MATHEMATICS

M. ROSENBLATT-ROT

1957

SovietRxiv

Abstract

Full Text

MATHEMATICS

M. ROSENBLATT-ROT

ENTROPY OF STOCHASTIC PROCESSES

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

1. Entropy of fields. Let there be a certain space with measure $(\mathfrak{A}, \mathcal{L}, \mu)$; with the aid of the density $p(x)$ $(x \in \mathfrak{A})$ with respect to μ, a field of probabilities A is defined. By means of a simple and reasonable axiomatics one can show that the measure of uncertainty created by the field A is defined uniquely as

$$H(A) = -\int_{\mathfrak{A}} p(x) \log p(x) \mu(dx)$$

(we shall assume that this integral exists). This entropy has all the elementary properties of the entropy of a finite field [2,3]. In the notation of [2,3] the following theorem holds.

Theorem 1. *If $H(A)$ and $H(AB)$ exist and are finite, then $H_A(B)$ also exists and is finite, and moreover $H(AB) = H(A) + H_A(B)$.*

If $P(\mathfrak{A}) < 1$, then we have an incomplete field; its entropy is defined by means of the same integral. This entropy possesses a number of elementary properties very close to the properties of the entropy of complete fields. If

$$H_A(B) = \int_{\mathfrak{A}} H(B \mid x) p(x) \mu(dx),$$

then Theorem 1 remains valid. Below we shall consider only complete fields.

2. Entropy of processes. Let spaces with measures $(\mathfrak{A}_t, \mathcal{L}_t, \mu_t)$ be given $(t \in I,\ I = \{t\}$ is the set of all integers). Let there be a stochastic process A with states $x_t \in \mathfrak{A}_t$. Suppose that the process is specified by the densities of the chains $x^{[t,t+n-1]} = (x_t, \dots, x_{t+n-1})$ $(t \in I,\ n = 1, 2, \dots)$ with respect to the measure $\mu^{[t,t+n-1]} = \mu_t \times \dots \times \mu_{t+n-1}$; let these be $\pi^{[t,t+n-1]}(x^{[t,t+n-1]})$, and denote by $A^{[t,t+n-1]}$ the field of these chains. Let $x = (\dots, x_t, \dots, x_{t+n-1}, \dots)$ and

$$f^{[t,t+n-1]}(x) = -n^{-1} \log \pi^{[t,t+n-1]}(x^{[t,t+n-1]}).$$

Definition. The **entropy of the process A at the moment t** is the quantity

$$H_t(A) = \lim_{n\to\infty} Mf^{[t,t+n-1]}(x) = \lim_{n\to\infty} n^{-1} H(A^{[t,t+n-1]})*,$$

(if this limit exists).

Theorem 2. *For the existence of $H_t(A)$ it is necessary and sufficient that the sequence of conditional entropies $H(A_{t+n} \mid A^{[t,t+n-1]})$ be Cesàro $C(1)$-summable, and $H_t(A)$ is the limit of these sums (convergence is understood in the sense of tending to a finite number or to $\pm\infty$). There always exists the (finite or infinite) limit*

$$\widetilde{H}_t^{(m)}(A) = \lim_{n\to\infty} H(A^{[t,t+m-1]} \mid A^{[t+m,t+n]}) \qquad (t \in I,\ m \geq 1).$$

* M is mathematical expectation, D is variance.

Let

$$\lambda_t^{(m)}(A) = \lim_{n\to\infty} n^{-1} H\left(A^{[t,t+m-1]} \mid A^{[t+m,t+n]}\right)$$

(if this limit exists).

Theorem 3. If one of the quantities $H_t(A)$, $H_{t+m}(A)$ exists and is finite, then, in order that the second quantity also exist, be finite, and $H_t(A) = H_{t+m}(A)$, it is necessary and sufficient that $\lambda_t^{(m)}(A) = 0$. For this it is sufficient that $\left|\widetilde{H}_t^{(m)}(A)\right| < \infty$*.

In what follows we shall exclude those processes for which $\widetilde{H}_t^{(m)}(A) = +\infty$ for at least one t and one m**.

Theorem 4. For processes with discrete sets of states***, the entropy, if it exists, does not depend on time, i.e. $H_t(A) = H(A) = \text{const}$ $(t \in I)$.

3. Properties $\mathcal{E}_t(A)$, $\mathcal{E}(A)$.

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

Let

$$g^{[t,t+n]}(x) = -\log\left\{\pi^{[t,t+n]}(x^{[t,t+n]})/\pi^{[t,t+n-1]}(x^{[t,t+n-1]})\right\} \quad (t \in I,\ n \geqslant 1).$$

Theorem 5. In order that the process A have the property $\mathcal{E}_t(A)$, it is necessary and sufficient that the sequence of random variables $g^{[t,t+n]}(x)$ $(n = 1, 2, ...)$ obey the law of large numbers. For this it is sufficient that

$$\lim_{n\to\infty} Df^{[t,t+n-1]}(x) = 0.$$

Suppose there is a Markov chain A with ergodicity coefficients[4] $\alpha_{i,i+1}$.

