ON CHAINS OF INFINITE ORDER

T. E. HARRIS

1. **Introduction.** We consider stationary[1] stochastic processes Z_n, $n=0, \pm 1, \cdots$, where Z_n can take D distinct values, $D \geq 2$. It is convenient to let the values be $Z_n = 0, 1, \cdots, D-1$. Let u be any sequence of integers, $u = (u_1, u_2, \cdots)$. Then the transitions of the process are described by the functions $Q_i(u)$,

$$(1.1) \qquad Q_i(u) = P(Z_n = i \mid Z_{n-1} = u_1, Z_{n-2} = u_2, \cdots) , \quad i = 0, 1, \cdots, D-1 .$$

Our aim will be to relate some stochastic properties of the Z_n-process to functional properties of the $Q_i(u)$. Because of the fact that the future behavior of Z_n depends in general on its complete past history, we shall refer to these processes as stationary infinite-order chains.

The first systematic study of such chains was made by Onicescu and Mihoc [13], and was carried on in further papers [14], [15], and [16] by Onicescu and Mihoc, and [12] by Onicescu. These authors considered chains of a somewhat more special type which they called *chaînes à liaisons complètes*. Further results were obtained by Doeblin and Fortet [6], who applied the term *chaîne à liaisons complètes* to any chain for which the relations

$$P(Z_n = i \mid Z_{n-1} = u_1, \cdots, Z_{n-k} = u_k)$$

are specified for every sequence $u_1, \cdots, u_k, k = 1, 2, \cdots, \infty$. See also Fortet [8] and Ionescu Tulcea and Marinescu [17].

The authors cited prove, under various hypotheses on the functions Q_i of (1.1), that $P(Z_n = i \mid Z_{-1} = u_1, \cdots, Z_{-k} = u_k)$ has a limit as $n \to \infty$, and obtain various other generalizations of the limit theorems for Markov chains. Also, in [6] the case of cyclic motions is considered. We shall not treat this case. The case of infinitely many states, stronger hypotheses, is treated in [17].

Our point of view is somewhat different. We introduce the random variables X_n, $n = 0, \pm 1, \cdots$, defined by

[1] Throughout this paper a "stationary" process, Markov or not, will be a process which not only has transition laws independent of time but also has a stationary absolute distribution.

$$(1.2) \qquad\qquad X_n = \sum_{j=1}^{\infty} Z_{n-j}/D^j \ .$$

That is, X_n is the number whose representation in the D-ary numeral system is $.Z_{n-1}Z_{n-2}\cdots$. (If we make the proper conventions, ambiguities will have total probability 0.) Thus X_n contains the complete past history of the Z_n process, and is a Markov process whose transition probabilities are defined in the following way. Let $0 \leq x \leq 1$ be a number whose D-ary expansion is

$$x = .u_1 u_2 \cdots \ .$$

Observe that $X_n = x$ is equivalent to $Z_{n-1} = u_1$, $Z_{n-2} = u_2$, $\cdots$. Now if $Z_n = i$, then $X_{n+1} = .i u_1 u_2 \cdots = (i+x)/D$. Now let $f_i(x)$ be functions of x defined by

$$(1.3) \qquad\qquad f_i(x) = Q_i(u) \ , \qquad\qquad i = 0, \cdots, D-1$$

where $.u_1 u_2 \cdots$ is the D-ary expansion of x and the Q_i are defined by (1.1). If $X_n = x$, then X_{n+1} is formed by applying with probability $f_i(x)$ the transformation $[i + (\quad)]/D$ to X_n; that is

$$(1.4) \qquad\qquad P\Big(X_{n+1} = \frac{i+x}{D}\ \Big|\ X_n = x\Big) = f_i(x) \ .$$

The representation (1.2) was used by Borel [3] for the case where the Z_n are independent and equidistributed. Apparently it has not been systematically exploited for other cases, although an abstract analogue of (1.2) is used in [17]. The representation (1.2) has the advantage that Fourier and Laplace transform methods can be used to deal with the distribution of the complete past history of the Z_n-process.

After making precise the relation between the Z_n- and X_n-processes, we show the existence of a unique stationary Z_n-process whose conditional probabilities

$$P(Z_n = i \mid Z_{u-1} = u_1, \cdots)$$

are equal to specified functions Q_i, provided the latter satisfy certain conditions. This extends a result of Doeblin and Fortet. Next we study the distribution $G(x)$ of X_n. It is shown that this has one of three forms, provided certain general conditions of mixing behavior hold. (1) $G(x)$ has a single jump of magnitude 1 at one of the points $i/(D-1)$, $i = 0, \cdots, D-1$. This is true if and only if $P(Z_n = i) = 1$. (2) $G(x) = x$, $0 \leq x \leq 1$. This is true if and only if the Z_n are independent and equidistributed on $0, 1, \cdots, D-1$. (3) $G(x)$ is continuous and purely singular.[2]

---

[2] The fact that G is singular if the Z_n are independent and not equidistributed was pointed out to the author by Henry Scheffé.

Next we consider processes which we shall call *grouped Markov chains*. Let Y_n, $n=0$, ± 1, $\cdots$, be the variables of a stationary Markov chain whose states are divided into D mutually exclusive and exhaustive nonempty subsets B_0, B_1, $\cdots$, B_{D-1}. Define $Z_n=i$ when $Y_n \in B_i$, $i=0, 1$, $\cdots$, $D-1$. We shall refer to this type of Z_n-process as a grouped Markov chain; it is in general not Markovian. We study such chains for the case where Y_n has a finite number of possible states and where each element of the transition matrix of the Y_n-process is positive. Using the Laplace transform, we show how to determine the functions

$$P(Z_n=i \mid Z_{n-1}=u_1, \cdots, Z_{n-k}=u_k)$$

and

$$P(Z_n=i \mid Z_{n-1}=u_1, Z_{n-2}=u_2, \cdots) ,$$

as well as the corresponding functions of a real variable $f_i(x)$ given by (1.3). This may be considered a solution of the prediction problem for grouped Markov chains.

