APPROXIMATIONS IN INFORMATION THEORY

MILLU ROSENBLATT-ROTH
BUCHAREST UNIVERSITY

1. Introduction

In the papers [8]–[11], [14] the author studied stochastic processes and channels, stationary, or nonstationary, with discrete time and arbitrary sets of states. In these papers, for regular processes and channels, two basic theorems of Shannon type [15] are proved for the case in which the states of the process and the channel input states are discrete and the output states arbitrary.

In this study, the essential role of the differential entropy of probability fields, processes, and channels appears. Obviously, if the sets of states are discrete, instead of the differential entropy, the correspondent entropy appears. Here we study the problem of approximation of processes with continuous sets of states by discrete processes and also of channels with continuous input-sets by channels with discrete input-sets.

In this study an essential role is played by the concept of ϵ-entropy of a set, of a probability field, of a channel, and of a complex source-channel. We may observe also the role played by differential entropy in the approximation problem. The constructions used here in the approximation problems are such that the essential properties of the given object are preserved.

2. The differential entropy of probability fields

Let us consider the measure space (X, S, μ) where X is a set of elements x and S a σ-algebra of subsets of X and μ a measure in S. Over X let us consider a probability field A, defined by the probability density $p(x)$ with respect to μ. By M we denote the expectation.

DEFINITION 2.1. *The value* $h(A) = -M \log p(x)$ *is the differential entropy of* A *with respect to* μ.

Obviously, $h(A)$ exists only if $M |\log p(x)| < +\infty$, and from $|h(A)| < M |\log p(x)|$ it follows that in this case it is finite.

Let (X, S, μ), (Y, Σ, ν) be measure spaces, $\pi(x, y)$ the probability density of some field C over their product, A the field defined by the probability density $p(x)$ induced by $\pi(x, y)$ in X, and $q_x(y) = \pi(x, y)/p(x)$ the conditional probability density of some probability field B_x over Y. We denote $C = AB$ (the union).

DEFINITION 2.2. *The conditional differential entropy of B with respect to A, for a given measure ν is* $h_A(B) = h(B|A) = Mh(B_x)$.

THEOREM 2.1. *If from* $h(A)$, $h_A(B)$, $h(AB)$, *two exist, then the third of them exists also and* $h(AB) = h(A) + h_A(B)$.

The proof is analogous with that of the corresponding theorem of the entropy case, using Fubini's theorem.

Different properties of the differential entropy of a field may be found in [8]–[11], [14]. In another publication by the same author will be given an axiomatic approach to the differential entropy of probability fields.

3. The approximation of probability fields

3.1. *The ϵ-entropy of a set.* Let us consider the measure space (X, S, μ), separable with respect to the distance $\rho(x, y)$, $x \in X$, $Z \in S$.

DEFINITION 3.1.1. *The sequence θ of measurable sets $Z_i \in S$, $(1 \leq i \leq n)$ is a* cover *of X if* (a) *these sets are nonoverlapping, and* (b) *X is their sum.*

DEFINITION 3.1.2. *The sequence θ_ϵ of measurable sets $Z_i^\epsilon \in S$, $(1 \leq i \leq n)$ is an* ϵ-cover *of X if it is a cover of X, and if $d(Z_i^\epsilon) \leq 2\epsilon$, $(1 \leq i \leq n)$, (d = the diameter).*

DEFINITION 3.1.3. *The space X is* centering *if in it, for every set $Z \subset X$ with $d(Z) = 2r$, there exists an element x, the center of Z, for which $\rho(x, y) \leq r$ for any $y \in Z$.* (See [6], p. 8.)

We may proceed as if every separable metric space were centering. Indeed, in [6] by means of the known theorem of Mazur-Banach ([1], chapter XI, section 8, theorem 10) and of theorem VI from ([6], section 1), it is proved that *every separable metric space X may be imbedded in a centering space X_** [16]. For totally bounded spaces let us denote by $N_\epsilon(X)$ the minimal number of elements in any ϵ-cover θ_ϵ.

DEFINITION 3.1.4. *The number $K'_\epsilon(X) = \log N_\epsilon(X)$ is the (minimal) ϵ-entropy of the set X.* (See [6], [16].)

DEFINITION 3.1.5. *The number* $K_\epsilon(X) = \log [N_\epsilon(X)/\mu(X)] = K'_\epsilon(X) - \log \mu(X)$ *is the normed (minimal) ϵ-entropy of the set X.*

3.2. *The discrete ϵ-entropy of a probability field.*

We shall use the following symbols:

(i) $D(X)$ will denote the totality of probability fields over (X, S); A and A' will be elements of $D(X)$;

(ii) if x and x' are elements of X, the probability density of A with respect to μ will be denoted by $p(x)$, and the corresponding conditional probability will be written as $p(x|x')$;

(iii) $D^0(X)$ will denote the totality of discrete fields with states $x_i \in X$, and $I(A, A') = h(A) - h(A|A')$.

DEFINITION 3.2.1. *If $Z_{x_i}^\epsilon$ is the sphere in X with center x_i and radius ϵ, let $W_\epsilon(AA')$ denote the property that for every state x_i of the discrete field $A' \in D^0(X)$ the condition*

(3.2.1) $$P_{AA'}\{Z^{\epsilon}_{x_i}|x_i\} = 1$$

is satisfied.

If the property $W_\epsilon(AA')$ is satisfied, the conditional field A_{x_i} possesses a set of states $Z_i \subset Z^{\epsilon}_{x_i}$; obviously, we may consider $Z_i \in \theta_\epsilon$, where θ_ϵ is any ϵ-cover of X.

Definition 3.2.2. *The quantity $H_\epsilon(A) = \inf I(A, A')$, where the lower bound is considered for all pairs AA' for which the property $W_\epsilon(AA')$ is satisfied for the given A, is the discrete ϵ-entropy of the field A.*

Theorem 3.2.1. *The discrete ϵ-entropy $H_\epsilon(A)$ is equal to $H(A) + K_\epsilon(X)$.*

Proof. (a) From the definition of $H_\epsilon(A)$ it follows that $H_\epsilon(A) = h(A) - \sup h(A|A')$ where the upper bound is taken over all pairs AA' for which the property $W_\epsilon(AA')$ is satisfied for the given A; consequently, we must prove only that $\sup h(A|A') = -K_\epsilon(X)$. We shall prove that for any given probability field A' we may construct another probability field A^0 so that $h(A|A') \leq h(A|A^0) = -K_\epsilon(X)$.

(b) If θ_ϵ is any ϵ-cover of X, $Z_i \in \theta_\epsilon$ $(1 \leq i \leq n)$, and x_i the center of $Z_i (1 \leq i \leq n)$, let us consider the probability field A' with elementary events x_i and any arbitrarily determined probabilities $P(x_i)$, $(1 \leq i \leq n)$. Obviously, if $p(x|x_i)$ does not vanish only for $x \in Z_i$, the condition $W_\epsilon(AA')$ is satisfied. In this case $h(A|x_i) \leq \log \mu(Z_i)$; $h(A|A') \leq \sum_{i=1}^{n} P(x_i) \log \mu(Z_i)$.

(c) In the same conditions as above, if we consider $p(x|x_i) = 1/\mu(Z_i)$ for $x \in Z_i$ and zero in the rest, we define the field A'' so that $h(A|x_i) = \log \mu(Z_i)$;

(3.2.2) $$h(A|A'') = \sum_{i=1}^{n} P(x_i) \log \mu(Z_i).$$

(d) If θ'_ϵ is any ϵ-cover of X, $Z'_i \in \theta'_\epsilon$, $(1 \leq i \leq n')$, with all elements Z'_i of the same μ-measure defined by

(3.2.3) $$u = \sum_{i=1}^{n} P(x_i)\mu(Z_i)$$

and n' given by the entire part of $\mu(X)/u$, let us define the probability field A''' with elementary events x'_i, $(1 \leq i \leq n')$ the centers of Z'_i, and any arbitrarily determined probabilities $P(x'_i)$, $(1 \leq i \leq n')$. If we consider that $p(x|x'_i) = 1/u$ for $x \in Z'_i$ and zero in the rest, then obviously the condition $W_\epsilon(AA''')$ is satisfied and

(3.2.4) $$h(A|x'_i) = \log u; \; h(A|A''') = \log u.$$

From the convexity of the function $\log x$ we obtain the inequality

(3.2.5) $$\sum_{i=1}^{n} P(x_i) \log \mu(Z_i) \leq \log u$$

so that $h(A|A') \leq h(A|A'') \leq h(A|A''') \leq \log u \leq \log(\mu(X)/n')$.

(e) Let us consider any ϵ-cover θ^0_ϵ of X, $Z^0_i \in \theta^0_\epsilon$, $(1 \leq i \leq n_0 = N_\epsilon(X))$, with $\mu(Z^0_i) = \mu(X)/n_0$, $(1 \leq i \leq n_0)$, x^0_i the centers of Z^0_i, $(1 \leq i \leq n_0)$, and $P(x^0_i)$ any arbitrarily given probabilities. We define the probability field A^0 by $p(x|x_i^0) = 1/\mu(Z_i^0)$ for $x \in Z_i^0$ and zero in the rest. Obviously,

$$(3.2.6) \qquad h(A|A^0) = \sum_{i=1}^{n} P(x_i^0) \log \mu(Z_i^0) = \log \frac{\mu(X)}{n_0} = -K_\epsilon(X).$$

Because $n_0 < n'$, it follows that $h(A|A''') \leq h(A|A^0) = -K_\epsilon(X)$ so that

$$(3.2.7) \qquad h(A|A') \leq h(A|A'') \leq h(A|A''') \leq h(A|A^0) = -K_\epsilon(X).$$

Consequently, the upper bound of $h(A|A')$ for all A' which satisfies the condition $W_\epsilon(AA')$ is equal to the upper bound of $h(A|A^0)$, that is to $-K_\epsilon(X)$, and our theorem is proved.

THEOREM 3.2.2. *Taking the upper bound for $A \in D(X)$ one has* $\sup H_\epsilon(A) = K'_\epsilon(X)$.

