

# ASYMPTOTIC PROPERTIES OF MARKOFF TRANSITION PROBABILITIES

BY

J. L. DOOB

Let  $X$  be any space of abstract elements, and let  $\mathfrak{F}_x$  be any Borel field of  $X$  sets, including  $X$  itself. Let  $P^{(t)}(x, A)$  be a function of  $x \in X$ ,  $A \in \mathfrak{F}_x$ , and the positive real variable  $t$ , with the following properties:

- (a)  $0 \leq P^{(t)}(x, A) \leq P^{(t)}(x, X) = 1$ ;
- (b) for fixed  $x$  and  $t$ ,  $P^{(t)}(x, A)$  is a completely additive function of sets  $A$ ; for fixed  $A$  and  $t$ ,  $P^{(t)}(x, A)$  is measurable in  $x$  with respect to the field  $\mathfrak{F}_x$ .
- (c) with a self-explanatory notation for integration with respect to a set function,

$$(0.1) \quad P^{(s+t)}(x, A) = \int P^{(s)}(y, A) P^{(t)}(x, dG_y), \quad s, t > 0^{(1)}.$$

A function with these properties can be considered the transition probability function of a Markoff process; it is the probability of a transition into  $A$  from initial position  $x$ , in time  $t$ . The Markoff process is determined for  $t > t_0$  if a distribution of  $x$  at time  $t = t_0$  is assigned. The process is called stationary, or temporally homogeneous, if the initial distribution reproduces itself, that is, if the set function  $\Phi(A)$  defining the initial distribution satisfies the equation

$$(0.2) \quad \Phi(A) = \int P^{(t)}(x, A) \Phi(dG_x).$$

The probability formalization of a stochastic process is now well known. In the present case the initial distribution and the transition probabilities are used to define a probability measure in the space of all functions  $x(t)$ , where  $t \geq t_0$ , and  $x(t)$  is a function which takes on values in  $X$ . For example, to the set of functions satisfying the condition  $x(s) \in A$  ( $s > t_0$  is fixed) is assigned as measure

$$(0.3) \quad P\{x(s) \in A\} = \int P^{(s)}(x, A) \Phi(dG_x),$$

where  $\Phi(G)$  defines the initial distribution. If  $\Phi$  is self-reproducing, the definition in (0.3) and the corresponding definitions for more complicated sets be-

---

Presented to the Society, November 25, 1945, under the title *Markoff chains—denumerable case*; received by the editors February 27, 1947.

(1) In the following, the domain of integration will not be indicated explicitly when it is the whole space.

come independent of  $t_0$ , and the probability measure is then defined on the space of all functions  $x(t)$ ,  $-\infty < t < \infty$ ; in this case  $x(s)$ , for each  $s$ , has the  $\Phi$  distribution. The transformation  $T_s$  taking  $x(t)$  into  $x(t+s)$  is then a measure preserving transformation in function space, and the theory of measure preserving transformations is applicable.

In the following, it will be supposed that  $t$  runs through integral values only. Function space then reduces to sequence space. The results obtained will however be applicable, with the obvious changes, when  $t$  runs through all real numbers; for example the average  $(1/n) \sum_{j=1}^n x(j)$  is replaced by the average  $(1/t) \int_0^t x(s) ds$ .

If certain restrictions discussed below are imposed on the given transition probabilities, it follows that

$$(0.4) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t P^{(s)}(x, A) = Q(x, A)$$

exists for all  $x$ . Moreover:

(a) This limit is uniform in  $x$  and  $A$ .

(b) For fixed  $x$ ,  $Q(x, A)$  defines a self-reproducing probability distribution.

(c)  $X$  can be expressed as a sum of a finite number of disjunct sets,  $N, A_1, A_2, \dots$ , where

$$(0.5) \quad \begin{aligned} P^{(i)}(x, A_j) &= \begin{cases} 1, & x \in A_j, \\ 0, & x \in A_k, k \neq j; \end{cases} \\ Q(x, A) &= Q_j(A), & x \in A_j, \\ Q(x, N) &\equiv 0. \end{aligned}$$

The sets  $A_1, A_2, \dots$  are called final sets. If a simple type of cyclic transition is excluded, the Cesaro limit in (0.4) becomes an ordinary limit<sup>(2)</sup>.

In order to explain the significance of the conditions imposed on the transition probabilities to obtain these results, a familiar example will be recalled. Let  $X$  be any abstract space, and let  $\mathfrak{F}_x$  be a Borel field of  $X$  sets (including  $X$  itself). Let  $Tx$  be any one-to-one transformation of the points of  $X$  which together with its inverse takes  $\mathfrak{F}_x$  sets into  $\mathfrak{F}_x$  sets. Define  $P^{(i)}(x, A)$  as follows:

$$(0.6) \quad P^{(i)}(x, A) = \begin{cases} 1 & \text{if } T^i x \in A, \\ 0 & \text{if } T^i x \in X - A. \end{cases}$$

This function satisfies the three conditions (a), (b), (c), stated above, and therefore defines a Markoff process (in conjunction with an initial distribution). The average in (0.4) becomes the average number of times  $T^i x \in A$  as  $t$

<sup>(2)</sup> See [7] for a more complete statement and for details of the proof. Numbers in brackets refer to the bibliography at the end of the paper.

increases. It is thus hopeless to expect that without further hypotheses on the transition probabilities averages like that in (0.4) can behave in any more regular fashion than the corresponding averages for point transformations. If some probability measure, that is, a measure with value 1 on the whole space, has been defined on the sets of  $\mathfrak{F}_x$ , and if the transformation  $T$  is measure preserving, the measure of  $X$  sets defines a self-reproducing distribution. Even the hypothesis that a Markoff process has a self-reproducing distribution can however at best make the averages in (0.4) behave like the corresponding averages for measure preserving point transformations. The behavior detailed above, beginning with the existence of the limit in (0.4) for *all*  $x$ , is actually far more regular than the corresponding behavior for measure preserving point transformations. Now the transition probability  $P^{(t)}(x, A)$  defined in (0.6) is for fixed  $x$  and  $t$  a highly singular set function; it defines a distribution concentrated at a point. The restrictions imposed on the transition probabilities of Markoff processes by various authors can be considered as smoothing restrictions imposed to make the representation of  $P^{(t)}(x, A)$  in terms of a point transformation impossible.

If there is a self-reproducing distribution  $\Phi$ , it will be shown that under a mild restriction on the field  $\mathfrak{F}_x$ ,  $X$  can be expressed as the sum of at most a continuum of sets  $A_i$  and a remainder of  $\Phi$  measure 0, such that the  $A_i$ 's have essentially the properties (0.5). The present paper can be considered as an examination of the strengthening of this decomposition theorem which is made possible by the imposition of weak additional restrictions on the transition probability functions.

The theorems of the present paper will be stated, for the most part, in terms of stationary Markoff processes. Note that this is the same as stating theorems on Markoff transition probabilities for which it is known that there exist self-reproducing distributions. Although it is not customary for writers on Markoff processes to presuppose the existence of a self-reproducing distribution, the existence is implied in their hypotheses, since each  $Q_j(A)$  in (0.5) defines such a distribution.

The weakest conditions imposed on the transition probabilities to insure the truth of (0.4) and the subsequent paragraph are those of Doeblin and of Kryloff and Bogoliouff.

*Condition of Doeblin.* The space  $X$  is the interval (0.1). The field  $\mathfrak{F}_x$  is the field of Borel subsets of this interval. It is supposed that there is an  $s$  and a positive  $\epsilon$  such that  $P^{(s)}(x, A) \leq 1 - \epsilon$  if  $A$  has Lebesgue measure less than  $\epsilon$ .

*Condition of Kryloff and Bogoliouff.* The space  $X$  and the field  $\mathfrak{F}_x$  are as in Doeblin's condition. Let  $P^{(s)}$  be the linear operator defined by the kernel  $P^{(s)}(x, A)$  taking the Banach space  $\mathfrak{M}$  of completely additive set functions defined on  $\mathfrak{F}_x$  into itself,

$$(0.7) \quad P^{(s)}\Phi = \int P^{(s)}(x, A)\Phi(dG_x).$$

(The norm of an element of  $\mathfrak{M}$  is its variation.) It is supposed that there is an  $s$  and a completely continuous linear operator  $V$ , also taking  $\mathfrak{M}$  into itself, such that  $P^{(s)} - V$  has norm less than 1. Note that  $P^{(s)}$  is the  $s$ th power of  $P^{(1)}$ <sup>(3)</sup>.

Although the conditions of Doeblin and of Kryloff and Bogoliouboff are very weak, they are insufficiently weak to cover some of the simplest cases commonly encountered. In these cases the uniformity in  $x$  of approach to the limit in (0.4) is not true. One such example is furnished by the Markoff process encountered in renewal theory (see the following paper in this volume). Another example is the Gaussian stationary process with correlation function  $\rho^t$ , one of the simplest Markoff processes. Here  $X$  is the real line,  $\mathfrak{F}_x$  the field of Borel sets, and

$$(0.8) \quad P^{(t)}(x, A) = \frac{1}{[2\pi(1 - \rho^{2t})]^{1/2}} \int_A \exp \frac{-(y - \rho^t x)^2}{2(1 - \rho^{2t})} dy,$$

where  $\rho$  is a constant between 0 and 1<sup>(4)</sup>. Evidently

$$(0.9) \quad \lim_{t \rightarrow \infty} P^{(t)}(x, A) = \frac{1}{(2\pi)^{1/2}} \int_A \exp \frac{-y^2}{2} dy.$$

In this case (0.4) holds as an ordinary limit. It is clear however that the limit is not uniform in  $x$ . While it is hardly necessary to invoke the general theory of Markoff processes to treat this simple example, it is unsatisfactory that the general theory as developed up to the present does not already cover it.

There is a self-reproducing distribution in this Gaussian process, the distribution defined by the integral in (0.9), and with this distribution the Gaussian process can be made stationary. The renewal process is another example of a Markoff process, not covered by the hypothesis stated above, which has a self-reproducing distribution. Under the conditions of Doeblin and of Kryloff and Bogoliouboff,  $Q_j(A)$  in (0.5) is a self-reproducing distribution. It is therefore natural to study those Markoff processes which have such distributions, that is to study stationary Markoff processes, in order to extend the Doeblin-Kryloff-Bogoliouboff results. The author, Yosida, and Kakutani have made the first steps in this direction; references to this work will be given below.

The notation used is, in summary:

$X$  is an abstract space of points  $x, y, \dots$ .

(\*) The space  $X$  and field  $\mathfrak{F}_x$  can be taken considerably more generally in both these conditions. The condition of Kryloff and Bogoliouboff is somewhat weaker than that of Doeblin as stated, but it will be seen below that the two conditions become equivalent if other measures of Borel sets than Lebesgue measure are allowed in Doeblin's condition.

(4) Cf. [4] for a detailed discussion of this process. The discussion covers the continuous parameter case, but the change to  $t$  integral is trivial.

