UDC 519.21

MARKOV PROCESSES OVER DENUMERABLE PRODUCTS OF SPACES, DESCRIBING LARGE SYSTEMS OF AUTOMATA*

L. N. Vasershtein

Problemy Peredachi Informatsii, Vol. 5, No. 3, pp. 64–72, 1969

A criterion is obtained for the uniqueness of the stationary probabilities of a Markov operator. This criterion is useful for operators related to homogeneous automata games. Similar methods are applicable to certain problems in statistical physics, biology, etc.

§1. Let a system of a denumerable number of automata function in discrete time according to the following probabilistic law: the state of the i-th automaton at time t + 1 depends stochastically on the states of some finite set of its "neighbors" at time t, and for given states of all automata at time t their states at time t + 1 are independent. In other words, we are considering a Markov process over the set $X = \Pi X_i$, where X_i is the set of states of the i-th automaton.

Markov processes of this kind arise in the study of systems with local interactions, consisting of a "large number" of identical elements. Similar problems can be found in certain models of statistical physics (see e. g., [1]), in the description of neural networks, and so on.

We are basically interested in the uniqueness of the stationary probabilities (i. e., in a probability measure, or, in another terminology, in the probability distributions), and in the convergence of any initial probabilities to the stationary ones. We are also interested in the continuous dependence of the stationary probabilities (when they exist and are unique) on the transition function.

§2. Let the denumerable space X be a product of a denumerable number of measurable (e.g., finite discrete) spaces $X_i : X = \prod_{i \in N} X_i$, N be denumerable. We denote by C(X) the space of all bounded functions over X which can be uniformly approximated (with any degree of accuracy) by measurable functions, depending only on a finite number of coordinates. Let P be transition function over X mapping the space C(X) into itself. Below we give a criterion (see Theorem 2), which guarantees the uniform convergence $P^t f \to$ const for any $f \in C(X)$. Such convergence immediately implies the uniqueness of the stationary probabilities (i. e., of the probabilistic measure over X) for P. The criterion for a more general situation will be discussed in §4.

We remark that in this problem the transition probabilities $P^t(x, \cdot)$ and $P^t(y, \cdot)$ are usually mutually singular for any t (this is shown by the trivial example in which the transition function P is a product of independent transition functions over X_i), therefore the classical criterion of ergodicity is not applicable here.

Notation (to be used everywhere except in §4). The measurable subsets a, b, $c \subset X$; the points x, $y \in X$; I, $S \subset N$, where S is finite; i, $k \in N$; p and q are probabilities over X; x_I (corresponding to p_I) is the projection of x (or of p) over $X_I = \prod_{i \in I} X_i$.

The measurable $a \subset X_I$ are identified by their images in the canonical projection of X on X_I. The measurable $a \subset X_S$ (S is arbitrary and finite) are called elementary. Among the other subsets of the set X the elementary ones are distinguished by the fact that their indicators (in an older terminology, the characteristic functions) lie in C(X). Any function from C(X) is uniformly approximated by a finite linear combination of the indicators of elementary sets. Therefore the condition $PC(X) \subset C(X)$ imposed on the transition function P is equivalent to the following: the function $P(\cdot, a) \in C(X)$ for any elementary a. Any probability p over X is determined by its values over the elementary sets, i. e., by their "finite-dimensional" projections p_S.

Theorem 1. Let all X_i be finite. Then there is a stationary probability p for P. The uniqueness of the stationary p is equivalent to the convergence $1/n \sum_{t=1}^{n} qP^t \to p$ for any probability q. (We consider over the space of all probabilities over X the topology of the convergence over each elementary set.)

*The author's attention was drawn to this problem by Professor I. I. Pyatetskii-Shapiro.

Proof. Let q be any probability over X and $q^{(n)} = \frac{1}{n}\sum^{n} qP^t$. From the sequence of probabilities $q^{(n)}$ we can select a subsequence converging to some probability p, and then $p^P = p$. If the stationary p is unique, then $q^{(n)} \to p$ for any q, which is what we wanted to prove.