PROOF. Analogously, as for the entropy, it is easy to see that the upper bound of $h(A)$ for $A \in D(X)$ is $\log \mu(X)$; from theorem 3.2.1 it follows that

$$(3.2.8) \qquad \sup H_\epsilon(A) = \log \mu(X) + K_\epsilon(X) = K'_\epsilon(X).$$

Let us suppose that $p(x)$ is a uniformly continuous function.

THEOREM 3.2.3. *For any $\epsilon > 0$, for a given probability field $A \in D(X)$ which possesses finite differential entropy $h(A)$ there exists a discrete probability field $A_\epsilon \in D^0(X)$ with states not depending on A, such that* (a) *the property $W_\epsilon(AA_\epsilon)$ is satisfied,* (b) $H(A_\epsilon) = H_\epsilon(A) + o(1)$, *and* (c) $I(A, A_\epsilon) = H(A_\epsilon) + o(1)$.

PROOF. Let us consider any ϵ-cover θ_ϵ^0 of X, $Z_i^0 \in \theta_\epsilon^0$, $(1 \leq i \leq n_0 = N_\epsilon(X))$, with $\mu(Z_i^0) = \mu(X)/n_0$. We define the field $A_\epsilon \in D^0(X)$ with the elementary events x_i^0 (the centers of Z_i^0) and $P_{A_\epsilon}(x_i^0) = P_A(Z_i^0) = p_i\mu(Z_i^0)$, where

$$(3.2.9) \qquad p_i \in [\inf p(x), \sup p(x)],$$

and the lower and upper bounds are considered for $x \in Z_i^0$. We define the conditional probability field $(A|x_i^0)$ by means of $p_{A|A_\epsilon}(x|x_i^0) = 1/\mu(Z_i^0)$ for $x \in Z_i^0$ and zero in the rest. Obviously, $W_\epsilon(AA_\epsilon)$ is satisfied and $h(A|x_i^0) = h(A|A_\epsilon) = -K_\epsilon(X)$,

$$(3.2.10) \qquad \begin{aligned} H(A_\epsilon) &= -\sum_{i=1}^{n_0} p_i\mu(Z_i^0) \log [p_i\mu(Z_i^0)] \\ &= -\sum_{i=1}^{n} (p_i \log p_i)\mu(Z_i^0) + K_\epsilon(X) \\ &= h(A) + K_\epsilon(X) + o(1). \end{aligned}$$

Using theorem 3.2.1, (b) follows and

$$(3.2.11) \qquad \begin{aligned} I(AA_\epsilon) = h(A) - h(A|A_\epsilon) = h(A) + K_\epsilon(X) &= H_\epsilon(A) \\ &= H(A_\epsilon) + o(1). \end{aligned}$$

4. The differential entropy of stochastic processes

4.1. *Generalities.* Let us denote: (i) I = the set of all entire numbers; (ii) I^+ = the set of all natural numbers; (iii) $(X_\tau, S_\tau, \mu_\tau)$ = a measure space $(\tau \in I)$,

$x_\tau \in X_\tau$, $Z_\tau \in S_\tau$; (iv) $\alpha \subset I$ = a finite set of $|\alpha|$ numbers $\tau_i \in I$; (v) $(X^\alpha, S^\alpha, \mu^\alpha) = \times_{\tau\in\alpha}(X_\tau, S_\tau, \mu_\tau)$; (vi) $x^\alpha = (x_{\tau_1}, x_{\tau_2}, \cdots, x_{\tau|\alpha|}) = \{x_\tau, \tau \in \alpha\}$, $Z^\alpha \in S^\alpha$; (vii) $\alpha^* = I - \alpha$, $(X^{\alpha^*}, S^{\alpha^*}) = \times_{\tau\in\alpha^*}(X_\tau, S_\tau)$; (viii) $(X, S) = \times_{\tau\in I}(X_\tau, S_\tau)$, $x \in X$, $Z \in S$.

Let us consider that in the spaces X^α there exists a stochastic process A, that is, a consistent system of probability measures $P^\alpha(Z^\alpha)$, and let us denote by $P(Z)$ the extension of the measures $P^\alpha(Z^\alpha)$ in X.

We shall suppose that the measures P^α are μ^α-absolutely continuous, and let us denote by $\pi^\alpha(x^\alpha)$ the probability density; we also denote by $\pi^{\{\beta|\alpha\}}(x^\beta|x^\alpha)$ the probability density of $P^{\{\beta|\alpha\}}(Z^\beta|x^\alpha)$.

If $|\alpha|\cdot f^\alpha(x) = -\log \pi^\alpha(x^\alpha)$, it follows that $|\alpha|\cdot Mf^\alpha(x) = h(A^\alpha)$ with $A^\alpha = [\pi^\alpha(x^\alpha), X^\alpha, \mu^\alpha]$. If $\alpha_n = [t, t+n-1]$, let us denote

$$
\begin{aligned}
p^{\alpha_{n+1}}(x) &= \pi^{\{t+n|\alpha_n\}}(x_{t+n}|x^{\alpha_n}) = \pi^{\alpha_{n+1}}(x^{\alpha_{n+1}})/\pi^{\alpha_n}(x^{\alpha_n});\\
g^{\alpha_n}(x) &= -\log p^{\alpha_n}(x); \ \beta_m = [t+m, t+n-1];\\
\pi^{\{\alpha_m|\beta_m\}}(x^{\alpha_m}|x^{\beta_m}) &= \frac{\pi^{\alpha_n}(x^{\alpha_n})}{\pi^{\beta_m}(x^{\beta_m})};\\
|\beta_m|\cdot\varphi_{t,n}^{(m)}(x) &= -\log \pi^{\{\alpha_m|\beta_m\}}(x^{\alpha_m}|x^{\beta_m}).
\end{aligned} \tag{4.1.1}
$$

It follows that

$$|\beta_m|\cdot M\varphi_{t,n}^{(m)}(x) = h(A^{\alpha_m}|A^{\beta_m}). \tag{4.1.2}$$

Let us also denote

$$\lambda_t^{(m)}(A) = \lim_{n\to\infty} M\varphi_{t,n}^{(m)}(x), \qquad (t \in I,\ m \in I^+) \tag{4.1.3}$$

if this limit exists and is finite.

DEFINITION 4.1.1. *The limit* $h_t(A) = \lim_{n\to\infty} n^{-1}h(A^{\alpha_n})$ *(if it exists) is the differential entropy of the process A at the instant t.*

In [14] are given different properties of $h_t(A)$.

4.2. *The entropy stability.*

DEFINITION 4.2.1. *The stochastic process A possesses* (a) *the weak,* (b) *the strong, and* (c) *in the norm the entropy stability property at the instant t, if* $f^{\alpha_n}(x)$ *converges to* $h_t(A)$, *respectively,* (a) *in probability,* (b) *almost everywhere, and* (c) *in the norm in the Banach space* L_1.

We shall denote these properties by $E_t^{(i)}(A)$, $(i = 1, 2, 3)$.

THEOREM 4.2.1. *In order that the stochastic process A possesses the property* $E_t^{(i)}(A)$, *it is necessary and sufficient that the sequence* $\{g^{\alpha_k}(x)\}$ *satisfies the law of large numbers, respectively in* (a) *the weak sense* $(i = 1)$, (b) *the strong sense* $(i = 2)$, *and* (c) *in the norm* $(i = 3)$.

The proof is the same as in ([14], theorem 1.2).

THEOREM 4.2.2. *If one of the properties* $E_t^{(i)}(A)$, $E_{t+m}^{(i)}(A)$, $(m \in I^+)$ *is satisfied, then in order that the other property be satisfied also, it is necessary and sufficient that the convergence of* $\varphi_{t,n}^{(m)}(x)$ *to* $\lambda_t^{(m)}(A)$ *holds* (a) *in the probability* $(i = 1)$, (b) *almost everywhere* $(i = 2)$, *and* (c) *in the norm* $(i = 3)$.

The proof is the same as in ([14], theorem 3.4.).

DEFINITION 4.2.2. *If the property $E_t^{(i)}(A)$ is satisfied for all $t \in I$, then the property $E^{(i)}(A)$ is satisfied for* $(i = 1, 2, 3)$.

Obviously, from theorem 4.2.2 we may immediately obtain the necessary and sufficient conditions for $E^{(i)}(A)$, as in ([14], theorem 3.5). If A possesses only discrete sets of states X_t and finite $H(A_t)$, $(t \in I)$ (for example, if X_t is finite, $(t \in I)$), from the property $E_{t_0}^{(1)}(A)$ for an arbitrary $t_0 \in I$, property $E^{(1)}$ follows ([7], theorem 3.2). In [14] are given different properties of $E_t^{(3)}(A)$.

DEFINITION 4.2.3. *If $h_t(A)$ exists and has the same finite value for all $t \in I$ and the property $E^{(1)}(A)$ is satisfied, then A is regular.*

DEFINITION 4.2.4. *If A, B are two stochastic processes, $I_t(A, B) = \lim_{n\to\infty} n^{-1}\cdot I(A^{\alpha_n}, B^{\alpha_n})$ (if it exists) is the common quantity of information of A, B, at the instant t.*

For a stationary A, we denote $g_n(x) = g^{\alpha_n}(x)$.

THEOREM 4.2.3. *In order that the stationary stochastic process A possess the property $E^{(i)}(A)$, it is necessary and sufficient that the sequence $\{g_n(U^n x)\}$ (U is the shift operator) verify the law of large numbers, respectively in* (a) *the weak sense* $(i = 1)$, (b) *the strong sense* $(i = 2)$, *and* (c) *in the norm* $(i = 3)$.

In ([14], theorem 3.9 and 3.10) are given different sufficient conditions for $E^{(i)}$ $(i = 1, 3)$. Analogous results may be obtained for $E^{(2)}$. (The particular case of discrete sets of states was studied in ([2], [3]).)

