SEQUENTIAL METHODS IN STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 16
PWN–POLISH SCIENTIFIC PUBLISHERS
WARSAW 1985

NOTES ON DEPENDENCE WITH COMPLETE CONNECTIONS

MARIUS IOSIFESCU

National Institute of Metrology, Centre of Mathematical Statistics, Bucharest, Romania

These notes are intended to present the fundamentals of the theory of dependence with complete connections. This type of stochastic dependence has been introduced and developed mainly by Romanian probabilists.

Various examples involving dependence with complete connections are discussed at length. One of them is constituted by stochastic approximation procedures.

1

We start with a special case that motivates the general definition we shall give later.

Let us consider an initial urn U_0 containing a_j balls of colour j, $1 \leqslant j \leqslant m+1$, and denote by $a_j^{(n)}$, $1 \leqslant j \leqslant m+1$, the structure of the urn U_n, $n \geqslant 0$, $a_j^{(0)} = a_j$, $1 \leqslant j \leqslant m+1$, given by the following rule: if the structure of the urn U_{n-1} was $a_j^{(n-1)}$, $1 \leqslant j \leqslant m+1$, and on trial n (which is a drawing from U_{n-1}) a ball of colour i was drawn, then the structure of the urn U_n is $a_j^{(n)} = a_j^{(n-1)} + \delta_{ij} d_j$, where d_j are nonnegative integral numbers, $1 \leqslant j \leqslant m+1$. This amounts to the fact that if on trial n a ball of colour i was drawn, then this ball is replaced while d_i balls of colour i are added to the urn.

Remark. The above urn scheme was devised by the Romanian mathematicians Onicescu and Mihoc in the middle thirties. The special case $m = 2$, $d_1 + d_2 = d$ is known as the Pólya urn (and in turn as the Markov urn for $d = 1$).

Define

$$X_n = \text{the colour of the ball drawn on trial } n,$$

and set

$$p_j^{(n)} = \mathrm{P}(X_{n+1} = j \mid X_1, \ldots, X_n), \qquad n \geqslant 1,$$

$$p_j^{(0)} = \mathrm{P}(X_1 = j) = \frac{a_j}{M}, \qquad 1 \leqslant j \leqslant m+1,$$

where $M = \sum_{j=1}^{m+1} a_j$. (Please note the assumption we make here.)

We want to find the relationship between $\boldsymbol{p}_n = (p_j^{(n)})_{1 \leqslant j \leqslant m+1}$ and $\boldsymbol{p}_{n-1} = (p_j^{(n-1)})_{1 \leqslant j \leqslant m+1}$ conditional on $X_n = i$. Clearly,

$$p_j^{(n-1)} = \frac{a_j^{(n-1)}}{M^{(n-1)}}, \qquad p_j^{(n)} = \frac{a_j^{(n-1)} + \delta_{ij} d_j}{M^{(n-1)} + d_i}, \qquad 1 \leqslant j \leqslant m+1, \tag{1}$$

where $M^{(n-1)} = \sum_{j=1}^{m+1} a_j^{(n-1)} (M^{(0)} = M)$. Now, denoting by $v_j^{(n)}$ the number of balls of colour j that occurred on the first n drawings ($v_j^{(0)} = 0$), we can write

$$a_j^{(n)} = a_j + v_j^{(n)} d_j, \qquad 1 \leqslant j \leqslant m+1, \; n \geqslant 0. \tag{2}$$

Assuming all the $d_j > 0$ (please discuss the general case), the first equation of (1), equation (2) and the obvious equation $\sum_{j=1}^{m+1} v_j^{(n)} = n$ yield

$$M^{(n-1)} = \frac{\sum_{k=1}^{m+1} a_k/d_k + n - 1}{\sum_{k=1}^{m+1} p_k^{(n-1)}/d_k}, \qquad a_j^{(n-1)} = M^{(n-1)} p_j^{(n-1)};$$

hence, by means of the second equation of (1), the probability $p_j^{(n)}$ is a function $f_{ij}^{(n)}$ of $\boldsymbol{p}_{n-1} = (p_j^{(n-1)})_{1 \leqslant j \leqslant m+1}$:

$$p_j^{(n)} = f_{ij}^{(n)}(p_1^{(n-1)}, \ldots, p_{m+1}^{(n-1)}), \qquad 1 \leqslant i, j \leqslant m+1.$$

Therefore the evolution of the probabilistic structure of the urn scheme we have considered can be described as follows. Starting with a probability vector $\boldsymbol{p}_0 = (p_1^{(0)}, \ldots, p_{m+1}^{(0)})$, we select a colour $X_1 = i$ according to $\boldsymbol{p}_0$, and then construct the new probability vector

$$\boldsymbol{p}_1 = \left(f_{ij}^{(1)}(p_1^{(0)}, \ldots, p_{m+1}^{(0)})\right)_{1 \leqslant j \leqslant m+1}.$$

Next, a colour $X_2 = k$ is selected according to $\boldsymbol{p}_1$, then the new probability vector

$$\boldsymbol{p}_2 = \left(f_{kj}^{(2)}(p_1^{(1)}, \ldots, p_{m+1}^{(1)})\right)_{1 \leqslant j \leqslant m+1}$$

is constructed, and so on.

Onicescu and Mihoc (1935) say that the random variables $X_1, X_2, \ldots$ (the colours successively observed) are connected into a *chain with complete connections*.

2

The special case considered above motivates the following general framework (Iosifescu (1963)): Assume that we are given two measurable spaces $(W, \mathscr{W})$ and $(X, \mathscr{X})$ and, for any $n = 1, 2, \ldots$, a mapping $u^{(n)}$ of $W \times X$ into W which is measurable with respect to $\mathscr{W} \times \mathscr{X}$ and $\mathscr{W}$, and a probability kernel $P^{(n)}$ on $W \times \mathscr{X}$ (i.e., $P^{(n)}(w, \cdot)$ is a probability measure on $\mathscr{X}$ for any $w \in W$ and $P^{(n)}(\cdot, A)$ is a $\mathscr{W}$-measurable function for any $A \in \mathscr{X}$).

The collection $((W, \mathscr{W}), (X, \mathscr{X}), u^{(n)}, P^{(n)}, n = 1, 2, \ldots)$ is called a *random system with complete connections* (RSCC for short).

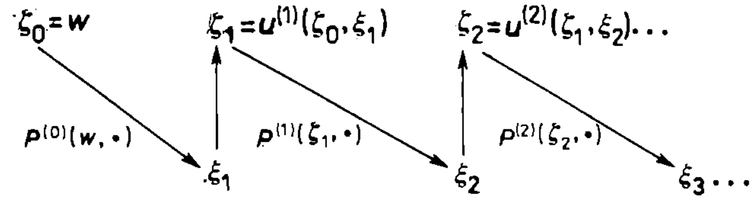
With any RSCC one can associate two sequences of random variables, $(\zeta_n)_{n \geq 0}$ with values in W and $(\xi_n)_{n \geq 1}$ with values in X, such that $\zeta_0 = w$ (arbitrarily fixed),

$$\zeta_{n+1} = u^{(n)}(\zeta_n, \xi_{n+1}), \qquad n \geq 0,$$

$$\mathrm{P}(\xi_{n+1} \in A \mid \xi_n, \zeta_n, \ldots) = P^{(n)}(\zeta_n, A), \qquad n \geq 0,$$

P-a.s. whatever $A \in \mathscr{X}$.