We identify in some fashion the denumerable set N with the set of natural numbers and denote by $\rho k(p, q)$ the smallest α such that if $a \subset X_k$ and $b,\ c \subset X_{[1, k-1]} = \prod_{i=1}^{k-1} X_i$, then $|p(a \cap b)q(c) - q(a \cap c)p(b)| \leq \alpha p(b)q(c)$. We denote by $\rho_{i,k}(P)$ the number $\sup_{x_{N\setminus i} = ''_{N\setminus i}} \rho_k(P(x, \cdot),\ P(y, \cdot))$, characterizing the "dependence of the k-th coordinate on the i-th."

Theorem 2. Let $A = (\rho_{i,k}(P))_{i,k\in N}$ and $u\ \sum_{i\in N}(A^t)_{i,k} \to 0$, then $t \to \infty$ for any $k \in N$, where A^t is the degree of the matrix A. Then $|P^t(x, a) - P^t(y, a)| \leqslant \sum_{k\in S}\sum_{i\in N}(A^t)_{i,k} \to 0$ for any $a \subset X_S$. Consequently, $P^t f \to$ const for any function $f \in C(X)$ and $\mathrm{var}((P^t(x, \cdot))_S - p_S) \to 0$ uniformly with respect to $x \in X$ for any stationary probability p. In particular, the stationary probability (if it exists) is unique.

The proof is given in §5.

Corollary 1. Let $\sum_{i\in N} \rho_{i,k}(P) \leqslant \theta$ for some $\theta < 1$. Then $\mathrm{var}((P(x, \cdot))_S - (P(y, \cdot))_S) \leq \theta^t |S| \to 0$ for S, where $|S|$ is the power of S.

Note. Theorem 2 and Corollary 1 can obviously be generalized, assuming $A = (\rho_{i,k}(P^{t_0}))_{i,k\in N}$.

Theorem 3. Let $P_{(\lambda)}$ be some network (e.g., a sequence) of transition functions over X and $P_{(\lambda)} \to P$ (over the set of transition functions we consider the topology of the uniform convergence in $x \in X$ over each elementary set). Then

a) if $P^t \to p$,* then $P^t_{(\lambda)} \to p$ in the sense of a double limit, and consequently $p^{(\lambda)} \to p$ for any probabilities $p^{(\lambda)}$, stationary for $P_{(\lambda)}$;

b) if $p^{(\lambda)} \to p$, where $p^{(\lambda)}P_{(\lambda)} = p^{(\lambda)}$, then $pP = p$;

c) if all X_i are finite and the stationary probability p for P is unique, then $p^{(\lambda)} \to p$ for any $p^{(\lambda)}$, stationary with respect to $P_{(\lambda)}$.

Proof. a) Let $\varepsilon > 0$ and a be any elementary set. We find t_0 such that $(P^t(x, a) - p(a)) \leq \varepsilon/2$ when $t \geq t_0$, $x \in X$. Then we find λ_0 such that $|P^0_{(t\lambda)}(x, a) - p^{t_0}(x, a)| \leq \varepsilon/2$ when $\lambda > \lambda_0$, $x \in X$. Then $|p^{t_0}_{(\lambda)}(x, a) - p(a)| \leq \varepsilon$ when $\lambda > \lambda_0$ $x \in X$, whence $|P^t_{(\lambda)}(x, a) - p(a)| \leq \varepsilon$ when $(\lambda, t) > (\lambda_0, t_0)$.

The proof of b) and c) is even simpler.

§3. In problems concerning homogeneous games of automata (see [2-4]) X_i is the finite set of states of the i-th automaton, and normally the following conditions are satisfied:

1) for $a \subset X_S$ the function $P(\cdot, a)$ depends only on the finite set $\Gamma(S)$ of coordinates ("local finiteness");

2) $P(x, \cdot) = \prod_{i\in N} (P(x, \cdot))_i$ for any points $x \in X$ ("independence of transitions");

3) over N, X, and P some group operates jointly; over N transitively, for example 3', an N — n dimensional mesh, $X_i = \{0, 1\}$ and P is invariant with respect to all parallel translations (homogeneity).