Theorem 6. Sufficient conditions for the Markov process A to have the property $\mathcal{E}_t(A)$ are:

a)

$$\lim_{n\to\infty} n^{\beta-2} \sum_{k=0}^{n-1} Dg^{[t,t+k]}(x) = 0, \quad \text{if } \alpha_{i,i+1} > 0 \ (1 \leqslant i < \infty),$$

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

b)

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Dg^{[t,t+k]}(x) = 0 \quad \text{in all cases.}$$

Let

$$\varphi_{t,m,n}(x) = -n^{-1} \log \left\{ \pi^{[t,t+n-1]}(x^{[t,t+n-1]}) / \pi^{[t+m,t+n-1]}(x^{[t+m,t+n-1]}) \right\}.$$

Theorem 7. If the process A has one of the two properties $\mathcal{E}_t(A)$, $\mathcal{E}_{t+m}(A)$ $(m \geqslant 1)$, then, in order that it also have the other property, it is necessary and sufficient that the sequence $\varphi_{t,m,n}(x)$ converge in probability to $\lambda_t^{(m)}(A)$ as $n \to \infty$.

Theorem 8. In order that the process A have the property $\mathcal{E}(A)$, it is necessary and sufficient that it have the property $\mathcal{E}_{t_0}(A)$ for some t_0 and that the sequence $\varphi_{t,m,n}(x)$ converge in probability to $\lambda_t^{(m)}(A)$ for all $t \in I$, $m > 0$ $(n \to \infty)$.

Let $L(P)$ be the space of all real functions $f(x)$ of the variable $x \in \mathfrak{X}$ such that $M|f(x)| < \infty$.

---

* This condition is fulfilled for all $t \in I$, $m \geqslant 0$, if the sets $\mathfrak{X}_\tau$ $(\tau \in I)$ are finite and $\mu_\tau(x_\tau) = 1$, $x_\tau \in \mathfrak{X}_\tau$. Consequently, in this case either $H_t(A) \equiv H(A)$ $(t \in I)$, or $H_t(A)$ does not exist for any $t \in I$.

** Under these conditions, if $H_t(A)$, $H_{t+m}(A)$ exist and are finite, then $H_t(A) \leqslant H_{t+m}(A)$ $(m = 1, 2, \ldots)$.

*** $\mu_\tau(x_\tau) = 1,\ x_\tau \in \mathfrak{X}_\tau,\ \tau \in I.$

Theorem 9. The sequence of functions $f^{[t,t+n-1]}(x)$ $(n = 1, 2, ...)$ cannot converge in the mean (in $L(P)$) to any constant except $H_t(A)$. If the process A does not have finite entropy, then the sequence of functions $f^{[t,t+n-1]}(x)$ cannot converge in the mean to any function from $L(P)$.

Example. Let there be a Markov chain with two states, such that if $p_{ij}^{(k)}$ is the probability of transition during the time $(k-1, k)$ from state i to state j, then $p_{11}^{(k)} = p_{22}^{(k)} = 1 - \alpha_k$; $p_{12}^{(k)} = p_{21}^{(k)} = \alpha_k$, with $\lim_{k\to\infty} \alpha_k = 2\alpha$ $(0 < \alpha < 1)$. From Theorems 2 and 3 one can obtain that $H_t(A) = H(A) = -\alpha \log \alpha - (1 - \alpha)\log(1-\alpha)$. From Theorems 6-8, bearing in mind that $\beta = 0$, it follows that $\mathcal{E}_t(A)$ and even $\mathcal{E}(A)$ exist.

4. Estimate of the volume (number) of standard chains; application to coding theory. Let λ $(0 < \lambda < 1)$ be some given constant number, and let $N^{[t,t+n-1]}(\lambda)$ be some part of

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

such that: 1) $P[N^{[t,t+n-1]}(\lambda)] \geq \lambda$; 2) $\mu^{[t,t+n-1]}[N^{[t,t+n-1]}(\lambda)]$ has the smallest value subject to the first condition. The existence of $N^{[t,t+n-1]}(\lambda)$ is easy to prove.

Theorem 10. If the process A has the property $\mathcal{E}_t(A)$, then there exists a limit independent of λ $(0 < \lambda < 1)$

$$\lim_{n\to\infty} n^{-1} \log \mu^{[t,t+n-1]} \left[N^{[t,t+n-1]}(\lambda)\right] = H_t(A).$$

Suppose that there is a certain text, which is a sequence of symbols (letters) belonging to some group (alphabet) with r elements. We shall regard this text as a certain (nonstationary or stationary) stochastic process with finite entropy $H_t(A)$ and with the property $\mathcal{E}_t(A)$. Let the question be posed of coding the given text in the same alphabet, so that decoding is possible. Each n-term chain $x^{[t,t+n-1]}$ of the given text has a certain probability; let $\sigma^{[t,t+n-1]}(x^{[t,t+n-1]})$ be the length of the chain of the coded text into which the chain $x^{[t,t+n-1]}$ passes after coding. Let

$$\rho^{(t)} = \limsup_{n\to\infty} n^{-1} M\sigma^{[t,t+n-1]}(x^{[t,t+n-1]})$$

be the compression coefficient of the given text at the moment t by the coding.