The X_n-process is closely related to models which have recently been used for learning and decision processes by Bush and Mosteller [4], Bales and Householder [1], Flood [7], and others. The author wishes to thank these men for stimulating the present line of work.

Theorem 3 can be extended to certain types of these "learning models." A discussion of certain learning models has been given by Bellman, Harris, and Shapiro [2] and by Karlin [11]. Karlin's work has points of contact with ours.[3]

2. **Relation of the Z_n- and X_n-processes.** In this section we make explicit the relation between the Z_n- and X_n-processes and give a general condition which implies the existence of a Z_n-process with prescribed Q_i. Later sections will show that this condition is satisfied in many instances.[4]

Let $D \geq 2$ be an integer and let $u=(u_1, u_2, \cdots)$ represent a sequence of integers with $0 \leq u_j \leq D-1$. Let $Q_i(u)$ be functions of u, $i=0, 1, \cdots$, $D-1$, with

(2.1) $$Q_i(u) \geq 0 , \qquad\qquad i=0, \cdots, D-1 ,$$

(2.2) $$\sum_{i=0}^{D-1} Q_i(u) \equiv 1 .$$

Now if x is a real number, $0 \leq x \leq 1$, we adopt the following convention about the D-ary expansion of x in the ambiguous cases. The

---

[3] See § 4.

[4] Further discussion of the relationship follows Theorem 6.

D-ary expansion of $x=1$ will be taken as $x=.(D-1)(D-1)\cdots$. In all other ambiguous cases, an expansion terminating in 0's will be preferred to one terminating in $(D-1)$'s. Thus in the decimal system the expansion of $x=1$ will be $.999\cdots$ while the expansion of $x=1/2$ will be $.5000\cdots$ rather than $.499\cdots$. Thus the D-ary expansion of x is unambiguously defined.

Now define functions $f_i(x)$, $0\leq x\leq 1$, by

$$(2.3) \qquad\qquad f_i(x)=Q_i(u)$$

where $x=.u_1u_2\cdots$.

THEOREM 1. *Suppose functions $Q_i(u)$ are given satisfying (2.1) and (2.2) and such that the $f_i(x)$ defined by (2.3) are Borel-measureable; suppose there exists a distribution $G(x)$, $G(0-)=0$, $G(1)=1$, which satisfies the functional equation*

$$(2.4) \qquad\qquad G(x)=\sum_{j=0}^{D-1}\int_0^{Dx-j} f_j(y)dG(y) , \qquad\qquad 0\leq x\leq 1 .$$

Then there exists a stationary process $\cdots Z_0, Z_1, \cdots$, such that Z_n has possible values $0, 1, \cdots, D-1$, and such that

$$(2.5) \qquad P(Z_n=i \,|\, Z_{n-1}, Z_{n-2}, \cdots)=Q_i(Z_{n-1}, Z_{n-2}, \cdots)$$

with probability 1.

Proof. We consider a real-valued Markov process $\cdots X_n, X_{n+1}, \cdots$ whose transition probabilities are given by

$$(2.6) \qquad P\!\left(X_n=\frac{i+x}{D}\,\middle|\,X_{n-1}=x\right)=f_i(x) , \qquad\qquad 0\leq x\leq 1 ,$$

where the $f_i(x)$ are the functions defined by (2.3). It can be verified that if G satisfies (2.4), then G is a stationary absolute distribution for this Markov process; we shall suppose that X_n has this distribution.

Define the function $h(x)$, $0\leq x\leq 1$, by

$$(2.7) \qquad\qquad h(x)=\text{1st digit in } D\text{-ary expansion of } x.$$

Now define random variables Z_n by

$$(2.8) \qquad\qquad Z_{n-1}=h(X_n) , \qquad\qquad n=0, \pm 1, \cdots$$

The Z_n then form a stationary process, whose nature is clearly completely determined by $G(x)$. It can be shown that

$$(2.9) \qquad P\!\left[X_n=\frac{Z_{n-1}}{D}+\frac{Z_{n-2}}{D^2}+\cdots\right]=1 ,$$

since $P[X_{n-1}=DX_n-h(X_n)]=1$ for all n. Also

$$(2.10) \qquad Q_i(Z_{n-1}, Z_{n-2}, \cdots) = f_i\left(\frac{Z_{n-1}}{D} + \frac{Z_{n-2}}{D^2} + \cdots\right)$$

holds with probability 1. The only sequences $(Z_{n-1}, Z_{n-2}, \cdots)$ for which the two sides of (2.10) might be different are sequences other than $(D-1, D-1, \cdots)$ which terminate in unbroken $(D-1)$'s, and these can be shown to have probability 0.

It can be shown from this that $Q_i(u)$ is a permissible version of $P(Z_n = i \mid Z_{n-1} = u_1, Z_{n-2} = u_2, \cdots)$.

3. Continuity properties of the Q_i. We assume that functions Q_i are given satisfying (2.1) and (2.2) and that functions f_i are then defined by (2.3).

We shall refer to a point x whose D-ary expansion terminates in an unbroken sequence of 0's as a *lattice point*.

If $u^n = (u_1^n, u_2^n, \cdots)$ is a sequence for each $n = 1, 2, \cdots$, then $u^n \to u$ will mean that for each k, $u_k^n = u_k$ for all n sufficiently large.

CONDITION A. *For each i and u, $u^n \to u$ implies $Q_i(u^n) \to Q_i(u)$ as $n \to \infty$.*

THEOREM 2. *Under Condition A the $f_i(x)$ are continuous to the right for each x, $0 \leq x < 1$, and continuous to the left except possibly at lattice points. Left-continuity holds at $x = 1$.*

COROLLARY. *Under Condition A the $f_i(x)$ are Borel-measurable (in fact, belong to Baire class 1.).*

The proof follows from the definition of the $f_i(x)$. The corollary follows from the well-known fact that a function with only countably many discontinuities belongs to Baire class 1.