It follows from condition 1) that $PC(X) \subset C(X)$. From condition 2) we have $\rho_{i,k}(P) = \sup_{x_{N\setminus i} = y_{N\setminus i}} \sup_{a\subset X_k}(P(x, a) - P(y, a))$. When 2) and 3) are satisfied, the condition of Theorem 2 is equivalent to $\sum_{i\in N} \rho_{i,k}(P) < 1$. In the following examples we shall assume that condition 2) is satisfied, and $X_i = \{0, 1\}$ (in addition we shall assume that 1) is satisfied and, except for example 1, also condition 3)). Under these assumptions from Theorems 1-3 we obtain

*We are considering the probability p as a transition function which is independent of x.

$p^t \to p$, where p is the unique stationary probability for P, which is a continuous function of P. If $\sum_{i\in N} A_{i,k} \leqslant \theta < 1$, then $|p^t(x, a) - p(a)| \leq |S|\theta^t$ for any $a \subset X_S$.

Example 1. Let N be the set of natural numbers, $\Gamma(i) = \{i + 1\}$, $P(x, 0_k) = \alpha_k + (\beta_k - \alpha_k)x_{k+1}$, $(0 \leq \alpha_i \beta_i \leq 1)$. Then according to Corollary 2, $p^t \to p$, where p is the unique stationary probability, if $\prod_{i=k}^{k+t} |\beta_i - \alpha_i| \to 0$ for any k, i.e., the product $\Pi|\beta_i - \alpha_i|$ converges to zero. It is easy to verify that the stationary probability is not unique if this condition is not satisfied. We remark that a time-inhomogeneous Markov process has been examined for a set of two points.

Example 2. Let $\Gamma(i)$ consist of three elements for any $i \in N$, and the set $\Gamma^t(i)$ not intersect with $\Gamma^t(k)$ when $i \neq k$ (the graph (N, Γ) is a tree; Γ^t denotes the degree of the graph). Let us assume $P(x, 0_k) = \frac{1}{2} + \frac{\theta}{2}\operatorname{sign}\left(\frac{3}{2} - \sum_{i\in\Gamma(k)} x_i\right)$, $0 \leq \theta \leq 1$. It is obvious that the probability p, determined by the equations $p(\bigcap_{i\in S} 0_i) = (1/2)^{|S|}$ (a Bernoulli scheme), is stationary. According to Corollary 2 we have $|P^t(x, a) - p(a)| \leq |S|(3\theta)^t$, $(a \subset X_S)$. A more detailed analysis of this simple example shows that $|P^t(x, a) - p(a)| \leq |S|((3/2)\theta)^t$, and in the case $\theta = (2/3)$ $|P^t(x, a) - p(a)| \leq |S|/(t + 1)^{1/2}$. When $\theta > 2/3$ there are other stationary probabilities apart from p.

In the following examples condition 3_1) is assumed to be satisfied, $|\theta| \leq 1$, $a \subset X_S$, $1_i = X \setminus 0_i = \{x \in X | x_i = 1\}$.

Example 3. Let $P(x, 0_k) = 1/2 + \theta(-1)^{x_i + x_{i+1}}/2$. It is easy to verify that the probability p of Example 2 (the Bernouilli scheme) is stationary. When $\theta = 1$, there is another stationary probability concentrated at the point $x_i \equiv 0$. According to Corollary 2, $p^t \to p$ when $|\theta| < 1/2$. We shall show that indeed $|P^t(x, a) - p(a)| \leq C_1|\theta|^{C_2 + \log_2 3} \to 0$ when $|\theta| < 1$, where C_1 and C_2 are positive numbers depending on S. We define the function f_S over X by the equations $f_S(x) = (-1)^{\sum_{i\in S} x_i}$. It is easy to verify that $Pf_S = \theta^{|S|} f_S f_{S+1} = f_{S'}$, where S' is the symmetric difference of the sets S and S + 1. Iterating, we obtain $P^t f_S = \theta^{|S| + |S'| + \ldots + |S^{(t)}|} f_{S_{(t)}}$. We notice that $|S| + |S'| + \ldots + |S^{(t)}| \geq C_2 t^{\log_2 3}$, where $C_2 > 0$, and that the finite linear combinations of the functions f_S are everywhere dense in C(X).