5. The approximation of stochastic processes

5.1. *Notations.* Let us consider the sequence of measure spaces $(X_\tau, S_\tau, \mu_\tau)$, separable for the respective distances $\rho_\tau(x_\tau, y_\tau)$, and let us retain all the notations in 4.1. Further, let $\rho_\alpha(x^\alpha, y^\alpha) = \max_{\tau\in\alpha} \rho_\tau(x_\tau, y_\tau)$; $\rho(x, y) = \sup_{\tau\in I} \rho_\tau(x_\tau, y_\tau)$; $D(X)$ be the totality of stochastic processes over (X, S); $\theta_{(\tau)\epsilon}$, θ_ϵ^α, θ_ϵ, be ϵ-covers of the spaces X_τ, X^α, X, respectively. Obviously, $\theta_\epsilon^\alpha = \times_{\tau\in\alpha} \theta_{(\tau)\epsilon}$, $\theta_\epsilon = \times_{\tau\in I} \theta_{(\tau)\epsilon}$; that is, if $i^\alpha = \{i_\tau, \tau \in \alpha\}$, $i = \{i_\tau, \tau \in I\}$, $Z^\alpha \in \theta_\epsilon^\alpha$, $Z \in \theta_\epsilon$, there exist i^α and i such that $Z^\alpha = Z_{i^\alpha} = \times_{\tau\in\alpha} Z_{i_\tau}$, $Z = Z_i = \times_{\tau\in I} Z_{i_\tau}$, $Z_{i_\tau} \in \theta_{(\tau)\epsilon}(\tau \in I)$.

Let $D^0(X)$ denote the totality of discrete stochastic processes with states in $Z_\tau(\tau \in I)$. If $Z^\epsilon(x_{i_\tau})$, $Z^\epsilon(x_{i^\alpha})$, $Z^\epsilon(x_i)$ are spheres in X_τ, X^α, X, respectively, with centers x_{i_τ}, x_{i^α}, x_i and radius ϵ (for the distances ρ_τ, ρ^α, ρ), obviously

$$Z^\epsilon(x_{i^\alpha}) = \times_{\tau\in\alpha} Z^\epsilon(x_{i_\tau}), \qquad Z^\epsilon(x_i) = \times_{\tau\in I} Z^\epsilon(x_{i_\tau}). \tag{5.1.1}$$

Let us denote by $W_\epsilon(AA')$ the property that for every sample x_i of the discrete process $A' \in D^0(X)$ the condition $P_{AA'}\{Z^\epsilon(x_i)|x_i\} = 1$ holds; that is, the property $W_\epsilon(A^\alpha A'^\alpha)$ is satisfied for any $\alpha \subset I$.

5.2. *The ϵ-entropy of a sequence of sets.*

LEMMA 5.2.1. *The normed ϵ-entropy $K_\epsilon(X^\alpha)$ is equal to $\sum_{\tau\in\alpha} K_\epsilon(X_\tau)$.*

The proof follows from the definition of the distance $\rho_\alpha(x^\alpha, y^\alpha)$.

DEFINITION 5.2.1. *The quantity $K_{t,\epsilon}(X) = \lim_{n\to\infty} n^{-1}K_\epsilon(X^{\alpha_n})$, if it exists, is the normed ϵ-entropy of the sequence of sets $\{X_\tau\}$ at the instant t.*

DEFINITION 5.2.2. *If $K_{t,\epsilon}(X)$ exists and has the same finite value for all $t \in I$, then the sequence $\{X_\tau\}$ is regular.*

5.3. *The discrete ϵ-entropy of a stochastic process.*

DEFINITION 5.3.1. *The quantity $H_{t,\epsilon}(A) = \lim_{n\to\infty} n^{-1}H_\epsilon(A^{\alpha_n})$, if it exists, is the discrete ϵ-entropy of the stochastic process A at the instant t.*

We shall suppose that $H_{t,\epsilon}(A)$, $h_t(A)$, $K_{t,\epsilon}(X)$ exist and are finite for a fixed $t \in I$.

THEOREM 5.3.1. *The discrete ϵ-entropy $H_{t,\epsilon}(A)$ is equal to $h_t(A) + K_{t,\epsilon}(X)$.*

The proof follows from theorem 3.2.1.

THEOREM 5.3.2. *Let us assume that the stochastic process A possesses finite differential entropy $h_t(A)$, the property $E_t^{(i)}(A)$, $(i = 1, 2, 3)$, and that the normed ϵ-entropy $K_{t,\epsilon}(X)$ exists and is finite.*

Then, for any $\epsilon > 0$, there exists a discrete stochastic process $A_\epsilon \in D^0(X)$ with states not depending on A, such that

(a) *the property $W_\epsilon(AA_\epsilon)$ is satisfied;*

(b) *$H_{t,\epsilon}(A)$ and $H_t(A_\epsilon)$ exist, are finite, and $H_t(A_\epsilon) = H_{t,\epsilon}(A) + o(1)$;*

(c) *$I_t(A, A_\epsilon) = H_t(A_\epsilon) + o(1)$;*

(d) *A_ϵ possesses the corresponding property $E_t^{(i)}$, $(i = 1, 2, 3)$;*

(e) *if $\{X_\tau\}$ is regular, from the regularity of A follows that of A_ϵ;*

(f) *from the stationarity of A follows that of A_ϵ.*

PROOF. (a) In every X_τ let us consider an ϵ-cover $\theta_{(\tau)\epsilon}$ with $Z^{i_\tau} \in \theta_{(\tau)\epsilon}$ such that $\mu_\tau(Z_{i_\tau}) = \mu_\tau(X_\tau)/N_\epsilon(X_\tau)$. From the definition of ρ_α it follows that in this manner is generated an ϵ-cover θ_ϵ^α with $Z_{i^\alpha}^\epsilon = \times_{\tau\in\alpha} Z_{i_\tau}^\epsilon$, and $\mu^\alpha(Z_{i^\alpha}^\epsilon) = \mu^\alpha(X^\alpha)/N_\epsilon(X^\alpha)$ for all i^α.

Let us denote by $x_{i_\tau}^\epsilon$ the center of $Z_{i_\tau}^\epsilon$ and by $x_{i^\alpha}^\epsilon = \{x_{i_\tau}^\epsilon, \tau \in \alpha\}$ the center of $Z_{i^\alpha}^\epsilon$. We define the probability field $A_\epsilon^\alpha \in D^0(X^\alpha)$ with the elementary events $x_{i^\alpha}^\epsilon$ and

$$P_{A_\epsilon^\alpha}(x_{i^\alpha}^\epsilon) = P_{A^\alpha}(Z_{i^\alpha}^\epsilon) = p_{i^\alpha}^\epsilon \cdot \mu(Z_{i^\alpha}^\epsilon) \tag{5.3.1}$$

where $p_{i^\alpha}^\epsilon \in [\inf p^\alpha(x^\alpha), \sup p^\alpha(x^\alpha)]$, with the lower and upper bounds taken for $x^\alpha \in Z_{i^\alpha}^\epsilon$.

For any $\tau \in \alpha$, we define the union $A_\tau A_{(\tau)\epsilon}$ by means of the probability density

$$p_{A_\tau|A_{(\tau)\epsilon}}(x_\tau|x_{i_\tau}^\epsilon) = N_\epsilon(X_\tau)/\mu_\tau(X_\tau) \tag{5.3.2}$$

when $x_\tau \in Z_{i_\tau}^\epsilon$ $(\tau \in \alpha)$ and by zero in the remainder, and the union $A^\alpha A_\epsilon^\alpha$ by means of the probability density

$$p_{A_\alpha|A_\epsilon^\alpha}(x^\alpha|x_{i^\alpha}^\epsilon) = \prod_{\tau\in\alpha} p_{A_\tau|A_{(\tau)\epsilon}}(x_\tau|x_{i_\tau}^\epsilon) = \frac{N_\epsilon(X^\alpha)}{\mu^\alpha(X^\alpha)} = \prod_{\tau\in\alpha} \frac{N_\epsilon(X_\tau)}{\mu_\tau(X_\tau)} \tag{5.3.3}$$

for $x^\alpha \in Z_{i^\alpha}^\epsilon$, and by zero in the rest.

Obviously, the properties $W_\epsilon(A_\tau A_{(\tau)\epsilon})$, $W_\epsilon(A^\alpha A_\epsilon^\alpha)$, and $W_\epsilon(AA_\epsilon)$ are satisfied.

(b) We obtain immediately, as in theorem 3.2.3, that

$$\begin{aligned} h(A^\alpha|x_{i^\alpha}^\epsilon) &= h(A|A_\epsilon^\alpha) = -K_\epsilon(X^\alpha), \\ H(A_\epsilon^\alpha) &= h(A^\alpha) + K_\epsilon(X^\alpha) + o(1), \\ I(A^\alpha, A_\epsilon^\alpha) &= H(A_\epsilon) + o(1). \end{aligned} \tag{5.3.4}$$

We obtain immediately the results (b) and (c) if we recall the definitions of $h_t(A)$, $H_t(A_\epsilon)$, $H_{t,\epsilon}(A)$, $I_t(A, A_\epsilon)$.

(c) Obviously for $x \in Z^\epsilon_{i^\alpha}$,

$$\begin{aligned}(5.3.5)\qquad f^\alpha_{A_\epsilon}(x^\epsilon_{i^\alpha}) &= -n^{-1}\cdot\log P_{A^\alpha_\epsilon}(x^\epsilon_{i^\alpha}) = -n^{-1}\cdot\log p^\epsilon_{i^\alpha} + n^{-1}K_\epsilon(X^\alpha)\\ &= -n^{-1}\log p^\alpha(x) + n^{-1}K_\epsilon(X^\alpha) + o(1)\\ &= -n^{-1}\log p^\alpha(x) + K_{t,\epsilon}(X) + o(1),\end{aligned}$$

and consequently,

$$(5.3.6)\qquad f^\alpha_{A_\epsilon}(x^\epsilon_{i^\alpha}) - H_t(A_\epsilon) = f^\alpha_A(x) - h_t(A) + o(1).$$

Because P_{A_ϵ} is derived from P_A, from $E^{(i)}_t(A)$ follows $E^{(i)}_t(A_\epsilon)$, $(i = 1, 2, 3)$.

