}(b/\bar{a}) = \sum_{x_1, \dots, x_k} \bar{g}(x_1, \dots, x_k) p(y_1, \dots, y_n / x_1, \dots, x_n),$$

and

$$\tilde{p}(B/\bar{a}) = \sum_{b \in B} \tilde{p}(b/\bar{a}),$$

where $B \subset Y^n$, for any sets $A \in \mathfrak{B}_{0,j}$ and $B \in \mathfrak{B}_{k+1,n}$ (the definition of the σ -algebras \mathfrak{B}_{jk} is given analogously to (1), but with U replaced by Y) the inequality

$$|\tilde{p}(A \cap B/\bar{a}) - \tilde{p}(A/\bar{a})\tilde{p}(B/\bar{a})| \leq ce^{-\chi(k-j)}$$

holds.

The intuitive meaning of this rather weak restriction is that we assume that, for arbitrary random signals at the input at the moments $1, \dots, j$, it is possible to choose signals at the input at the intermediate moments $j+1, \dots, k$ so that (consistency condition b)) transmission at the remaining time moments $k+1, \dots, n$ proceeds almost as in some segment of the system T^{n-k} , and so that (strong ergodicity condition), for a given input signal at the moments $k+1, \dots, n$, the output signal at the moments $1, \dots, j$ is only weakly dependent on the output signal at the moments $k+1, \dots, n$.

Theorem 2. *For any system of channels $T = \{T^n\}$ for which the condition of consistency and strong ergodicity is satisfied,*

$$C(T) = \lim_{n \rightarrow \infty} \frac{1}{n} c(|T^n|). \quad (3)$$

For particular cases see (4, 7).

Suppose now that U, V are arbitrary measurable spaces with σ -algebras of measurable sets \mathfrak{B}_U and \mathfrak{B}_V . The definition of a message with independent components, quantization, mean loss, etc., is given by obvious analogy with what was done in the discrete case. The **information corresponding to the joint distribution $\pi(A)$ on the product of measurable spaces $(E \times F, \mathfrak{B}_E \times \mathfrak{B}_F)$** is defined as

$$I(\{\pi(A)\}) = \sup I(\{\tilde{\pi}(\tilde{e}, \tilde{f})\}), \quad (4)$$

where the supremum is taken over all possible partitions $E = \tilde{e}_1 \cup \dots \cup \tilde{e}_r$, $F = \tilde{f}_1 \cup \dots \cup \tilde{f}_s$, $\tilde{e}_i \cap \tilde{e}_j = 0$, $\tilde{f}_i \cap \tilde{f}_j = 0$, $i \neq j$, $\tilde{e}_i \in \mathfrak{B}_E$, $\tilde{f}_j \in \mathfrak{B}_F$, where

$\tilde{\pi}(\tilde{e}_i, \tilde{f}_j) = \pi(\tilde{e}_i \times \tilde{f}_j)$ (for more detail see (4)), and, analogously to (13) from (1), we introduce the quantities

$h(\varepsilon, D)$, corresponding to any set D of message segments. We shall regard the set of all probability distributions $Q(A)$, $A \in \mathfrak{B}_U^n$, where \mathfrak{B}_U^n is the n -th power of the σ -algebra \mathfrak{B}_U , as a metric space with distance

$$\gamma(Q_1, Q_2) = \sup_{A \in \mathfrak{B}_U^n} |Q_1(A) - Q_2(A)|, \quad (5)$$

and it is precisely in the sense of this distance that we shall understand the maximal closed convex system of distributions $[D]$ containing all distributions D . We shall call a δ -net of cardinality K for a system of distributions D on U^n a set Q_1, \dots, Q_K of distributions on \mathfrak{B}_U^n such that, for all $Q \in D$, $\min_{i=1, \dots, K} \gamma(Q, Q_i) \leq \delta$. We shall denote by $\Gamma(\delta, D)$ the least size of a δ -net for the system D .

Theorem 3. *For any system of messages with independent components $S = \{S^n\}$, for which the consistency conditions are satisfied, and for any $\chi > 0$ and $\delta > 0$,*

$$\Gamma\left(\frac{\delta}{n}, S^1\right) = o(2^{2^{\chi n}}) \quad (n \rightarrow \infty), \quad (6)$$

the relation

$$H(\varepsilon, S) = \lim_{n \rightarrow \infty} \frac{1}{n} h(\varepsilon, [S^n]) \quad (7)$$

holds.

Results analogous to those formulated in remarks 4, 5, 6 from ⁽¹⁾ are valid. The result corresponding to remark 4 was proved earlier by the author ⁽⁴⁾. Condition (6), which is a generalization of the compactness requirement, is not very burdensome. It is always true if the space U is finite; and if (U, \mathfrak{B}_U) is a Euclidean space and the distributions $Q \in S^1$ are given by densities, then it is sufficient that these densities, together with their first partial derivatives, be uniformly integrable with respect to S^1 and $u \in U$, and uniformly bounded. In the case where S^1 consists of Gaussian distributions, it is sufficient that the determinant of the covariance matrix and its reciprocal be uniformly bounded.