4. Existence of stationary Z_n- and X_n-processes. Our procedure will be as follows. We consider a Markov process X_n with transition probabilities defined by (2.6), where the $f_i(x)$ are given functions. We give conditions on the $f_i(x)$ which insure that the probabilities $P(X_n \leq x \mid X_0 = y)$ are C-1 summable to a distribution $G(x)$ which is independent of y. The distribution $G(x)$ satisfies (2.4) and is the only stationary distribution for the X_n-process.

Now let functions $Q_i(u)$ be given satisfying (2.1) and (2.2). Making use of Theorem 1 we show that under certain restrictions on the Q_i there is a uniquely determined stationary process Z_n satisfying (2.5) with probability 1. This process is ergodic. It is discussed in Theorem 6.

Under somewhat stronger conditions Doeblin and Fortet proved essentially that

$$\lim_{n \to \infty} P(Z_n = i \mid Z_{-1}, Z_{-2}, \cdots)$$

exists with probability 1 and is independent of Z_{-1}, Z_{-2}, $\cdots$.[5] We shall show the C-1 anologue of this under the weaker conditions.

The method we use is a development of one used originally by Doeblin in [5].

Now consider given functions $f_i(x) \geq 0$, $i=0, \cdots, D-1$; $0 \leq x \leq 1$; $\sum f_i(x) \equiv 1$. We use the notation

$$(4.1) \qquad\qquad (x \equiv y)_m$$

to mean that the first m digits in the D-ary expansion of x are the same as the first m in the expansion of y. We define

$$(4.2) \qquad\qquad \varepsilon_m = \sup_{i,\,(x \equiv y)_m} |f_i(x) - f_i(y)| , \qquad\qquad m = 0, 1, \cdots .$$

Doeblin and Fortet used a condition which would be equivalent in the present context to

$$(4.3) \qquad\qquad \sum_{m=0}^{\infty} \varepsilon_m < \infty .$$

We shall use Condition B, expressed by the requirements

$$(4.4) \qquad\qquad \lim_{n \to \infty} \varepsilon_m = 0 ,$$

$$(4.5) \qquad\qquad \sum_{m=0}^{\infty} \prod_{k=0}^{m} \left(1 - \frac{1}{2} D \varepsilon_k \right) = \infty .$$

We shall understand that any of the factors $\left(1 - \frac{1}{2} D \varepsilon_k \right)$ in (4.5) which

is zero or negative will be replaced by 1. As an example, Condition B is satisfied provided we have for sufficiently large k

$$\varepsilon_k \leq \frac{2}{Dk} .$$

In addition to Condition B, some sort of condition of positivity will be required. We shall choose the simplest one.

CONDITION C. For some i, $f_i(x) \geq \Delta > 0$, $\qquad\qquad\qquad 0 \leq x \leq 1$.

It is easy to see how C can be replaced by weaker conditions. For example, in the case $D=2$, $f_0(x)=x$, Condition C is not satisfied but it will be clear from the subsequent arguments that a condition sufficiently like C is satisfied.

THEOREM 3. *Let $f_i(x)$, $i=0, \cdots, D-1$, be nonnegative functions with*

[5] Simple examples show that the existence of limiting probabilities does not, in general, imply the existence of a stationary distribution. The existence of at least one stationary Z_n-process can be shown under quite weak conditions. The difficulty is to show uniqueness.

$\sum f_i \equiv 1$. *Let* X_n, $n=0, 1, \cdots$, *be the variables of a Markov process with* $X_0=y$, $0 \le y \le 1$ *and with transition law defined by* (2.6).

Define

$$(4.6) \qquad G_n(y\,;\,x)=P(X_n \le x \mid X_0=y) \ .$$

Then Conditions B *and* C *imply that* $G_n(y\,;x)$ *is summable C–1 to a distribution* $G(x)$ *which is independent of* y. *If* (4.3) *holds then the ordinary limit exists. In either case the limit is uniform in* y.[6]

For the proof of Theorem 3 we require the following lemma about sums of (not necessarily independent) random variables.

LEMMA 1. *Let* x_1, x_2, $\cdots$, *be positive integer-valued random variables. Let* $s_n=x_1+\cdots+x_n$ *and let* u_m *be the probability that for some* j *we have* $s_j=m$, $m=1, 2, \cdots$. *Suppose*

$$(4.7) \qquad P(x_n > i \mid x_1, x_2, \cdots, x_{n-1}) \ge R_i$$

where the R_i *are nonnegative numbers which are independent of* $x_1, \cdots,$ x_{n-1} *and* n *and satisfy*

$$(4.8) \qquad \sum_{i=1}^{\infty} R_i = \infty \ .$$

Then

$$(4.9) \qquad \lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} u_m = 0 \ .$$

The proof of the lemma, which is closely related to a standard renewal theorem, is simple, and is omitted.

Proof of Theorem 3. The method is related to an idea of Doeblin [5], who proved the ergodic theorem for Markov chains with a finite number of states by considering two particles starting in different states, which move independently until they simultaneously occupy the same state, after which they merge. An idea similar to Doeblin's original one is used in [6], and a related device has been used by Hodges and Rosenblatt [9].

In order not to obscure the main idea by details we give the proof for the case $D=2$. Since Condition C holds we can just as well take $f_0(x) \ge \varDelta > 0$. Then

$$\left| \frac{1}{2} D\varepsilon_k \right| = |\varepsilon_k| < 1 \ , \qquad\qquad k=0, 1, \cdots \ .$$

Let $t_0, t_1, \cdots$, be independent random variables uniformly distributed

_____
[6] We use conditional probabilities $G_n(y;x)$, etc., to mean those probabilities which are uniquely determined by the Markov transition operator, starting from a given value y. They are thus uniquely defined for all y.

on $(0, 1)$. Define processes X_n and X'_n as follows: $X_0=y$, $X'_0=y'$, $0 \leq y, y'$ ≤ 1. Suppose X_n and X'_n are determined. Then

$$X_{n+1} = \begin{cases} \dfrac{1}{2} X_n & \text{for } t_n \leq f_0(X_n) \\[2mm] \dfrac{1}{2} + \dfrac{1}{2} X_n & \text{for } t_n > f_0(X_n) \,; \end{cases}$$

while

$$X'_{n+1} = \begin{cases} \dfrac{1}{2} X'_n & \text{for } t_n \leq f_0(X'_n) \\[2mm] \dfrac{1}{2} + \dfrac{1}{2} X'_n & \text{for } t_n > f_0(X'_n) \,. \end{cases}$$

It is convenient to let U_n (U'_n) designate the transformation applied to X_n (X'_n). That is, $U_n = i$ $[U'_n = i]$ if $X_{n+1} = (i + X_n)/2$ $[X'_{n+1} = (i + X')/2]$. Then

(4.10) $P(U_n \neq U'_n \mid X_n, X'_n) \leq |f_0(X_n) - f_0(X'_n)|$.