One may imagine the following diagram representing both sequences associated with an RSCC:



The Onicescu–Mihoc case is obtained by taking

$$X = \{1, 2, \ldots, m+1\},$$

$$W = \{p = (p_j)_{1 \leq j \leq m+1} : 0 \leq p_j \leq 1, \sum_{j=1}^{m+1} p_j = 1\},$$

$$u^{(n)}\big((p_j)_{1 \leq j \leq m+1}, i\big) = \big(f_{ij}^{(n)}(p_1, \ldots, p_{m+1})\big)_{1 \leq j \leq m+1},$$

$$P^{(n)}\big((p_j)_{1 \leq j \leq m+1}, i\big) = p_i, \qquad 1 \leq i \leq m+1.$$

A formal existence theorem for the sequences associated with an RSCC will be proved later.

3

We are now going to give a few important examples of RSCC's. Most of them are *homogeneous*, i.e., they are RSCC's for which $u^{(n)}$ and $P^{(n)}$ do not depend on n (to be denoted u and P).

a) Markov chains. It is easily seen that Markovian dependence is obtained by taking $X = W$, $u(w, x) = x$. (The nonhomogeneous case of Markovian dependence clearly appears as a special case of a nonhomogeneous RSCC). You may show how *multiple* Markovian dependence is also a special case of RSCC's.

b) Learning models. By mathematical learning theory we mean the body of research methods and results concerned with the conceptual representation of learning phenomena, the mathematical formulation of hypotheses about learning, and the derivation of testable theorems.

The purpose of mathematical learning theory is to provide simple quantitative descriptions of processes which are basic to behaviour modifications. Learning models are in fact devices for providing simple description of basic learning processes. Prior to 1950 learning models were concerned with predicting, at most, the mean learning curves*) obtained from experiments. Although a few of the earlier models involved probability measures and processes, no properties other than the mean learning curve were deduced. The main feature of the modern formulations initiated since 1950 by Estes (1950) and Bush and Mosteller (1951, 1955) is that they imply random processes, i.e., it is assumed that there is some random process to which the behaviour in simple learning experiments conforms. Another important feature is the step-by-step nature of the learning process.

All stochastic models for learning studied up to now enter the following general theoretical scheme. The behaviour of the subject on trial n is determined by his *state* S_n (an indicator of the subject's response tendencies) at the beginning of the trial. S_n is a random variable taking on values in a measurable *state space* $(S, \mathscr{S})$. On trial n an *event* E_{n+1} occurs that results in a change of state. E_{n+1} is a random variable taking on values in a measurable *event space* $(E, \mathscr{E})$ and specifies those occurrences on trial n that affect subsequent behaviour. Typically E_{n+1} includes a specification of the subject's response and its observable outcome or payoff. To represent the fact that the occurrence of an event effects a change of state we consider a measurable mapping v of $S \times E$ into S and postulate that $S_{n+1} = v(S_n, E_{n+1})$, $n \geqslant 0$. Finally we assume that the probability distribution of E_{n+1} given $E_n, S_n, \ldots$ depends only on the state S_n and we denote it by $Q(S_n, \cdot)$.

By a *general learning model* we mean a collection $((S, \mathscr{S}), (E, \mathscr{E}), v, Q)$, i.e., (trivially) an RSCC. (Notice we have just changed the notation!) Various special learning models are obtained by simply particularizing S, E, v, and Q.

References for stochastic learning models: Iosifescu and Theodorescu (1969), Lakshmivarahan (1981), Norman (1972), Rouanet (1967).

*) The notion of a mean learning curve measures the "performance" of an individual as a function of training time or trials. It was proposed in 1919 by L. L. Thurstone.

c) Continued fractions. Each irrational number y in the unit interval has a unique infinite continued fraction expansion of the form

$$y = \cfrac{1}{a_1(y) + \cfrac{1}{a_2(y) + \dots}},$$

where the $a_n(y)$ are natural numbers determined as follows. Put $Ty = 1/y (\text{mod } 1)$. Then $a_1(y) = 1/y - Ty$, and $a_{n+1}(y) = a_n(Ty) = a_1(T^n y)$, $n \geqslant 1$. It is obvious that, when the unit interval is endowed with the σ-algebra of Lebesgue sets, the a_n are random variables defined almost everywhere with respect to any probability measure assigning probability 0 to the set of rational numbers (in particular with respect to Lebesgue measure λ).

The metric theory of continued fractions is concerned with the study of the random sequence $(a_n)_{n \geqslant 1}$. The first problem of this theory was raised in 1812 by Gauss, who in a letter to Laplace stated that

$$\lim_{n \to \infty} \lambda(r_n^{-1} < x) = \frac{\log(1+x)}{\log 2} \tag{3}$$

for each x in the unit interval, and asked for an estimate of the error

$$E_n(x) = \lambda(r_n^{-1} < x) - \frac{\log(1+x)}{\log 2}.$$

Here $r_n(y) = 1/(T^n y)$, $n \geqslant 1$, i.e.,

$$r_n = a_n + \cfrac{1}{a_{n+1} + \cfrac{1}{a_{n+2} + \dots}}.$$

Kuzmin (1928) first proved (3), giving an error estimate $E_n(x) = O(q^{\sqrt{n}})$ as $n \to \infty$ with $0 < q < 1$. One year later Lévy (1929) gave a different proof, allowing to improve the error estimate to $E_n(x) = O(q^n)$ with $q < 0.7$. It has been recently shown that the optimal value of q is 0.303663002898732658 60... See Wirsing (1974).