The results (e) and (f) follow immediately from the construction of the stochastic process A_ϵ.

6. The approximation of stochastic transition functions

6.1. *The metric space of stochastic transition functions*

Let us denote: (X, S), (X', S') two measurable spaces; $x \in X$, $x' \in X'$, $Z \in S$, $T \in S'$, $R(X', S')$ the totality of probability measures $P'(T)$ with the domain of definition (X', S'), $R(X, S, X', S')$ the totality of stochastic transition functions $P(x, T)$ with the domain of definition (X, S, X', S').

If P' and P'_1 are elements of $R(X', S')$, let us denote by $\beta'(P', P'_1)$ the total variation of $P' - P'_1$.

DEFINITION 6.1.1. *If P and P_1 are elements of $R(X, S, X', S')$ and if for a given $x \in X$ we denote by $P(x, \cdot)$, $P_1(x, \cdot)$ the corresponding measures, elements in $R(X', S')$, we define*

$$(6.1.1)\qquad \beta(P, P_1) = \sup_{x\in X} \beta'[P(x, \cdot), P_1(x, \cdot)] = \sup_{x\in X, T\in S'} |P(x, T) - P_1(x, T)|.$$

DEFINITION 6.1.2. (See [4].) *The ergodic coefficient of $P \in R(X, S, X', S')$ may be defined by*

$$(6.1.2)\qquad \alpha(P) = 1 - \sup_{x, x_1\in X} \beta'[P(x, \cdot), P(x_1, \cdot)].$$

Obviously, $0 \le \beta(P, P_1) \le 1$.

DEFINITION 6.1.3. *Two stochastic transition functions $P, P_1 \in R(X, S, X', S')$ are mutually almost singular, if for each $\epsilon > 0$ there exist some elements $x_\epsilon \in X$, $T_{x_\epsilon} \in S'$ such that $P(x_\epsilon, T_{x_\epsilon}) < \epsilon$, $P_1(x_\epsilon, T_{x_\epsilon}) < \epsilon$ where * denotes the complement.*

LEMMA 6.1.1. (a) *In order that $\beta(P, P_1) = 0$, it is necessary and sufficient that* $P \equiv P_1$; (b) *in order that $\beta(P, P_1) = 1$, it is necessary and sufficient that P, P_1 be mutually almost singular.*

PROOF. The proof of (a) is obvious; therefore, we shall prove only (b).

Necessity. If $\beta(P, P_1) = 1$, for any $\epsilon > 0$ there exist some $x_\epsilon \in X$ and some $T_{x_\epsilon} \in S'$, such that

$$(6.1.3)\qquad 1 - \epsilon < |P(x_\epsilon, T_{x_\epsilon}) - P_1(x_\epsilon, T_{x_\epsilon})| < 1.$$

From the equality

$$(6.1.4) \qquad P(x, T) - P_1(x, T) = -[P(x, T^*) - P_1(x, T^*)]$$

it follows that we may limit ourselves to the case where

$$(6.1.5) \qquad P(x_\epsilon, T_{x_\epsilon}) - P_1(x_\epsilon, T_{x_\epsilon}) < 0,$$

so that

$$(6.1.6) \qquad 1 - \epsilon < 1 - \epsilon + P(x_\epsilon, T_{x_\epsilon}) < P_1(x_\epsilon, T_{x_\epsilon}); \quad P(x_\epsilon, T_{x_\epsilon}) < P_1(x_\epsilon, T_{x_\epsilon}) - 1 + \epsilon < \epsilon$$

that is $P(x_\epsilon, T_{x_\epsilon}) < \epsilon$, $P_1(x_\epsilon, T^*_{x_\epsilon}) < \epsilon$.

Sufficiency. If P and P_1 are mutually almost singular, for any $\epsilon > 0$ there exist some $x_\epsilon \in X$, $T_{x_\epsilon} \in S'$ such that the inequalities in definition 6.1.3 are satisfied, and consequently,

$$(6.1.7) \qquad 1 - 2\epsilon < 1 - \epsilon - P(x_\epsilon, T_{x_\epsilon}) < P_1(x_\epsilon, T_{x_\epsilon}) - P(x_\epsilon, T_{x_\epsilon}) < P_1(x_\epsilon, T_{x_\epsilon}) < 1,$$

that is, $\beta(P, P_1) = 1$.

LEMMA 6.1.2. *If P and P_1 belong to $R(X, S, X', S')$, then $|\alpha(P) - \alpha(P_1)| \leq 2\beta(P, P_1)$.*

PROOF. Let us suppose that $\alpha(P) \leq \alpha(P_1)$. Obviously

$$(6.1.8) \qquad |P(x, T) - P(x_1, T)| \leq |P(x, T) - P_1(x, T)| + |P_1(x, T) - P_1(x_1, T)| + |P_1(x_1, T) - P(x_1, T)|.$$

Taking everywhere the upper bound for all $x \in X$, $x_1 \in X$, $T \in S'$, it follows immediately that $\alpha(P_1) - \alpha(P) \leq 2\beta(P, P_1)$, which proves the theorem.

THEOREM 6.1.1. *The space $R(X, S, X', S')$ is a complete metric space for the distance $\beta(P, P_1)$.*

PROOF. (a) *The function $\beta(P, P_1)$ is a distance.* The function β is symmetric, and in lemma 6.1.1 we have seen that from $\beta(P, P_1) = 0$ it follows that $P = P_1$. Let us consider $P_i(x, T) \in R(X, S, X', S')$, $(i = 1, 2, 3)$ and $|P_i(x, T) - P_j(x, T)| = u_{i,j}(x, T)$, $(i = 1,\ j = 2;\ i = 2,\ j = 3;\ i = 3,\ j = 1)$. From $u_{1,3} \leq u_{1,2} + u_{2,3}$, if we take everywhere the upper bound for $x \in X$, $T \in S'$, the triangular inequality follows for β.

(b) *The space $R(X, S, X', S')$ is complete.* Let $P_n(x, T)$, $(n \in I^+)$ be a β-fundamental sequence in this space, that is, $\beta(P_n, P_m) \to 0$, $(n, m \to \infty)$.

(b_1) From the definition of β it follows that the numerical sequence $P_n(x, T)$, $(n \in I)$ is fundamental for each pair of *fixed elements* $x \in X$, $T \in S'$, so that from the completeness of the real line there exists a limit $P(x, T)$ to which $P_n(x, T)$ converges as $n \to \infty$. From $P_n(x, T) \in R(X, S, X', S')$ it follows that $P(x, T) \in R(X, S, X', S')$.

(b_2) Because $P_n(x, T)$, $(n \in I)$ is a β-fundamental sequence, it follows that for any fixed $\epsilon > 0$ we may find a number $N = N(\epsilon)$ such that $\beta(P_n, P_m) < \epsilon$ for any $m, n \geq N(\epsilon)$, that is, $|P_n(x, T) - P_m(x, T), \leq \epsilon$ for all $x \in X$, $T \in S'$, and for all $m, n \geq N(\epsilon)$.

If m increases to infinity, from (b_1) it follows that $|P_n(x, T) - P(x, T)| \leq \epsilon$ for every fixed x, T with $n \geq N(\epsilon)$; that is, the convergence of P_n to P is uniform with respect to all $x \in X$, $T \in S'$ so that $\beta(P_n, P) \leq \epsilon$ for $n \geq N(\epsilon)$; that is, $\beta(P_n, P) \to 0$; in other words, the space $R(X, S, X', R')$ is complete.

THEOREM 6.1.2. *For β-convergence, the ergodic coefficient is continuous; that is, from $\beta(P_n, P) \to 0$ as $n \to \infty$ it follows that $\alpha(P_n) \to \alpha(P)$.*

The proof follows from lemma 6.1.2.

THEOREM 6.1.3. *The β-convergence is equivalent to convergence in distribution uniformly in $x \in X$, $T \in S'$.*

The proof follows from the definition of the distance β.

6.2. *The metric space of stochastic transition operators.* Let us consider some measurable space (X, S), and let us denote by V_X the Banach space of all real-valued generalized measures μ on the σ-algebra S, with norm $\|\mu\|$ one half of the total variation of μ. Obviously, for any probability measure μ, it follows that $\|\mu\| = \frac{1}{2}$.

Let us consider (see [4]) the subspace $L_X \subset V_X$ of all functions $\lambda \in V_X$ for which $\lambda(X) = 0$. In [4] it is proved that

$$\|\lambda\| = \sup_{Z \in S} |\lambda(Z)|. \tag{6.2.1}$$

If $P, P_1 \in R(X, S, X', S')$, it follows for any fixed $x \in X$ that $P(x, \cdot)$, $P_1(x, \cdot) \in V_{X'}$ and $\nu(x, \cdot) = P(x, \cdot) - P_1(x, \cdot) \in L_{X'}$ so that

$$\|\nu(x, \cdot)\| = \sup_{T \in S'} |\nu(x, T)|, \tag{6.2.2}$$

and consequently,

$$\beta(P, P_1) = \sup_{x \in X} \|\nu(x, \cdot)\|. \tag{6.2.3}$$

DEFINITION 6.2.1. *We define the stochastic transition operator Q which corresponds to the stochastic transition function $P(x, T)$ as a map from V_X to $V_{X'}$: $\mu' = Q\mu$, by means of the equality*

$$\mu'(T) = \int_X P(x, T)\, \mu(dx) \tag{6.2.4}$$

with $\mu \in V_X$, $\mu' \in V_{X'}$, $T \in S'$.