From (4.10) we then have

(4.11) $P(U_n = U'_n, U_{n+1} = U'_{n+1}, \cdots, U_{n+k} = U'_{n+k}) \geq (1 - \varepsilon_0)(1 - \varepsilon_1) \cdots (1 - \varepsilon_k)$,

independently of X_n, X'_n.

Now the event $\{U_n = U'_n, \cdots, U_{n+k} = U'_{n+k}\}$ implies[7]

(4.12) $(X_{n+k+1} \equiv X'_{n+k+1})_{k+1}$

which in turn implies

(4.13) $|X_{n+k+1} - X'_{n+k+1}| \leq 2^{-k-1}$.

Let us say that an "engagement" occurs on the nth step if $U_{n-1} \neq U'_{n-1}$, $U_n = U'_n$. If we interpret the random variables $x_1, x_2, \cdots$ of Lemma 1 as the intervals between successive engagements, we see from (4.11), (4.12), Conditions B and C, and Lemma 1 that

(4.14) $\lim\limits_{N \to \infty} \dfrac{\text{Expected no. engagements in 1st } N \text{ steps}}{N} = 0$,

the limit in (4.14) being uniform in the starting points y and y'.

It can be shown from (4.14) that for any $\varepsilon > 0$, we have,

(4.15) $\lim\limits_{N \to \infty} \dfrac{1}{N} \sum\limits_{n=0}^{N-1} P(|X_n - X'_n| > \varepsilon) = 0$,

---

[7] A slight modification is necessary if $y = 1$ or $y' = 1$.

uniformly in y and y'. The argument is roughly as follows. Whenever $U_n \approx U_n'$, the length of time till an engagement occurs is small, with uniformly high probability, because of Condition C. Therefore, if only a small number of engagements have occurred in the first N steps, where N is large, then the probability is high that $U_N = U_N'$, $U_{N-1} = U_{N-1}'$, $\cdots$, $U_{N-k} = U_{N-k}'$ where k is large. Thus (4.15) follows from (4.13). (It is easy to make this argument precise.) A simple type of argument then shows that $P(X_n \leq x | X_0 = y)$ is C-1 summable to a distribution $G(x)$ which is independent of y. Moreover the difference

$$(4.16) \qquad \left| \frac{1}{N} \sum_{n=0}^{N-1} P(X_n \leq x \mid X_0 = y) - G(x) \right|$$

goes to zero uniformly in y at all points of continuity of $G(x)$.

If the stronger condition (4.3) holds, as well as Condition C, we can replace (4.15) by the stronger statement

$$(4,17) \qquad P(|X_n - X_n'| > \varepsilon) \to 0 \; .$$

In fact, with probability 1 we have $U_n = U_n'$ for all sufficiently large n in this case. We then get actual convergence, rather than just C-1 summability, of the distributions to $G(x)$.

THEOREM 4. *Assume that Conditions* B *and* C *hold. Then* $G(x)$ *of Theorem* 3 *either has a single discontinuity of magnitude* 1 *at one of the points* 0, $1/(D-1)$, $2/(D-1)$, $\cdots$, 1 *or is continuous.*

Proof. First let i in Condition C be 0. If $f_0(0) = 1$ it is clear that $G(x)$ has a jump of magnitude 1 at $x = 0$, and conversely. If $f_0(0) < 1$, $G(x)$ is everywhere continuous. First, $G(x)$ must be continuous at 0. For let K and n be integers, $0 < K < n$. Consider an X_n-process with an arbitrary starting point $X_0 = y$. If the D-ary expansion of X_n begins with K 0's then $U_{n-1} = U_{n-2} = \cdots = U_{n-K} = 0$. Hence

$$(4.18) \qquad P(X_n < D^{-K}) \leq P(U_{n-1} = U_{n-2} = \cdots = U_{n-K} = 0) \; .$$

Now no matter what is the value of $X_{n-K} = \bar{x}$, we have

$$(4.19) \; P(U_{n-1} = \cdots = U_{n-K} = 0 | X_{n-K} = \bar{x}) = f_0(\bar{x}) f_0(\bar{x}/2) f_0(\bar{x}/2^2) \cdots f_0(\bar{x}/2^{K-1}) \; .$$

Because $f_0(0) < 1$ and $f_0(x)$ is continuous at 0, the right side of (4.19) $\to 0$ as $K \to \infty$, uniformly in $\bar{x}$. Using (4.18) and (4.19), we have continuity of $G(x)$ at 0.

Similar arguments show continuity of $G(x)$ at other points x. The argument is almost the same if the i of Condition C is not 0.

THEOREM 5. *Under the conditions of Theorem* 3, $G(x)$ *is a stationary absolute distribution for the* X_n-*process and satisfies* (2.4). *It is the only stationary distribution.*

Simple examples show that if the $f_i(x)$ do not satisfy the proper continuity conditions, there can be a limiting distribution independent of the starting point which is nevertheless not a stationary distribution.

The uniqueness of the stationary distribution, once its existence is known, is an immediate consequence of the existence of a C-1 limiting distribution for X_n uniformly independent of X_0.

In the case $f_i(i/(D-1))=1$ it is readily verified that $P(X_n=i/(D-1)) =1$ is a stationary distribution satisfying (2.4). We can thus limit ourselves below to the case where $G(x)$ is continuous. (Theorem 4.)