It is not difficult to show that

$$\lambda(r_1 > t) = 1/t,$$

$$\lambda(r_{n+1} > t \mid a_1, \dots, a_n) = \frac{s_n + 1}{s_n + t}, \qquad t \geqslant 1, n \geqslant 1,$$

where

$$s_n = \cfrac{1}{a_n + \ddots + \cfrac{1}{a_1}}$$

This is known as the Borel–Lévy formula. (Try to prove it. You may consult Ibragimov and Linnik (1971)). It implies at once

$$\lambda(a_1 = k) = \frac{1}{k(k+1)},$$

$$\lambda(a_{n+1} = k \mid a_1, \ldots, a_n) = \frac{s_n+1}{(s_n+k)(s_n+k+1)}, \qquad n \geqslant 1,$$

for any natural integer k. Noting that

$$s_{n+1} = \frac{1}{a_{n+1}+s_n}, \qquad n \geqslant 0,$$

with $s_0 = 0$, we are led to consider the RSCC $\big((W, \mathscr{W}), (X, \mathscr{X}), u, P\big)$ for which

W = the unit interval $[0, 1]$, $\mathscr{W}$ = the σ-algebra of Borel sets in $[0, 1]$;

$X = \{1, 2, \ldots\}$, $\mathscr{X}$ = the σ-algebra of all subsets of X;

$$u(w, x) = \frac{1}{w+x}, \qquad P(w, x) = \frac{w+1}{(w+x)(w+x+1)}, \qquad w \in W,\ x \in X.$$

Clearly, starting with $w = 0$, the ξ and ζ sequences associated with the above RSCC are just $(a_n)_{n \geqslant 1}$ and $(s_n)_{n \geqslant 0}$ under $\mathrm{P} = \lambda$. More generally, starting with a rational

$$w = \cfrac{1}{x_p + \ddots + \cfrac{1}{x_1}}$$

the ξ and ζ sequences are $(a_{n+p})_{n \geqslant 1}$ and $(s_{n+p})_{n \geqslant 0}$ under $\mathrm{P} = \lambda(\cdot \mid a_k = x_k,\ 1 \leqslant k \leqslant p)$.

References for continued fractions: Iosifescu (1974, 1978).

d) Stochastic approximation. Let F_a, $a \in \boldsymbol{R}$ (= the real line), be a family of distribution functions on the real line. Assume that

$$M(a) = \int_{\boldsymbol{R}} y \, dF_a(y)$$

exists and the equation $M(a) = 0$ has a unique root θ. Robbins and Monro (1951) gave a statistical method for estimating θ. They chose an arbitrary $X_0 \in \boldsymbol{R}$ and defined recursively random variables X_n by

$$X_{n+1} = X_n - \frac{c}{n+1} Y_{n+1}, \qquad n \geqslant 0,\ c \geqslant 0,$$

where Y_{n+1} is a random variable the conditional distribution of which, given $X_0, \ldots, X_n$, is F_{X_n}. Under suitable conditions the convergence (either in probability or almost sure) of X_n to θ as $n \to \infty$ can be proved.

It is clear that the sequence $(X_n)_{n \geq 0}$ is the ζ-chain associated with the (nonhomogeneous) RSCC $((W, \mathscr{W}), (X, \mathscr{X}), (u^{(n)})_{n \geq 0}, P)$ for which

$$(W, \mathscr{W}) = (X, \mathscr{X}) = \text{the real line with the Borel } \sigma\text{-algebra};$$

$$u^{(n)}(w, x) = w - \frac{c}{n+1} x, \qquad n \geq 0, w \in W, x \in X;$$

$$P(w, A) = \int_A dF_w(Y).$$

e) Other examples of RSCC's can be found in Macchi (1976) (these pertain to the theory of stochastic approximation and the theory of adaptive automata), Blackwell (1976) and Rieder (1975) (these pertain to gambling and the theory of dynamic programming), Smith and Wilkinson (1969) (these pertain to branching processes in a random environment). See also Section 1.2 in Grigorescu and Iosifescu (1982).

4

In what follows we shall limit ourselves to considering just homogeneous RSCC's.

Before proving a formal existence theorem for the sequences of random variables associated with a (homogeneous) RSCC let us make a few notational conventions. For any $n \in N^* = \{1, 2, \ldots\}$ the n-fold product measurable space $(X, \mathscr{X}) \times \ldots \times (X, \mathscr{X})$ (n times) will be denoted by $(X^n, \mathscr{X}^n)$. Clearly, $(X^1, \mathscr{X}^1) = (X, \mathscr{X})$. The element $(x_1, \ldots, x_n) \in X^n$ will be denoted by $x^{(n)}$. If in the same formula there appear $x^{(n)}$ and $x^{(n-1)}$, say then the first $n-1$ coordinates of $x^{(n)}$ are precisely the coordinates of $x^{(n-1)}$. (In fact, $x^{(n)} = (x^{(n-1)}, x_n)$). Remember that we defined recursively the random variables ζ_n by $\zeta_{n+1} = u(\zeta_n, \xi_{n+1})$, $n \geq 0$. It is therefore convenient to introduce simple notation for the composition of the mappings $u(\cdot, x)$, $x \in X$, of W into itself. Thus, we define

$$u_n(w, x^{(n)}) = \begin{cases} u(u_{n-1}(w, x^{(n-1)}), x_n) & \text{if } n > 1, \\ u_1(w, x^{(1)}) = u(w, x_1) & \text{if } n = 1, \end{cases}$$

and with no possible confusion we shall simply write $wx^{(n)}$ for $u_n(w, x^{(n)})$. With this convention ζ_n is nothing but $w\xi^{(n)}$ for all $n \in N^*$ ($\xi^{(n)} = (\xi_1, \ldots, \xi_n)$).

Now we can prove the following

EXISTENCE THEOREM. *For a given* RSCC $((W, \mathscr{W}), (X, \mathscr{X}), u, P)$ *and an arbitrarily fixed* $w \in W$ *there exist a probability space* $(\Omega, \mathscr{K}, \mathrm{P}_w)$ *and a* *sequence of X-valued random variables* $(\xi_n)_{n\geq 1}$ *defined on* Ω *such that*

$$\mathrm{P}_w(\xi_1 \in A) = P(w, A),$$
$$\mathrm{P}_w(\xi_{n+1} \in A \mid \xi_1, \ldots, \xi_n) = P(w\xi^{(n)}, A) \tag{4}$$

P_w*-almost surely for any* $A \in \mathscr{X}$, $n \in N^*$.

Proof. Before proceeding further, let us notice that, since the σ-algebra generated by $\zeta^{(n)} = (\zeta_1, \ldots, \zeta_n)$ is contained in the σ-algebra generated by $\xi^{(n)} = (\xi_1, \ldots, \xi_n)$, the conditional probability occurring in (4) is equal to (1) $\mathrm{P}_w(\xi_{n+1} \in A \mid \zeta^{(n)}, \xi^{(n)})$. Therefore no new element has been introduced with respect to the (somewhat) heuristic discussion on page 247.

For the proof (in a similar manner to the proof of the existence theorem for Markov chains) we define

$$\Omega = X^{N^*} = \{(x_1, x_2, x_3, \ldots): x_i \in X, i \in N^*\},$$

$$\mathscr{K} = \text{the infinite product } \sigma\text{-algebra } \mathscr{X} \times \mathscr{X} \times \ldots$$

(the smallest σ-algebra containing all cylindrical sets $A_1 \times \ldots \times A_l \times X^{N^*}$ in Ω, where $A_k \in \mathscr{X}$, $1 \leqslant k \leqslant l$, $l \in N^*$),

$$\xi_n(\omega) = x_n, \quad n \in N^*, \qquad \text{if} \quad \omega = (x_n)_{n \in N^*},$$

and using the prescribed transition probabilities we put

$$\mathrm{P}_w(\xi_1 \in A_1) = \left(\mathrm{P}_w(A_1 \times X^{N^*})\right) = P(w, A_1),$$

$$\mathrm{P}_w(\xi_1 \in A_1, \ldots, \xi_l \in A_l) = \left(\mathrm{P}_w(A_1 \times \ldots \times A_l \times X^{N^*})\right)$$
$$= \begin{cases} \int_{A_1} P(w, dx_1) P(wx_1, A_2) & \text{if } l = 2 \\ \int_{A_1} \cdots \int_{A_{l-1}} P(w, dx_1) P(wx_1, dx_2) \ldots P(wx^{(l-2)}, dx_{l-1}) P(wx^{(l-1)} \; A_1) & \\ & \text{if } l > 3. \end{cases}$$

Now the possibility of extending P_w to the whole of $\mathscr{K}$ is ensured by the Ionescu Tulcea theorem (see, e.g., Neveu (1964)).

