

THE THEORY OF INFORMATION TRANSMISSION THROUGH STOCHASTIC COMMUNICATION CHANNELS

We shall denote the channel specified by these elements by

1957

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-195701.07683>

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

Abstract

Full Text

MATHEMATICS

M. Rosenblatt-Roth

THE THEORY OF INFORMATION TRANSMISSION THROUGH STOCHASTIC COMMUNICATION CHANNELS

(Presented by Academician A. N. Kolmogorov on 20 VII 1956)

1. The concept of a stochastic source; the concept of a stochastic channel; connection of a channel with a feeding source. In the statistical theory of transmission, the output of any information source A is understood as a certain random process; the source itself is defined by means of the stochastic structure of this process and, consequently, is characterized as in ⁽²⁾. A stochastic channel is characterized by: a) the elements that can enter the channel; at each instant τ one and only one element x_τ may enter, and these elements form a space with measure $(\mathfrak{A}_\tau, \mathcal{F}_\tau, \mu_\tau)$ ($\tau \in I$); b) the elements that can leave the channel; at each instant τ one and only one element y_τ may leave the channel, and these elements form a space with measure $(\mathfrak{B}_\tau, \Sigma_\tau, \nu_\tau)$; c) the law of transmission, which is specified by means of the probability density of the exit, during the time $[t, t+n]$, from the channel of the element $y^{[t, t+n]} = (y_t, \dots, y_{t+n}) \in \mathfrak{B}^{[t, t+n]}$, if it is known that elements have entered the input which form the sequence* $x = (\dots x_{-1}, x_0, x_1, \dots) \in \mathfrak{A}$, $x_\tau \in \mathfrak{A}_\tau$, $\tau \in I$, i.e., by means of a system of functions $\pi_{B|A}^{[t, t+n]}(y^{[t, t+n]} | x)$. For a given $x \in \mathfrak{A}$ there exists ⁽³⁾ (and only one) probability measure $P_{B|A}(| x)$ on $\mathfrak{B} = \dots \times \mathfrak{B}_{-1} \times \mathfrak{B}_0 \times \mathfrak{B}_1 \times \dots$, which is an extension of the given distributions.

We shall denote the channel specified by these elements by

$$\Delta = [\mathfrak{A}, P_{B|A}(| x), \mathfrak{B}].$$

If

$$\pi_{B|A}^{[t, t+n]}(y^{[t, t+n]} | x) \equiv \pi_{B|A}^{[t, t+n]}(y^{[t, t+n]} | x^{[-\infty, t+n]})$$

for all $y^{[t, t+n]} \in \mathfrak{B}^{[t, t+n]}$, $x \in \mathfrak{A}$, $t \in I$, $n \geq 0$, we shall say that there is a **channel without anticipation**. If

$$\pi_{B|A}^{[t, t+n]}(y^{[t, t+n]} | x) = \pi_{B|A}^{[t, t+n]}(y^{[t, t+n]} | x^{[t-m, t+n]})$$

for all $y^{[t, t+n]} \in \mathfrak{B}^{[t, t+n]}$, $x \in \mathfrak{A}$, $t \in I$, $n \geq 0$, we shall say that there is a **channel with finite memory**. The smallest number m for which what has been described holds is called the **memory of the channel**; if it is equal to

zero, one says that there is a **channel with independent noises**. Let there be a source A and a channel Δ such that the set of elements \mathfrak{A}_τ which the given source can produce at the instant τ coincides with the set of elements that can enter the given channel at the instant τ , and this for all $\tau \in I$. In this case one says that the given source **feeds** the given channel. If the given source A feeds the channel Δ , then at the output of the channel a source B is defined; the double source AB is also defined.

* We shall use the notation from (2), sometimes indicating by a subscript the source to which it refers.