Theorem 11. If the incoming text has the statistical structure of a (nonstationary or stationary) process A with r states, possessing the property $\mathcal{E}_t(A)$, then the lower bound of the compression coefficient $\rho^{(t)}$ of the given text over all codes is equal to $H(A)/\log r$ for all t.

5. Stationary processes.

Theorem 12. For stationary processes the entropy $H(A)$ (finite or infinite) always exists.

Let T be the shift operator, i.e. if $x \in \mathfrak{A} = ... \times \mathfrak{A}_{-1} \times \mathfrak{A}_0 \times \mathfrak{A}_1 \times ...$, then $x' = Tx \in \mathfrak{A}$, and $x'_\tau = x_{\tau+1}$ $(\tau \in I)$. Let also $x^{[0,n-1]} = x^{(n)}$; $\pi^{[0,n-1]}(x^{[0,n-1]}) = \pi^{(n)}(x^{(n)})$; $g^{[-n,0]}(x) = g_n(x)$; $f^{[0,n-1]}(x) = f_n(x)$; $\mu^{[0,n-1]} = \mu^{(n)}$; $N^{[0,n-1]}(\lambda) = N^{(n)}(\lambda)$.

Theorem 13. In order that the stationary process A possess the property $\mathcal{E}(A)$, it is necessary and sufficient that the sequence of random variables $g^{[0,n]}(x) = g_n(T^n x)$ $(n = 0, 1, 2, ...)$ obey the law of large numbers.

Consequently, ergodicity is not necessary for $\mathcal{E}(A)$.

Theorem 14. Let there be a stationary process with an arbitrary set of states, such that: a) $g_n(x) \in L(P)$; b) there exists some function $g(x) \in L(P)$ such that the sequence $g_n(x)$ as $n \to \infty$ converges in the mean (in $L(P)$) to $g(x)$. Then the sequence

$f_n(x)$ converges in mean to some invariant function $h(x)$. In the case of ergodicity, property $\mathcal{E}(A)^*$ holds.

Theorem 15. Let there be a stationary, ergodic process for which conditions a), b) of Theorem 14 are satisfied. Then the sequence of random variables $g_n(T^n x)$ $(n = 0, 1, 2, ...)$ obeys the law of large numbers.

Let A be a stationary simple Markov chain with a set of states $\mathfrak{A}$ (where a σ-algebra $\mathcal{S}$ and a measure μ are given on $\mathfrak{A}$), stationary density $p(x)$ (with respect to μ), and density (with respect to μ) of transition probabilities $q(x, y)$.

Theorem 16. If A is a stationary, simple, uniformly ergodic[6] Markov chain and, for some $\delta > 0$, $M|\log q(x, y)|^{2+\delta} < \infty$, and the entropy is finite, then the distribution of the random variable $n^{-1/2}[\log \pi^{(n)}(x^{(n)}) + nH(A)]$ converges to the normal distribution with parameters $(0, \sigma^2)$, where σ^2 is determined by the probabilities (unconditional and transition) of the chain A.

Let u_λ be determined from

$$\lambda = (2\pi)^{-1/2} \int_{-\infty}^{u_\lambda} \exp\left(-\frac{1}{2}x^2\right) dx.$$

Theorem 17. Under the conditions of Theorem 16,

$$\log \mu^{(n)}[N^{(n)}(\lambda)] = nH(A) + \sqrt{n}\,\sigma u_\lambda + o(\sqrt{n})^{**}.$$

The author expresses deep gratitude to 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, Moscow, 1956, **2**, p. 132.
[2] C. E. Shannon, Bell. Syst. Techn. J., **27**, 379, 623 (1948).
[3] A. Ya. Khinchin, Uspekhi Mat. Nauk, **8**, 3, 55 (1953).
[4] R. L. Dobrushin, DAN, **102**, No. 1 (1955).
[5] B. McMillan, Ann. Math. Stat., **24**, 2 (1953).
[6] E. B. Dynkin, Ukr. Mat. Zhurn., **6**, 1 (1954).
[7] A. A. Yushkevich, Uspekhi Mat. Nauk, **8**, 5, 57 (1953).

* If the set of states is finite, then in [(5)] it is proved that $H(A)$ exists, is finite, and that conditions a), b) are satisfied.

** Under the assumption that the set of states is finite, Theorems 16 and 17 are proved in [(7)].

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

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