Equation (4) is clearly an immediate consequence of the very definition of P_w on cylindrical sets. ■

Let us now give a formal proof of the fact that $(\zeta_n)_{n\geq 0}$ is a W-valued Markov chain. For any bounded real valued W-measurable function f defined on W we can write (2)

$$\mathrm{E}_w\left(f(\zeta_{n+1}) \mid \zeta^{(n)}\right) = \mathrm{E}_w\left(f(\zeta_n \xi_{n+1}) \mid \zeta^{(n)}\right) = \int_X P(\zeta_n, dx) f\left(u(\zeta_n, x)\right) \tag{5}$$

---

(1) On account of (4) it is also equal to $\mathrm{P}_w(\xi_{n+1} \in A \mid \zeta_1, \ldots, \zeta_n)$. Please justify this assertion.

(2) E_w denotes the mean value operator with respect to P_w.

(Here we use the remark made in footnote 1 on p. 252). It follows that $(\zeta_n)_{n\geq 0}$, $\zeta_0 \equiv w$, is a Markov chain with initial probability distribution concentrated at w and transition probability function

$$\mathrm{P}_w(\zeta_{n+1}\in B\,|\,\zeta_n) = P(\zeta_n,\, B(\zeta_n)), \qquad n \geq 0,$$

for any $B\in \mathscr{W}$, where

$$B(v) = \{x\colon\ u(v,\, x)\in B\}, \qquad v\in W.$$

This is obtained by simply taking in (5) $f = \delta_B$, where

$$\delta_B(v) = \begin{cases} 1 & \text{if } v\in B, \\ 0 & \text{if } v\notin B, \end{cases}$$

$B\in \mathscr{W}$, $v\in W$. In fact (5) exhibits a more general thing, namely that the transition operator [3] of the Markov chain $(\zeta_n)_{n\geq 0}$ is

$$(Uf)(w) = \int_X P(w,\, dx) f(u(w,\, x)).$$

It can be shown in the same manner that the sequences $(\zeta_n,\, \xi_n)_{n\geq 1}$ and $(\zeta_n,\, \xi_{n+1})_{n\geq 0}$ are also Markov chains. The transition operators of these Markov chains are

$$(U'f)(w,\, x) = \int_X P(w,\, dx') f(u(w,\, x'),\, x') \quad (= f'(w))$$

and

$$(U''f)(w,\, x) = \int_X P(u(w,\, x),\, dx') f(u(w,\, x),\, x') \quad (= f''(u(w,\, x))),$$

respectively.

5

Turning to other problems, in what follows we shall first be interested in the asymptotic behaviour as $n\to\infty$ of the probabilities $\mathrm{P}_w((\xi_n,\, \ldots,\, \xi_{n+l-1})\in A)$ for $A\in \mathscr{X}^l$, $l\in N^*$. Let us write for $w\in W$, $l,\, n\in N^*$, $A\in \mathscr{X}^l$,

$$\mathrm{P}_w((\xi_n,\, \ldots,\, \xi_{n+l-1})\in A) = P_l^n(w,\, A).$$

---

(3) For an arbitrary Markov chain with state space $(Y,\, \mathscr{Y})$ and transition probability function p its transition operator is defined as $(Uf)(y) = \int_Y p(y,\, dz) f(z)$ for any bounded real valued $\mathscr{Y}$-measurable function f defined on Y.

Clearly, $P_l^n(w, \cdot)$ is a probability measure on $\mathscr{X}^l$ and

$$P_l^1(w, A) = P_l(w, A) = \begin{cases} P(w, A) & \text{if } l = 1 \\ \int_{X^l} P(w, dx_1) P(wx_1, dx_2) \dots P(wx^{(l-1)}, dx_l)\, \delta_A(x^{(l)}) & \\ & \text{if } l > 1. \end{cases}$$

Equation (4) is then easily generalized to

$$\mathrm{P}_w\big((\zeta_{n+1}, \dots, \zeta_{n+l}) \in A \mid \xi^{(n)}\big) = P_l(w\xi^{(n)}, A) \tag{6}$$

P_w-almost surely for any $A \in \mathscr{X}^l$, $l, n \in N^*$. Further, it is obvious that

$$P_l^n(w, A) = P_{l+n-1}(w, X^{n-1} \times A), \tag{7}$$

for any $n \in N^*$ with the convention $X^0 \times A = A$. Now for $n \geqslant 2$ and $1 \leqslant r \leqslant n$

$$\begin{aligned} P_l^n(w, A) &= \mathrm{P}_w\big((\zeta_{r+1}, \dots, \zeta_n, \dots, \zeta_{n+l-1}) \in X^{n-r-1} \times A\big) \\ &= \mathrm{E}_w\big(\delta_{X^{n-r-1} \times A}(\zeta_{r+1}, \dots, \zeta_n, \dots, \zeta_{n+l-1})\big) \\ &= \mathrm{E}_w\big(\mathrm{E}_w(\delta_{X^{n-r-1} \times A}(\zeta_{r+1}, \dots, \zeta_n, \dots, \zeta_{n+l-1}) \mid \xi^{(r)})\big) \\ &= \mathrm{E}_w\big(\mathrm{P}_w((\zeta_{r+1}, \dots, \zeta_n, \dots, \zeta_{n+l-1}) \in X^{n-r-1} \times A \mid \xi^{(r)})\big) \\ &= \mathrm{E}_w\big(P_{n-r+l-1}(w\xi^{(r)}, X^{n-r-1} \times A)\big) \\ &= \mathrm{E}_w\big(P_l^{n-r}(w\xi^{(r)}, A)\big) = \int_{Y^r} P_r(w, dx^{(r)})\, P_l^{n-r}(wx^{(r)}, A). \end{aligned}$$

(We used both (6) and (7)). Therefore we proved the equation

$$P_l^n(w, A) = \int_{X^r} P_r(w, dx^{(r)})\, P_l^{n-r}(wx^{(r)}, A), \qquad 1 \leqslant r \leqslant n. \tag{8}$$

This is clearly an extension of the Chapman–Kolmogorov equation occurring in Markov chains.