Instead of starting with a fixed value for X_0 it is now convenient to give X_0 an arbitrary continuous distribution $G_0(x)$ assigning probability 1 to the interval (0, 1). Letting $G_n(x)=P(X_n\leq x)$ we have

$$(4.20) \qquad G_{n+1}(x)=\sum_{j=0}^{D-1}\int_0^{Dx-j} f_j(y)dG_n(y) , \qquad n=0, 1, \cdots.$$

Then $G_n(x)$ is continuous for each n, and we know from Theorem 3 that $G_n(x)$ is summable to $G(x)$. It follows from Condition B that it is justified to pass to the limit under the integral sign in (4.20) (C-1 limit if necessary), and Theorem 5 follows.

We can now give the main results of the present section. As before u and u' will denote sequences of integers between 0 and $D-1$ inclusive. For convenience we let V denote the set of all sequences which terminate in unbroken $(D-1)$'s, with the single exception of the sequence, each of whose members is $D-1$. For any stationary process whatever it can be shown that

$$\text{Prob}\,[(Z_{-1}, Z_{-2}, \cdots)\in V]= \text{Prob}\,[(Z_1, Z_2, \cdots)\in V]=0 .$$

We use the notation $(u\equiv u')_m$ to mean that the first m elements in the u sequence are the same as the first m elements in the u' sequence.

Now let $Q_i(u)$ be nonnegative functions of u with $\sum_{i=0}^{D-1}Q_i(u)\equiv 1$ and define

$$(4.21) \qquad \varepsilon_m= \sup_{\substack{i, (u\equiv u')_m \\ u\notin V, u'\notin V}} |Q_i(u)-Q_i(u')|.$$

Then the quantities ε_m defined by (4.21) are identical with those defined by (4.2) if functions $f_i(x)$ are defined by (2.3).

We shall say that the Q_i satisfy Condition B if (4.4) and (4.5) are satisfied. These are requirements that the future is conditioned only slightly by the remote past. Condition C will mean that for some j

$$(4.22) \qquad Q_j(u)\geq \Delta > 0 , \qquad u\notin V.$$

Now let I be a finite sequence of integers,

$$I = (i_0, i_1, \cdots, i_k)$$

and let

(4.23) $$Q_I^n(u) = P(Z_n = i_0, \cdots, Z_{n+k} = i_k \mid Z_{-1} = u_1, \cdots) .$$

The quantities $Q_I^n(u)$ are to be interpreted as defined, relative to the "past" u, by means of the $Q_i(u)$; they thus have meaning even before it is known that there is a stationary absolute distribution.

THEOREM 6. *Let the functions $Q_i(u)$ satisfy Conditions* B *and* C. *Then*

    a) *there exists a stationary process Z_n such that (1.1) holds with probability* 1;

    b) *this is the only stationary process for which (1.1) holds*;

    c) *the C-1 limit of $Q_I^n(u)$ exists for every u (except those in the set V defined above) for every I, and is equal to the stationary measure of I. The C-1 limit is approached uniformly in u.*

    d) *For every u not in V we have, for each $i = 0, 1, \cdots, D-1$,*

(4.24) $$\lim_{k \to \infty} P(Z_0 = i \mid Z_{-1} = u_1, \cdots, Z_{-k} = u_k) = Q_i(u) ,$$

provided the left side of (4.24) is defined for each k.

Proof. Define functions $f_i(x)$ by (2.3). From Theorem 3 there is a unique distribution $G(x)$ satisfying (2.4). From Theorem 1 there exists a stationary Z_n-process for which (1.1) holds with · probability 1. As remarked in § 2, the nature of the Z_n-process is determined by the distribution G. Hence, since G is uniquely determined, so is the Z_n-process. This proves (a) and (b) above.

The proof of (c) is an immediate consequence of the relation between the Z_n- and X_n-processes, together with Theorems 3 and 5. A slight modification is required if $u = (1, 1, \cdots)$.

The relation in (d) above is, it is well known, true for almost all u. A simple argument shows that it holds for every u not in V.

5. Further properties of $G(x)$. We now change our point of view somewhat. Suppose we are given a stationary infinite-order chain Z_n as defined in the introduction. Define

(5.1) $$X_n = Z_{n-1}/D + Z_{n-2}/D^2 + \cdots , \qquad n = 0, \pm 1, \cdots .$$

Then X_n is a stationary process.

We shall further suppose throughout § 5 that the Z_n-process is of the mixing type.[8] The X_n-process then is likewise.

Let the functions $Q_i(u)$ be defined by

$$Q_i(u) = P(Z_n = i \mid Z_{n-1} = u_1, Z_{n-2} = u_2, \cdots) .$$

---

[8] See [**10**, p. 36]. Roughly, if A and B are events, and $B(n)$ is the event B translated n units in time, then $P[AB(n)] \to P(A)P(B)$.

As in §2 we then define functions $f_i(x)$ by

(5.2) $f_i(x) = Q_i(u)$, $i = 0, \cdots, D-1$,

where $.u_1 u_2 \cdots$ is the D-ary expansion of x. The functions $Q_i(u)$ are defined at least for almost all u-sequences (" almost all " in the sense of the measure on sequences in the Z_n-process.)

Let $G(x)$ be the distribution of X_n. It is then clear that the functions $f_i(x)$ are defined for almost all x (G-measure). It is also readily seen that the X_n-process is Markovian and that $G(x)$ satisfies (2.4) with the $f_i(x)$ defined by (5.2).

Remark on uniqueness. Let G^* be a distribution satisfying (2.4), with $G^*(0-) = 0$, $G^*(1) = 1$, and suppose G^* is absolutely continuous with respect to G. Then G and G^* are identical. This follows from the general theory of Markov processes.

LEMMA 2. *Let Z_n be a stationary infinite-order chain as defined in the introduction. Suppose Z_n is mixing. Let*

$$X_n = \sum_{j=1}^{\infty} Z_{n-j}/D^j$$

and let $G(x)$ be distribution of X_n. Then G either has a single discontinuity of magnitude 1 at one of the points $0, 1/(D-1), \cdots, 1$ or is continuous.

The proof is similar to that of Theorem 4 and is omitted.