Let us further suppose that X and Y are arbitrary measurable spaces with σ -algebras of measurable sets \mathfrak{B}_X and \mathfrak{B}_Y . The definitions of a segment of a channel without memory, the consistency condition, correction, and error probability are given in the same way as in the discrete case. In addition, suppose (cf. ⁽⁴⁾) that a measurable function $\tau(x) \geq 0$, $x \in X$, is given. Let, for $\alpha = (x_1, \dots, x_n)$, $\tau(\alpha) = \tau(x_1) + \dots + \tau(x_n)$. We shall call the **mean weight for a correction** \mathfrak{A}

$$k(\mathfrak{A}) = \frac{1}{N} \sum_{j=1}^N \sum_{a \in X^n} q_j(a) \tau(a).$$

For any $\varepsilon > 0$ we shall put

$$r(\varepsilon, n, N) = \inf_{k(\mathfrak{A}) \leq \varepsilon} \sup_{P \in T^n} r(\mathfrak{A}, P),$$

where \mathfrak{A} ranges over all corrections of volume N and length n such that $k(\mathfrak{A}) \leq \varepsilon$. We shall call the **capacity** $C(\varepsilon, T)$ of the system of channels T , for a given weight ε , the least upper bound of the numbers R such that, for all $\tilde{\varepsilon} > \varepsilon$,

$$r(\tilde{\varepsilon}, n, [2^{nR}]) \rightarrow 0$$

as $n \rightarrow \infty$. The definition of capacity is given analogously to (5) from ⁽¹⁾, with the only difference that the definition of information (4) is used and the upper bound is taken not over all distributions $g(A)$ on $A \in \mathfrak{B}_X^n$, but only over such distributions $g(A)$ that

$$\int_{X^n} \tau(a) g(da) \leq \varepsilon.$$

The convex closure $[G]$ of the aggregate G is defined in the same way as in the discrete case, but using the metric specified by condition (5). For any two segments of the channel $P_1(A/x)$ and $P_2(A/x)$, $A \in \mathfrak{B}_y$, $x \in X$, we shall denote

$$\gamma(P_1, P_2) = \sup_{x \in X, A \in \mathfrak{B}_y} |P_1(A/x) - P_2(A/x)|.$$

We shall call a set P_1, \dots, P_K of such segments a δ -net of cardinality K for the system of channel segments G of length n if for all $P \in G$

$$\min_{i=1, \dots, K} \gamma(P, P_i) \leq \delta.$$

We shall denote by $\Delta(\delta, G)$ the smallest size of a δ -net for the system G .

Theorem 4. For any system of memoryless channels $T = \{T^n\}$, for which the consistency conditions hold, and for any $\chi > 0$, $\delta > 0$,

$$\Delta\left(\frac{\delta}{n}, T^1\right) = o(2^{2^{xn}}) \quad (n \rightarrow \infty), \quad (8)$$

the relation

$$C(\varepsilon, T) = \lim_{n \rightarrow \infty} \frac{1}{n} c(\varepsilon, [T^n]). \quad (9)$$

holds.

Results analogous to those formulated in remarks 1, 2, 3 of ⁽¹⁾ are valid. The result corresponding to remark 1 was proved earlier by the author ⁽⁴⁾. Condition (9) is fairly general and is satisfied in “good cases” if the space X is finite or compact. At the same time, if, for example, X is a Euclidean space, then it is usually false. Here the following easily proved remark is helpful.

Let $\tilde{X} \in \mathfrak{B}_x$. For any channel segment P of length n , let $P_{\tilde{X}}$ denote the segment obtained if the domain of definition of the transition function of the channel $P(B/a)$ is restricted to values $a \in \tilde{X}^n$. Let $G_{\tilde{X}}$ be the aggregate of segments $P_{\tilde{X}}$, where $P \in G$. Theorem 4 is valid if there exists a sequence $\tilde{X}_k \in \mathfrak{B}_x$ such that this theorem is applicable to the system

$$T_{\tilde{X}_k} = (T_{\tilde{X}_k}^n, n = 1, 2, \dots)$$

and

$$\lim_{k \rightarrow \infty} c(\varepsilon, T_{\tilde{X}_k}) = c(\varepsilon, T).$$

In this form the theorem turns out to be widely applicable. For example, it is valid for a broad class of Gaussian memoryless channels (the case of homogeneous memoryless channels with additive Gaussian noise was studied by Wolfowitz ⁽⁸⁾).

Finally, we note that the main theorem is also valid for messages with dependent values and an arbitrary set of states, if the condition of uniform strong mixing and the condition

$$\Gamma\left(\frac{\delta}{n}, S^{[\log n]^2}\right) = o(2^{2^{X^n}}),$$

which generalizes (6), hold. An analogous generalization is also valid for channels.

Moscow State University
named after M. V. Lomonosov

Received
4 VIII 1962

REFERENCES CITED

- ¹ R. L. Dobrushin, DAN, **148**, No. 6 (1963).
- ² M. Rosenblatt, Proc. Nat. Acad. Sci. USA, **42**, 43 (1956).
- ³ R. L. Dobrushin, Theory of Probability and Its Applications, **1**, 72 (1956).
- ⁴ R. L. Dobrushin, UMN, **14**, No. 6, 3 (1959).
- ⁵ C. E. Shannon, IRE Conv. Rec., **7**, 142 (1959).
- ⁶ A. Feinstein, *Foundations of Information Theory*, N. Y., 1959; A. Feinstein, *Foundations of Information Theory*, Moscow, 1960.
- ⁷ D. Blackwell, L. Breiman, A. J. Thomassian, Ann. Math. Stat., **29**, 12 (1959).
- ⁸ J. Wolfowitz, Inform. and Decision Proc., N.Y.—Toronto, 1960, p. 178.

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

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