Remark. Equation (8) shows that $P_l^n(\cdot, A)$ (as a function on W) is the $(n-1)$th iterate of $P_l(\cdot, A)$ under the operator U defined on p. 253:

$$P_l^n(\cdot, A) = U^{n-1}\big(P_l(\cdot, A)\big), \qquad n \geqslant 1.$$

To study the asymptotic behaviour of the $P_l^n(w, A)$ we need some prerequisites.

For a finite completely additive signed measure μ on a σ-algebra $\mathscr{M}$ in a space M its norm $\|\mu\|$ is defined as

$$\|\mu\| = \tfrac{1}{2} \operatorname{var} \mu = \tfrac{1}{2}\big(\sup_{A \in \mathscr{M}} \mu(A) - \inf_{A \in \mathscr{M}} \mu(A)\big).$$

If $\mu(M) = 0$, then

$$\|\mu\| = \tfrac{1}{2} \operatorname{var} \mu = \mu(M^+),$$

where $M^+ \cup M^- = M$ is the Hahn decomposition of M with respect to μ. (This means that $\mu(M^+ \cap A) \geqslant 0$ and $\mu(M^- \cap A) \leqslant 0$ for any $A \in \mathscr{M}$.)

LEMMA. *Consider a measure space $(M, \mathscr{M}, \mu)$, the signed measure μ being supposed finite and completely additive and such that $\mu(M) = 0$. If f is a bounded real and $\mathscr{M}$-measurable function defined on M, then*

$$\left|\int_M f(x)\,\mu(dx)\right| \leqslant \|\mu\|\,(\sup_{x\in M} f(x) - \inf_{x\in M} f(x)).$$

Proof. Let $M^+ \cup M^- = M$ be the Hahn decomposition of M with respect to μ. The condition $\mu(M) = 0$ implies that $\mu(M^+) = -\mu(M^-) = \|\mu\|$. Consequently,

$$\int_M f(x)\,\mu(dx) = \int_{M^+} + \int_{M^-} \leqslant (\sup f)\,\mu(M^+) + (\inf f)\,\mu(M^-)$$
$$= \|\mu\|\,(\sup f - \inf f).$$

On the other hand, taking into account that

$$\sup(-f) = -\inf f, \qquad \inf(-f) = -\sup f,$$

we have also

$$-\int_M f(x)\,\mu(dx) \leqslant \|\mu\|\,(\sup f - \inf f). \qquad \blacksquare$$

Next, let us define

$$b_n = \sup|P(w' x^{(n)}, A) - P(w'' x^{(n)}, A)|, \qquad n \geqslant 1,$$

the upper bound being taken over all w', $w'' \in W$, $x^{(n)} \in X^n$, $A \in \mathscr{X}$.

PROPOSITION. *We have*

$$|P_l(w' x^{(n)}, A) - P_l(w'' x^{(n)}, A)| \leqslant \sum_{j=n}^{n+l-1} b_j$$

for any l, $n \in N^*$, w', $w'' \in W$, $x^{(n)} \in X^n$, $A \in \mathscr{X}^l$.

Proof. We proceed by induction. For $l = 1$ the inequality is true by the very definition of b_n. Assume the statement holds for l and let us prove it for $l+1$. Putting $w_1 = w' x^{(n)}$, $w_2 = w'' x^{(n)}$ we can write

$$P_{l+1}(w' x^{(n)}, A) - P_{l+1}(w'' x^{(n)}, A)$$
$$= \int_{X^{l+1}} P_l(w' x^{(n)}, dy^{(l)})\, P(w_1 y^{(l)}, dz)\, \delta_A(y^{(l)}, z) -$$
$$- \int_{X^{l+1}} P_l(w'' x^{(n)}, dy^{(l)})\, P(w_2 y^{(l)}, dz)\, \delta_A(y^{(l)}, z)$$
$$= \int_{X^l} [P_l(w_1, dy^{(l)}) - P_l(w_2, dy^{(l)})] \int_X P(w_1 y^{(l)}, dz)\, \delta_A(y^{(l)}, z) +$$
$$+ \int_{X^l} P(w_2, dy^{(l)} \int_X [P(w_1 y^{(l)}, dz) - P(w_2 y^{(l)}, dz)]\, \delta_A(y^{(l)}, z).$$

By making use of the above lemma twice we get

$$|P_{l+1}(w'x^{(n)}, A) - P_{l+1}(w''x^{(n)}, A)| \leqslant \sum_{j=n}^{n+l-1} b_j + b_{n+l} = \sum_{j=n}^{n+l} b_j,$$

which completes the proof. ∎

Now we can prove the following

ERGODIC THEOREM (Ionescu Tulcea (1959)). *Assume that the series* $\sum b_n$ *converges and there exist a probability measure p on* $\mathscr{X}$ *and a positive constant* $0 < a \leqslant 1$ *such that* $P(w, A) \geqslant ap(A)$ *for any* $w \in W$, $A \in \mathscr{X}$. *Let* $r \in N^*$ *denote the smallest natural integer such that* $\sum_{j \geqslant r} b_j \leqslant 1/4$. *Then for any* $l \in N^*$ *there exists a probability measure* P_l^∞ *on* $\mathscr{X}^l$ *such that*

$$|P_l^n(w, A) - P_l^\infty(A)| \leqslant \inf_{1 \leqslant s < n} \left(\left(1 - \frac{a^r}{4}\right)^{n/s - 1} + \frac{4}{a^n} \sum_{j \geqslant s} b_j \right)$$

for any l, $n \in N^*$, $w \in W$, $A \in \mathscr{X}^l$.

Proof. Put $p^n = p \times \ldots \times p$ (n times) (direct product of measures), $n \in N^*$, and define for $w \in W$, $A \in \mathscr{X}^l$,

$$p_l(w, A) = \begin{cases} p^l(A) & \text{if } l \leqslant r, \\ \int_{X^r} p^r(dx^{(r)}) \int_{X^{l-r}} P_{l-r}(wx^{(r)}, dy^{(l-r)}) \delta_A(x^{(r)}, y^{(l-r)}) & \text{if } l > r. \end{cases}$$

On account of the above lemma and proposition we get

$$|p_l(w', A) - p_l(w'', A)| \leqslant \sum_{j \geqslant r} b_j \leqslant \tfrac{1}{4} \tag{9}$$

for any $l \in N^*$, w', $w'' \in W$, $A \in \mathscr{X}^l$.

By the very definition of p_l and P_l we can write

$$P_l(w, A) \geqslant a^r p_l(w, A) \tag{10}$$

for any $l \in N^*$, $w \in W$, $A \in \mathscr{X}^l$. Now (9) and (10) imply that

$$P_l(w', A) \geqslant a^r \left(p_l(w'', A) - \tfrac{1}{4}\right) \tag{11}$$

for any $l \in N^*$, w', $w'' \in W$, $A \in \mathscr{X}^l$. Keeping w'' fixed in (11) it easily follows that either

$$P_l(w', A) \geqslant \tfrac{1}{4} a^r \quad \text{for any} \quad w' \in W$$

or

$$P_l(w', X^l - A) \geqslant \tfrac{1}{4} a^r \quad \text{for any} \quad w' \in W.$$

Therefore

$$|P_l(w', A) - P_l(w'', A)| = |P_l(w', X^l - A) - P_l(w'', X^l - A)| \leqslant 1 - \tfrac{1}{4} a^r \tag{12}$$

for any $l \in N^*$, w', $w'' \in W$, $A \in \mathscr{X}^l$.