LEMMA 3. *Under the conditions of Lemma 2, $G(x)$, if it is continous, is either purely singular or purely absolutely continuous.*

Proof. Suppose we have the continuous case. To obtain a contradiction let us suppose

$$G = cG_1 + (1-c)G_2 ,   0 < c < 1 ,$$

where G_1 and G_2 are the singular and the absolutely continuous parts of G respectively, neither being identically zero.

If we write (2.4) in the operator form $G = TG$, then we have

(5.3) $c(G_1 - TG_1) = -(1-c)(G_2 - TG_2)$.

Now it is easily seen from the nature of T that TG_1 is singular and TG_2 is absolutely continuous. Moreover, neither $G_1 - TG_1$ nor $G_2 - TG_2$ can vanish identically. This follows from the remark above on uniqueness. Thus (5.3) is a contradiction.

LEMMA 4. *Under the conditions of Lemma 2, the f_i are determined uniquely by G up to a set of G-measure 0.*

For from (2.4) we have

(5.4) $G(x) - G\left(\dfrac{i}{D}\right) = \displaystyle\int_0^{Dx-1} f_i(y)\,dG(y)$, $\dfrac{i}{D} \leq x < \dfrac{i+1}{D}$; $i = 0, \cdots, D-1$.

Uniqueness of the f_i follows from (5.4) and the Nikodym-Radon theorem.

THEOREM 7. *Let Z_n be a stationary infinite-order chain of the mixing type. Let*

$$X_n = \sum_{j=1}^{\infty} Z_{n-j}/D^j$$

and let $G(x)$ be the distribution of X_n. Then $G(x)$ is one of the three following types.

(a) *$G(x)$ has a single jump of magnitude 1 at one of the points $i/(D-1)$, $i=0, \cdots, D-1$. This is true if and only if $P(Z_n=i)=1$.*

(b) *$G(x)=x$, $0 \leq x \leq 1$. This is true if and only if the Z_n are independent, each being equidistributed on $0, 1, \cdots, D-1$.*

(c) *$G(x)$ is continuous and purely singular.*

Proof. If $G(x)$ has any discontinuities, then (a) follows from Lemma 2. Next we introduce the moment-generating function (s is any complex number)

$$\phi(s) = \int_0^1 e^{sx} dG(x) .$$

From (2.4) it follows that ϕ satisfies

(5.5) $$\phi(Ds) = \phi(s) + \sum_{j=1}^{D-1} (e^{js}-1) \int_0^1 e^{sx} f_j(x) dG(x) .$$

Setting $s=2\pi ki$, $i=\sqrt{-1}$, we have

(5.6) $$\phi(2\pi kDi) = \phi(2\pi ki) , \qquad k = \pm 1, \pm 2, \cdots .$$

First suppose $\phi(2\pi ki)=0$, $k=\pm 1, \pm 2, \cdots$. Since $\phi(it)$ is the characteristic function of a distribution on $(0, 1)$, it is uniquely determined by its values at the points $2\pi ki$; hence in this case

$$\phi(it) = \frac{e^{it}-1}{it}$$

and $G(x)=x$. It can be verified directly that (2.4) is satisfied with $G(x)=x$ and $f_j(x)=1/D$. From Lemma 4, this is the only case where $G(x)=x$ can occur.

Next suppose that for some integer k we have $\phi(2\pi ki) \neq 0$. Iteration of (5.6) shows that $\phi(it)$ does not$\rightarrow 0$ as $t \rightarrow \infty$ or $t \rightarrow -\infty$ and hence G is not purely absolutely continuous. Thus Lemma 3 shows that G, if continuous and not of type (b), is purely singular.[9]

_____

[9] The fact that G has in general no absolutely continuous component can be seen from a simple argument not involving Fourier transforms.

6. Grouped Markov chains. Let Y_n be the variables of a Markov chain with a finite number of states, which we shall call 1, 2, $\cdots$, K. Let the transition matrix be $M=(p_{ij})$, i, $j=1$, $\cdots$, K. We assume $p_{ij}>0$. Otherwise, even if some power of M has all positive elements, there may be complications. We also assume $K>1$. Now let the states of the chain be divided into D mutually exclusive and exhaustive nonempty subsets B_0, $\cdots$, B_{D-1}. We can define an infinite-order chain Z_n by

$$(6.1) \qquad\qquad Z_n=i \rightleftarrows Y_n \in B_i \ .$$

We shall call such a process a *grouped Markov chain*. We shall be particularly interested in the case where the Y_n-process, and hence the Z_n-process, is stationary. We show that Conditions B and C are satisfied, determine the distribution of the "past" of the Z_n-process, and show how the functions $Q_i(u)$ and the corresponding $f_i(x)$, can be determined. The Q_i or f_i give the solution to the problem of predicting the future values of Z_n, given the past.

We first give a result about Markov chains.

THEOREM 8. *Let $M=(p_{ij})$ be the transition matrix of a Markov chain, $i,j=1$, $\cdots$, $K>1$; $p_{ij}>0$. Let Y_n be the variables of the chain. Let*

$$(6.2) \qquad\qquad \lambda = \min_{i,\,j,\,k,\,l} \frac{p_{kj}p_{il}}{K^2\, p_{ij}p_{kl}}$$

(Note that $0<\lambda<1$.) For each $n=1, 2, \cdots$, let A_n be a nonempty subset of states of the chain. Let g and h be two states. Then

$$(6.3) \qquad \left| P(Y_{n+1}\in A_{n+1}|Y_0=g,\ Y_1\in A_1,\ \cdots,\ Y_n\in A_n) \right.$$
$$\left. -P(Y_{n+1}\in A_{n+1}|Y_0=h,\ Y_1\in A_1,\ \cdots,Y_n\in A_n) \right| \leq (1-\lambda)^n\ ,$$
$$n=1, 2, \cdots\ .$$

The proof is omitted. It can be carried out with Doeblin's "two-particle" method.