2. Entropy of a channel with a source feeding it.

By $B^{[t,t+n-1]}$ we denote the field of chains $y^{[t,t+n-1]}$ of the source B . For a given source feeding the channel, $\pi_{A|B}^{[t,t+n-1]}(x^{[t,t+n-1]} | y^{[t,t+n-1]})$ is determined and, consequently, $H(A^{[t,t+n-1]} | B^{[t,t+n-1]})$ as the mean value of the random variable $H(A^{[t,t+n-1]} | y^{[t,t+n-1]})$. Let $y = (\dots y_t \dots y_{t+n-1} \dots)$ and

$$f_{A|B}^{[t,t+n-1]}(x, y) = -n^{-1} \log \pi_{A|B}^{[t,t+n-1]}(x^{[t,t+n-1]} | y^{[t,t+n-1]}).$$

Definition. The **entropy of the channel** Δ , fed by the source A at time t , is the quantity

$$\begin{aligned} H_t(A | B) &= \lim_{n \rightarrow \infty} M_{AB} f_{A|B}^{[t,t+n-1]}(x, y) \\ &= \lim_{n \rightarrow \infty} n^{-1} H(A^{[t,t+n-1]} | B^{[t,t+n-1]}), \end{aligned}$$

(if this limit exists).

Theorem 1. For the existence of $H_t(A | B)$ it is necessary and sufficient that the sequence

$$H(A_{t+n} \times B_{t+n} | A^{[t,t+n-1]} \times B^{[t,t+n-1]}) - H(B_{t+n} | B^{[t,t+n-1]}) \quad (n = 1, 2, \dots),$$

be Cesàro summable $C(1)$, and $H_t(A | B)$ is the limit of these sums. A sufficient condition for $H_t(A | B)$ to exist and be finite is the existence and finiteness of $H_t(B)$ and $H_t(AB)$, and in this case

$$H_t(AB) = H_t(B) + H_t(A | B)$$

(convergence is understood in the sense of convergence to a finite number or to $\pm\infty$).

It is not hard to prove for $H_t(A | B)$ a theorem analogous to Theorem 3 of (2). If we exclude the case where $|\widetilde{H}_t^{(m)}(B)| = \infty$, $|\widetilde{H}_t^{(m)}(AB)| = \infty$ for at least one $t \in I$, $m > 0$, the following theorem holds.

Theorem 2. For all channels with their feeding sources,

$$H_t(A | B) \equiv H(A | B) = \text{const} \quad (t \in I)$$

(if this entropy exists)*.

3. Properties $\mathcal{E}_t(A | B)$ and $\mathcal{E}(A | B)$.

Definition. If $f_{A|B}^{[t,t+n-1]}(x, y)$ converges in probability to $H_t(A | B)$, we shall say that the property $\mathcal{E}_t(A | B)$ holds. If this property holds for all $t \in I$, we shall say that the property $\mathcal{E}(A | B)$ holds.

Let

$$g_{A|B}^{[t,t+n]}(x, y) = g_{AB}^{[t,t+n]}(x, y) - g_B^{[t,t+n]}(y).$$

Theorem 3. In order that a channel Δ with source A have the property $\mathcal{E}_t(A | B)$, it is necessary and sufficient that the sequence of random variables

$$g_{A|B}^{[t,t+n]}(x, y) \quad (n = 1, 2, \dots)$$

obey the law of large numbers. For this it is sufficient that the properties $\mathcal{E}_t(AB)$ and $\mathcal{E}_t(B)$ hold simultaneously, or that

$$\lim_{n \rightarrow \infty} D_{AB} f_{A|B}^{[t,t+n-1]}(x, y) = 0.$$

Theorem 4. If there is a channel Δ with memory of unit length and the source A is a simple Markov chain, then sufficient conditions for the property $\mathcal{E}_t(A | B)$ to hold are:

a)

$$\lim_{n \rightarrow \infty} n^{\beta-2} \sum_{k=0}^{n-1} D_{AB} g_{A|B}^{[t,t+n]}(x, y) = 0, \quad \text{if } \alpha_{i,i+1} > 0 \quad (1 \leq i < \infty);$$

$$\eta_n = \max_{1 \leq i \leq n-1} (1 - \alpha_{i,i+1}); \quad 1 - \eta_n^{1/2} = O(n^{-\beta}) \quad (0 \leq \beta < 1)**;$$

b)

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{k=0}^{n-1} D_{AB} g_{A|B}^{[t,t+k]}(x, y) = 0 \quad \text{in all cases.}$$

It is not hard to prove for $\mathcal{E}_t(A | B)$, $\mathcal{E}(A | B)$ theorems analogous to Theorems 7 and 8 of (2). Let $L(P_{AB})$ be the space of all real functions $f(x, y)$ of the variable $(x, y) \in \mathfrak{A} \times \mathfrak{B}$ such that $M_{AB}|f(x, y)| < \infty$.

* In particular, the theorem is true for channels with finite sets of states \mathfrak{A}_τ at the input and \mathfrak{B}_τ at the output ($\tau \in I$).

** Here $\alpha_{i,i+1}$ denote the ergodicity coefficients (4) of the Markov chain AB .

Theorem 5. The sequence of functions $f_{A|B}^{[t,t+n-1]}(x, y)$ ($n = 1, 2, \dots$) cannot converge in mean to any constant other than $H_t(A | B)$. If the channel Δ with source A does not have finite entropy, then the sequence of functions $f_{A|B}^{[t,t+n-1]}(x, y)$ ($n = 1, 2, \dots$) cannot converge in mean to any element of $L(P_{AB})$.

4. Stationary channels fed by stationary sources

Theorem 6. If the channel Δ is stationary and is fed by a stationary source A , then the sources AB and B are also stationary.*

Theorem 7. If the channel Δ is stationary and the source A feeding it is also stationary, then $H(A | B)$ exists and is finite.

Let

$$g_{A|B}^{[1,n]}(x, y) = g_{n(A|B)}(x, y).$$

Theorem 8. In order that a stationary channel Δ , fed by a stationary source A , possess the property $\mathcal{E}(A | B)$, it is necessary and sufficient that the sequence of random variables

$$g_{A|B}^{[0,n]}(x, y) = g_{n(A|B)}(T^n x, T^n y)$$

obey the law of large numbers. For this it is sufficient that the properties $\mathcal{E}(AB)$ and $\mathcal{E}(B)$ exist.

Theorem 9. Suppose there is a stationary channel Δ , fed by a stationary source A , such that: a) $g_{n(A|B)}(x, y) \in L(P_{AB})$; b) there exists some function $g_{A|B}(x, y) \in L(P_{AB})$ such that the sequence $g_{n(A|B)}(x, y)$, as $n \rightarrow \infty$, converges in mean (in $L(P_{AB})$) to $g_{A|B}(x, y)$. Under these conditions**, if the source AB is ergodic, the property $\mathcal{E}(A | B)$ holds.

It follows from Theorem 8 that, for $\mathcal{E}(A | B)$, ergodicity of the source is not necessary.

Theorem 10. Under the conditions of Theorem 9, if the source AB is ergodic, the sequence of random variables

$$g_{A|B}^{[0,n]}(x, y) = g_{n(A|B)}(T^n x, T^n y) \quad (n = 0, 1, \dots)$$

obeys the law of large numbers.

5. Feinstein's fundamental lemma ⁽⁶⁾

Definition. A source (stochastic process) A is **regular** if it has a finite entropy $H(A)$, independent of time, and the property $\mathcal{E}(A)$.

Definition. Let $F(\Delta)$ be the set of all sources A such that: 1) A can feed the channel Δ ; 2) A is regular; 3) the entropy of the channel Δ with source A , i.e. $H(A | B)$, exists, is finite, and is independent of time; 4) the property $\mathcal{E}(A | B)$ holds. Under these conditions we shall say that $F(\Delta)$ forms a **regular set of sources attached to the channel Δ** .

Definition. A channel Δ is **regular** if it has no anticipation and the regular set of sources attached to it is nonempty.

Definition. We shall call the **regular capacity** of a channel the quantity

$$C = \sup[H(A) - H(A | B)],$$

where the least upper bound is taken over all $A \in F(\Delta)$. We exclude from consideration the case $C = +\infty$.

In what follows we shall assume that the state sets \mathfrak{A}_t are discrete (finite or countable), while the \mathfrak{B}_τ are arbitrary ($\tau \in I$). Let the channel Δ be regular, with finite memory m . By $x_i^{[t-m, t+n-1]}$ ($1 \leq i \leq N_t$) we shall denote N_t chains of the set

$$\mathfrak{A}^{[t-m, t+n-1]} = \mathfrak{A}_{t-m} \times \dots \times \mathfrak{A}_{t+n-1},$$

and by $\mathfrak{B}_i^{[t, t+n-1]}$, N_t -measurable sets of the space $\mathfrak{B}^{[t, t+n-1]}$.

* For finite sets of input and output states, see (7). The author became acquainted with work (5) when all the results contained in (1, 2) and in the present note had already been obtained.

** These conditions are satisfied for a channel with finite sets of input and output states; moreover, if the source A is ergodic and the channel Δ has finite memory, the property $\mathcal{E}(A | B)$ holds.

Definition. Let λ be any constant number ($0 < \lambda < 1/2$). A group $\{x_i^{[t-m, t+n-1]}\}$, ($1 \leq i \leq N_t$), of chains of the space $\mathfrak{A}^{[t-m, t+n-1]}$ shall be called **λ -distinguishable** if there exists a group of measurable sets $\mathfrak{B}_i^{[t, t+n-1]}$ of the space $\mathfrak{B}^{[t, t+n-1]}$ such that: 1) $\mathfrak{B}_i^{[t, t+n-1]}$ ($1 \leq i \leq N_t$) do not intersect one another; 2) $P_{AB}(\mathfrak{B}_i^{[t, t+n-1]} | x_i^{[t-m, t+n-1]}) > 1 - \lambda$ ($1 \leq i \leq N_t$).

Feinstein's fundamental lemma. *If the channel Δ is regular and has finite memory m , then, however small $\lambda > 0$ may be, for sufficiently large n there exists a λ -distinguishable group $\{x_i^{[t-m, t+n-1]}\}$ ($1 \leq i \leq N_t$) of chains of the space $\mathfrak{A}^{[t-m, t+n-1]}$ with number of members $N_t > 2^{n(C-\lambda)}$, where C is the regular capacity of the channel.*

6. Shannon's fundamental theorems (7). Let the production of some regular source $\overset{0}{A}$ be subject to transmission through some regular channel Δ with finite memory m . Let the entropy of the source $\overset{0}{A}$ be equal to H_0 , and the regular capacity of the channel Δ be equal to C , with $\overset{0}{A} \notin F(\Delta)$. Let $x_\tau \in \mathfrak{A}_\tau$ ($\tau \in I$) be the sets of states of the source $\overset{0}{A}$.

Shannon' s first theorem. Suppose there are given: 1) a regular channel Δ with regular capacity C and finite memory m ; 2) a regular source $\overset{0}{A} \notin F(\Delta)$ with entropy $H_0 < C$. Let $\varepsilon > 0$ be given. Then the production of the source $\overset{0}{A}$, beginning at time t_0 , can be encoded so that transmission through the channel is possible beginning at time $t - m$. Moreover, there exists a function $n'(\tau_0, \tau, \varepsilon)$ such that the production of the source is divided into partial chains

$$\alpha_i(t_0^{(k)}) = \left(\overset{0}{x}_{t_0^{(k)}}, \overset{0}{x}_{t_0^{(k)}+1}, \dots, \overset{0}{x}_{t_0^{(k)}+n^{(k)}-1} \right) \quad (k = 0, 1, \dots; t_0^{(0)} = t_0)$$

of length $n^{(k)} > n'(t_0^{(k-1)}, t^{(k-1)}, \varepsilon)$, each of which will be transformed into some chain

$$x_{t^{(k)}-m}, x_{t^{(k)}-m+1}, \dots, x_{t^{(k)}+n^{(k)}-1} \quad (k = 0, 1, \dots; t^{(0)} = t),$$

and, transmitting this chain through the channel, we can, from the chain obtained at the output of the channel, with probability exceeding $1 - \varepsilon$, correctly determine the chain $\alpha_i(t_0^{(k)})$. This determination consists in the fact that, as $\alpha_i(t_0^{(k)})$, we choose the chain most probable for the given $n^{(k)}$ last elements of the chain obtained at the output of the channel.

Shannon' s second theorem. Under the conditions of the first theorem, the code can be chosen in such a way that the transmission rate is arbitrarily close to H_0 .

The author expresses his deep gratitude to Academician A. N. Kolmogorov for his assistance in carrying out this work.

Moscow State University
named after M. V. Lomonosov

Received
20 VII 1956

CITED LITERATURE

1. M. Rosenblatt-Roth, Proceedings of the 3rd All-Union Mathematical Congress, 2, Moscow, 1956, pp. 132-133.
2. M. Rosenblatt-Roth, DAN, 112, No. 1 (1957).
3. A. N. Kolmogorov, *Basic Concepts of Probability Theory*, 1936.
4. R. L. Dobrušin, DAN, 102, No. 1 (1955).

5. A. Ya. Khinchin, *Uspekhi Mat. Nauk*, 11, 1(67) (1956).
6. A. Feinstein, Inst. Radio Eng., Trans of Prof. Group on Information Theory, 4, IX, 1954.
7. C. E. Shannon, Bell. Syst. Techn. J., 27 (1948).

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

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