Example 4. (see [4]). Let $P(x, 1_k) = \theta + (1 - \theta)x_k x_{k+1} (\theta > 0)$. It is obvious that the probability p concentrated at the point $x_i \equiv 1$ is stationary. According to the corollary, $2P^t \to p$ when $\theta > 1/2$. A more accurate estimate shows that this is true when $\theta \geq 1/3$. On the other hand, in [5] it is shown that for sufficiently small θ there exists another stationary probability.

Example 5. Let $P(x, 1_k) = 1/2 + (\theta/2)x_{k-1}x_k x_{k+1}$. Then Corollary 2 gives ergodicity when $|\theta| < 2/3$. We show that indeed for any $|\theta| < 1$ (uniformly with respect to $x \in X$) $P^t f \to$ const for any $f \in C(X)$; consequently there exists a unique stationary probability, continuously dependent according to Theorem 3 on the parameter θ. This follows from the following fact which can be proved by induction on t: $P^t(x, \bigcap_{i\in S} 1_i) = \sum_{T\subset\Gamma^t(S)} C(t, T, S) \prod_{i\in T} x_i$, where $\sum_{T=\varnothing} |C(t, T, S)| \leqslant$ $\leqslant \left(\frac{1}{2} + \frac{|\theta|}{2}\right)^{t+|S|} \to 0$ when $t \to \infty$.

(The finite linear combinations of the indicators $\prod_{i\in S} x_i$ of the sets $\bigcap_{i\in S} 1_i$ are everywhere dense in C(X), where S runs through all finite subsets N.) Here we used the equation $P\left(\prod_{i\in S} x_i\right) = \prod_{i\in S} Px_i$, which is valid in view of property 2).

Example 6. Let $P(x, 1_i) = \theta + (1 - 2\theta)x_i(1 - x_{i+1})$. Then, as in Example 5, it can be shown that $P^t f \to$ const for any $f \in C(X)$ if only $|\theta| < 1$. If $\theta = 1$ or $\theta = -1$, then there are many different stationary probabilities. This example is interesting, because it enables us to evaluate the stationary probabilities of certain sets, e.g., $p(0_i) = 3/4 - 1/2(\theta)$, $p(0_i \cap 0_{i+1}) = 1/2(1 - \theta)$, $p(0_i \cap 1_{i+1}) = 1/4$.

We remark that in Examples 5 and 6 an analytical relationship is obtained between the stationary probabilities p and the parameter θ over the whole segment $|\theta| < 1$.

§4. In order to clarify the proof of Theorem 2, we discuss the idea behind it in a more abstract form.

Let there exist over some measurable space X a uniform structure T, i.e., a system of subsets $\tilde{a} \subset X \times X$ containing the diagonal $X \subset X \times X$ and also satisfying some axioms (see [6]). Let P be the transition function over X which, transforms into itself the Banach space C(X) of all uniformly continuous bounded measurable functions (with the norm $\sup_{x\in X}|f(x)|$).

Theorem A. Let over $\tilde{X} = X \times X$ exist such a transition function V that, (a) $V^t((x,y),\tilde{a}) \to 1$, uniformly with respect to $(x,y) \in \tilde{X}$ for any measurable neighborhood of the diagonal $\tilde{a} \in T$; (b) the projections of the probabilities $V((x,y),\cdot)$ are equal to $P(x,\cdot)$ and $P(y,\cdot)$ for any point $(x,y) \in \tilde{X}$. Then $P^t f \to$ const for any $f \in C(X)$.