It is readily shown that for *every* u, $0\leq u_i\leq D-1$, the limit

$$\lim_{k\to\infty} P(Z_0=i|\ Z_{-1}=u_1,\ \cdots,\ Z_{-k}=u_k)$$

exists, for the grouped Markov chain. We may take this limit as a permissible version of $Q_i(u)$ for the Z_n-process defined by (6.1). It can also be seen that

$$(6.4) \qquad\qquad |Q_i(u)-Q_i(u')|\leq (1-\lambda)^{m-1}\ , \qquad\qquad m=1, 2, \cdots\ ,$$

whenever the first m terms of u and u' coincide. Thus Condition B is satisfied with[10]

---

[10] The stronger condition (4.3) is of course, also satisfied.

$$\varepsilon_m \leq (1-\lambda)^{m-1} , \qquad\qquad m=1, 2, \cdots .$$

Condition C is a consequence of the obvious fact that the $Q_i(u)$ are uniformly positive.

THEOREM 9. *Let Z_n be defined by* (6.1). *Let*

$$(6.5) \qquad\qquad X_n = \sum_{j=1}^{\infty} Z_{n-j}/D^j$$

and let $G(x)$ be the distribution of X_n. Then $G(x)$ is continuous, $0 \leq x \leq 1$, and strictly increasing, $0 \leq x < 1$.

Theorem 7 is applicable since Z_n is of the mixing type. Since Z_n has a positive probability of taking at least two distinct values (we are assuming $D>1$), continuity follows. The strictly increasing character of G follows from the fact that the event $(Z_{-1}=u_1, \cdots, Z_{-k}=u_k)$ has positive probability for every sequence $0 \leq u_1, \cdots, u_k \leq D-1$.

DEFINITIONS. Let Y_n and Z_n be as in (6.1) and let X_n be defined by (6.5). Define

$$(6.6) \qquad H_j(x) = P(X_n \leq x \,|\, Y_n = j) , \quad 0 \leq x \leq 1, j=1, 2, \cdots, K,$$

$$(6.7) \qquad \theta_j(s) = \int_0^1 e^{sx} dH_j(x) , \qquad\qquad j=1, 2, \cdots, K,$$

s any complex number.

Let $\pi_j, j=1, \cdots, K$ be the (unique) set of stationary probabilities satisfying

$$(6.8) \qquad\qquad \pi_j = \sum_{r=1}^{K} \pi_r p_{rj} , \qquad\qquad j=1, \cdots, K.$$

Let p_{ij}^* be the set of inverse probabilities

$$(6.9) \qquad\qquad p_{ij}^* = p_{ji}\pi_j/\pi_i .$$

Let $M(s)$ be the matrix defined as follows:

$$(6.10) \qquad\qquad M(s) = (p_{ij}^* e^{s\nu(j)})$$

where $\nu(j)=k$ when j belongs to the group of states B_k.

THEOREM 10. (See preceding definitions.) *The function $\theta_j(s)$, $j=1, \cdots, K$, is the sum of the elements in the jth row of the convergent matrix product*

$$(6.11) \qquad\qquad M(s/D) \, M(s/D^2) \, M(s/D^3) \cdots .$$

Proof. Let Y_n^* be the variables of a stationary inverse Markov chain with transition probabilities given by (6.9) and let $Z_n^* = i$ when $Y_n^* \in B_i$. It is clear that Z_n^* is inverse to the Z_n-process in the sense

that the process Z^*_{-n} obeys the same probabilistic laws as the Z_n-process. Define

$$X^*_n = \sum_{r=1}^{\infty} Z^*_{n+r}/D^r \ .$$

(Incidentally X_n and X^*_n follow the same law, not inverse laws.)

It is clear that $H_j(x)$, as defined in (6.6) above, is also given by

$$(6.12) \qquad\qquad H_j(x) = P(X^*_n \leq x \mid Y^*_n = j) \ .$$

We now use (6.12) to find the functions $H_j(x)$.

Suppose r is an integer, $0 \leq r \leq D-1$, and suppose $r/D \leq x < (r+1)/D$; that is, $x = .r u_2 u_3 \cdots$. Then

$$(6.13) \quad H_j(x) = P(Z_1^*/D + Z_2^*/D^2 + \cdots \leq .r u_2 \cdots \mid Y_0^* = j)$$
$$= P(Z_1^* < r \mid Y_0^* = j)$$
$$\qquad + P(Z_1^* = r, \ Z_2^*/D + Z_3^*/D^2 + \cdots \leq .u_2 u_3 \cdots \mid Y_0^* = j)$$
$$= P(Z_1^* < r \mid Y_0^* = j) + \sum_{m \in B_r} p^*_{jm} H_m(Dx - r) \ , \qquad r/D \leq x < (r+1)D.$$

Next we note that X^*_n has the same distribution $G(x)$ as X_n. Moreover

$$(6.14) \qquad\qquad G(x) = P(X^*_n \leq x) = \sum_{r=1}^{K} \pi_r H_r(x) \ .$$

Since $G(x)$ is continuous (Theorem 9), the $H_r(x)$ must also be continuous. Now (6.13) implies the differential relationship

$$(6.15) \qquad\qquad dH_j(x) = \sum_{m \in B_r} p^*_{jm} \, d[H_m(Dx - r)] \ ,$$

$$\frac{r}{D} \leq x < \frac{r+1}{D} \ , \qquad\qquad j = 1, \cdots, K.$$

Defining $\theta_j(s)$ by (6.7) and letting $\Theta(s)$ be the column vector whose components are the θ_j, we see that (6.15) implies (multiplying both sides of (6.15) by e^{Dsx} and integrating)

$$(6.16) \qquad\qquad \Theta(Ds) = M(s)\Theta(s) \ ,$$

where $M(s)$ is defined in (6.10). Iterating (6.16) and replacing s by s/D gives

$$\Theta(s) = M(s/D) \cdots M(s/D^n)\Theta(s/D^n) \ .$$

Since $\theta_j(0) = 1$, $\Theta(s/D^n)$ approaches the column vector each of whose components is 1 as $n \to \infty$, while $M(s/D^n)$ approaches the stochastic matrix (p^*_{ij}). The powers $(p^*_{ij})^n$ converge exponentially as $n \to \infty$, and it is readily seen that the elements of the difference $M(s/D^n) - (p^*_{ij})$ are

$O(D^{-n})$, where O is uniform in s for any bounded s-region. Hence the matrix product in (6.11) converges uniformly in any bounded s-region, and Theorem 10 follows.