Obviously, Q is linear and continuous. If G_X is the subspace of all probability measures in V_X, it is obvious that Q maps G_X into $G_{X'}$ and its norm is one. If Q_1 corresponds to P_1 in the same manner as Q to P, let us consider the linear continuous operator $Q - Q_1$ which maps V_X into $L_{X'}$. We denote by $N(Q - Q_1)$ the norm of $Q - Q_1$, that is,

$$N(Q - Q_1) = \sup_{\mu \in V_X} \{\|(Q - Q_1)\mu\| / \|\mu\|\} = 2 \sup_{\mu \in V_X} \|(Q - Q_1)\mu\|. \tag{6.2.5}$$

LEMMA 6.2.1. *For any $\mu \in G_X$, if $\mu' = Q\mu \in G_{X'}$, $\mu_1' = Q_1\mu \in G_{X'}$, the inequality $\beta'(\mu', \mu_1') \leq \beta(P, P_1)$ is satisfied.*

PROOF. From the definition of μ', μ_1' it follows that for any $T' \in S'$,

$$(6.2.6)\qquad |\mu'(T) - \mu_1'(T)| \le \int_X |P(x, T) - P_1(x, T)|\, \mu(dx) \le \sup_{x\in X} |\nu(x, T)|$$

so that

$$(6.2.7)\qquad \beta'(\mu', \mu_1') = \sup_{T\in S'} |\mu'(T) - \mu_1'(T)| \le \sup_{x\in X, T\in S'} |\nu(x, T)| = \beta(P, P_1).$$

THEOREM 6.2.1. *The norm $N(Q - Q_1)$ is equal to $2\beta(P, P_1)$.*

PROOF. (a) Let $e_x(Z)$, $(x \in X)$ denote the probability measure with $e_x(Z) = 1$, if $x \in Z$, and $e_x(Z) = 0$, if $x \in Z^*$, so that $e_x \in G_X$ for any fixed $x \in X$. If $e_x' = Q_{e_x}$, it follows that

$$(6.2.8)\qquad e_x'(T) = \int_X P(x_1, T)\, e_x(dx_1) = P(x, T).$$

Considering also the analogous relation corresponding to $e_{(1)x}' = Q_1 e_x$, it follows that

$$(6.2.9)\qquad [(Q - Q_1)e_x](T) = P(x, T) - P_1(x, T),$$

so that, using the definition of $N(Q - Q_1)$,

$$(6.2.10)\qquad N(Q - Q_1) \ge 2\cdot\|Qe_x - Q_1e_x\| = 2\|e_x' - e_{(1)x}'\| = 2\cdot\|\nu(x, \cdot)\|$$

for any $x \in X$. This implies the inequality

$$(6.2.11)\qquad N(Q - Q_1) \ge 2\cdot \sup_{x\in X} \|\nu(x, \cdot)\| = 2\cdot\beta(P, P_1).$$

(b) If $\mu \in G_X$, then for any $T \in S'$,

$$(6.2.12)\qquad |[(Q - Q_1)\mu](T)| = \left|\int_X \nu(x, T)\, \mu(dx)\right| \le \int_X \sup_{x\in X} |\nu(x, T)|\, \mu(dx) = \sup_{x\in X} |\nu(x, T)| \le \beta(P, P_1).$$

Obviously $(Q - Q_1)\mu \in L_{X'}$. Let us suppose that X_+', X_-' are respectively the positive and negative sets of a Hahn decomposition of X' for the function $(Q - Q_1)\mu$. From the above inequality it follows in particular that $[(Q - Q_1)\mu]\,(X_+') \le \beta(P, P_1)$. Consequently, it is easy to see that $[(Q - Q_1)\mu]\,(X_+') = -[(Q - Q_1)\mu]\,(X_-') = \|(Q - Q_1)\mu\|$ so that $\|(Q - Q_1)\mu\| \le \beta(P, P_1)$ for any $\mu \in G_X$ and $N(Q - Q_1) \le 2\beta(P, P_1)$, which proves our lemma. Obviously $N(Q - Q_1)$ is a distance in the metric space of all probability transition operators.

6.3. *Another expression of $\beta(P, P_1)$.* Let us consider the measurable space (X, S).

DEFINITION 6.3.1. *Between two measures, μ_1 and $\mu_2 \in V_X$, there exists the relation $\mu_1 \prec \mu_2$ if $\mu_1(Z) \le \mu_2(Z)$ for any $Z \in S$.*

For $\rho \in V_X$ let us denote

$$(6.3.1)\qquad \sigma(\mu_1, \mu_2) = \sup_{\rho\quad \mu_1, \rho \prec \mu_2} \rho(X).$$

A. N. Kolmogorov pointed out ([4], section 1) that $\sigma(\mu_1, \mu_2)$ may also be defined by

$$\sigma(\mu_1, \mu_2) = \inf \sum_{i=1}^{m} \min [\mu_1(Z_i), \mu_2(Z_i)] \tag{6.3.2}$$

where the lower bound is taken over all possible finite covers θ of X, and $Z_i \in \theta$, $(1 \leq i \leq m < \infty)$. It is known ([4], section 1) that $\|\mu_1 - \mu_2\| = 1 - \sigma(\mu_1, \mu_2)$. If $\mu_1 = P(x, \cdot)$, $\mu_2 = P_1(x, \cdot)$, we obtain $\|P(x, \cdot) - P_1(x, \cdot)\| = 1 - \sigma[P(x, \cdot), P_1(x, \cdot)]$ so that we obtain the following theorem.

THEOREM 6.3.1. *The following equalities hold:*

$$\begin{aligned}\beta(P, P_1) &= \sup_{x \in X} \|\nu(x, \cdot)\| \\ &= 1 - \inf_{x \in X} \sigma[P(x, \cdot), P_1(x, \cdot)].\end{aligned} \tag{6.3.3}$$

If X, X' are denumerable sets with the states x_i, $(i \in I)$, then $P(x, T)$ and $P_1(x, T)$ are given by means of the stochastic matrices Q, Q_1 with elements $p_{k,m}$, $p_{k,m}^{(1)}$.

THEOREM 6.3.2. *The distance $\beta(P, P_1)$ is equal to*

$$\beta(P, P_1) = 1 - \inf_{1 \leq k < \infty} \sum_{m=1}^{\infty} \min (p_{k,m}, p_{k,m}^{(1)}). \tag{6.3.4}$$

PROOF. It is easy to see [4] that the expression given by A. N. Kolmogorov for $\sigma(\mu_1, \mu_2)$ does not change if we consider not only finite covers of X but also denumerable covers of it.

Let us observe that the sum in this expression cannot decrease if instead of the cover θ we consider another cover θ', finer than θ, that is, in which each set in θ is a sum of certain sets in θ'. Because the cover θ_0, each set of which contains only one element $Z_i = x_i$, is finer than any arbitrary cover θ, from the expression of σ it follows that

$$\sigma(\{p_{k,m}\}, \{p_{k,m}^{(1)}\}) = \sum_{m=1}^{\infty} \min (p_{k,m}, p_{k,m}^{(1)}). \tag{6.3.5}$$

From theorem 6.3.1, the desired result follows.

6.4. *The discrete case.* Here we shall prove theorem 6.2.1 using the definition of $\beta(P, P_1)$ from theorem 6.3.1.

Let us consider [4] the linear space F of those infinite dimensional vectors $q = \{q_i\}$, $(i \in I^+)$ for which the sum of the components vanishes and the sum of their absolute values converges. If $U = \{u_{i,j}\}$, $U_1 = \{u_{i,j}^{(1)}\}$ are some stochastic matrices and $q \in F$, then $Uq \in F$, $U_1 q \in F$. Let us define the norm of q by

$$\|q\| = \sum_{i=1}^{\infty} |q_i| = 2 \sum_{i=1}^{\infty} (q_i)^+ = -2 \cdot \sum_{i=1}^{\infty} (q_i)^- \tag{6.4.1}$$

where $(a)^+ = \max (a, 0)$, $(a)^- = \min (a, 0)$. Obviously, $(a + b)^+ \leq (a)^+ + (b)^+$. Let us denote

$$\beta(U, U_1) = 1 - \inf_{i} \sum_{j=1}^{\infty} \min (u_{i,j}, u_{i,j}^{(1)}). \tag{6.4.2}$$

THEOREM 6.4.1. *The norm $N(U - U_1)$ is equal to*

$$N(U - U_1) = \sup_{q \in F} \frac{\|Uq - U_1 q\|}{\|q\|} = 2\beta(U, U_1). \tag{6.4.3}$$

PROOF. (a) First we shall prove that the number on the left side is not greater than the one on the right; for this it is sufficient to prove that for any $q \in F$, $\|Uq - U_1q\| \leq 2\|q\| \cdot \beta(U, U_1)$.

(a_1) If q has only two nonvanishing components $q_{i_1} = \|q\|/2 = \lambda$, $q_{i_2} = -\lambda$, using the same method as in ([4], p. 372) and the notation

$$\sum_{k=1}^{\infty}{}' = \sum_{l=i_1,i_2} \sum_{k=1}^{\infty}, \tag{6.4.4}$$

we obtain the inequalities

$$\begin{aligned} \|(U - U_1)q\| \\ &= 2 \sum_{k=1}^{\infty} [q_{i_1}(u_{i_1k} - u^{(1)}_{i_1k}) + q_{i_2}(u_{i_2k} - u^{(1)}_{i_2k})]^+ \\ &\leq 2 \sum_{k=1}^{\infty}{}' [q_l(u_{lk} - u^{(1)}_{lk})]^+ = \|q\| \sum_{k=1}^{\infty} [(u_{i_1k} - u^{(1)}_{i_1k})^+ + (u^{(1)}_{i_2k} - u_{i_2k})^+] \\ &= \|q\| \sum_{k=1}^{\infty} \{[u_{i_1k} - \min(u_{i_1k}, u^{(1)}_{i_1k})] + [u^{(1)}_{i_2k} - \min(u_{i_2k}, u^{(1)}_{i_2k})]\} \\ &= \|q\| \sum_{l=i_1,i_2} \left[1 - \sum_{k=1}^{\infty} \min(u_{lk}, u^{(1)}_{lk})\right] \leq 2\|q\| \cdot \beta(U, U_1). \end{aligned} \tag{6.4.5}$$

(a_2) If q is any vector in F, it is easy to see that it may be represented as an absolute convergent sum

$$q = \sum_{i=1}^{\infty} q^{(i)} \tag{6.4.6}$$

of vectors $q^{(i)} \in F$ in such a way that each vector $q^{(i)}$ has only two nonvanishing components, and also

$$\|q\| = \sum_{i=1}^{\infty} \|q^{(i)}\|. \tag{6.4.7}$$

From (a_1) we obtain the relations

$$\begin{aligned} \|(U - U_1)q\| &\leq \sum_{i=1}^{\infty} \|(U - U_1)q^{(i)}\| \leq 2\beta(U, U_1) \cdot \sum_{i=1}^{\infty} \|q^{(i)}\| \\ &= 2\beta(U, U_1)\|q\|. \end{aligned} \tag{6.4.8}$$

(b) We shall prove the inverse inequality.