Now on account of equation (8) we can write for $1 \leqslant s < n$

$$P_l^n(w', A) - P_l^n(w'', A) =$$
$$= \int_{X^s} P_s(w', dx^{(s)}) P_l^{n-s}(w' x^{(s)}, A) - \int_{X^s} P_s(w'', dx^{(s)}) P_l^{n-s}(w'' x^{(s)}, A)$$
$$= \int_{X^s} \big(P_s(w', dx^{(s)}) - P_s(w'', dx^{(s)})\big) P_l^{n-s}(w' x^{(s)}, A) +$$
$$+ \int_{X^s} P_s(w'', dx^{(s)}) [P_l^{n-s}(w' x^{(s)}, A) - P_l^{n-s}(w'' x^{(s)}, A)]$$
$$= \mathrm{I} + \mathrm{II}.$$

Putting

$$\underline{P}_l^n(A) = \inf_{w \in W} P_l^n(w, A), \qquad \bar{P}_l^n(A) = \sup_{w \in W} P_l^n(w, A),$$

remark that on account of equation (8) the sequence $(\underline{P}_l^n(A))_{n \geqslant 1}$ is nondecreasing and the sequence $(\bar{P}_l^n(A))$ is nonincreasing. Thus both sequences are convergent. We shall prove that they have a common limit.

By virtue of the above lemma and inequality (12) we have

$$|\mathrm{I}| \leqslant (1 - \tfrac{1}{4} a^r)\big(\bar{P}_l^{n-s}(A) - \underline{P}_l^{n-s}(A)\big),$$

and the above proposition and lemma yield

$$|\mathrm{II}| \leqslant \sum_{j \geqslant s} b_j.$$

(Remember that $P_l^{n-s}(w, A) = P_{l+n-s-1}(w, X^{n-s-1} \times A)$.)

Summing up we can write

$$\bar{P}_l^n(A) - \underline{P}_l^n(A) \leqslant (1 - \tfrac{1}{4} a^r)\big(\bar{P}_l^{n-s}(A) - \underline{P}_l^{n-s}(A)\big) + \sum_{j \geqslant s} b_j;$$

hence

$$\bar{P}_l^n(A) - \underline{P}_l^n(A) \leqslant (1 - \tfrac{1}{4} a^r)^k \big(\bar{P}_l^{n-ks}(A) - \underline{P}_l^{n-ks}(A)\big) + \sum_{j \geqslant s} b_j \Big(\sum_{h=0}^{k-1} (1 - \tfrac{1}{4} a^r)^h\Big)$$

for any $k \in N^*$ such that $n > ks$. It follows that the difference $\bar{P}_l^n(A) - \underline{P}_l^n(A)$ approaches 0 as $n \to \infty$, that is, the sequences $(\bar{P}_n^l(A))_{n \geqslant 1}$ and $(\underline{P}_l^n(A))_{n \geqslant 1}$ have a common limit $P_l^\infty(A)$. It is not difficult to deduce the domination claimed. The fact that P_l^∞ is a probability measure on $\mathscr{X}^l$ follows from the Vitali–Hahn–Saks theorem (see Neveu (1964)). ∎

We say that an RSCC is *uniformly ergodic* iff for any $l \in N^*$ there exists a probability measure P_l^∞ on $\mathscr{X}^l$ such that

$$\varepsilon_n = \sup_{\substack{w \in W, l \in N^* \\ A \in \mathscr{X}^l}} |P_l^n(w, A) - P_l^\infty(A)| \to 0 \quad \text{as} \quad n \to \infty..$$

Therefore the ergodic theorem we have just proved gives a set of sufficient conditions for uniform ergodicity together with the estimates of the coefficients ε_n, $n \in N^*$.

Remark 1. In the Markov case $W = X$, $\mathscr{W} = \mathscr{X}$, $u(w, x) = x$ the existence of a limiting probability measure $P_1^\infty = P^\infty$ on $\mathscr{X}$ such that

$$\lim_{n \to \infty} P^n(x, A) = P^\infty(A) \tag{13}$$

uniformly with respect to $x \in X$ and $A \in \mathscr{X}$ ensures the existence of probability measures P_l^∞, $l \in N^*$, such that

$$\lim_{n \to \infty} P_l^n(x, A) = P_l^\infty(A)$$

uniformly with respect to $x \in X$, $l \in N^*$, $A \in \mathscr{X}^l$. It is a simple exercise to show that for $l > 1$

$$P_l^\infty(A) = \int_{X^l} P^\infty(dx_1) P(x_1, dx_2) \ldots P(x_{l-1}, dx_l) \delta_A(x^{(l)}).$$

A necessary and sufficient condition for (13) to hold (in fact for the uniform ergodicity of a Markov chain) is the existence of an $n_0 \in N^*$ such that

$$\sup_{\substack{x', x'' \in X, \\ A \in \mathscr{X}}} |P^{n_0}(x', A) - P^{n_0}(x'', A)| < 1,$$

and the corresponding ε_n, $n \in N^*$, tend exponentially to zero as $n \to \infty$.

Remark 2. A detailed study of uniform ergodicity of an RSCC can be found in Grigorescu and Iosifescu (1982) and Iosifescu and Theodorescu (1969).

Remark 3 (Exercise!). Show that the assumptions of the ergodic theorem proved before are fulfilled by the RSCC associated with the continued fraction expansion. (Hint: take p as the probability measure concentrated at 2.)

6

Now we shall prove uniform ergodicity implies some mixing property of the sequence $(\xi_n)_{n \in N^*}$ under any P_w, $w \in W$. Denote by $\mathscr{K}_{(m,n)}$ the σ-algebra generated by the random variables $\xi_m, \ldots, \xi_n$, $m \leqslant n$. We have

$$\mathrm{P}_w\big((\xi_{t+n}, \ldots, \xi_{t+n+l-1}) \in A \mid \xi^{(t)}\big) = P_l^n(w\xi^{(t)}, A),$$

$$\mathrm{P}_w\big((\xi_{t+n}, \ldots, \xi_{t+n+l-1}) \in A\big) = P_l^{t+n}(w, A)$$

for every $w \in W$, l, n, $t \in N^*$, $A \in \mathscr{X}^l$. On account of uniform ergodicity we get

$$\left|\mathrm{P}_w\left((\xi_{t+n}, \ldots, \xi_{t+n+l-1}) \in A \mid \xi^{(t)}\right) - \mathrm{P}_w\left((\xi_{t+n}, \ldots, \xi_{t+n+l-1}) \in A\right)\right| \leqslant \varepsilon_n + \varepsilon_{n+t}.$$