Proof. It is obvious that $(P^t f)(x) - (P^t f)(y) = (V^t \tilde{f})(x,y) \to 0$, where $\tilde{f}(x,y) = f(x) - f(y)$, as required.

If the uniform topology over X is given by the measurable distance r, then condition (a) will be satisfied if $V^t r \to 0$ uniformly over $\tilde{X}$. We now assume that the distance R is bounded. We determine the distance between the probabilities over X using the formula $r(p,q) = \inf \int_X r(\tilde{x})v(d\tilde{x})$, where inf is taken with respect to all probabilities v over $\tilde{X}$ with the projections p and q.

Theorem B. Let $|P^{t_0}|_r < 1$ for some t_0, where $|P|_r = \sup_{x\neq y}\frac{r(P(x,\cdot),P(y,\cdot))}{r(x,y)}$. Then $P^t f \to$ const for any $f \in C(X)$.

Note. If the space X is finite, then Theorem B obviously follows from Theorem A. Technical difficulties arise in the general case, when attempting to derive B from A; these are related to the requirement that the function $V(\cdot,\tilde{a})$ be measurable for any measurable $\tilde{a}$. Therefore we give the direct proof of B.

Proof. Let $f \in C(X)$. For any $\varepsilon > 0$ there exists a $\delta > 0$ such that $|f(x) - f(y)| \leq \varepsilon + (1/\delta)r(x,y)$ for all $x, y \in X$. It is obvious that $(P^{t_0}f)(x) - (P^{t_0}f)(y) = \int_X (f(x) - f(y))v(d\tilde{x})$, where $\tilde{x} = (x,y)$ and v is any probability over $\tilde{X}$ with the projections $P^{t_0}(x,\cdot)$ and $P^{t_0}(y,\cdot)$. Since $\inf\int_{\tilde{X}} r(\tilde{x})v(d\tilde{x})$ is according to the condition, not greater than θ for such v. therefore we obtain $(P^{t_0}f)(x) - (P^{t_0}f)(y) \leq \varepsilon + (\theta/\delta)r(x,y)$. Iterating, we obtain $(P^{t_0t}f)(x) - (P^{t_0t}f)(y) \leq \varepsilon + (\theta^t/\delta)r(x,y)$. Consequently, $\varlimsup_{t\to\infty}\sup_{x,y\in X}|(P^t f)(x) - (P^t f)(y)| \leqslant \varepsilon$. Since ε is arbitrary and positive, we obtain the assertion of Theorem B.

As an example, we derive from Theorem A (we could also use Theorem B) the following well-known statement: let $|P(x,a) - P(y,a)| \leq \theta < 1$ for any $x, y \in X$ and a measurable $a \subset X$, then $P^t f \to$ const for any bounded measurable function f over X. For this it is sufficient to consider over X a discrete topology, and to put in Theorem A $V((x,y),\cdot) = P(x,\cdot) * P(y,\cdot)$, where the operation * is defined as follows (this operation is also used in §5 in proving Lemma 1).

The operation *. Let p and q be two probabilities over some measurable space Y. We determine the probability $p * q$ over $Y \times Y$ with the projections p and q such that $(p * q)(\tilde{a}) \geq 1 - \mathrm{var}(p - q)/2$ for any measurable $\tilde{a}$ containing a diagonal $\left(\text{where } \frac{1}{2}\mathrm{var}(p-q) = \sup_a (p(a) - q(a))\right)$. Let $p - q = (p-q)^+ - (p-q)^-$ be a Jordan expansion. We determine the measure $p \circ q$ over $Y \times Y$ by the formula $(p \circ q)(\tilde{a}) = (p - (p-q)^+)(a)$, where $a = \{y \in Y | (y,y) \in \tilde{a}\}$. We put $p*q = p\circ q + \frac{(p-q)^+ \times (p-q)^-}{C}$, where $C = (p-q)^+(Y) = (p-q)^-(Y) = (1/2)(\mathrm{var}(p-q))$ (if $C = 0$, then $p * q = p \circ q$).