(b_1) From the given definition of $\beta(U, U_1)$ it follows that for any $\epsilon > 0$ there exist two different numbers i_1, i_2 such that for $l = i_1, i_2$, the inequality

$$\left| \sum_{j=1}^{\infty} \min(u_{l,j}, u^{(1)}_{l,j}) - [1 - \beta(U, U_1)] \right| < \epsilon \tag{6.4.9}$$

is satisfied.

The existence of one value i_1 with the indicated property follows from the definition of the lower bound; in the case where another value $i_2 \neq i_1$ with the indicated property does not exist, we may use the following method. Instead of the matrices U, U_1 with the states $\{x_i\}$, $(i = 1, 2, \cdots)$, we consider the matrices T, T_1 with the states $\{x_i\}$, $(i = 0, 1, 2, \cdots)$ where $t_{i,j} = u_{i,j}$, $t_{0,j} = u_{i_1,j}$, $(i, j = 1, 2, \cdots)$, $t_{i,0} = 0$ $(i = 0, 1, 2, \cdots)$, and analogously for T_1. Obviously, $\beta(U, U_1) = \beta(T, T_1)$; here $i_2 = 0$ and $i_1 \neq i_2 = 0$ have the desired property. Consequently, from the beginning we may suppose that U, U_1 possesses this property.

(b_2) For a fixed vector q which possesses only two components $q_{i_1} = \|q\|/2$, $q_{i_2} = -q_{i_1}$, from the inequalities in (a_1), using the inequality in (b), we obtain $\|(U - U_1)q\| \geq 2\|q\|[\beta(U, U_1) - \epsilon]$, or

$$N(U - U_1) \geq \frac{\|(U - U_1)q\|}{\|q\|} \geq 2\beta(U, U_1) - 2\epsilon, \tag{6.4.10}$$

and consequently, $N(U - U_1) \geq 2\beta(U, U_1)$.

THEOREM 6.4.2. *The following equalities hold:*

$$N(U - U_1) = 2\cdot\beta(U, U_1) = \sup_i \sum_{k=1}^{\infty} |u_{i,k} - u_{i,k}^{(1)}|. \tag{6.4.11}$$

PROOF. We may observe that the next to last inequality in (b_2) shows that the upper bound in the last inequality in (b_2) is attained for vectors q which possess only two nonvanishing components. If F_1 is the totality of these vectors, it follows that

$$\begin{aligned}\|(U - U_1)q\| &= \|q\| \cdot \sum_{k=1}^{\infty} \{(u_{i_1,k} - u_{i_1,k}^{(1)})^+ + (u_{i_2,k}^{(1)} - u_{i_2,k})^+\} \\ &= \|q\| \cdot \sum_{k=1}^{\infty}{}' |u_{l,k} - u_{l,k}^{(1)}|,\end{aligned} \tag{6.4.12}$$

and consequently,

$$\begin{aligned}\sup_{q\in F_1} \frac{\|(U - U_1)q\|}{\|q\|} &= 2\beta(U, U_1) = \tfrac{1}{2} \cdot \sup_{i_1,i_2} \sum_{k=1}{}' |u_{l,k} - u_{l,k}^{(1)}| \\ &= \sup_{i\in I} \sum_{k=1} |u_{i,k} - u_{i,k}^{(1)}|.\end{aligned} \tag{6.4.13}$$

6.5. *The approximation theorems of stochastic transition functions.* Let us suppose that $P(x, T) \in R(X, S, X', S')$, that θ is a cover of X, and that x_i^0 is an arbitrarily fixed element in $Z_i \in \theta$. We define the stochastic transition function $P_1(x, T)$ equal to $P(x_i^0, T)$ for any $x \in Z_i \in \theta$.

LEMMA 6.5.1. *The distance* $\beta(P, P_1)$ *satisfies the inequality* $\beta(P, P_1) \leq 1 - \alpha(P)$.

PROOF. One can write

$$\begin{aligned}\beta(P, P_1) &= \sup |\nu(x, T)| = \sup |P(x, T) - P(x_i^0, T)| \\ &\leq \sup |P(x, T) - P(x_1, T)| \leq \sup |P(x, T) - P(x_1, T)| = 1 - \alpha(P)\end{aligned} \tag{6.5.1}$$

where the first upper bound is taken for $x \in X$, $T \in S'$, the second for $i \in I^+$,

$x \in Z_i$, $T \in \sum'$, the third for $i \in I$, $x \in Z_i$, $x_1 \in Z_i$, $T \in S'$, the fourth for x, $x_1 \in X$, $T \in S'$.

Let us consider that (X, S) is a separable metric space with the distance $\rho(x, x_1)$, and $P(x, T)$ is uniformly continuous in $x \in X$, uniform for all $T \in S'$. That is, for any $\delta > 0$ there exists a number $\epsilon = \epsilon(\delta) > 0$ such that

$$|P(x, T) - P(x_1, T)| < \delta \tag{6.5.2}$$

for all $x, x_1 \in X$ for which $\rho(x, x_1) < \epsilon$ and for all $T \in S'$. If θ_ϵ is an ϵ-cover of X, $Z_i^\epsilon \in \theta_\epsilon$, x_i^ϵ the center of Z_i^ϵ, $(i \in I)$, let us define the stochastic transition function $P_\epsilon(x, T)$ equal to $P(x_i^\epsilon, T)$ for $x \in Z_i^\epsilon$, $(i \in I)$, $T \in S'$.

THEOREM 6.5.1. *For any* $\delta > 0$, *there exists a number* $\epsilon = \epsilon(\delta)$ *such that* $\beta(P, P_\epsilon) < \delta$.

The *proof* follows from the first two equalities of the proof of lemma 6.5.1 letting $Z_i = Z_i^\epsilon$, $x_i^0 = x_i^\epsilon$, $P_1 = P_\epsilon$ if we observe that for all $T \in S'$, $x \in Z_i^\epsilon$, $i \in I^+$, the following inequality is satisfied:

$$|P(x, T) - P(x_i^\epsilon, T)| < \delta. \tag{6.5.3}$$

Here we shall study the simultaneous approximation of a probability field and of a stochastic transition function which transforms it.

Let us denote by $p(x, x')$ the conditional probability density of $P(x, T)$; if we consider also the probability distribution $P_A(Z)$ of the field A, then the conditional distribution $P_{A_{x'}}(Z|x')$ of the field $A_{x'} \in D(X)$ is completely defined for any $x' \in X'$. Let us denote by $p(x|x')$ the probability density of $A_{x'}$. By P_A and $P(x, T)$, a field $B \in D(X')$ is completely defined also.

Let us denote by $R^0(X, S, X', S')$ the totality of probability transition functions with domain of definition (X_1, S_1, X', S') where X_1 is any discrete subset of X.

THEOREM 6.5.2. *Let us consider* $\delta > 0$, $A \in D(X)$, $P(x, T) \in R(X, S, X', S')$ *uniformly continuous in* $x \in X$, *uniformly for* $T \in S'$.

There exists a number $\epsilon = \epsilon(\delta)$, *discrete probability fields* $A_\epsilon \in D^0(X)$, $(A_{x'})_\epsilon \in D^0(X)$, $(x' \in X')$, *and a discrete stochastic transition function* $P_\epsilon(x, T) \in R^0(X, S, X', S')$ *such that*

(a) *the properties* $W_\epsilon(AA_\epsilon)$, $W_\epsilon[A_{x'}(A_{x'})_\epsilon]$ *are satisfied,*

(b) *if* Q, Q_ϵ *are stochastic transition operators defined by* P, P_ϵ *respectively, and* $P_B = Q \cdot P_A$, $P_{B\epsilon} = Q \cdot P_{A\epsilon} = Q_\epsilon \cdot P_{A\epsilon}$, *then*

$$\beta'(P_B, P_{B\epsilon}) \leq \beta(P, P_\epsilon) < \delta, \tag{6.5.4}$$

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

PROOF. (a) From theorem 3.2.3 it follows that for any $x' \in X'$ there is a discrete probability field $(A_{x'})_\epsilon$ such that the condition $W_\epsilon[A_{x'}(A_{x'})_\epsilon]$ is satisfied. Obviously, if θ_ϵ is an ϵ-cover, then the states of $(A_{x'})_\epsilon$ and those of A_ϵ are the centers x_i^ϵ of $Z_i^\epsilon \in \theta_\epsilon$. We may observe that x_i^ϵ does not depend on $x' \in X'$. Let us define the probability in $(A_{x'})_\epsilon$ by

$$P_{(A_{x'})_\epsilon}(x_i^\epsilon|x') = P_{A_{x'}}(Z_i^\epsilon|x') = \int_{Z^\epsilon} p(x|x')\,dx = p_i(x') \cdot \mu(Z_i^\epsilon) \tag{6.5.5}$$

where $p_i(x')$ is a number between the lower and the upper bounds of $p(x|x')$, for $x \in Z_i^\epsilon$. Let us also define the union $A_{x'}(A_{x'})_\epsilon$ by the density $p_{A|(A_{x'})_\epsilon}(x|x_i^\epsilon, x') = p_{A|A_\epsilon}(x|x_i^\epsilon)$ equal to $N_\epsilon(X)/\mu(X)$ for $x \in Z_i^\epsilon$ and zero in the rest, for any arbitrary $x' \in X'$. Consequently,

$$p_{A_{x'}(A_{x'})_\epsilon}(x, x_i^\epsilon|x') = p_{A|(A_{x'})_\epsilon}(x|x_i^\epsilon, x') \cdot P_{(A_{x'})_\epsilon}(x_i^\epsilon|x'), \tag{6.5.6}$$

which is equal to $p_i(x')$ for $x \in Z_i^\epsilon$ and to zero in the rest.