$A, B, \dots$  are  $X$  sets;  $\mathfrak{F}_x$  is a given Borel field of these sets. In practice  $X$  is usually a Borel set in Euclidean space with the topology determined by Euclidean distance. For this reason no special notation is used to distinguish points and sets of the abstract space  $X$  from points and sets of numbers.

$\Omega$  is function (sequence) space as already defined;  $\omega$  is a point of  $\Omega$ , that is, a function (sequence)  $x(t)$ .

$\Lambda, M, \dots$  are  $\Omega$  sets.

$\mathfrak{F}_\omega$  is the Borel field of  $\Omega$  sets generated by sets of the form

$$\{x(t_j) \in A_j, j = 1, \dots, n\}, \quad A_j \in \mathfrak{F}_x.$$

$P(\Lambda)$  is the probability measure on the sets of  $\mathfrak{F}_\omega$  determined by a given self-reproducing distribution and the given transition probabilities. For this reason it is convenient to call the sets of  $\mathfrak{F}_\omega$  the  $P$ -measurable sets. (The parameter  $t$  was chosen to be integral-valued to avoid certain complications in  $P$ -measure. In the continuous parameter case a topology is introduced in  $X$  and regularity properties are imposed on the transition probabilities to insure measurability of the stochastic process.)

$\alpha(\omega), \beta(\omega), \dots$  are complex-valued functions of  $\omega$ , that is, chance variables.

$E\{\alpha\} = \int \alpha(\omega) P(d\omega)$  is the expectation of  $\alpha$ .

$P\{\alpha; \Lambda\}$  and  $E\{\alpha; \beta\}$  are respectively the probability of  $\Lambda$  and the expectation of  $\beta$  conditioned by assigned  $\alpha$ ; they are functions of  $\alpha$ . It will sometimes be convenient to write  $P\{\alpha=x; \Lambda\}$  and  $E\{\alpha=x; \beta\}$  to stress a particular value assigned to  $\alpha$ . The conditional expectation  $E\{\alpha; \beta\}$  can be obtained as the integral of  $\beta$  in terms of the probability measure  $P\{\alpha; \Lambda\}$ .

$T_s$  is the transformation of function space already defined. We shall sometimes write  $T_s\alpha$  for  $\alpha(T_s\omega)$ .

$\Phi$  measure is the measure of  $X$  sets defined by a preassigned self-reproducing probability measure  $\Phi(A)$ . This will be used, together with the transition probabilities, to define a stationary Markoff process, so that for every  $s$

$$(0.10) \quad \Phi(A) = P\{x(s) \in A\}.$$

Alternatively a stationary Markoff process will be presupposed and  $\Phi(A)$  will then be defined by (0.10). In this approach the transition probabilities must be defined indirectly, and the fact that the definition is ambiguous on sets of measure zero will cause complications. These will be discussed below.

A  $P$ -measurable set  $\Lambda$  will be called invariant if  $T\Lambda$  and  $\Lambda$  differ by at most a set of  $P$ -measure zero. A set  $A$  of  $\mathfrak{F}_x$  will be called invariant if

$$(0.11) \quad P\{x(0); x(1) \in A\} = 1$$

almost everywhere ( $P$ -measure), on the  $\Omega$  set  $\{x(0) \in A\}$ , that is, if

$$(0.11') \quad P\{x(0) = x; x(1) \in A\} = 1$$

for almost all  $x$  ( $\Phi$  measure) in  $A$ . If  $A$  is invariant (0.11) and (0.11') are also true with  $x(1)$  replaced by  $x(s)$ , for every  $s > 0$ . The set  $A$  is invariant if and only if the  $\Omega$  set  $\{x(0) \in A\}$  is invariant. Moreover it is known that if a set  $A$  is invariant it differs by at most a set of  $P$ -measure zero from a set of the form  $\{x(0) \in A\}$ . This fact makes it possible to consider only invariant sets of this simple type. More generally [3, pp. 118–119] it is known that if any  $P$ -measurable function  $\alpha(\omega)$  satisfies the condition  $T\alpha = e^{i\lambda}\alpha$  for some constant  $\lambda$ , neglecting a possible exceptional set of  $P$ -measure zero, then  $\alpha$  differs on at most a set of  $P$ -measure zero from a  $P$ -measurable function of the form  $f[x(0)]$ . A function satisfying this equation with  $\lambda = 0$  is called an invariant function; if  $e^{i\lambda} \neq 1$ ,  $\log \alpha$  when suitably defined is called an angle variable.

A stationary process (Markoff or not) is said to be metrically transitive if there is no invariant  $P$ -measurable function not constant with probability 1. In the present case, the  $x(t)$  process is a Markoff process, and it has already been noted that in this case, in discussing both metric transitivity and angle variables, it is sufficient to consider functions of  $x(0)$ . In all cases, when investigating metric transitivity, it is sufficient to consider functions taking on only the values 0 and 1, that is, characteristic functions of point sets. A Markoff process is thus metrically transitive if and only if every invariant  $\Phi$  measurable set  $A$  has  $\Phi$  measure either 0 or 1, that is if and only if every  $\Omega$  set of the form  $\{x(s) \in A\}$  (fixed  $s$ ,  $A$   $\Phi$  measurable) which is unchanged, neglecting sets of zero  $P$  measure, by a change of  $s$ , has  $P$  measure 0 or 1<sup>(6)</sup>.

The theorems of the present paper concern conditional expectations of the form  $E\{x(0) = x; T_t\alpha\}$ . These conditional expectations are considered at first as functions of  $x$  for fixed  $\alpha$ , and the almost everywhere properties of the functions for  $t \rightarrow \infty$  are investigated. These results would remain valid if for each  $s$  the conditional expectations were changed on a set of probability 0, and in fact these conditional expectations are ordinarily so defined that there is ambiguity on sets of zero probability. Sometimes, however, individual values of  $x$  will be considered and such an ambiguity will be inadmissible. It will then be supposed that the conditional probabilities are *regular*. By this is meant that transition probabilities

$$P\{x(s) = x; x(s+t) \in A\} = P^{(t)}(x, A)$$

are given, as uniquely defined functions, satisfying conditions (a), (b), and (c) of the first paragraph of this paper. If  $\alpha$  is a chance variable depending on the functions  $x(t)$  for  $t \geq s$ ,  $E\{x(s) = x; \alpha\}$  is then unambiguously defined as the integral of  $\alpha$  in terms of the conditional probability measure for  $x(s) = x$ .

The hypothesis of regularity of the transition probabilities is not so strong a restriction as might appear at first sight. In the first place Markoff processes are most frequently given in terms of uniquely defined transition prob-

<sup>(6)</sup> The general background of ergodic theory, including concepts like metric transitivity, and so on, is contained in [5]. The specific application to Markoff processes is given in [3].

abilities; in this case regularity is intrinsic in the definition. In the second place even if the transition probabilities are not so given, it is usually possible to redefine them to be regular. In fact, *this is always possible if the field  $\mathfrak{F}_x$  is strictly separable*<sup>(6)</sup>. This hypothesis is always satisfied, for example, when  $\mathfrak{F}_x$  consists of the Borel subsets of a Borel set in a Euclidean space. The proof of this possibility of redefinition when  $\mathfrak{F}_x$  is strictly separable goes as follows. Although for a given  $A \in \mathfrak{F}_x$ ,  $P\{x(0)=x; x(1) \in A\}$  is not a uniquely defined function of  $x$ , it can be so defined for each  $A \in \mathfrak{F}_x$  that, excluding an  $X$  set  $N$  independent of  $A$ , with

$$(0.12) \quad N \in \mathfrak{F}_x, \quad P\{x(0) \in N\} = \Phi(N) = 0,$$

$P\{x(0)=x; x(1) \in A\}$  is for fixed  $x$  a probability measure in  $A$  [3, p. 96]. Define  $P^{(1)}(x, A)$  as this function for  $x \in X - N$ , and as any probability measure in  $A$  for  $x \in N$ . The function  $P^{(t)}(x, A)$  defined inductively by

$$(0.13) \quad P^{(t)}(x, A) = \int P^{(t-1)}(y, A) P^{(2)}(x, d\mathbf{y}_y)$$

is a regular transition probability of the process.

The above discussion made essential use of the hypothesis that  $\mathfrak{F}_x$  was strictly separable. In the following it will frequently be desirable to perform integrations of the type

$$(0.14) \quad \int f(y) P\{x(0) = x; x(1) \in d\mathbf{y}_y\}$$

without any assumption of regularity imposed on the transition probabilities. It is known that this is permissible and that the usual rules of integration, taking the limit under the integral sign, and so on, apply, and the definition of (0.14) is in terms of the usual Lebesgue sums [3, p. 98]. This convention will be used below, for integration with respect to  $P\{x(0)=x; x(t) \in A\}$  measure or any other conditional probability measure with no further comment.

We shall need the following fact below. *If the transition probabilities are regular and if  $A_1, A_2, \dots$  are invariant, there is a set  $N$  of  $\Phi$  measure 0 such that*

$$(0.15) \quad P\{x(0) = x; x(t) \in A_j - A_j N\} = 1, \text{ if } x \in A_j - A_j N, j \geq 1, t \geq 1.$$

---

(6) A Borel field of sets is called strictly separable if it is the Borel field generated by a finite or denumerable subcollection of its sets. *Note added in proof:* The italicized statement is incorrect; the proof outlined is based on Theorem 3.1 [3, p. 96] which is incorrect. (It was tacitly assumed in the proof of this theorem that the  $t$  image of  $\Omega$  had outer  $F$  measure 1; this is not necessarily true, although it is true if the  $t$  image is a Borel set.) The italicized statement is however correct if, for example,  $\mathfrak{F}_x$  consists of the Borel subsets of a Borel set in a Euclidean space. The discussion below of (0.14) is correct, with a slightly modified justification.

In fact there is a set  $N_1$  of  $\Phi$  measure 0, independent of  $j$  and  $t$ , such that

$$P\{x(0) = x; x(t) \in A_j\} = 1 \quad \text{if } x \in A_j - A_j N_1,$$

and a set  $N_2 \supset N_1$  of  $\Phi$  measure 0, independent of  $j$  and  $t$ , such that

$$P\{x(0) = x; x(1) \in A_j - A_j N_1\} = 1 \quad \text{if } x \in A_j - A_j N_2, \\ \dots \dots \dots$$

Then (0.15) is true with  $N = \sum_j N_j$ .

**THEOREM 1 (STATIONARY PROCESS).** (a) *Let  $\alpha(\omega)$  be any  $P$ -measurable function, and suppose that for some  $p \geq 1$*

$$(1.1) \quad E\{|\alpha|^p\} < \infty.$$

*Then the average*

$$(1.2) \quad \frac{1}{t} \sum_{s=1}^t E\{x(0); T_s \alpha\}$$

*converges in the mean with the same index  $p$  to a function  $\alpha^*$ :*

$$(1.3) \quad \lim_{t \rightarrow \infty} E\left\{\left|\frac{1}{t} \sum_{s=1}^t E\{x(0); T_s \alpha\} - \alpha^*\right|^p\right\} = 0.$$

(b) *If  $p > 1$ , or more generally if*

$$(1.4) \quad E\{|\alpha| \log^+ |\alpha|\} < \infty \text{ (7),}$$

*the limit exists pointwise:*

$$(1.5) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t E\{x(0) = x; T_s \alpha\} = \alpha^*$$

*except for an  $x$  set of  $\Phi$  measure 0. If  $A \in \mathfrak{F}_x$  the averages in (1.2), (1.3), and (1.5) can be specialized to*

$$(1.6) \quad \frac{1}{t} \sum_{s=1}^t P\{x(0); x(s) \in A\}.$$

(c) *If the process is metrically transitive the limit  $\alpha^*$  is  $E\{\alpha\}$  which reduces to  $P\{x(0) \in A\} = \Phi(A)$  when  $t \rightarrow \infty$  in (1.6). Conversely if  $\alpha^* = E\{\alpha\}$  for every  $\alpha$  the process is metrically transitive.*

(d) *Suppose that the process is a Markoff process and let  $Q(x, A)$  be the limit in (1.6) when  $t \rightarrow \infty$ . Then  $Q(x, A)$  is the conditional probability of  $\{x(0) \in A\}$  relative to the field of invariant sets; that is*

$$(1.7) \quad \int_B Q(x, A) \Phi(dG_x) = \Phi(A \cdot B)$$

---

(7)  $\log^+ |\alpha| = \log |\alpha|$  if  $|\alpha| \geq 1$ ,  $\log^+ |\alpha| = 0$  if  $|\alpha| < 1$ .



if  $B$  is invariant. More generally, if  $\alpha$  depends only on  $x(0), x(1), \dots$ ,  $\alpha^*$  is the conditional expectation of  $\alpha$  relative to the field of invariant sets. The function  $Q(x, A)$  satisfies the equations:

$$(1.8) \quad Q(x, A) = \int P\{x(0) = y; x(t) \in A\} Q(x; dG_y),$$

$$(1.9) \quad Q(x, A) = \int Q(y, A) P\{x(0) = x; x(t) \in dG_y\},$$

$$(1.10) \quad Q(x, A) = \int Q(y, A) Q(x, dG_y),$$

for almost all values of  $x$  ( $\Phi$  measure). The exceptional set varies with  $A$  and the choice of the conditional probabilities involved. There is metric transitivity if and only if  $Q(x, A) = \Phi(A)$  for almost all  $x$  ( $\Phi$  measure) for each  $A$ .

(e) Suppose that the process is a Markoff process with regular transition probabilities and suppose that  $C = C(\alpha)$  is the  $x$  set on which (1.5) is true, where  $\alpha$  is bounded and depends only on  $x(0), x(1), \dots$ . Then  $C$  includes every  $x$  satisfying

$$(1.11) \quad P\left\{x(0) = x; \sum_{t=0}^{\infty} [x(t) \in C]\right\} = 1.$$

Note that (1.6) is the counterpart of (0.4). Although the limit when  $t \rightarrow \infty$  in (1.6) may not exist on an  $x(0)$  set of  $\Phi$  measure 0, in many applications the transition probabilities are regular and the possibility of exceptional points of nonconvergence can be eliminated by the last statement of the theorem.

Various special cases of this theorem have been proved by the author [3, p. 114], by Yosida [8], and by Kakutani [6]. It has always been supposed hitherto, however, in proving (1.3) and (1.5) that the process was Markoff, and, in finding pointwise limits, that  $\alpha$  was bounded. It will be clear from the proof that the conditional expectations for preassigned  $x(0)$  can be replaced in (a), (b), and (c) by any other conditional expectations without altering either the validity of the theorem or its proof. It would be interesting to prove the existence of the pointwise limit in (1.5) for any  $\alpha$  with  $E\{|\alpha|\} < \infty$ , or to obtain a counter example.

**Proof of (a).** Under the stated hypotheses,  $E\{|\alpha|\} < \infty$ . Hence according to the point ergodic theorem (strong law of large numbers for stationary processes)

$$(1.12) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t T_s \alpha = \alpha_1$$

exists with probability 1, and according to the mean ergodic theorem if  $E\{|\alpha|^p\} < \infty$  for some  $p \geq 1$ , it follows that  $E\{|\alpha_1|^p\} < \infty$  and there is con-

vergence in the mean of order  $p$ . Hence, in this case, when  $t \rightarrow \infty$ ,

$$\begin{aligned}
 (1.13) \quad & E \left\{ \left| \frac{1}{t} \sum_{s=1}^t E\{x(0); T_s \alpha\} - E\{x(0); \alpha_1\} \right|^p \right\} \\
 &= E \left\{ \left| E \left\{ x(0); \left[ \frac{1}{t} \sum_{s=1}^t T_s \alpha - \alpha_1 \right] \right\} \right|^p \right\} \\
 &\leq E \left\{ E \left\{ x(0); \left| \frac{1}{t} \sum_{s=1}^t T_s \alpha - \alpha_1 \right|^p \right\} \right\} \\
 &= E \left\{ \left| \frac{1}{t} \sum_{s=1}^t T_s \alpha - \alpha_1 \right|^p \right\} \rightarrow 0.
 \end{aligned}$$

Thus (1.3) holds, with  $\alpha^* = E\{x(0); \alpha_1\}$ .

**Proof of (b).** To prove (1.5) it will be sufficient to show that if  $p=1$  and if (1.4) is true then (1.12) can be integrated to the limit ( $t \rightarrow \infty$ ) with respect to the conditional measure  $E\{x(0); \Lambda\}$ . This will be shown by showing that (in terms of this conditional measure) excluding perhaps an  $x(0)$  set of zero measure there is convergence in (1.12) almost everywhere, and that the averages are dominated by an integrable function. (It has already been noted that this procedure is correct even though strictly speaking the conditional probability may not define a measure.) According to (1.12) the bracket converges to  $\alpha_1$  except on a set  $M$  with  $P\{M\} = 0$ . Since

$$(1.14) \quad \int_{\Omega} P\{x(0) = x; M\} \Phi(dG_x) = P\{M\} = 0,$$

it follows that  $P\{x(0); M\} = 0$  for almost all  $x(0)$  ( $\Phi$  measure). That is to say, excluding the exceptional values of  $x(0)$  the limit in (1.12) exists almost everywhere ( $P\{x(0); \Lambda\}$  measure). Moreover according to a theorem of Wiener [9, p. 27] (1.4) implies that

$$(1.15) \quad E \left\{ \text{L.U.B.}_t \left| \frac{1}{t} \sum_{s=1}^t T_s \alpha \right| \right\} < \infty.$$

Hence, excluding an  $x(0)$  set of  $\Phi$  measure 0,

$$(1.16) \quad E \left\{ x(0); \text{L.U.B.}_t \left| \frac{1}{t} \sum_{s=1}^t T_s \alpha \right| \right\} < \infty.$$

This means that, excluding an  $x(0)$  set of  $P$ -measure 0, the average in (1.12) is dominated by an integrable function, as was to be proved.

**Proof of (c).** According to the ergodic theorem there is metric transitivity if and only if  $\alpha_1 = E\{\alpha\}$  for all  $\alpha$ , in which case, then,  $\alpha^* = E\{\alpha\}$  also. If  $\alpha$  is defined to be 1 when  $x(0) \in A$  and 0 otherwise, the average in (1.2) becomes that in (1.6). The limit is then  $E\{\alpha\} = \Phi(A)$  when there is metric transitivity.

**Proof of (d).** Let  $Q(x, A)$  be the limit in (1.6). To prove that  $Q(x, A)$  is the stated conditional expectation, that is, that (1.7) is true, we note that since the process is a Markoff process, and since  $B$  is invariant,

$$\begin{aligned}
 (1.17) \quad & \int_B \frac{1}{t} \sum_{s=1}^t P\{x(0) = x; x(s) \in A\} \Phi(dG_x) \\
 &= \frac{1}{t} \sum_{s=1}^t P\{x(0) \in B, x(s) \in A\} \\
 &= \frac{1}{t} \sum_{s=1}^t P\{x(s) \in B, x(s) \in A\} \\
 &= P\{x(0) \in A \cdot B\} = \Phi(A \cdot B).
 \end{aligned}$$

When  $t \rightarrow \infty$ , this becomes (1.7). The identification of  $\alpha^*$  as a conditional probability is made in the same way. The equalities (1.8), (1.9) and (1.10) are "almost everywhere" equalities. Hence to show their truth it is sufficient to show that integration over an arbitrary  $\mathfrak{F}_x$  set  $B$  yields equality in each case. This is clear in (1.9), using the fact that  $\Phi$  is self-reproducing. It is also clear in (1.8) and (1.10), using (1.7), if  $B$  is an invariant set. Now it is clear from the definition of  $Q(x(0), A)$  that this function of  $x(0)$  is invariant, and therefore that it is permissible to restrict  $B$  to be invariant in testing (1.8) and (1.10). Thus the three equalities are completely proved. Finally, if there is metric transitivity we have already seen that  $Q(x, A) = \Phi(A)$  for almost all  $x$  ( $\Phi$  measure). Conversely, if this equation is true, and if  $A$  is invariant,  $\Phi(A) = 1$  or 0, since the conditional probability  $Q(x, A)$  is the characteristic function of  $A$  if  $A$  is invariant. Hence the process is metrically transitive. Moreover it is clearly sufficient if  $Q(x, A) = \Phi(A)$  on a collection of sets so large that the Borel field they generate coincides with  $\mathfrak{F}_x$ .

**Proof of (e).** Suppose that  $|\alpha| \leq K$ , that  $\alpha$  depends only on  $x(0), x(1), \dots$ , that (1.5) is true for  $x \in C$ , and that the process is a Markoff process with regular transition probabilities. Let  $x$  satisfy (1.11), that is to say it is supposed that starting from  $x(0) = x$ ,  $x(s)$  takes on a value in  $C$  for some  $s \geq 0$  with probability 1. We show that then  $x \in C$ . Let  $\nu$  be the first value of  $s$  with  $x(s) \in C$ . Then if we use the fact that the process is a Markoff process, and that the equality  $\nu = s$  is a condition on  $x(0), \dots, x(s)$ ,

$$(1.18) \quad E\{x(0) = x; T_s \alpha\} = \sum_{j=0}^{\infty} \int_{\{\nu=j\}} T_s \alpha P\{x(0) = x; d\Lambda\},$$

where, if  $s \geq j$ , the  $j$ th integral in (1.18) can be written in the form

$$(1.19) \quad \int_{\{\nu=j\}} E\{x(j); T_s \alpha\} P\{x(0) = x; d\Lambda\}.$$

Hence

$$\begin{aligned}
 (1.20) \quad & \frac{1}{t} \sum_{s=1}^t E\{x(0) = x; T_s \alpha\} \\
 &= \sum_{j=1}^t \int_{\{\nu=j\}} \frac{1}{t} \sum_{s=j}^t E\{x(j); T_s \alpha\} P\{x(0) = x; d\Lambda\} \\
 &+ \frac{1}{t} \sum_{s=1}^t \int_{\{\nu>s\}} T_s \alpha P\{x(0) = x; d\Lambda\}.
 \end{aligned}$$

Now the second sum is at most

$$(1.21) \quad \frac{K}{t} \sum_{s=1}^t P\{x(0) = x; \nu > s\}$$

which goes to 0 when  $t \rightarrow \infty$ , since  $P\{x(0) = x; \nu > t\} \rightarrow 0$ . The integrand in the  $j$ th integral of the first sum in (1.20) is bounded by  $K$  and, since  $x(j) \in C$  when  $\nu = j$ , converges to a limit when  $t \rightarrow \infty$ . The integral, and hence the sum, therefore also converges to a limit; this finishes the proof.

It is a remarkable fact, due in the first place to von Neumann, and proved later more generally by Ambrose, Halmos, and Kakutani [1], that any stationary process (with mild restrictions on the field of measurable sets) can be decomposed into metrically transitive processes. Although the various decomposition-limit theorems on Markoff processes are merely reflections of this fact (the proofs vary with the special character of the hypotheses and the predilections of the authors) the exact adaptation of the decomposition theorem to Markoff processes has never been stated explicitly. This is done in the following theorem, whose proof will be given in detail; the theorem is not a corollary of the general theorem because stronger results are desired in this special case.

**THEOREM 2 (STATIONARY MARKOFF PROCESS WITH REGULAR TRANSITION PROBABILITIES).** *Let  $Q(x, A)$  be the conditional probability of  $A (\in \mathfrak{F}_x)$  with respect to the field of invariant sets, so that (by Theorem 1)*

$$(2.1) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t P\{x(0) = x; x(s) \in A\} = Q(x, A)$$

for almost all  $x$  ( $\Phi$  measure).

*If the field  $\mathfrak{F}_x$  is strictly separable,  $Q(x, A)$  can be so defined that the following is true<sup>(\*)</sup>. There is a collection of sets  $\{A_\mu\}$  where  $A_\mu \in \mathfrak{F}_x$  and  $\mu$  varies through*

(\*) *Added in proof:* (Cf. footnote 6.) It is necessary to strengthen this hypothesis; it will be supposed that  $\mathfrak{F}_x$  consists of the Borel subsets of a set in a Euclidean space. With this restriction, the discussion of the unique definition of  $Q(x, A)$ , involved in the definition of  $A_\mu$ , is correct.

some set of real numbers, with the following properties:

(a) Every set  $A \in \mathfrak{F}_x$  which is a sum of sets  $A_\mu$  is invariant and (neglecting sets of  $\Phi$  measure 0) every invariant set can be expressed in this way.

$$(b) \quad \Phi(N) = 0, \quad \text{where } N = X - \sum_{\mu} A_{\mu}.$$

If  $x \in X - N$ ,  $Q(x, A)$  is a probability measure in  $A \in \mathfrak{F}_x$  satisfying

$$(2.2) \quad Q(x, A) = Q_{\mu}(A), \quad x \in A_{\mu},$$

$$(2.3) \quad Q_{\mu}(A_{\mu}) = 1,$$

$$(2.4) \quad \int Q(x, A) \Phi(dG_x) = \Phi(A).$$

(c) Each probability measure  $Q_{\mu}(A)$  is self-reproducing,

$$(2.5) \quad \int P\{x(0) = y; x(1) \in A\} Q_{\mu}(dG_y) = Q_{\mu}(A),$$

and more generally if  $\Psi(d\mu)$  is any probability measure in  $\mu$  space,

$$(2.6) \quad \Phi_1(A) = \int Q_{\mu}(A) \Psi(d\mu)$$

is a self-reproducing probability measure. Conversely if  $\Phi_1(A)$  is self-reproducing and absolutely continuous with respect to  $\Phi(A)$  it can be put in the form (2.6). If, for some  $\mu$ ,  $\Phi(A_{\mu}) > 0$ , then  $Q_{\mu}(A)$  is proportional to  $\Phi(A \cdot A_{\mu})$ ,  $Q_{\mu}(A) = \Phi(A \cdot A_{\mu}) / \Phi(A_{\mu})$ .

(d)  $P\{x(0) = x; x(t) \in A_{\mu}\} = 1$  everywhere on  $A_{\mu}$  and for each  $\mu$  and each  $A \in \mathfrak{F}_x$ , (2.1) is true for almost all  $x$  ( $Q_{\mu}$  measure). The Markoff process with  $\Phi$  replaced by  $Q_{\mu}$  is metrically transitive. The original process is metrically transitive if and only if some  $A_{\mu}$  has  $\Phi$  measure 1.

The analogues of (1.9) and (1.10) have been omitted here, since they reduce to trivial identities. Conditions are given below which exclude the possibility that any  $A_{\mu}$  have zero  $\Phi$  measure. Under these conditions there can be at most denumerably many of these sets. The stronger conditions of Doeblin and of Kryloff and Bogoliouboff imply that there can be at most a finite number.

**Definition of  $A_{\mu}$  and proof of (a).** It has already been remarked that  $Q(x(0), A)$  is an invariant function, for fixed  $A$ . This implies that the  $x$  set defined by  $Q(x, A) < k$  is an invariant set. In the following we shall suppose that  $\mathfrak{F}_x$  is strictly separable, specifically that it is the Borel field generated by the sets  $C_1, C_2, \dots, C_j \in \mathfrak{F}_x$ . It can then be supposed that the conditional expectation  $Q(x, A)$  is so defined that, for fixed  $x$ ,  $Q(x, A)$  is a probability measure in  $A$ . It will be supposed below that  $Q(x, A)$  is so defined, and

that this function is then a uniquely given function of  $x$  and  $A$ . Let  $\mathfrak{F}_x$  be the Borel field of invariant sets generated by the denumerable collection of  $X$  sets of the form

$$(2.7) \quad \{Q(x, C_n) < r\}, \quad n = 1, 2, \dots; r \text{ rational.}$$

If  $A$  is an invariant  $X$  set, we have seen that  $Q(x, A)$  is the characteristic function of  $A$ , neglecting sets of  $\Phi$  measure 0. Hence, neglecting sets of  $\Phi$  measure 0,  $A$  is the set defined by the equality  $Q(x, A) = 1$ . Since  $A$  is in the Borel field generated by  $C_1, C_2, \dots$  this equality defines a set in  $\mathfrak{F}_x$ , neglecting sets of  $\Phi$  measure 0. Hence  $\mathfrak{F}_x$ , although it may not include  $A$ , includes a set differing from it by at most a set of  $\Phi$  measure 0.

The sets of  $\mathfrak{F}_x$  which contain no proper non-null subsets in  $\mathfrak{F}_x$  are called the atoms of  $\mathfrak{F}_x$ . They can be exhibited explicitly in the form

$$(2.8) \quad \prod_i D_i^{\epsilon_i}, \quad \epsilon_i = \pm 1, \quad D_i^1 = D_i, \quad D_i^{-1} = X - D_i,$$

where  $D_1, D_2, \dots$  are the sets in (2.7). The non-null intersections in (2.8) are the atoms. The cardinal number of these non-null intersections is at most that of the continuum, and we can therefore denote them by  $\{A_\mu\}$  where  $\mu$  ranges through some set of real numbers. In the following we shall find it necessary to drop some of these atoms, putting them in a set  $N$  of  $\Phi$  measure 0. Since every set of  $\mathfrak{F}_x$  is composed of atoms, every invariant set differs from a sum of  $A_\mu$ 's by at most a set of  $\Phi$  measure 0. The converse, that every  $A_\mu$  sum is invariant if it is a set in  $\mathfrak{F}_x$  is a consequence of (b), to be proved below. In fact if  $A \in \mathfrak{F}_x$  is a sum of  $A_\mu$ 's, where  $\mu$  ranges over some set  $M$ ,  $Q_\mu(A) = 1$  when  $\mu \in M$  and  $Q_\mu(A) = 0$  otherwise. Hence the set  $A$  is defined by the equality  $Q_\mu(A) = 1$ , that is, neglecting sets of  $\Phi$  measure 0,  $A$  is defined by the equality  $Q(x, A) = 1$ . This definition has already been seen to define an invariant set.

**Proof of (b).** Since  $Q(x, A)$  is  $\mathfrak{F}_x$  measurable for fixed  $A$ , it depends only on the atom in which  $x$  lies. Equation (2.2) is merely the analytic formulation of this fact. If  $A \in \mathfrak{F}_x$ , the conditional probability  $Q(x, A)$  is 1 almost everywhere in  $A$  ( $\Phi$  measure); that is to say, using (2.2),  $Q(x, A) = 1$  in every  $A_\mu$  except in some constituting a set of  $\Phi$  measure 0. Drop these atoms from the  $A_\mu$  for every set  $A$  defined by (2.7). If this is done, (2.3) will be true. Equation (2.4) is obtained by integrating (2.1) with respect to  $\Phi$  measure.

**Proof of (c).** According to Theorem 1, (1.8) is true for almost all  $x$  ( $\Phi$  measure), that is, (2.5) is true except possibly for a set of  $\mu$  corresponding to atoms of total  $\Phi$  measure 0. Absorb these in  $N$  for all sets  $A = C_1, C_2, \dots$ . Then (2.5) will be true for all  $A \in \mathfrak{F}_x$ . If (2.6) is used to define a probability measure  $\Phi_1(A)$ , integration of (2.5) in  $\mu$  ( $\Psi$  measure) gives the equation showing that  $\Phi_1$  is self-reproducing. Conversely suppose that  $\Phi_1$  is any self-reproducing probability measure. Then

$$(2.9) \quad \frac{1}{t} \sum_{s=1}^t \int P\{x(0) = y; x(s) \in A\} \Phi_1(dG_y) = \Phi_1(A).$$

When  $t \rightarrow \infty$  the integrand converges to  $Q(y, A)$ , except possibly on a set of  $\Phi$  measure 0. If  $\Phi_1$  is absolutely continuous with respect to  $\Phi$ , the exceptional set has at most zero  $\Phi_1$  measure also; when  $t \rightarrow \infty$ , (2.9) becomes

$$(2.10) \quad \int Q(y, A) \Phi_1(dG_y) = \Phi_1(A),$$

which is a special case of (2.6), since  $\Phi_1(dG_y)$  induces a measure in  $\mu$  space. In particular, suppose that  $\Phi(A_\mu) > 0$ . Then by (2.4),  $\Phi(A \cdot A_\mu) = Q_\mu(A) \cdot \Phi(A_\mu)$ .

**Proof of (d).** It was proved in the introductory remarks of this paper that (since the sets  $D_1, D_2, \dots$  defined by (2.7) are invariant) there is a set  $D_0$  of  $\Phi$  measure 0 such that  $P\{x(0) = x; x(1) \in D_j - D_j D_0\} = 1$  for  $x \in D_j - D_j D_0$ . We shall now suppose that  $\mathfrak{F}_x$  is the field of sets generated by  $D_0, D_1, \dots$  (rather than by  $D_1, D_2, \dots$ ). None of the previous arguments require any change because of this. It is now true, however, that  $P\{x(0) = x; x(1) \in A\} = 1$  for  $x \in A$  if  $A \in \mathfrak{F}_x$  and if  $A \subset X - D_0$ . Then the first statement of (d) is true if  $D_0$ , which is a sum of atoms, is incorporated in  $N$ . Equation (2.1) is true for each  $A$  for almost all  $x$  ( $\Phi$  measure) and hence, using (2.10), if  $N_1$  is the exceptional set,  $Q(y, N_1) = 0$  for almost all  $y$  ( $\Phi$  measure). Hence  $Q_\mu(N_1) = 0$  for all  $\mu$  except some corresponding to  $A_\mu$ 's of total  $\Phi$  measure 0. Absorb these in  $N$  for each  $A = C_1, C_2, \dots$ . We shall show below that then, for each  $\mu$  and each  $A \in \mathfrak{F}_x$ , (2.1) is true for almost all  $x$  ( $Q_\mu$  measure). According to Theorem 1, the process with  $\Phi$  replaced by  $Q_\mu$  will be metrically transitive if

$$(2.11) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t P\{x(0) = x; x(s) \in A\} = Q_\mu(A)$$

is true for almost all  $x$  ( $Q_\mu$  measure), for every  $A$ . In fact, according to a remark made in proving Theorem 1, it is sufficient that (2.12) hold for a collection of sets  $A$  determining the field  $\mathfrak{F}_x$ . This is precisely what we have shown, with  $A = C_1, C_2, \dots$ . Hence the reduced process is metrically transitive. It then follows from Theorem 1 that (2.11), that is (2.1), holds for almost all  $x$  ( $Q_\mu$  measure) for each  $A \in \mathfrak{F}_x$ . If the given process is metrically transitive, the invariant  $X$  sets have  $\Phi$  measure either 0 or 1. It is then clear from (2.8) that one atom must have  $\Phi$  measure 1. This atom can never be absorbed in  $N$ . Conversely if some  $A_\mu$  has  $\Phi$  measure 1, say  $A_*$ , then  $Q_*(A) = \Phi(A)$ , and the "reduced" process, which has been shown to be metrically transitive, is precisely the original process.

In general, it is not difficult to show that the  $\mu$  can be so chosen that any Borel set of  $\mu$ 's corresponds to a set of atoms which constitute a set of  $\mathfrak{F}_x$ , and conversely any  $\mathfrak{F}_x$  set can be expressed in this way (ignoring the atoms

in  $N$ ). This gives an elegant parametrization of the invariant sets.

The following theorems make Theorem 2 more precise by imposing restrictions on the transition probabilities which make it possible to discuss the truth of (2.1) for individually specified values of  $x$ .

**THEOREM 3 (STATIONARY MARKOFF PROCESS WITH REGULAR TRANSITION PROBABILITIES).** *Let  $P_1(x, A, t)$  be the singular component of the measure  $P\{x(0)=x; x(t)\in A\}$  with respect to  $\Phi$  measure. In the following,  $\alpha(\omega)$  will be any  $P$  measurable function depending on  $x(t)$  for  $t\geq 0$ , and either bounded or at least with*

$$(3.1) \quad E\{x(0) = x; |T_{t_0}\alpha| \} < \infty, \quad x \in X, t \geq 1$$

and

$$(3.2) \quad \text{L.U.B.}_y E\{x(0) = y; |T_{t_0}\alpha| \} < \infty$$

for some  $t_0 > 0$ .

(a)  $P_1(x, X, t)$  is monotone non-increasing in  $t$ , for fixed  $x$ .

(b) With the above convention on  $\alpha$ ,

$$(3.3) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t E\{x(0) = x; T_s\alpha\}$$

exists for all  $x$  satisfying

$$(3.4) \quad \lim_{t \rightarrow \infty} P_1(x, X, t) = 0.$$

If the process is metrically transitive, the limit in (3.4) is  $E\{\alpha\}$ .

(c) If  $C$  is the  $x$  set on which (3.4) is true,  $C$  includes every  $x$  satisfying

$$(3.5) \quad P\left\{x(0) = x; \sum_{t \geq 0} [x(t) \in C]\right\} = 1.$$

If (3.4) is satisfied on an  $X$  set of positive  $\Phi$  measure, and if the process is metrically transitive, (3.4) is satisfied by almost all values of  $x$  ( $\Phi$  measure).

(d) If  $x$  satisfies (3.4), and if  $A \in \mathfrak{F}_x$ ,

$$(3.6) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t P\{x(0) = x; x(s) \in A\} = Q(x, A)$$

exists. The set function  $Q(x, A)$  is a probability measure which is absolutely continuous with respect to  $\Phi$  measure and is self-reproducing,

$$(3.7) \quad \int P\{x(0) = y; x(t) \in A\} Q(x, dG_y) = Q(x, A).$$

If the process is metrically transitive,  $Q(x, A) = \Phi(A)$ .



(e) If almost all  $x$  ( $\Phi$  measure) satisfy (3.4),  $Q(x, A)$  (for those values of  $x$ ) satisfies

$$(3.8) \quad \int Q(y, A)Q(x, dG_y) = Q(x, A)^{(9)}$$

and also satisfies

$$(3.9) \quad \int Q(y, A)P\{x(0) = x; x(t) \in dG_y\} = Q(x, A)$$

for almost all  $x$  ( $\Phi$  measure)<sup>(10)</sup>. If all  $x$  satisfy (3.4), (3.8) and (3.9) are true for all  $x$ .

**Proof of (a).** For each  $t$  let  $A_t$  be a set of  $\Phi$  measure 0 for which

$$(3.10) \quad P\{x(0) = x; x(t) \in A\} = P_1(x, X, t).$$

The set  $A_t$  depends on  $x$  and  $t$ , but  $x$  in the following will be held fast. Then

$$(3.11) \quad \begin{aligned} &P_1(x, X, s+t) \\ &= \int P\{x(t) = y; x(s+t) \in A_{s+t}\}P\{x(0) = x; x(t) \in dG_y\}, \quad s, t > 0. \end{aligned}$$

Since  $\Phi(A_{s+t}) = 0$ , the integrand vanishes for almost all  $y$  ( $\Phi$  measure); if  $B$  is the set on which it does not vanish,

$$(3.12) \quad \begin{aligned} P_1(x, X, s+t) &\leq \int_B P\{x(0) = x; x(t) \in dG_y\} \\ &= P\{x(0) = x; x(t) \in B\} \leq P_1(x, X, t), \end{aligned}$$

as was to be proved. It is interesting to note, although we shall not use the fact explicitly, that  $P_1$  obeys the same functional equation (0.1) as the transition probabilities themselves.

**Proof of (b).** The problem is that of justifying going to the limit under an integral sign. In fact in

$$(3.13) \quad \frac{1}{t} \sum_{s=1}^t E\{x(0) = x; T_s \alpha\} = E\left\{x(0) = x; \frac{1}{t} \sum_{s=1}^t T_s \alpha\right\},$$

the average  $(1/t) \sum_{s=1}^t T_s \alpha$  converges with probability 1. According to Theorem 1 integration to the limit is justified for almost all  $x$  ( $\Phi$  measure) if  $E\{|\alpha| \log^+ |\alpha|\} < \infty$ . The present hypotheses are of a different type. If

(9) The integration is meaningful even though  $Q(y, A)$  is not defined on a  $y$  set  $N$  of  $\Phi$  measure 0 because, according to (d),  $N$  is also of  $Q(x, dG_y)$  measure 0.

(10) In this integral the integrand is again undefined on a set of  $\Phi$  measure 0, but this has integration measure 0 for almost all ( $\Phi$  measure) values of  $x$ .

$$0 < s \leq t - t_0,$$

$$\begin{aligned}
 (3.14) \quad & \frac{1}{t} \sum_{j=1}^t E\{x(0) = x; T_j \alpha\} \\
 &= \int E\left\{x(s) = y; \frac{1}{t} \sum_{j=s+t_0}^t T_j \alpha\right\} P\{x(0) = x; x(s) \in dG_y\} \\
 &+ E\left\{x(0) = x; \frac{1}{t} \sum_{j=1}^{s+t_0-1} T_j \alpha\right\}.
 \end{aligned}$$

Now by hypothesis (3.2) is true, and it follows easily that (3.2) is true with  $t_0$  replaced by any  $t > t_0$ , with the same bound, say  $K$ . Then the  $j$ th integrand in (3.14) is bounded by  $K$ . According to Theorem 1 this integrand converges in the mean of order 1, and therefore in probability, that is, in  $\Phi$  measure. Unfortunately  $\Phi$  measure is not the relevant integration measure. However if  $A_t$  is defined as in the proof of (a), the integrand converges on  $X - A_s$  in the integration measure. Integration to the limit on  $X - A_s$  is therefore allowed and the maximum oscillation of (3.14) when  $t \rightarrow \infty$  is thus at most  $KP_1(x, X, s)$ , an upper bound to the integration over  $A_s$ . By hypothesis this can be made arbitrarily small by choosing  $s$  large. Hence the left side of (3.14) converges to a limit when  $t \rightarrow \infty$ . If the process is metrically transitive, the limit is  $E\{\alpha\}$ , since in that case the almost everywhere limit in (3.13) when  $t \rightarrow \infty$  is  $E\{\alpha\}$ , by Theorem 1.

**Proof of (c).** If (3.4) is satisfied on an  $X$  set  $C$  and if  $x$  satisfies (3.5), that is, if starting from  $x(0) = x$ ,  $x(s)$  takes on a value in  $C$  for some  $s \geq 0$  with probability 1, we shall now show that  $x$  also satisfies (3.4). Let  $\nu$  be the first value of  $s$  with  $x(s) \in C$ . Then using the fact that the process is a Markoff process, and that  $\nu = s$  is a condition on  $x(0), \dots, x(s)$ ,

$$(3.15) \quad P\{x(0) = x; x(t) \in A\} = \sum_{s=0}^{\infty} P\{x(0) = x; \nu = s, x(t) \in A\};$$

where, if  $t \geq s$ ,

$$\begin{aligned}
 (3.16) \quad & P\{x(0) = x; \nu = s, x(t) \in A\} \\
 &= \int_{\{\nu=s\}} P\{x(s); x(t) \in A\} P\{x(0) = x; d\Lambda\}.
 \end{aligned}$$

In particular, if  $\Phi(A) = 0$

$$\begin{aligned}
 (3.17) \quad & P\{x(0) = x; \nu = s, x(t) \in A\} \\
 &\leq \int_{\{\nu=s\}} P_1(x(s), X, t-s) P\{x(0) = x; d\Lambda\}.
 \end{aligned}$$

If  $A$  is chosen so that the left side of (3.15) becomes  $P_1(x, X, t)$ , it then fol-

lows that if  $t \geq n$ ,

$$\begin{aligned}
 P_1(x, X, t) &\leq \sum_{s=0}^n \int_{\{\nu=s\}} P_1(x(s), X, t-s) P\{x(0) = x; d\Lambda\} \\
 &\quad + \sum_{s=n+1}^{\infty} P\{x(0) = x; \nu = s, x(t) \in A\} \\
 (3.18) \quad &\leq \sum_{s=0}^n \int_{\{\nu=s\}} P_1(x(s), X, t-s) P\{x(0) = x; d\Lambda\} \\
 &\quad + \sum_{s=n+1}^{\infty} P\{x(0) = x; \nu = s\}.
 \end{aligned}$$

Now the last series in (3.18) is the remainder of a convergent series:

$$(3.19) \quad \sum_{s=0}^{\infty} P\{x(0) = x; \nu = s\} = 1,$$

and can be made arbitrarily small by choosing  $n$  large. When  $\nu = s$ ,  $x(s) \in C$ . Hence the integrands in (3.18) converge to 0 ( $t \rightarrow \infty$ ) by the definition of  $C$ . Since the series of integrals is dominated term by term by the series (3.19), it follows that the left side of (3.18) converges to 0. We have thus shown that  $C$  includes the values of  $x$  satisfying (3.5). If the process is metrically transitive,  $\Phi(C) = 0$  or  $\Phi(C) = 1$ . In fact if  $\Phi(C) > 0$ , and if  $n(t)$  is the number of times  $x(s) \in C$  for  $0 \leq s \leq t$ ,  $n(t)/t \rightarrow \Phi(C)$ , with probability 1, according to the ergodic theorem. Thus for almost all values of  $x(0) = x$  ( $\Phi$  measure), (3.5) is satisfied and in fact  $x(t)$  will even take on values in  $C$  for infinitely many values of  $t$ .

**Proof of (d).** If  $x$  satisfies (3.4) and if  $\alpha(\omega)$  is defined to be 1 when  $x(0) \in A$  and 0 otherwise, the average in (3.3) reduces to that in (3.6). Since there is a limit in (3.6) for every  $A$ ,  $Q(x, A)$  is a probability measure in  $A$ . According to (b),  $Q(x, A) = E\{\alpha\} = \Phi(A)$  if the process is metrically transitive. The equation

$$\begin{aligned}
 \frac{1}{\tau} \sum_{t=s+1}^{s+\tau} P\{x(0) = x; x(t) \in A\} &= \frac{1}{\tau} \sum_{t=1}^{\tau} P\{x(0) = x; x(s+t) \in A\} \\
 (3.20) \quad &= \frac{1}{\tau} \sum_{t=1}^{\tau} \int P\{x(t) = y; x(s+t) \in A\} P\{x(0) = x; x(t) \in dG_y\} \\
 &= \int P\{x(0) = y; x(s) \in A\} \cdot \frac{1}{\tau} \sum_{t=1}^{\tau} P\{x(0) = x; x(t) \in dG_y\}
 \end{aligned}$$

leads (when  $\tau \rightarrow \infty$ ) to (3.7). This means that  $Q(x, A)$  is a self-reproducing distribution which may or may not coincide with the given one,  $\Phi(A)$ . If  $\Phi(A) = 0$ ,

$$\begin{aligned}
 Q(x, A) &= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t P\{x(0) = x; x(s) \in A\} \\
 (3.21) \quad &\leq \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t P_1(x, X, s) = 0.
 \end{aligned}$$

Hence  $Q(x, A)$  is absolutely continuous with respect to  $\Phi$  measure, for fixed  $x$ . The absolute continuity may not be uniform in  $x$ .

**Proof of (e).** If almost all  $x$  ( $\Phi$  measure) satisfy (3.4), (3.7) must hold for almost all  $x$ . Hence for those  $x$ ,

$$(3.22) \quad \int \frac{1}{t} \sum_{s=1}^t P\{x(0) = y; x(s) \in A\} Q(x, dG_y) = Q(x, A).$$

When  $t \rightarrow \infty$  the integrand converges boundedly to  $Q(y, A)$  for almost all  $y$  ( $\Phi$  measure) and therefore for almost all  $y$  in the integration measure, because the latter is absolutely continuous in  $\Phi$  measure. Integration to the limit is therefore permissible, and yields (3.8). It has already been seen in Theorem 1 that (3.9) holds for almost all  $x$  ( $\Phi$  measure). If all  $x$  satisfy (3.4), (3.8) must hold for all  $x$ . Moreover when  $\sigma \rightarrow \infty$  in

$$\begin{aligned}
 (3.23) \quad &\frac{1}{\sigma} \sum_{s=t+1}^{t+\sigma} P\{x(0) = x; x(s) \in A\} = \frac{1}{\sigma} \sum_{s=1}^{\sigma} P\{x(0) = x; x(s+t) \in A\} \\
 &= \int \frac{1}{\sigma} \sum_{s=1}^{\sigma} P\{x(0) = y; x(s) \in A\} P\{x(0) = x; x(t) \in dG_y\},
 \end{aligned}$$

the integrand converges boundedly to  $Q(y, A)$  when  $\sigma \rightarrow \infty$ . Integration to the limit yields (3.9).

The following theorem gives conditions under which the Cesàro limits used up to the present can be replaced by ordinary limits.

**THEOREM 4 (STATIONARY MARKOFF PROCESS WITH REGULAR TRANSITION PROBABILITIES AND NO ANGLE VARIABLES).** *In the following  $\alpha$  is a  $P$  measurable function satisfying (3.2) for all large  $t_0$  and depending only on  $x(t)$  for  $t \geq \text{const.}$  (the constant may vary with  $\alpha$ ).*

(a) *There is a  $t$  set  $I$  of relative measure 1<sup>(11)</sup>, independent of  $x$  and of  $\alpha$  such that if  $x$  satisfies (3.4)*

$$(4.1) \quad \lim_{t \rightarrow \infty, t \in I} E\{x(s) = x; T_t \alpha\}$$

*exists and is independent of  $s$ . If the process is metrically transitive, the limit is  $E\{\alpha\}$ .*

<sup>(11)</sup> A  $t$  set of relative measure 1 is a set with the property that (number of integers in the set between  $n$  and  $-n$ )/ $2n \rightarrow 1$  when  $n \rightarrow \infty$ .

(b) If (3.4) is satisfied by almost all  $x$  ( $\Phi$  measure), (4.1) can be strengthened:

$$(4.2) \quad \lim_{t \rightarrow \infty} E\{x(s) = x; T_t \alpha\}$$

exists for those  $x$ , and is  $E\{\alpha\}$  if the process is metrically transitive.

With proper choice of  $\alpha$ , (4.1) and (4.2) become

$$(4.1') \quad \lim_{t \rightarrow \infty, t \in I} P\{x(s) = x; x(t) \in A\} = Q(x, A)$$

and

$$(4.2') \quad \lim_{t \rightarrow \infty} P\{x(s) = x; x(t) \in A\} = Q(x, A).$$

The specialization from (4.1) and (4.2) to (4.1') and (4.2') is clear, and will receive no further comment.

**Proof of (a).** According to a theorem of von Neumann and Koopman [5, pp. 36-37] if the process has no angle variables, there is a  $t$  set  $I$  of relative measure 1, with the property that

$$(4.3) \quad \lim_{t \rightarrow \infty, t \in I} E\{\beta(\omega) T_t \alpha(\omega)\}$$

exists for all chance variables  $\alpha, \beta$  with  $E\{|\alpha|^2\} < \infty$  and  $E\{|\beta|^2\} < \infty$ . If the process is metrically transitive, the limit is  $E\{\alpha\} \cdot E\{\beta\}$ . Moreover the limit is unchanged if  $\alpha$  is replaced by  $T_s \alpha$  or if  $\beta$  is replaced by  $T_s \beta$ . It will be convenient to write  $P\{x(0) = x; x(t) \in A\}$  explicitly as the sum of its absolutely continuous and singular components with respect to  $\Phi$  measure,

$$(4.4) \quad P\{x(0) = x; x(t) \in A\} = \int_A p(x, y, t) \Phi(dG_y) + P_1(x, A, t).$$

The conditional expectation in (4.1) can be written (for large  $t$ ) in the form

$$\begin{aligned} E\{x(s) = x; T_t \alpha\} &= \int E\{x(u) = y; T_t \alpha\} P\{x(s) = x; x(u) \in dG_y\} \\ (4.5) \quad &= \int E\{x(u) = y; T_t \alpha\} p(x, y, u - s) \Phi(dG_y) + \eta \\ &= E\{p(x, x(u), u - s) \cdot T_t \alpha\} + \eta, \end{aligned}$$

where

$$\begin{aligned} |\eta| &= \left| \int E\{x(u) = y; T_t \alpha\} P_1(x, dG_y, u - s) \right| \\ (4.6) \quad &\leq \text{L.U.B.}_v E\{x(0) = y; |T_{t-u} \alpha| \} P_1(x, X, u - s). \end{aligned}$$

If  $t \rightarrow \infty$ ,  $t \in I$ , the last expectation in (4.5) converges to a limit (put  $\beta = p(x, x(u), u-s)$  in (4.3)); in fact although the integration hypotheses imposed by Koopman and von Neumann are not satisfied here, a simple approximation extends their result as required. Moreover it was remarked in the proof of Theorem 3 that our hypotheses on  $\alpha$  imply that the L.U.B. in (4.6) is bounded in  $t$  when  $t \rightarrow \infty$ , say with bound  $K$ . Hence when  $t \rightarrow \infty$ , with  $t \in I$ , the left side of (4.6) has oscillation at most  $KP_1(x, X, u-s)$ . By hypothesis this can be made arbitrarily small by choosing  $u$  large. Hence the left side of (4.5) converges to a limit when  $t \rightarrow \infty$ ,  $t \in I$ . If the process is metrically transitive the limit is  $E\{\alpha\}$ . This can be seen from the fact that the limit in (4.3) is  $E\{\alpha\} \cdot E\{\beta\}$  in that case, or from the fact that the Cesàro limit is  $E\{\alpha\}$  in that case, according to Theorem 3.

**Proof of (b).** Suppose now that (3.4) is satisfied by almost all  $x$  ( $\Phi$  measure). For large  $t$  we can write

$$\begin{aligned} E\{x(0) = x; T_t\alpha\} &= \int E\{x(s) = y; T_t\alpha\} P\{x(0) = x; x(s) \in dG_y\} \\ (4.7) \quad &= \int E\{x(0) = y; T_{t-s}\alpha\} P\{x(0) = x; x(s) \in dG_y\}. \end{aligned}$$

According to (a) the integrand converges for almost all  $y$  when  $t \rightarrow \infty$  with  $t-s \in I$ . We let  $t \rightarrow \infty$  with no restriction, choosing  $s = s(t)$  so that

$$t-s \in I, \quad s \in I, \quad s \rightarrow \infty.$$

Now if  $A_s$  is a set of  $\Phi$  measure 0 satisfying the equation  $P_1(x, A_s, s) = P_1(x, X, s)$ , and containing the points of nonconvergence of the integrand in (4.7), the integrand converges boundedly everywhere on  $X - \sum_s A_s$  and the integration measures converge. Hence

$$(4.8) \quad \lim_{t \rightarrow \infty} \int_{X-A} E\{x(s) = y; T_t\alpha\} P\{x(0) = x; x(s) \in dG_y\} \quad \left( A = \sum_s A_s \right)$$

exists. Moreover if  $K$  has the same significance as in the proof of (a),

$$(4.9) \quad \int_A E\{x(s) = y; T_t\alpha\} P\{x(0) = x; x(s) \in dG_y\} \leq KP_1(x, X, s) \rightarrow 0$$

so that the left side of (4.7) converges to a limit when  $t \rightarrow \infty$ , as was to be proved. As in (a), if the process is metrically transitive, the limit is  $E\{\alpha\}$ .

Theorems 3 and 4 were devoted to the asymptotic character of  $E\{x(0) = x; T_t\alpha\}$  as conditioned by the validity of (3.4). If (3.4) is assumed true on an  $X$  set of positive  $\Phi$  measure, the decomposition theorem (Theorem 2) becomes essentially simplified. Only two cases will be discussed: when (3.4) is true for all  $x$  (Theorem 5) and when (3.4) is true for almost all  $x$  (Theorem 6). The

paper is then concluded by a detailed analysis of the hypotheses of Doeblin and of Kryloff and Bogoliouboff as compared with the hypothesis that (3.4) is true for all  $x$ .

**THEOREM 5 (STATIONARY MARKOFF PROCESS WITH REGULAR TRANSITION PROBABILITIES; EVERY VALUE OF  $x$  SATISFIES (3.4)).** *There are finitely or denumerably infinitely many disjunct sets  $A_1, A_2, \dots$ , of positive  $\Phi$  measure, and corresponding self-reproducing probability measures  $Q_1(A), Q_2(A), \dots$  with the following properties:*

$$(5.1) \quad \Phi(N) = Q_i(N) = 0 \quad \left( N = X - \sum_i A_i \right),$$

$$(5.2) \quad Q_i(A_i) = 1.$$

*A probability measure  $\Phi_1(A)$  is self-reproducing if and only if it can be put in the form*

$$(5.3) \quad \Phi_1(A) = \sum_i a_i Q_i(A), \quad a_i \geq 0, \sum_i a_i = 1.$$

*In particular every  $a_i$  is positive for  $\Phi_1 = \Phi$  and more generally these constants are all positive if and only if  $\Phi$  is absolutely continuous with respect to  $\Phi_1$ . Every  $\Phi_1$  is absolutely continuous with respect to  $\Phi$ .*

*Each  $Q_i$  process ( $\Phi$  replaced by  $Q_i$ ) is metrically transitive and the  $Q_i$ 's are the only self-reproducing distributions giving metrically transitive processes.*

*The transition probabilities satisfy*

$$(5.4) \quad P\{x(0) = x; x(t) \in A_i\} = 1, \quad x \in A_i,$$

*and, if  $A \in \mathfrak{F}_x$ ,*

$$(5.5) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t P\{x(0) = x; x(s) \in A\} = Q(x, A)$$

*exists for all  $x$ , with*

$$(5.6) \quad \begin{aligned} Q(x, A) &= Q_i(A) & (x \in A_i) \\ &= \sum a_i(x) Q_i(A) & \left( x \in N, a_i(x) \geq 0, \sum_i a_i(x) = 1 \right). \end{aligned}$$

*For  $x \in A_j$ , (5.5) can be strengthened as follows. To each  $A_j$  corresponds a positive integer  $d_j$ . If  $d_j = 1$ ,*

$$(5.7) \quad \lim_{t \rightarrow \infty} P\{x(0) = x; x(t) \in A\} = Q_j(A), \quad x \in A_j.$$

*If  $d_j > 1$ ,  $A_j$  is the sum of  $d_j$  disjunct sets  $A_{j1}, \dots, A_{jd_j}$  and there are probability measures  $Q_{j1}, \dots, Q_{jd_j}$  satisfying*

$$(5.8) \quad Q_i(A) = \frac{1}{d_i} [Q_{i1}(A) + \cdots + Q_{id_i}(A)],$$

$$(5.9) \quad Q_{ik}(A_{jk}) = 1, \quad Q_i(A_{jk}) = 1/d_j,$$

$$(5.10) \quad P\{x(0) = x; x(t) \in A_{jk}\} = \delta_{k+l+t}, \quad x \in A_i^{(12)}.$$

Finally

$$(5.11) \quad \lim_{s \rightarrow \infty} P\{x(0) = x; x(ds + t) \in A\} = Q_{i+l+t}(A),$$

$$x \in A_{il}; t = 0, \cdots, d_i - 1,$$

$$(5.12) \quad Q_{ik+l}(A) = \int P\{x(0) = y; x(t) \in A\} Q_{ik}(dG_y).$$

If (3.4) holds for all  $x$ ,  $Q(x, A)$  is defined by (3.6) for all  $x$  and all  $A \in \mathfrak{F}_x$ , and satisfies (3.8). According to a theorem of Blackwell [2, p. 567] since  $Q(x, A)$  is absolutely continuous with respect to  $\Phi(A)$  and satisfies (3.8),  $X$  is the sum of disjoint sets  $B_0, B_1, A_1, A_2, A_3, \cdots$ , finite or denumerably infinite in number, with the  $A_j$  of positive  $\Phi$  measure, and there are probability measures  $Q_1(A), Q_2(A), \cdots$  satisfying

$$(5.13) \quad \begin{aligned} Q(x, A) &= Q_j(A), & x \in A_j, \\ Q_j(A_j) &= 1, \\ Q_j(A) &> 0 & \text{if } A \subset A_j, \Phi(A) > 0, \\ Q(x, B_0) &\equiv 0, \\ \Phi(B_1) &= 0. \end{aligned}$$

We shall only touch on those points which are not obvious in the light of Blackwell's theorem and the preceding theorems of this paper. Equation (0.2) implies that  $\Phi(B_0) = 0$ . We set  $N = B_0 + B_1$  and shall absorb other sets of  $\Phi$  measure into  $N$  also, as required below. Evaluating (3.8), and using (5.13) and the fact that  $Q(x, A) = 0$  when  $\Phi(A) = 0$ ,

$$(5.14) \quad Q(x, A) = \sum_j Q_j(A) Q(x, A_j).$$

This is (5.6), with an explicit evaluation of the  $a_j(x)$ . If  $\Phi_1(A)$  is self-reproducing,

$$(5.15) \quad \begin{aligned} \Phi_1(A) &= \int P\{x(0) = y; x(t) \in A\} \Phi_1(dG_y) \\ &= \int \frac{1}{t} \sum_{i=1}^t P\{x(0) = y; x(t) \in A\} \Phi_1(dG_y). \end{aligned}$$

---

(12) Here and in the following the second subscript  $i$  in  $A_{il}$  is to be interpreted modulo  $d_i$ .



When  $t \rightarrow \infty$  this becomes

$$\begin{aligned} \Phi_1(A) &= \int Q(y, A) \Phi_1(dG_y) = \sum_j Q_j(A) \int Q(y, A_j) \Phi_1(dG_y) \\ (5.16) \qquad &= \sum_j Q_j(A) \Phi_1(A_j) \end{aligned}$$

which is (5.3). In particular if  $\Phi_1 = \Phi$ ,  $a_j = \Phi(A_j) > 0$ . Since every  $Q_j(A)$  is absolutely continuous with respect to  $\Phi$ , every  $\Phi_1$  must be also. Any two self-reproducing distributions whose corresponding constants of combination  $a_j$  are simultaneously zero define the same sets of measure 0. The only special feature about  $\Phi$  is that all of the corresponding  $a_j$ 's are positive. If  $\Phi_1$  did not have this property, (3.4) would hold only for almost all  $x$ , using  $\Phi_1$  as a base measure.

Equation (3.6) becomes for  $x \in A_j$

$$(5.17) \qquad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t P\{x(0) = x; x(s) \in A\} = Q_j(A).$$

Hence by Theorem 1(d) the  $Q_j$  process is metrically transitive. It is clear that the  $Q_j$ 's are the only self-reproducing distributions making the process metrically transitive. For example this follows from the fact that the  $A_j$ 's and sums of  $A_j$ 's are the only invariant sets, neglecting sets of  $\Phi$  measure 0, that is, with every  $Q_j$  vanishing. Since  $A_j$  is invariant, (5.4) certainly holds for almost all  $x$  in  $A_j$  ( $Q_j$  measure), and we have seen at the beginning of this paper that  $A_j$  can be replaced by a subset of the same  $Q_j$  measure for which (5.4) holds. The difference is put into  $N$ . We shall assume that this has been done.

Define  $d_j = 1$  if the  $Q_j$  process has no angle variables. Then (5.7) follows from Theorem 4. Now suppose that the  $Q_j$  process has an angle variable, that is, that there is a  $\Phi$  measurable function  $f(x)$  (not identically zero on a set of  $\Phi$  measure 1) and a real constant  $\lambda$  satisfying

$$(5.18) \qquad f[x(s)] = e^{i\lambda s} f[x(0)], \qquad s = 1, 2, \dots,$$

neglecting  $x(s)$  sets of  $Q_j$  measure 0. Since  $|f|$  is invariant, and since the  $Q_j$  process is metrically transitive,  $|f| \equiv \text{const.}$  on a set of  $Q_j$  measure 1. It is no restriction to take this constant as 1. Then

$$(5.19) \qquad f = e^{ig}, \qquad 0 \leq g < 2\pi.$$

It follows that the  $\Omega$  sets

$$(5.20) \qquad \begin{aligned} \{x(0) \in B\}: \quad &\{a < \lambda s + g[x(0)] < b \pmod{2\pi}\}, \\ \{x(s) \in C\}: \quad &\{a < g[x(s)] < b \pmod{2\pi}\} \end{aligned}$$

are identical, neglecting sets of probability 0. Thus

$$(5.21) \quad \begin{aligned} Q_j(B) = Q_j(C) = P\{a < g[x(s)] < b \pmod{2\pi}\}, \\ P\{x(0) = x; x(s) \in C\} = 1, \end{aligned} \quad x \in B,$$

neglecting an  $x$  set  $N(a, b)$  of  $Q_j$  measure 0. Let  $N_1$  be the sum of the  $N(a, b)$  for all rational  $a$  and  $b$ . Unless  $g$  is a function which takes only a finite or denumerably infinite set of values (aside from values taken on on a set of  $Q_j$  measure 0), there is a sequence of rational pairs  $(a_j, b_j)$ , shrinking to a point, corresponding to a sequence of pairs of sets  $(B_j, C_j)$  satisfying (5.21) and

$$(5.22) \quad B_1 \supset B_2 \supset \cdots, \quad C_1 \supset C_2 \supset \cdots, \quad Q_j(B_n) = Q_j(C_n) > 0,$$

such that  $\bigcap_j B_n$  contains points in  $A_j - N_1$ . Let  $x$  be such a point, and choose  $s$  so large that

$$(5.23) \quad P_1(x, X, s) < 1.$$

Then

$$(5.24) \quad P\left\{x(0) = x; x(s) \in \prod_n C_n\right\} = 1, \quad Q_j\left(\prod_n C_n\right) = 0,$$

which contradicts (5.23). Thus there is a constant  $c$  which  $f$  assumes on a set of positive  $Q_j$  measure. The constants  $ce^{i\alpha}, ce^{2i\alpha}, \dots$  are also assumed on sets of the same  $Q_j$  measure. This is impossible unless  $e^{i\alpha}$  is a root of unity. We can suppose then that  $e^{i\alpha}$  is a  $d_j$ th root of unity and that

$$f(x) = ce^{i\alpha}, ce^{2i\alpha}, \dots$$

on disjunct sets  $A_{j1}, \dots, A_{jd_j}$  respectively, all having the same  $Q_j$  measure. The sum of these sets is invariant, and therefore includes all of  $A_j$  except possibly a set of  $Q_j$  measure 0. Hence  $Q_j(A_{jk}) = 1/d_j$ . The remainder  $A_j - \sum_k A_{jk}$  is absorbed in  $N$ , and  $A_j$  is then reduced again as above, so that (5.4) will still hold. If  $x(s) \in A_{jk}$ , it follows that  $x(s+1) \in A_{jk+1}$ , that is, (5.21) becomes (5.10), valid for almost all points of  $A_{jk}$ . The sets  $A_{jk}$  and  $A_j$  are then modified to ensure the simultaneous validity of (5.4) and (5.10) without exceptional points. Unless there were a maximum value to the possible integers  $d_j$ , (5.10) would show that for almost all  $x$  ( $Q$  measure) and every  $n$  there are sets  $B_n$  with  $P\{x(0) = x; x(s) \in B_n\} = 1, Q_j(B_n) \leq 1/n$ . This contradicts (5.23). Hence there is a maximum  $d_j$  and it is assumed from now on that this maximum has been chosen above.

The transformation  $T$  considered on the  $Q_j$  process will be denoted by  $T_j$ . It has a point spectrum<sup>(13)</sup> consisting of 0 (corresponding to the fact that 1 is a characteristic function) and  $2\pi n/d_j$  (corresponding to the fact that  $f$  and its integral powers are characteristic functions) for  $n=1, \dots, d_j-1$ . These

<sup>(13)</sup> Cf. [5] for a discussion of the spectrum of a measure preserving transformation, that is, of the corresponding unitary transformation.

points all have multiplicity 1, since  $T_j$  is metrically transitive. Then  $T^{d_j}$  has a point spectrum consisting of the single point 0, which has multiplicity  $d_j$ . Corresponding to this fact, the transition probabilities

$$P\{x(0) = x; x(d, t) \in A\}, \quad t = 1, 2, \dots,$$

determine (together with  $Q_j$  measure) a stationary Markoff process which is not metrically transitive, because  $A_{j1}, \dots, A_{jd_j}$  are invariant, but which has no angle variables, and which becomes metrically transitive when reduced still further from  $A_j$  to any  $A_{jk}$ . Then the limit  $Q_{jt}(A)$  defined by (5.11) with  $t=0$  must exist, according to Theorem 4. Equations (5.11) and (5.12) are obtained when  $s \rightarrow \infty$  in

$$\begin{aligned} & P\{x(0) = x; x(d, s + t) \in A\} \\ (5.25) \quad &= \int P\{x(0) = y; x(t) \in A\} P\{x(0) = x; x(d, s) \in dG_y\} \\ &= \int P\{x(0) = y; x(d, s) \in A\} P\{x(0) = x; x(t) \in dG_y\}. \end{aligned}$$

**THEOREM 6 (STATIONARY MARKOFF PROCESS WITH REGULAR TRANSITION PROBABILITIES).** *If almost all  $x$  ( $\Phi$  measure) satisfy (3.4), there is a set  $N_0 \in \mathfrak{F}_x$  of  $\Phi$  measure 0 such that*

$$(6.1) \quad P\{x(0) = x; x(t) \in X - N_0\} = 1, \quad x \in X - N_0,$$

and that if  $X$  is replaced by  $X - N_0$ , and the field of sets  $A \in \mathfrak{F}_x$  by the field of sets  $A_0 = A - AN_0$ , the Markoff process with this space, these  $X$  measurable sets, the transition probabilities (6.1), and the same  $\Phi$  measure satisfies the hypotheses of Theorem 5. The decomposition of that theorem is therefore applicable to  $X - N_0$ .

Before proving Theorem 6, we present the following example. In Theorem 5 choose a self-reproducing measure  $\Phi_1$  as defined by (5.3), and suppose that  $a_1=0$ . The present hypotheses are then satisfied, using the basic  $\Phi_1$  measure, but Theorem 5 is no longer applicable since (3.4) is no longer true on  $A_1$ . The present theorem then gives the decomposition of  $X - A_1$  but gives no information on  $A_1$ .

To prove Theorem 6 we need only note that if (3.4) is true on  $X - N_0$ , by a process already used repeatedly in the proof of Theorem 5  $N_0$  can be increased slightly to make (6.1) true. The theorem is now obvious.

*Comparison between the condition that (3.4) be valid for all  $x$  and the conditions of Doeblin and of Kryloff and Bogoliouboff.* In the following it will be supposed that  $X$  is a Borel set of finite measure in Euclidean space and that  $\mathfrak{F}_x$  is the field of its Borel subsets. The Lebesgue measure of a set  $A$  will be denoted by  $mA$ . Doeblin's hypothesis is that for some  $\epsilon > 0$  and  $t_0$

$$(7.1) \quad P\{x(0) = x; x(t_0) \in A\} \leq 1 - \epsilon \quad \text{if } mA \leq \epsilon,$$

uniformly in  $x$ . This condition is shown to imply the existence of the limit  $Q(x, A)$  in (3.6) uniformly in  $x$  and  $A$ . The use of Lebesgue measure here is somewhat confusing, since the proofs are equally valid for many other measures of Borel sets. In order to clarify this point, his condition will be put in a somewhat different form. In his case  $Q(x, A)$  is a self-reproducing probability measure for each  $x$ , and the decomposition of Theorem 5 is valid; in fact only a finite number of sets  $A_j$  can be present. Moreover, using the notation of Theorem 5, if  $mA > 0$ , and if  $A \subset A_j$ ,  $Q_j(A) > 0$  also. Define  $\Phi(A)$  by

$$\Phi(A) = \sum_i a_i Q_i(A)$$

where the  $a_j$ 's are positive, have sum 1, and are otherwise unrestricted. Then  $\Phi(A)$  is a self-reproducing probability distribution, and  $\Phi(A) > 0$  if  $mA > 0$ , that is,  $mA$  is absolutely continuous with respect to  $\Phi(A)$ . Condition (7.1) can therefore be put in the following form:

$$(7.1') \quad P\{x(0) = x; x(t_0) \in A\} \leq 1 - \epsilon \quad \text{if } \Phi(A) \leq \epsilon.$$

Now if this condition is satisfied, and if  $P_1(x, A, t)$  is the singular component of  $P\{x(0) = x; x(t) \in A\}$  with respect to  $\Phi(A)$ , then certainly

$$(7.2) \quad P_1(x, X, t_0) \leq 1 - \epsilon$$

uniformly in  $x$ . The proof that  $P_1(x, X, t)$  is monotone nonincreasing in  $t$  now shows (using (7.2) in (3.12) with  $s = t_0$ ) that

$$(7.3) \quad P_1(x, X, t_0 + t) \leq (1 - \epsilon)P_1(x, X, t).$$

Hence

$$(7.4) \quad P_1(x, X, nt_0) \leq (1 - \epsilon)^n$$

and thus (3.4) is true, and the convergence is even uniform in  $x$  and exponentially fast. It is this uniformity of convergence to 0 which precludes the possibility of infinitely many sets  $A_j$ , and the still stronger condition (7.1') implies the uniformity of the limit equation (3.6).

The formulation of Doeblin's condition in (7.1') explains the rather arbitrary use of Lebesgue measure in the original formulation. Lebesgue measure was a measure of Borel sets with the property that  $mA > 0$  implied  $Q_j(A) > 0$  for some  $j$ . Any other measure of Borel sets with this property would serve as well. In the same way the  $\Phi(A)$  distribution in Theorem 5 was rather arbitrary (although it had to be self-reproducing); any self-reproducing distribution with the property that if any  $Q_j(A) > 0$ , then  $\Phi(A) > 0$  would have served as well.

The condition of Doeblin is thus stronger than the hypothesis that there

is a self-reproducing distribution and that (3.4) is true for all  $x$ . Doeblin's condition has the advantage that it does not require a priori the existence of a self-reproducing distribution. In some examples, however, the existence of such distributions is already known. This is true in the renewal process (see the following paper in this volume). Doeblin's condition is however essentially stronger and leads to stronger results where it is applicable. The condition of Kryloff and Bogoliouboff is somewhat weaker than that of Doeblin, as stated, but since it leads to exactly the same decomposition theorem, the two are essentially equivalent. In both cases (7.1') is true. The point is that the various possible choices for  $X$  measure in Doeblin's condition bring it up to the generality of the Kryloff-Bogoliouboff condition.

As a somewhat trivial example which may however be illuminating, consider the Gaussian Markoff process described in the Introduction. In this case, as already remarked, there is a self-reproducing distribution, and  $P_1(x, X, t) \equiv 0$ , so that Theorem 5 is applicable. Actually (0.9) shows that there is metric transitivity and no angle variables; there is only one set  $A_j$ , and there are no cyclic transitions. The convergence in (0.9) is not uniform in  $x$ ; this means that neither the Doeblin nor the Kryloff-Bogoliouboff conditions are satisfied.

#### BIBLIOGRAPHY

1. W. Ambrose, P. Halmos, and S. Kakutani, *The decomposition of measures*, II, Duke Math. J. vol. 9 (1942) pp. 43-47.
2. D. Blackwell, *Idempotent Markoff chains*, Ann. of Math. vol. 43 (1942) pp. 560-567.
3. J. L. Doob, *Stochastic processes with an integral-valued parameter*, Trans. Amer. Math. Soc. vol. 44, (1938) pp. 87-150.
4. ———, *The Brownian movement and stochastic equations*, Ann. of Math. vol. 43 (1942) pp. 351-369.
5. E. Hopf, *Ergodentheorie*, Ergebnisse der Mathematik, vol. 5, no. 2.
6. S. Kakutani, *Ergodic theorems and the Markoff process with a stable distribution*, Proc. Imp. Acad. Tokyo vol. 16 (1940) pp. 49-54.
7. K. Yosida and S. Kakutani, *Operator-theoretical treatment of Markoff's process and mean ergodic theorem*, Ann. of Math. vol. 42 (1941) pp. 188-228.
8. K. Yosida, *The Markoff process with a stable distribution*, Proc. