MATHEMATICS

M. ROSENBLATT-ROT

1960

SovietRxiv

Abstract

Full Text

MATHEMATICS

M. ROSENBLATT-ROT

NORMALIZED ε-ENTROPY OF SETS AND THE TRANSMISSION OF INFORMATION FROM CONTINUOUS SOURCES THROUGH CONTINUOUS COMMUNICATION CHANNELS

(Presented by Academician A. N. Kolmogorov, 10 VIII 1959)

1. Approximation of probability fields and transition probability functions. Let $(\mathfrak{A}, S, \mu)$ be a certain space with measure; $\rho_0(x, y)$ a metric in $\mathfrak{A}$; $D_0(\mathfrak{A})$ the collection of probability fields with elementary events $x \in \mathfrak{A}$; $A, B \in D_0(\mathfrak{A})$, AB their union; $\sigma_0^2(AB) = M_{AB}\rho_0^2(x, y)$ ($^{1-3}$).

For what follows it is important to note that in (^{5}) it is in fact proved that any metric separable space $\mathfrak{A}$ can be embedded in a centered space $\mathfrak{A}'$ (see (^{12})). A system θ_ε of sets $\mathcal{A}_i \in S$ forms an ε-covering of the space $\mathfrak{A}$ if: a) $\mathfrak{A} = \bigcup_i \mathcal{A}_i$; b) $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ $(i \neq j)$; c) $\sup_i d(\mathcal{A}_i) \leq 2\varepsilon$ ($d(\mathcal{A}_i)$ is the diameter of $\mathcal{A}_i$).

Let θ_ε^0 be a certain θ_ε in which all $\mathcal{A}_i$ have the greatest possible identical measure ω, and let $N_\varepsilon(\mathfrak{A})$ be the minimal number of elements in any θ_ε, i.e., the number of elements in θ_ε^0. In (^{4,5}) the quantity $*$
$\mathcal{H}'_\varepsilon(\mathfrak{A}) = \log N_\varepsilon(\mathfrak{A})$ was considered for totally bounded $\mathfrak{A}$; evidently $\omega = \mu(\mathfrak{A})/N_\varepsilon(\mathfrak{A})$, and ω remains finite also in the case of non-totally bounded $\mathfrak{A}$.

Definition 1. The quantity

$$\mathcal{H}_\varepsilon(\mathfrak{A}) = \log \frac{N_\varepsilon(\mathfrak{A})}{\mu(\mathfrak{A})} = \mathcal{H}'_\varepsilon(\mathfrak{A}) - \log \mu(\mathfrak{A})$$

will be called the **normalized (minimal) ε-entropy** of the space $\mathfrak{A}$. Let $(\mathfrak{A}, S)$, $(\mathfrak{B}, \Sigma)$ be measurable spaces; $x \in \mathfrak{A}$, $y \in \mathfrak{B}$, $\mathcal{A} \in S$, $\mathcal{B} \in \Sigma$; $\mathcal{P}_i(x, \mathcal{B}) \in R_0(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$ the collection of transition probability functions with domain of definition $(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$ $(i = 1, 2)$; $\alpha(\mathcal{P}_i)$ the coefficient of ergodicity of the transition probability function $\mathcal{P}_i(x, \mathcal{B})$ ($^{6-8}$); $\beta(\mathcal{P}_1, \mathcal{P}_2) = \sup |\mathcal{P}_2(x, \mathcal{B}) - \mathcal{P}_2(x, \mathcal{B})|$, where the least upper bound is taken over all $x \in \mathfrak{A}$, $\mathcal{B} \in \Sigma$.

Lemma 1. $|\alpha(\mathcal{P}_1) - \alpha(\mathcal{P}_2)| \leq 2\beta(\mathcal{P}_1, \mathcal{P}_2)$.

Lemma 2. $R_0(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$ is a complete metric space with metric $\beta(\mathcal{P}_1, \mathcal{P}_2)$. Under β-convergence the coefficient of ergodicity is continuous, i.e., from $\lim_{n\to\infty} \beta(\mathcal{P}, \mathcal{P}^n) = 0$ it follows that $\lim_{n\to\infty} \alpha(\mathcal{P}^n) = \alpha(\mathcal{P})$.

Let $M_{\mathfrak{A}}$ be the space of all countably additive finite functions defined on the σ-algebra of a measurable set $(\mathfrak{A}, S)$; let in it $\|\mu\|$ be equal to half the total variation of the generalized measure μ; $L_{\mathfrak{A}}$ the subspace of all $\lambda \in M_{\mathfrak{A}}$ for which $\lambda(\mathfrak{A}) = 0$. Let $\mathcal{P}_i$ be the operator corresponding to the transition probability function $\mathcal{P}_i(x, \mathcal{B})$, so that

In ([4,5]) it is denoted by $\mathcal{H}_\varepsilon(\mathfrak{A})$.

$\mathcal{P}_i\mu = \mu_i'$ for $\mu \in M_{\mathfrak{A}}$, $\mu_i' \in M_{\mathfrak{B}}$,

$$\mu_i'(\mathfrak{B}) = \int_{\mathfrak{A}} \mathcal{P}_i(x, \mathfrak{B})\mu(dx), \quad \mathfrak{B} \in \Sigma \quad (i = 1, 2),$$

and $\mathcal{N}(\mathcal{P}_1 - \mathcal{P}_2)$ is the norm of the operator $\mathcal{P}_1 - \mathcal{P}_2$ mapping $M_{\mathfrak{A}}$ into $L_{\mathfrak{B}}$. Let $G_{\mathfrak{A}}$ be the subspace of probability measures in $M_{\mathfrak{A}}$.

Lemma 3. $\mathcal{N}(\mathcal{P}_1 - \mathcal{P}_2) = 2\beta(\mathcal{P}_1 - \mathcal{P}_2)$.

Lemma 4. $\beta(\mu_1', \mu_2') \leqslant \beta(\mathcal{P}_1, \mathcal{P}_2)$ for $\mu \in G_{\mathfrak{A}}$, $\mu_i' = \mathcal{P}_i\mu \in G_{\mathfrak{B}}$ $(i = 1, 2)$.

Lemma 5*. $\beta(\mathcal{P}_1, \mathcal{P}_2) = 1 - \inf_{x\in\mathfrak{A}} \tilde{\alpha}[\mathcal{P}_1(x, \cdot), \mathcal{P}_2(x, \cdot)]$.

Let $(\mathfrak{A}, S)$ be a separable space with metric $\rho_0(x, x_1)$, and let $\mathcal{P}(x, \mathfrak{B}) \in R_0(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$. If $\mathcal{A}_i \in \theta_\varepsilon^0$ and x_i is the center of $\mathcal{A}_i$, set $\mathcal{P}_\varepsilon(x, \mathfrak{B}) = \mathcal{P}(x_i, \mathfrak{B})$ for $x \in \mathcal{A}_i$ $(i = 1, 2, \ldots)$. In what follows all probability densities are assumed to be uniformly continuous.

Theorem 1. Let $\delta > 0$ be an arbitrarily small number. If a probability field $A \in D_0(\mathfrak{A})$ and a uniformly continuous transition probability function $\mathcal{P}(x, \mathfrak{B}) \in R_0(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$ are given, then there exists a number $\varepsilon = \varepsilon(\delta)$ such that it is possible to define discrete fields $A_\varepsilon \in D_0(\mathfrak{A})$, $(A_\varepsilon \mid y) \in D_0(\mathfrak{A})$ $(y \in \mathfrak{B})$, and a discrete transition probability function $\mathcal{P}_\varepsilon(x, \mathfrak{B}) \in R_0(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$ so that, if $P_B = \mathcal{P} \cdot P_A$, $P_{B_\varepsilon} = \mathcal{P} \cdot P_{A_\varepsilon}$, then

$$\sigma_0(AA_\varepsilon) < \varepsilon, \quad \sigma_0[(A \mid y)(A_\varepsilon \mid y)] < \varepsilon \; (y \in \mathfrak{B}), \quad H(A_\varepsilon) = h(A) + \mathcal{H}_\varepsilon(\mathfrak{A}) + o(1),$$

$$I(A, A_\varepsilon) = H(A_\varepsilon) + o(1), \quad \beta(B, B_\varepsilon) \leqslant \beta(\mathcal{P}, \mathcal{P}_\varepsilon) < \delta,$$

$$I(A_\varepsilon, B_\varepsilon) = I(A, B) + o(1).$$

2. Approximation of stochastic processes and channels. Let $\alpha = [t, t + n - 1]$; let J be the set of all integers; let $\rho_\tau(x_\tau, y_\tau)$ and μ_τ be a metric and a measure in $\mathfrak{A}_\tau$ $(\tau \in J)$, $x^\alpha \in \mathfrak{A}^\alpha = \prod_{\tau \in \alpha} \mathfrak{A}_\tau$,

$$\rho_\alpha(x^\alpha, y^\alpha) = \max_{\tau \in \alpha} \rho_\tau(x_\tau, y_\tau), \quad \mu^\alpha = \prod_{\tau \in \alpha} \mu_\tau, \quad x \in \mathfrak{A} = \prod_{\tau \in J} \mathfrak{A}_\tau, \quad \rho(x, y) = \sup_{\tau \in J} \rho_\tau(x_\tau, y_\tau).$$

Let $D_0(\mathfrak{A}^\alpha)$ be the collection of fields A^α with state set $\mathfrak{A}^\alpha$, and let $D(\mathfrak{A})$ be the collection of processes A with state sets $\mathfrak{A}_\tau$ $(\tau \in J)$,

$$\sigma_\alpha^2(A^\alpha B^\alpha) = M_{A^\alpha B^\alpha} \rho_\alpha^2(x^\alpha, y^\alpha), \quad \sigma^2(AB) = M_{AB} \rho^2(x, y).$$

Definition 2. The normalized ε-entropy of the sequence of spaces $\mathfrak{A}_\tau$ $(\tau \in J)$ at time t is the quantity

$$\mathcal{H}_{t,\varepsilon}(\mathfrak{A}) = \lim_{n \to \infty} \frac{1}{n} \mathcal{H}_\varepsilon(\mathfrak{A}^\alpha)$$

(if this limit exists). If $\mathcal{H}_{t,\varepsilon}(\mathfrak{A})$ exists for all $t \in J$, is finite, and does not depend on t, the sequence $\mathfrak{A}_\tau$ is **regular**.

It is obvious that

$$\mathcal{H}_\varepsilon(\mathfrak{A}^\alpha) = \sum_{\tau \in \alpha} \mathcal{H}_\varepsilon(\mathfrak{A}_\tau);$$

let

$$I_t(A, B) = \lim_{n \to \infty} \frac{1}{n} I(A^\alpha, B^\alpha).$$

Let a stochastic nonanticipatory channel K with finite memory m be specified by means of spaces with measures $(\mathfrak{A}_\tau, S_\tau, \mu_\tau)$, $(\mathfrak{B}_\tau, \Sigma_\tau, \nu_\tau)$ $(\tau \in J)$, and transition probability functions

$$\mathcal{P}^\alpha(x^{\alpha'}, B^\alpha) \in R_0(\mathfrak{A}^{\alpha'}, S_{\alpha'}, \mathfrak{B}^\alpha, \Sigma^\alpha),$$

where $\alpha = [t, t + n - 1]$, $\alpha' = [t - m, t + n - 1]$, $x^{\alpha'} \in \mathfrak{A}^{\alpha'}$, $B^\alpha \in \Sigma^\alpha$. Let

$$(\mathfrak{A}, S) = \prod_{\tau \in J} (\mathfrak{A}_\tau, S_\tau), \quad (\mathfrak{B}, \Sigma) = \prod_{\tau \in J} (\mathfrak{B}_\tau, \Sigma_\tau).$$

The channel may be regarded as specified by the transition probability function $\mathcal{P}(x, \mathfrak{B}) \in R_0(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$, $x \in \mathfrak{A}$, $\mathfrak{B} \in \Sigma$, where for $\bar{\alpha} = J - \alpha$

$$\mathcal{P}(x, B^\alpha \times \mathfrak{B}^{\bar{\alpha}}) = \mathcal{P}(x^{\alpha'}, B^\alpha)$$

for all $B^\alpha \in \Sigma^\alpha$. Let $R(\mathfrak{A}, S; \mathfrak{B}, \Sigma)$ be the collection of all channels K with the same spaces with measures.

Lemma 6. $\beta(\mathcal{P}_1^\alpha, \mathcal{P}_2^\alpha) \leqslant \beta(\mathcal{P}_1^{\alpha_1}, \mathcal{P}_2^{\alpha_1})$ for $\alpha \subset \alpha_1$.

---

* For the definition of $\tilde{\alpha}(\mu_1, \mu_2)$, see (8).

Lemma 7. $R(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$ is a complete metric space with metric

$$\gamma(K_1, K_2) = \sup_{\alpha \subset J} \beta(\mathfrak{P}_1^\alpha, \mathfrak{P}_2^\alpha).$$

For any $u \subset J$, let $\mathcal{A}_i^u$ be the sets of some ε-covering θ_u^ε of the space $\mathfrak{A}_u$ and let x_i^u be the corresponding centers; then, for a given stochastic channel $K \in R(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$, the channel $K_\varepsilon \in R(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$ is defined by means of the transition probability function

$$\mathfrak{P}_\varepsilon^\alpha(x^\alpha, \mathfrak{B}^\alpha) = \mathfrak{P}^\alpha(x_i^\alpha, \mathfrak{B}^\alpha) \quad \text{for } x^{\alpha'} \in \mathcal{A}_i^{\alpha'} \ (i = 1, 2, \ldots).$$

Theorem 2. Suppose: 1) $\delta > 0$ is an arbitrarily small number; 2) the stochastic source $A \in D(\mathfrak{A})$ possesses finite differential entropy $h_t(A)$ and the property $\mathcal{E}_t(A)$; 3) the stochastic channel $K \in R(\mathfrak{A}, S, \mathfrak{B}, \Sigma)$ is characterized by uniformly continuous transition probability functions $\mathfrak{P}^\alpha(x^\alpha, \mathfrak{B}^\alpha)$ (uniformly also with respect to $\alpha \subset J$, $\mathfrak{B}^\alpha \in \Sigma^\alpha$), so that $h_t(A \mid B)$ exists and is finite and the property $\mathcal{E}_t(A \mid B)$ holds.

Then it is possible to choose $\varepsilon = \varepsilon(\delta)$ so that: 1) there exists a discrete stochastic source $A_\varepsilon \in D(\mathfrak{A})$ with states independent of the process A, such that $\sigma(AA_\varepsilon) < \varepsilon$, possessing finite entropy

$$H_t(A_\varepsilon) = h_t(A) + \mathcal{H}_{t,\varepsilon}(\mathfrak{A}) + o(1)$$

and the property $\mathcal{E}_t(A_\varepsilon)$, with

$$I_t(A, A_\varepsilon) = H_t(A_\varepsilon) + o(1);$$

2) there exists a discrete stochastic channel K_ε with input states independent of K, such that $\gamma(K, K_\varepsilon) < \delta$, possessing finite entropy

$$H_t(A_\varepsilon \mid B_\varepsilon) = h_t(A \mid B) + \mathcal{H}_{t,\varepsilon}(\mathfrak{A}) + o(1)$$

and the property $\mathcal{E}_t(A_\varepsilon \mid B_\varepsilon)$, when fed by the source A_ε, with

$$I_t(A, B) = I_t(A_\varepsilon, B_\varepsilon) + o(1);$$

3) under the regularity conditions of the sequence $\mathfrak{A}_\tau$ ($\tau \in I$), if A, K are regular, then $A_\varepsilon, K_\varepsilon$ are also regular; if A, K are stationary, then $A_\varepsilon, K_\varepsilon$ are also stationary; 4)

$$C = C_\varepsilon + o(1),$$

where C, C_ε are the regular capacities of the channels K, K_ε.*

In what follows we shall assume that the sequence $\mathfrak{A}_\tau$ is regular.

3. Shannon' s fundamental theorems

Theorem 3. Suppose there are given: 1) $\delta > 0$, $\lambda > 0$, arbitrarily small numbers; 2) a regular channel K with continuous sets of input states, with uniformly continuous transition probability functions, with finite memory and with finite regular capacity C; 3) a regular source $\mathring{A}$ with continuous sets of states and with finite differential entropy

$$h(\mathring{A}) < C.$$

Then, if one takes $\varepsilon = \varepsilon(\delta)$, $\mathring{A}_\varepsilon, K_\varepsilon$ as in Theorem 2, one has

$$\sigma(\mathring{A}\mathring{A}_\varepsilon) < \varepsilon, \qquad \gamma(K, K_\varepsilon) < \delta;$$

if

$$H(\mathring{A}_\varepsilon) = h(\mathring{A}) + \mathscr{H}_\varepsilon(\mathfrak{A}) < C + o(1),$$

then Shannon' s first fundamental theorem on the possibility of transmitting the output of the source $\mathring{A}_\varepsilon$ through the channel K_ε with error probability less than λ is valid ([11]).

In the case where $\mathfrak{A}_\tau$ are totally bounded spaces, the source $\mathring{A}_\varepsilon$ has finite sets of elements; let n_τ be their number ($\tau \in J$). Let

$$\overline{\mathscr{H}}'_\varepsilon(\mathfrak{A}) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \log n_{t+k}.$$

Theorem 4. Under the hypotheses of Theorem 3, if the state sets of the source $\mathring{A}$ are totally bounded and

$$\overline{\mathscr{H}}'_\varepsilon(\mathfrak{A}) < \infty,$$

then Shannon' s second fundamental theorem is valid concerning the possibility of choosing a code so that the rate–

* The definitions and notation from ([10,11]) are used.

the rate of transmission of the output of the source A^0_ε through K_ε were arbitrarily close to

$$H\left(A^0_\varepsilon\right) = h\left(A^0\right) + \mathscr{H}_\varepsilon(\mathfrak{A}) + o(1) \tag{11}$$

The author expresses his deep gratitude to Academician A. N. Kolmogorov for posing the problem and for valuable consultation.

Parkhon University
Faculty of Mathematics and Physics

Mathematical Institute
of the Romanian Academy of Sciences
Bucharest, Romania

Received
27 VII 1959

REFERENCES CITED

1. A. N. Kolmogorov, Theory of transmission of information, Session of the Academy of Sciences of the USSR on scientific problems of automation of production, 15-20 X 1956, Plenary sessions, Publ. House of the Academy of Sciences of the USSR, 1957, p. 66.
2. I. M. Gel' fand, A. N. Kolmogorov, A. M. Yaglom, Proceedings of the 3rd All-Union Mathematical Congress, 3, Moscow, 1956, p. 300.
3. A. N. Kolmogorov, Inst. Radio Eng., Trans. on Inf. Theory, v. IT-2, No. 4, 102 (1956).
4. A. N. Kolmogorov, DAN, 108, No. 3, 385 (1956).
5. A. N. Kolmogorov, V. M. Tikhomirov, Uspekhi Mat. Nauk, 14, issue 2 (86), 3 (1959).
6. E. B. Dynkin, Ukr. Mat. Zhurn., 6, No. 1, 21 (1954).
7. R. L. Dobrushin, DAN, 102, No. 5 (1955).
8. R. L. Dobrushin, Theory of Probability and Its Applications, 1, issue 1, 72 (1956); 2, issue 4, 365 (1956).
9. M. Rosenblatt-Roth, Proceedings of the 3rd All-Union Mathematical Congress, 2, Moscow, 1956, p. 132.
10. M. Rosenblatt-Roth, DAN, 112, No. 1, 16 (1957).
11. M. Rosenblatt-Roth, DAN, 112, No. 2, 202 (1957).
12. A. G. Vitushkin, DAN, 117, No. 5, 745 (1957).

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