§5. Proof of Theorem 2. For simplicity we assume that all measurable spaces X_i are finite and discrete (this assumption is only used in the proof of Lemma 1).

Definition. We say that the probabilities p and q over X satisfy the estimate α_k if over $\tilde{X} = X \times X$ there is a probability v with the projections p, q and such that $v(\tilde{\Delta}_k) \leq \alpha_k$ for all $k \in N$, where $\tilde{\Delta}_i = \{x_i \neq y_i\} \subset X \times X$.

Lemma 1. Let $\rho_k(p,q) \leq \alpha_k$ for all $k \in N$, then the probabilities p and q satisfy the estimate α_k.

Proof. If $p = \prod_{k\in N} p_k$ and $q = \prod_{k\in N} q_k$, then we can put $v = \prod_{k\in N} v_k$, where $v_k = p_k * q_k$ (see §4 concerning the operation *). In the general case we denote by $p^{x[1,n-1]}(a)$, $a \subset X_n$ the conditional probability of Doob (see [7]) for the probability p. (For $a \subset X_n$ the measurable function $p^{x[1,n-1]}(a)$ over $X_{[1,n-1]}$ is defined over the set of the complete p-measure by the equations $\int_b p^{x_{[1,n-1]}}(a)\,p(dx_{[1,n-1]}) = p(a \cap b)$, where $b \subset X_{[1,n-1]}$. For almost all x (in the sense p) $p^{x[1,n-1]}$ is a probability over X_n).

The probability v will be defined if we construct a set of probabilities $v_{[1,n]}$ over $\tilde{X}_{[1,n]} = \prod_{i=1}^{n} X_i \times X_i$ such that the projection $v_{[1,n+1]}$ over $X_{[1,n]}$ coincides with $v_{[1,n]}$.

We put $v_1 = p_1 * q_1$. Let $v_{[1,n]}$ be already constructed. We put $v^{\tilde{x}[1,n-1]} = p^{x[1,n]} * q^{y[1,n]}$, where $\tilde{x} = (x, y)$. We determine the probability $v_{[1,n+1]}$ over the sets $\tilde{a} \cap \tilde{b}$, where $\tilde{a} \subset \tilde{X}_{n+1}$, and $\tilde{b} \subset \tilde{X}_{[1,n]}$ by the formula $\tilde{v}_{[1,n+1]}(\tilde{a} \cap \tilde{b}) = \int_b v^{\tilde{x}[1,n]}(\tilde{a})\, v_{[1,n]}(d\tilde{x}_{[1,n]})$.

Lemma 2. Let the probabilities p, q satisfy the estimate α_k and $f \in C(X)$. Then $\int_X f(x)p(dx) - \int_X f(x)q(dx) \leqslant \sum_{k\in N} \alpha_k \rho_k(f)$, where $\rho_k(f) = \sup_{x_{N\setminus k} = y_{N\setminus k}} |f(x) - f(y)|$.

Proof. Let f depend only on a finite set T of coordinates (i.e., $\rho_i(f) = 0$ when $i \notin T$). We put $\tilde{f}(x, y) = f(x) - f(y)$, $a(S) = \prod_{i\in S} \bar{\Delta}_i \bigcap_{i\in T\setminus S} (\tilde{X} \setminus \bar{\Delta}_i)$. Then the sets $a(S)$, $S \subset T$ form a finite partitioning of the space $\tilde{X}$, and $\int_X f(x) \times (p - q)(dx) = \int_{\tilde{X}} \tilde{f}(\tilde{x})\, v(d\tilde{x}) = \sum_{S\subset T} \int_{a(S)} \tilde{f}(\tilde{x})\, v(d\tilde{x}) \leqslant \sum_{S\subset T} v(a(S)) \sup_{\tilde{x}\in a(S)} |\tilde{f}(\tilde{x})| \leqslant \sum_{S\subset T} v(a(S)) \sum_{k\in S} \rho_k(f) = \sum_{k\in T} \rho_k(f)\, v(\bar{\Delta}_k) \leqslant \sum_{k\in T} \alpha_k \rho_k(f) = \sum_{k\in N} \alpha_k \rho_k(f)$.