Obviously the condition $W_\epsilon[A_{x'}(A_{x'})_\epsilon]$ is satisfied. We also obtain

$$h(A|x_i^\epsilon, x') = h(A|(A_{x'})_\epsilon) = K_\epsilon(X). \tag{6.5.7}$$

If $\varphi(t) = t \log t$, then

$$\begin{aligned} H[(A_{x'})_\epsilon] &= -\sum_i \varphi[P_{(A_{x'})_\epsilon}(x_i^\epsilon|x')] = -\sum_i \varphi[P_{(A_{x'})_\epsilon}(Z_i^\epsilon|x')] \\ &= -\sum_i \varphi[p_i(x')\mu(Z_i^\epsilon)] = -\sum_i \varphi[p_i(x')]\cdot\mu(Z_i^\epsilon) \\ &\quad -\sum_i P_{A_x}(Z_i^\epsilon|x')\log\mu(Z_i^\epsilon) = h(A_{x'}) + K_\epsilon(X) + o(1). \end{aligned} \tag{6.5.8}$$

Let us consider the probability fields defined by

$$P_B(T) = \int_X P_A(dx)\, P(x, T), \qquad P_{B_\epsilon}(T) = \int_X P_A(dx)\, P_\epsilon(x, T). \tag{6.5.9}$$

(b) From lemma 6.2.1 and from theorem 6.5.1, it follows that for any $\delta > 0$ there exists a number $\epsilon = \epsilon(\delta)$ such that $\beta'(P_B, P_{B_\epsilon}) \le \beta(P, P_\epsilon) < \delta$, and consequently, for any $Z \in S$, $P_{B_\epsilon}(Z) = P_B(Z)(1 + o(1))$.

(c) Consequently,

$$\begin{aligned} H(A_\epsilon|B_\epsilon) &= M_{B_\epsilon}H[(A_{x'})_\epsilon] = M_B H[(A_{x'})_\epsilon](1 + o(1)) \\ &= h(A|B) + K_\epsilon(X) + o(1), \end{aligned} \tag{6.5.10}$$

and using theorems 3.2.3 (b) and 3.2.1, it follows that $I(A_\epsilon, B_\epsilon) = H(A_\epsilon) - H(A_\epsilon|B_\epsilon) = I(A, B) + o(1)$.

7. The stochastic complex source-channel

7.1. *The differential entropy of* $[A, \Delta]$. The stochastic channel Δ is defined by (a) the input-elements $x_\tau \in X_\tau$, $(\tau \in I)$; (b) the output-elements y_τ which form the measure space $(Y_\tau, V_\tau, \nu_\tau)$, $(\tau \in I)$, $(Y^{\alpha_n}, V^{\alpha_n}, \nu^{\alpha_n}) = X_{\tau\in\alpha_n}(Y_\tau, V_\tau, \nu_\tau)$; $(Y, V) = X_{\tau\in I}(Y_\tau, V_\tau)$, where $y^{\alpha_n} \in Y^{\alpha_n}$, $y \in Y$; (c) the transmission law which is defined by the probability density $\pi_{B|A}^{\alpha_n'}(y^{\alpha_n}|x)$ (with respect to the ν^{α_n}-measure) of the realization of the element $y^{\alpha_n} \in Y^{\alpha_n}$ in the time $\alpha_n = [t, t + n - 1]$ by the output of the channel, if it is known that by the input, $x \in X$ is entered.

In this manner, for any $t \in I$, $n \in I^+$, $T^{\alpha_n} \in V^{\alpha_n}$ the measure $P_{B|A}^{\alpha_n}(T^{\alpha_n}|x)$ is defined, and consequently, their extension $P_{B|A}(T|x)$ for $T \in V$, $x \in X$. Let us denote the channel defined in this manner by $\Delta = [X, P_{B|A}(\cdot|x), Y]$.

We shall use the ordinary concept of a nonanticipative channel with finite memory. A channel is stationary if $\pi_{B|A}^{\alpha'_n}(Uy^{\alpha_n}|Ux) = \pi_{B|A}^{\alpha_n}(y^{\alpha_n}|x)$ for any $y^{\alpha_n} \in Y^{\alpha_n}$, $x \in X$, $t \in I$, $n \in I^+$. Here $\alpha'_n = \{\tau, t + 1 \leq \tau \leq t + n\}$, and U is the shift operator.

If X_τ is simultaneously the set of states of the input process A and of the input of the channel Δ at the instant $\tau(\tau \in I)$, we may consider the composite process AB with sets of states $X_\tau \times Y_\tau$ and also the output-process B with sets of states Y_τ. In this case, let us denote by $[A, \Delta]$ the complex of the input-process A and the channel Δ. We also denote

$$\begin{aligned} g_{A|B}^{\alpha_n}(x, y) &= g_{AB}^{\alpha_n}(x, y) - g_B^{\alpha_n}(x); & f_{A|B}^{\alpha_n}(x, y) &= f_{AB}^{\alpha_n}(x, y) - f_B^{\alpha_n}(y), \\ G_{AB}^{\alpha_n}(x, y) &= g_A^{\alpha_n}(x) - g_{A|B}^{\alpha_n}(x, y); & F_{AB}^{\alpha_n}(x, y) &= f_A^{\alpha_n}(x) - f_{A|B}^{\alpha_n}(x, y). \end{aligned} \tag{7.1.1}$$

Definition 7.1.1. *The differential entropy of the complex* $[A, \Delta]$ *is the quantity*

$$h_t(A|B) = \lim_{n\to\infty} n^{-1} \cdot h(A^{\alpha_n}|B^{\alpha_n}) \tag{7.1.2}$$

(if it exists). The rate of information transmission in the complex $[A, \Delta]$ *is the quantity*

$$I_t(A, B) = \lim_{n\to\infty} n^{-1} \cdot I(A^{\alpha_n}, B^{\alpha_n}) \tag{7.1.3}$$

(if it exists)

Different properties of these concepts are given in ([8]–[11], [14]).

7.2. *The entropy stability for the complex* $[A, \Delta]$.

Definition 7.2.1. *The complex* $[A, \Delta]$ *possesses* (a) *the weak,* (b) *the strong, and* (c) *the norm entropy stability* (resp. *information stability*) *at the instant t if* $f_{A|B}^{\alpha_n}(x, y)$ (resp. $F_{AB}^{\alpha_n}(x, y)$) *converges to* $h_t(A|B)$ (resp. $I_t(A, B)$) *respectively* (a) *in probability,* (b) *almost everywhere,* (c) *in the norm in the Banach space* L_1.

We shall denote these properties by $E_t^{(i)}(A|B)$, $J_t^{(i)}(AB)$, $(i = 1, 2, 3)$.

Theorem 7.2.1. *In order that the complex* $[A, \Delta]$ *possess the property* $E_t^{(i)}(A|B)$ (resp. $J_t^{(i)}(AB)$), *it is necessary and sufficient that the sequence* $\{g_{A|B}^{\alpha_n}(x, y)\}$ (resp. $\{G_{AB}^{\alpha_n}(x, y)\}$) *satisfy respectively* (a) *the weak* $(i = 1)$, *(b) the strong* $(i = 2)$, *and* (c) *the norm* $(i = 3)$ *law of large numbers.*

The proof is analogous to that of theorem 4.2.1.

Definition 7.2.2. *If the property* $E_t^{(i)}(A|B)$ (resp. $J_t^{(i)}(AB)$), $(i = 1, 2, 3)$ *is satisfied for all* $t \in I$, *then the property* $E^{(i)}(A|B)$, $(J^{(i)}(AB))$ *is satisfied.*

As in the case of the processes (see 4.2), here also results may be obtained concerning the existence of the properties $E^{(i)}(A|B)$, $J^{(i)}(AB)$ and also concerning the stationary complexes $[A, \Delta]$, (see [8]–[11], [14]).

Definition 7.2.3. *The regular set of sources* F_Δ *of the channel* Δ *is the totality of regular sources A with the same states* X_τ *as in the input of* Δ *at the same instant, for which* $I_t(A, B)$ *exists, is finite, does not depend on the time, and satisfies the property* $J^{(1)}(A, B)$.

Definition 7.2.4. *The channel* Δ *is regular if it is nonanticipative and* F_Δ *is not void.*

Definition 7.2.5. *The regular capacity of the channel* Δ *is* $C = \sup I(A, B)$ *where the upper bound is taken for* $A \in F_\Delta$.

8. The approximation of stochastic channels

8.1. *The metric space of stochastic channels.* Let us suppose that the nonanticipative channel Δ with finite memory m is given by means of the measure spaces $(X_\tau, S_\tau, \mu_\tau)$, $(Y_\tau, V_\tau, \nu_\tau)$, $(\tau \in I)$ and of the probability transition functions $P^{\alpha_n}(X^{\alpha'_n}, T^{\alpha_n}) \in R(X^{\alpha'_n}, S^{\alpha'_n}, Y^{\alpha_n}, V^{\alpha_n})$ where $\alpha_n = [t, t+n-1]$, $\alpha'_n = [t-m, t+n-1]$, $(t \in I, n \in I^+)$.

Let us denote by $R(X, S, Y, V)$ the totality of channels over (X, S, Y, V).

DEFINITION 8.1.1. *If $\Delta, \Delta_1 \in R(X, S, Y, V)$, and P^α, P^α_1 are the corresponding probability transition functions $(\alpha \subset I)$, we define $\gamma(\Delta, \Delta_1) = \sup \beta(P^\alpha, P^\alpha_1)$, where the upper bound is taken over all $\alpha \subset I$.*

LEMMA 8.1.1. *If $\alpha \subset \alpha_1$, then $\beta(P^\alpha, P^\alpha_1) \leq \beta(P^{\alpha_1}, P^{\alpha_1}_1)$.*

The proof follows immediately from the definition of the distance β.

THEOREM 8.1.1. *The space $R(X, S, Y, V)$ is a complete metric space with the distance $\gamma(\Delta, \Delta_1)$.*

PROOF. The function $\gamma(\Delta, \Delta_1)$ is a distance because $\beta(P^\alpha, P^\alpha_1)$ is a distance for any $\alpha \subset I$. The space $R(X, S, Y, V)$ is complete for the distance $\gamma(\Delta, \Delta_1)$, because $R(X^{\alpha'}, S^{\alpha'}, Y^\alpha, V^\alpha)$ is complete for the distance $\beta(P^\alpha, P^\alpha_1)$.

8.2. *The approximation of the system (A, Δ).* Let us consider a cover θ_τ of X_τ, $Z_{i_\tau} \in \theta_\tau$, and let x_{i_τ} be any arbitrarily given element in Z_{i_τ}, $(\tau \in I)$. In this manner is also determined a cover $\theta^\alpha = \times_{\tau \in \alpha} \theta_\tau$ in X^α such that $x_{i^\alpha} = \{x_{i_\tau}, \tau \in \alpha\} \in Z_{i^\alpha} \in \theta^\alpha$. For any given channel $\Delta \in R(X, S, Y, V)$ let us define another channel Δ_1 by means of the probability transition functions $P^\alpha_1(x^{\alpha'}, T^\alpha)$ equal to $P^\alpha(x_{i^{\alpha'}}, T^\alpha)$ for $x^\alpha \in Z_{i^\alpha}$, $\alpha' = [t-m, t+n-1]$, $\alpha = [t, t+n-1]$. If $a(P^\lambda)$ is the ergodic coefficient of P^λ, let us denote $\sigma(\Delta) = \inf a(P^\lambda)$ where the lower bound is taken for all $\lambda \subset I$.

LEMMA 8.2.1. *The distance $\gamma(\Delta, \Delta_1)$ is less than or equal to $1 - \sigma(\Delta)$.*

The proof follows from lemma 6.5.1.

With the hypotheses and notation of 5.1, let us suppose that $P^\alpha(x^{\alpha'}, T^\alpha)$ is uniformly continuous in $x^\alpha \in X^\alpha$, (uniformly for all $\alpha \subset I$, $T^\alpha \in V^\alpha$). Let us define the stochastic channel Δ_ϵ by means of the probability transition functions $P^\alpha_\epsilon(x^{\alpha'}, T^\alpha) = P^\alpha(x^\epsilon_{i^{\alpha'}}, T^\alpha)$ for $x^{\alpha'} \in Z^\epsilon_{i^{\alpha'}}$. From theorem 6.5.1 follows immediately theorem 8.2.1.

THEOREM 8.2.1. *For any $\delta > 0$ there exists a number $\epsilon = \epsilon(\delta)$ such that $\gamma(\Delta, \Delta_\epsilon) < \delta$. Let us denote: $R^0(X, S, Y, V)$ the totality of stochastic processes with domain of definition (X_1, S_1, Y, V) where X_1 is any discrete subset of X; $C_\epsilon = \sup_{A_\epsilon} I_t(A_\epsilon, B_\epsilon)$.*

THEOREM 8.2.2. *Let us consider* (1) *a stochastic process $A \in D(X)$, which possesses finite $h_t(A)$ and the property $E^{(t)}_t(A)$;* (2) *a stochastic channel $\Delta \in R(X, S, Y, V)$ which is defined by uniformly continuous $P^\alpha(x^{\alpha'}, T^\alpha)$ (uniformly in $\alpha \subset I$, $T^\alpha \in V^\alpha$), and possesses finite $h_t(A|B)$ and the property $E^{(t)}_t(A|B)$* (resp. *$I_t(A, B)$ and $J^{(t)}_t(A, B)$*)*.*

For any given $\delta > 0$, we may determine a number $\epsilon = \epsilon(\delta)$ such that

(1) *there exists a discrete stochastic process $A_\epsilon \in D^0(X)$ with finite entropy*

(8.2.1) $$H_t(A_\epsilon) = h_t(A) + K_{t,\epsilon}(X) + o(1)$$

and the property $E_t^{(i)}(A_\epsilon)$, $(i = 1, 2, 3)$;

(2) *there exists a discrete stochastic channel* $\Delta_\epsilon \in R^0(X, S, Y, V)$, *such that if connected with* A_ϵ, *there exists finite*

(8.2.2) $$H_t(A|B) = h_t(A|B) + K_{t,\epsilon}(X) + o(1),$$

and the property $E_t^{(i)}(A|B)$, *or respectively*, $I_t(A_\epsilon, B_\epsilon)$, $J_t^{(i)}(A, B)$, $(i = 1, 2, 3)$;

(3) *the property* $W_\epsilon(AA_\epsilon)$ *is satisfied, and* $\gamma(\Delta, \Delta_\epsilon) < \delta$;

(4) *from the regularity of* $\{X_\tau\}$, Δ, *the same thing follows for* Δ_ϵ;

(5) *from the stationarity of* Δ *the same thing follows for* Δ_ϵ;

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

(7) $C = C_\epsilon + o(1)$.

For the proof we may use the process A_ϵ constructed in theorem 5.3.2 and the channel Δ_ϵ constructed in theorem 8.2.1; the proof runs analogously to that of theorem 6.5.2.

9. The basic theorems of Shannon type

We shall suppose here that $\{X_\tau\}$ is regular.

THEOREM 9.1. *Let us consider* (1) *a regular channel* Δ *with uniformly continuous probability transition functions, with continuous input sets of states, with finite memory, and with finite regular capacity* C;

(2) *a regular process* $\mathring{A}$ *with continuous input sets of states and* $h(\mathring{A}) < C$.

For a given $\delta > 0$, *if we determine* $\epsilon = \epsilon(\delta)$, $\mathring{A}_\epsilon$, Δ_ϵ *as in theorems* 5.3.2, 8.2.2, *obviously* $W_\epsilon(\mathring{A}\mathring{A}_\epsilon)$ *is satisfied and* $\gamma(\Delta, \Delta_\epsilon) < \delta$. *If*

(9.1.1) $$H(\mathring{A}_\epsilon) = h(\mathring{A}) + K_\epsilon(\mathring{X}) < C + o(1),$$

then concerning the possibility of transmission of the production of the process A_ϵ *through the channel* Δ_ϵ *with the error probability not greater than a given* λ, *the first basic theorem of Shannon type is true* ([14], *p.* 243).

If $\mathring{X}_\tau$, the sets of states of $\mathring{A}$ are totally bounded, then $\mathring{A}_\epsilon$ has at each instant a finite number n_τ of states. Let

(9.1.2) $$\bar{K}_\epsilon(X) = \limsup_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \log \mathring{n}_{t+k}.$$

THEOREM 9.2. *Under the conditions of theorem* 9.1, *if the sets of states of* $\mathring{A}$ *are totally bounded and* $\bar{K}_\epsilon(X) < \infty$, *then concerning the possibility of the choice of a code such that the transmission rate in the system* $[\mathring{A}_\epsilon, \Delta_\epsilon]$ *is as close to* $H(\mathring{A}_\epsilon) = h(\mathring{A}) + K_\epsilon(\mathring{X}) + o(1)$ *as one wishes, the second basic theorem of Shannon type is true* ([14], *p.* 244).

The proofs of these two theorems follow from the fact that $\mathring{A}_\epsilon$, Δ_ϵ verifies the conditions of the basic theorems in ([14], pp. 243–244).

REFERENCES

[1] S. Banach, *Théorie des Opérations Linéaires*, Warszawa, 1932.
[2] L. Breiman, "The individual ergodic theorem of information theory," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 809–811.
[3] K. L. Chung, "A note on the ergodic theorem on information theory," *Ann. Math. Statist.*, Vol. 32 (1961), pp. 612–614.
[4] R. L. Dobrushin, "The central limit theorem for non-homogeneous Markov chains," *Theor. Probability Appl.*, Vol. 1 (1956), pp. 72–89 and 365–425.
[5] A. I. Khintchin, "On the basic theorems of information theory," *Uspehi Mat. Nauk*, Vol. 11 (1956), pp. 17–75.
[6] A. N. Kolmogorov and V. M. Tihomirov, "The ϵ-entropy and the ϵ-content of sets in the functional spaces," *Uspehi Mat. Nauk*, Vol. 14 (1959), pp. 3–86.
[7] A. N. Kolmogorov, "On some asymptotic characteristics of the totally bounded metric spaces," *Dokl. Akad. Nauk SSSR*, Vol. 108 (1956), pp. 385–388.
[8] M. Rosenblatt-Roth, "The concept of entropy in the probability theory and its applications in the theory of transmission through channels," *Transactions of the Third All-union Mathematical Congress*, Moscow, Vol. 2 (1956), pp. 132–133.
[9] ———, "The concept of entropy in the probability theory and its applications in the theory of transmission through channels," thesis, Moscow State University, 1956.
[10] ———, "The entropy of stochastic processes," *Dokl. Akad. Nauk SSSR*, Vol. 112 (1957), pp. 16–19.
[11] ———, "The theory of transmission of information through stochastic channels," *Dokl. Akad. Nauk SSSR*, Vol. 112 (1957), pp. 202–205.
[12] ———, "Normed ϵ-entropy and the transmission of the information of continuous sources through continuous channels," *Dokl. Akad. Nauk SSSR*, Vol. 130 (1960), pp. 265–268.
[13] ———, "Normed ϵ-entropy of sets and the theory of transmission of information," *Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Prague, 1960, pp. 569–577.
[14] ———, "The concept of entropy in the probability theory and its applications in the theory of transmission by channels," *Theor. Probability Appl.*, Vol. 9 (1964), pp. 238–261.
[15] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, Vol. 27 (1948), pp. 378–423 and 623–656.
[16] A. G. Vitushkin, "The absolute entropy of the metric spaces," *Dokl. Akad. Nauk SSSR*, Vol. 117 (1957), pp. 745–748.