Hence, putting $A_1 = \{\xi^{(t)} \in B\} \in \mathscr{K}_{(1,t)}$ for an arbitrary $B \in \mathscr{X}^t$, $A_2 = \{(\xi_{t+n}, \ldots, \xi_{t+n+l-1}) \in A\}$ and integrating over the event A_1, we obtain

$$|\mathrm{P}_w(A_1 \cap A_2) - \mathrm{P}_w(A_1)\,\mathrm{P}_w(A_2)| \leqslant (\varepsilon_n + \varepsilon_{n+t})\,\mathrm{P}_w(A_1). \tag{14}$$

Clearly, (14) is valid for any $w \in W$, $A_1 \in \mathscr{K}_{(1,t)}$, $A_2 \in \mathscr{K}_{(t+n,t+n+l-1)}$, t, n, $l \in N^*$. If we put

$$\varphi_w(n) = \sup |\mathrm{P}_w(A_2 \mid A_1) - \mathrm{P}_w(A_2)|,$$

the upper bound being taken over all $A_1 \in \mathscr{K}_{(1,t)}$ such that $\mathrm{P}_w(A_1) \neq 0$, $A_2 \in \mathscr{K}_{(t+n,t+n+l-1)}$, t, $l \in N^*$, it follows that

$$\varphi_w(n) \leqslant \varepsilon_n + \varepsilon_{n+1}.$$

Therefore *the sequence* $(\xi_n)_{n \in N^*}$ *is* φ*-mixing under* P_w *for any* $w \in W$, *i.e.*, $\lim_{n \to \infty} \varphi_w(n) = 0$. ∎

Our next step will be to show that *under uniform ergodicity there exists a probability* P_∞ *on* $\mathscr{K}$ *such that* $(\xi_n)_{n \in N^*}$ *is a strictly stationary sequence on* $(\Omega, \mathscr{K}, \mathrm{P}_\infty)$.

The proof of this assertion is as follows. For $B \in \mathscr{K}_{(1,l)}$ (which equals $\mathscr{X}^l \times X^{N^*}$ – see the proof of the existence theorem), $l \in N^*$, we set

$$\mathrm{P}_\infty(B) = P_l^\infty(B^{(l)})$$

if $B = B^{(l)} \times X^{N^*}$. Define as usual the shift operator T on Ω by

$$T\omega = (x_{n+1})_{n \in N^*}$$

if $\omega = (x_n)_{n \in N^*}$. By uniform ergodicity we have

$$|\mathrm{P}_w(T^{-n}B) - \mathrm{P}_\infty(B)| \leqslant \varepsilon_{n+1} \tag{15}$$

for all $n \in N^*$, $w \in W$, $B \in \mathscr{K}_{(1,l)}$, $l \in N^*$.

Clearly, P_∞ is a completely additive probability measure on the algebra

$$\mathscr{A} = \bigcup_{l \in N^*} \mathscr{K}_{(1,l)}$$

which generates the σ-algebra $\mathscr{K}$. To show that P_∞ extends uniquely to $\mathscr{K}$ it is sufficient to prove that for any decreasing sequence $B_1 \supset B_2 \supset \ldots$ with $\bigcap_{n \in N^*} B_n = \emptyset$, $B_n \in \mathscr{A}$, $n \in N^*$, one has $\lim_{n \to \infty} \mathrm{P}_\infty(B_n) = 0$ (see, e.g., Neveu (1964)). On account of (15) this is easily done. Indeed, if $\bigcap_{n \in N^*} B_n = \emptyset$, then we have also $\bigcap_{n \in N^*} T^{-k} B_n = \emptyset$ for any $k \in N^*$. But, P_w being a completely additive probability measure on $\mathscr{X}$, we have $\lim_{n\to\infty} P_w(T^{-k}B_n) = 0$ for any $w\in W$, $k\in N^*$. By virtue of (15) it follows that

$$\limsup_{n\to\infty} P_\infty(B_n) \leqslant \varepsilon_{k+1},$$

regardless of the value of $k\in N^*$, but of course this means that $\lim_{n\to\infty} P_\infty(B_n) = 0$, which completes the proof of the possibility of uniquely extending P_∞ to the whole of $\mathscr{X}$.

Finally, the strict stationarity of $(\xi_n)_{n\in N^*}$ under P_∞ also follows from uniform ergodicity. Indeed, we have

$$|P_{h+l}^\infty(X^h\times B^{(l)}) - P_l^\infty(B^{(l)})| \leqslant \varepsilon_n+\varepsilon_{n+h},$$

whence

$$P_{h+l}^\infty(X^h\times B^{(l)}) = P_l^\infty(B^{(l)}),$$

for arbitrary $h\in N^*$, and this means that

$$P_\infty\big((\xi_{h+1},\ldots,\xi_{h+l})\in B^{(l)}\big) = P_l^\infty(B^{(l)}), \qquad h\in N^*,$$

i.e., strict stationarity holds.

It is quite natural to ask whether $(\xi_n)_{n\in N^*}$ is φ-mixing under P_∞. The answer is affirmative and the proof rests on inequality (14). Let us take for $t_0\leqslant t$

$$A_l = \{(\xi_{t-t_0+1},\ldots,\xi_t)\in A^{(t_0)}\}, \qquad A_2 = \{(\xi_{t+n},\ldots,\xi_{t+n+l-1})\in A^{(l)}\}$$

for given $A^{(t_0)}\in\mathscr{X}^{t_0}$, $A^{(l)}\in\mathscr{X}^l$. Inequality (14) yields

$$\begin{aligned}&|P_w(A_1\cap A_2)-P_w(A_1)P_w(A_2)|\\&= |P_{t_0+n+l-1}^{t-t_0+1}(w, A^{(t_0)}\times X^{(n-1)}\times A^{(l)}) - P_{t_0}^{t-t_0+1}(w, A^{(t_0)})P_l^{t+n}(w, A^{(l)})|\\&\leqslant (\varepsilon_n+\varepsilon_{n+t})P_{t_0}^{t-t_0+1}(w, A^{(t_0)}).\end{aligned}$$

Letting $t\to\infty$, we obtain

$$|P_{t_0+n+l-1}^\infty(A^{(t_0)}\times X^{(n-1)}\times A^{(l)}) - P_{t_0}^\infty(A^{(t_0)})P_l^\infty(A^{(l)})| \leqslant \varepsilon_n P_{t_0}^\infty(A^{(t_0)})$$

for all l, n, $t_0\in N^*$, $A^{(t_0)}\in\mathscr{X}^{t_0}$, $A^{(l)}\in\mathscr{X}^l$, or, equivalently,

$$|P_\infty(A_1\cap A_2)-P_\infty(A_1)P_\infty(A_2)| \leqslant \varepsilon_n P_\infty(A_1)$$

for all $A_1\in\mathscr{X}_{(1,t_0)}$, $A_2\in\mathscr{X}_{(t_0+n,t_0+n+l-1)}$, n, t_0, $l\in N^*$. It follows that

$$\varphi_\infty(n)\leqslant\varepsilon_n, \qquad n\in N^*,$$

where φ_∞ is defined analogously to φ_w (with P_w replaced by P_∞). ∎

There exists an extensive theory of limiting properties (central limit theorem, law of large numbers, law of iterated logarithm) of *strictly station-*

ary φ-mixing sequences. It is therefore important to note that there is a close relationship between limiting properties of $(\xi_n)_{n\in N^*}$ under P_∞ and the same properties under P_w.