The $\theta_j(s)$ and the $H_j(x)$ can be calculated in various ways. One possibility is to determine the coefficients in the power-series expansions of the θ_j by differentiating (6.16) at $s=0$. The values of the θ_j on some interval near 0 on the imaginary axis can be calculated, and (6.16) can then be used to determine the θ_j on the rest of the imaginary axis.

We can now find the functions $Q_i(u)$ and $f_i(x)$ for grouped Markov chains. In theorem 11, Z_n is a grouped Markov chain as defined above.

THEOREM 11. *Let $u_1, \cdots, u_k$ be integers, $0 \leq u_j \leq D-1$. Then*

$$P(Z_n = i \mid Z_{n-1} = u_1, \cdots, Z_{n-k} = u_k) = \sum_{j \in B_i} \pi_j [H_j(x_2) - H_j(x_1)] / [G(x_2) - G(x_1)]$$

where

$$x_1 = u_1/D + u_2/D^2 + \cdots + u_k/D^k , \quad x_2 = u_1/D + \cdots + u_k/D^k + D^{-k} ,$$

and

$$G(x) = \sum_{j=1}^{K} \pi_j H_j(x) .$$

The proof is merely a reinterpretation of Theorem 10.

We thus have an expression for the conditional distribution of Z_n if a finite segment of the past is known.

Next we consider the situation when the complete past is known. Consider the X_n-process and the associated functions $f_i(x)$. Then, if $x = .u_1 u_2 \cdots$,

$$(6.17) \qquad f_i(x) = \frac{d}{dG(x)} \sum_{j \in B_i} \pi_j H_j(x) = P(Z_n = i \mid Z_{n-1} = u_1, Z_{n-2} = u_2, \cdots)$$

where (6.17) holds for every x, $0 \leq x < 1$, provided we take right-hand derivatives on the right side. Thus (6.17) gives the conditional distribution for Z_n if the complete past is known.

Example. Suppose

$$(p_{ij}) = \begin{pmatrix} \dfrac{1}{4} & \dfrac{1}{4} & \dfrac{1}{2} \\[2mm] \dfrac{1}{3} & \dfrac{1}{3} & \dfrac{1}{3} \\[2mm] \dfrac{1}{2} & \dfrac{1}{4} & \dfrac{1}{4} \end{pmatrix}$$

Then $\pi_1 = 4/11$, $\pi_2 = 3/11$, $\pi_3 = 4/11$, and in this case $p_{ij}^* = p_{ij}$. Take states 1 and 2 as B_0, state 3 as B_1, so that $D = 2$. The $\theta_j(s)$ then satisfy the equations

$$(6.18) \qquad \theta_1(2s) = \frac{1}{4}\theta_1(s) + \frac{1}{4}\theta_2(s) + \frac{1}{2}e^s\theta_3(s)$$

$$\theta_2(2s) = \frac{1}{3}\theta_1(s) + \frac{1}{3}\theta_2(s) + \frac{1}{3}e^s\theta_3(s)$$

$$\theta_3(2s) = \frac{1}{2}\theta_1(s) + \frac{1}{4}\theta_2(s) + \frac{1}{4}e^s\theta_3(s) ,$$

which, with the conditions $\theta_j(0) = 1$, determine them uniquely. The $H_j(x)$ can then be determined by Fourier inversion.

REFERENCES

1. R. F. Bales and A. S. Householder, unpublished.

2. R. Bellman, T. Harris, H. Shapiro, *Studies in functional equations occurring in decision processes*, RAND P–382 (1953).

3. E. Borel, *Sur les probabilités dénombrables et leurs applications arithmetiques*, Rend. Circ. Mat. Palermo, **26** (1909), 247–271.

4. R. Bush and F. Mosteller, *A stochastic model with applications to learning*, Ann. Math. Stat., **24** (1953), 559–585.

5. W. Doeblin, *Exposé de la théorie des chaines simples constantes de Markoff à un nombre fini d'états*, Rev. Math. de l'Union Interbalkanique, **2** (1938), 77–105.

6. W. Doeblin and R. Fortet, *Sur des chaînes à liaisons complètes*, Bull. Soc. Math. France, LXV (1937), 132–148.

7. M. M. Flood, *On stochastic learning theory*, RAND P–353 (1952).

8. R. Fortet, (première thèse), *Sur l'iteration des substitutions algebriques linéaires à une infinité de variables*, Revista de la Ciencias, No. **424**, Año XL, Lima, 1938.

9. J. L. Hodges and M. Rosenblatt, *Recurrence-time moments in random walks*, Pacific J. Math., **3** (1953), 127–136.

10. E. Hopf, *Ergodentheorie*, Chelsea, 1948.

11. S. Karlin, *Some random walks occurring in learning models*, Pacific J. Math., **3** (1953), 725–756.

12. O. Onicescu, *Théorie Générale des chaines à liaisons complètes*, Actualités Sci. Ind., **737**, IV, 4th part, section III, 29–41. Paris, Hermann, 1938.

13. O. Onicescu and G. Mihoc, *Sur les chaines de variables statistiques*, Bull. Sc. Math., **59** (1935), 174–192.

14. ————, *Sur les chaines statistiques*, C. R. Acad. Sci. Paris, **202** (1936), 2031–2033.

15. ————, *La dépendence statistique*, Actualités Sci. Ind. **503**, VII. Paris Hermann, 1937.

16. ————, *Comportement asymptotique des chaines à liaisons complètes*, Disqu. Math. Phys., **1** (1940), 61–62.

17. Ionescu Tulcea and G. Marinescu, *Sur certaines chaines à liaisons complètes*, C. R. Acad. Sci. Paris, **227** (1948), 667–669.

18. T. E. Harris, *A method for Markov chains*, (abstract). Ann. Math. Stat., **23** (1952), 141.

COLUMBIA UNIVERSITY
AND THE RAND CORPORATION