Let us now consider the general case. We fix the point $y \in X$ and determine the function f_S over X by the equations $f_S(x) = f(x')$, where $x'_S = x_S$, and $x'_{N\setminus S} = y_{N\setminus S}$. It is obvious that $\rho_k(f_S) \leq \rho_k(f)$ and $f_S \to f$ when $S \to N$. By going to the limit when $S \to N$ in the equation $\int_X f_S(x)(p - q)(dx) \leqslant \sum_{k\in N} \alpha_k \rho_k(f_S)$ (already proved), we obtain the assertion of the lemma.

Corollary of Lemma 2. If $x_{N\setminus i} = y_{N\setminus i}$, then let the probabilities $P(x, \cdot)$, $P(y, \cdot)$ satisfy the estimate $A_{i,k}$. Then $\rho(Pf) \leq A\rho(f)$ for any $f \in C(X)$, where $\rho(f)$ is a column vector $(\rho_k(f))_{k\in N}$, and the inequality is taken component by component.

Theorem 2'. Under the conditions of the previous corollary let $\sum_{i\in N} (A^t)_{i,k} \to 0$ when $t \to \infty$ for any $k \in N$, where A^t is the degree of the matrix A. Then the assertion of Theorem 2 is satisfied.

Proof. Iterating the inequality $\rho(Pf) \leq A\rho(f)$, we obtain $\rho(A^t f) \leq A^t \rho(f)$, whence $\sum_{k\in N} \rho_k(P^t f) \leqslant \sum_{k\in N} \sum_{i\in N} (A^t)_{i,k} \rho_k(f)$. Applying the last inequality to the case when f is an indicator of an elementary set a (then $P^t f = P^t(\cdot, a)$), we obtain the assertion of the theorem.

Theorem 2 (for finite X_i) obviously follows from Lemma 1 and Theorem 2'. In the general case, Lemma 2 can be directly proved under the conditions of Lemma 1.

Finally we remark that, by somewhat weakening Theorem 2 by replacing the condition $b, c \subset X_{[1,k-1]}$ by the condition $b, c \subset X_{N\setminus k}$ in the definition $\rho_k(p, q)$, we can obtain that the conditions of Theorem 2 will not depend on the choice of a numbering of the sets N, as happens for the conditions of Theorem 2.

§6. We shall now briefly consider the relationship of our work to [1]. In [1] the following problem is considered. Let a measurable space $X = \prod_{i\in N} X_i$ be the product of a denumerable number of finite discrete spaces, and for any $k \in N$ and $a \subset X_k$ over the measurable space $X_{N\setminus k}$ be given the function $p^{\cdot}(a) \in C(X_{N\setminus k})$. The question is whether the probability p over X for which the functions $p^{\cdot}(a)$ are conditional probabilities, i.e., $\int_b p^{x_{N\setminus k}}(a)\, p(dx_{N\setminus k}) = p(a \cap b)$ for any $k \in N$, $a \subset X_k$, of a measurable $b \subset X_{N\setminus k}$ is unique. Among the results of [1] are the following.

Theorem (Dobrushin). We put $\rho_{i,k} = \sup_{x_{N\setminus i} = y_{N\setminus i}} \sup_{a\subset X_k} (p^{x_{N\setminus k}}(a) - p_{y_{N\setminus k}}(a))$ ($\rho_{i,i} = 0$ for all $i \in N$).

Let $\sum_{i\in N} \rho_{i,k} \leqslant \theta$ for all $k \in N$ for a certain $\theta < 1$. Then the probability p over X with given conditional probabilities $p'(a)$, is unique.