For example, it is not difficult to prove that for any tail set Λ (this means that Λ belongs to the tail σ-algebra of $(\xi_n)_{n\in N^*}$, which is defined as the intersection of the σ-algebras $\mathscr{K}_{(n,\infty)}$(*), $n\in N^*$) we have $P_w(\Lambda) = P_\infty(\Lambda)$ for any $w\in W$. Next, if $(\alpha_n)_{n\in N^*}$ is a sequence of real numbers and $(\beta_n)_{n\in N^*}$ a sequence of positive numbers such that $\lim_{n\to\infty} \beta_n = \infty$, then by denoting

$$F_w^{(m,n)}(a) = P_w\left(\frac{f(\xi_m)+\ldots+f(\xi_{m+n-1})}{\beta_n} - \alpha_n < a\right),$$

$$F_\infty^{(n)}(a) = P_\infty\left(\frac{f(\xi_1)+\ldots+f(\xi_n)}{\beta_n} - \alpha_n < a\right),$$

where f is a real-valued $\mathscr{X}$-measurable function defined on X and a any real number, it can be shown that for arbitrary $m\in N^*$ and $w\in W$ the weak convergence of one of the sequences $(F_w^{(m,n)})_{n\in N^*}$ and $(F_\infty^{(n)})_{n\in N^*}$ to a distribution function F implies the weak convergence of the other sequence to the same F.

References for limiting properties of φ-mixing sequences (and RSCC's): Billingsley (1968), Grigorescu and Iosifescu (1982) Hall and Heyde (1980), Ibragimov and Linnik (1971), Iosifescu (1980), Iosifescu and Theodorescu (1969), Kesten and O'Brien (1976), Popescu (1977).

7

To conclude these notes we should remark that no topological assumptions were made on the measurable space $(W, \mathscr{W})$. In the case where W is a compact metric space a powerful operator-theoretical approach (remember the remark made on p. 254 concerning the probabilities $P^n(w, \cdot)$) has been developed. The details are to be found in the already cited books by Grigorescu and Iosifescu (1982), Iosifescu and Theodorescu (1969), and Norman (1972).

References

[1] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York 1968.
[2] D. Blackwell, *The stochastic process of Borel gambling and dynamic programming*, Ann. Statist. **4** (1976), 370–374.
[3] R. R. Bush and F. Mosteller, *A mathematical model for simple learning*, Psychol. Rev. **58** (1951), 313–323.
[4] —, —, *Stochastic Models for Learning*, Wiley, New York 1955.
[5] W. K. Estes, *Toward a statistical theory of learning*, Psychol. Rev. **57** (1950), 94–107.

(*) $\mathscr{K}_{(n,\infty)}$ is the σ-algebra generated by the random variables $\xi_n, \xi_{n+1}, \ldots$.

[6] S. Grigorescu and M. Iosifescu, *Dependence with Complete Connections and its Applications*, Scientific and Encyclopaedic Publishing House, Bucharest 1982, in Romanian.
[7] P. G. Hall and C. C. Heyde, *Martingale Limit Theory and Its Applications*, Academic Press, New York 1980.
[8] I. A. Ibragimov and Iu. V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, Groningen 1971.
[9] C. T. Ionescu Tulcea, *On a class of operators occurring in the theory of chains of infinite order*, Canad. J. Math. **11** (1959), 112–121.
[10] M. Iosifescu, *Random systems with complete connections with an arbitrary set of states*, Rev. Math. Pures Appl. **8** (1963), 611–645, addenda ibid. **9** (1964), 91–92, in Russian.
[11] –, *On the application of random systems with complete connections to the theory of f-expansions*, in *Progress in Statistics (European Meeting of Statisticians, Budapest, 1972)*, Colloq. Math. Soc. János Bolyai, vol. 9, North-Holland, Amsterdam 1974, 335–365.
[12] –, *Recent advances in the metric theory of continued fractions*, in: *Trans. 8th Prague Conf. Information Theory, Statist. Decision Functions, Random Processes*, Vol. A, Academia, Prague 1978, 27 40.
[13] –, *Recent advances in mixing sequences of random variables*, in: *3d Internat. Summer School Probab. Theory, Math. Statist. (Varna, 1978)*, Publishing House Bulgarian Acad. Sci., Sofia 1980, 111–138.
[14] M. Iosifescu and R. Theodorescu, *Random Processes and Learning*, Springer-Verlag, Berlin 1969.
[15] H. Kesten and G. L. O'Brien, *Examples of mixing sequences*, Duke Math. J. **43** (1976), 405–415.
[16] R. Kuzmin, *Sur un problème de Gauss*, in: *Atti Congr. Internaz. Mat. (Bologna, 1928)*, Tomo VI, Zanichelli, Bologna 1932, 83–89.
[17] S. Lakshmivarahan, *Learning Algorithms. Theory and Applications*, Springer-Verlag, New York 1981.
[18] P. Lévy, *Sur les lois de probabilités dont dépendent les quotients complets et incomplets d'une fraction continue*, Bull. Soc. Math. France **57** (1929), 178–194.
[19] C. Macchi, *Modèles d'apprentissage*, in: J. C. Simon (ed.), *Computer Oriented Learning Processes*, Noordhoff, Leyden 1976, 25–67.
[20] J. Neveu, *Bases mathématiques du calcul des probabilités*, Masson, Paris 1964.
[21] M. F. Norman, *Markov Processes and Learning Models*, Academic Press, New York 1972.
[22] O. Onicescu and G. Mihoc, *Sur les chaînes de variables statistiques*, Bull. Sci. Math (2) **59** (1935), 174–192.
[23] G. Popescu, *Functional limit theorems for random systems with complete connections*, in *Proc. Fifth Conference on Probability Theory (Brasov, 1974)*, Publishing House of the Romanian Academy, Bucharest 1977, 261–275.
[24] U. Rieder, *Bayesian dynamic programming*, Adv. in Appl. Probab. **7** (1975), 330–348.
[25] H. Robbins and S. Monro, *A stochastic approximation method*, Ann. Math. Statist. **22** (1951), 400–407.
[26] H. Rouanet, *Les modèles stochastiques d'apprentissage*, Gauthier–Villars, Paris 1967.
[27] W. L. Smith and K. Wilkinson, *On branching processes in random environments*, Ann. Math. Statist. **40** (1969), 814–827.
[28] E. Wirsing, *On the theorem of Gauss–Kusmin–Lévy and a Frobenius-type theorem for function spaces*, Acta Arith. **24** (1974), 507–528.

Presented to the semester
Sequential Methods in Statistics
September 7–December 11, 1981