It is obvious that the uniqueness of the stationary probabilities for the transition function P over X is equivalent to the uniqueness of the probability, invariant with respect to time shifts, over the space of trajectories X × Z (Z is the group of integers) with given conditional probabilities. However the direct application of Dobrushin's theorem to this situation gives a weaker estimate than Theorem 2.

Conversely, from given conditional probabilities it is possible (by several methods) to construct a transition function P over X, such that pP = p for any probability p, with given conditional probabilities. We construct such an operator P and prove Dobrushin's theorem (the proof will be very similar to that in [1]). Following [1], we denote by P_i the transition function over X defined by the formula $(pP_i)(a \cap b) = \int_b p^{x_{N\setminus i}}(a)\, p(dx_{N\setminus i})$ for any $a \subset X_i,\ b \subset X_{N/i}$. We want to prove the uniqueness of a probability p such that $p^{P_i} = p$ for all $i \in N$.

We identify in some way the set N with the set of natural numbers and put $P = \prod_{i=1}^{\infty} P_i$. This operator can be meaningfully considered over the space of probabilities over X, since $(pP_i)_{[1,n]} = p_{[1,n]}$ when $i > n$. Over C(X) the operator P acts according to the formula $Pf = \lim_{n\to\infty}\left(\prod_{i=1}^{n} P_i\right) f$ where the limit exists, since $P_i C(X) \subset C(X)$ and $P_i f = f$ when $i > n$ for functions f depending only on the first n coordinates.

It is easy to verify that if $x_{N\setminus i} = y_{N\setminus i}$, then the probabilities $P_n(x, \cdot)$ and $P_n(y, \cdot)$ satisfy the estimate $A^{(n)}_{i,k}$, where $A^{(n)}_{k,k} = 1$ and $A^{(n)}_{k,n} = \rho_{k,n}$ for the case $k \neq n$, and in the remaining cases $A^{(n)}_{i,k} = 0$. Applying the corollary from Lemma 2 we obtain $\rho\left(\prod_{i=1}^{n} P_i\right) f \leqslant \left(\prod_{i=1}^{n} A^{(i)}\right) \rho(f)$, whence $\rho(Pf) \leqslant \left(\prod_{i=1}^{\infty} A^{(i)}\right) \rho(f)$, where the limit $A = \prod_{i=1}^{\infty} A^{(i)} = \lim_{n\to\infty} \prod_{i=1}^{n} A^{(i)}$ exists, since in the matrix $A^{(i)}$ all the columns except the i-th are the same as in a unit matrix.

It is immediately verified that $\sum_{k \in N} A_{i,k} \leqslant \sum_{k \in N} \rho_{i,k} \leqslant \theta$, whence, just as in the proof of Theorem 2, we obtain $P^t(\cdot, a) \to$ const for any elementary a. Consequently a probability p, such that pP = p, is unique, which is what we wanted to prove.

REFERENCES

1. R. L. Dobrushin, "Description of a stochastic field by conditional probabilities, and conditions for its regularity," Teoriya veroyatn. i ee primen., 13, 2, 201–229, 1968.
2. M. L. Tsetlin, "Finite automata and the simulation of the simplest forms of behavior," Usepkhi matem. nauk, 18, 4, 3–28, 1953.
3. I. M. Gel'fand and I. I. Pyatetskii-Shapiro, "On homogeneous games and automaton games," Dokl. AN SSSR, 152, 4, 1963.
4. O. N. Stavskaya and I. I. Pyatetskii-Shapiro, "On homogeneous networks of spontaneously active elements," Problemy kibernetiki, Nauka, Moscow, 20, 91–106, 1968.
5. M. G. Shnirman, "On the ergodicity of a certain Markov chain," Problemy kibernetiki, Nauka, Moscow, 20, 115–122, 1968.
6. N. Bourbaki, General Topology, Basic Structures [Russian translation], Fizmatgiz, Moscow, 1956.
7. P. Halmos, Measure Theory [Russian translation], Izd. inostr. lit., Moscow, 1953.

24 February 1969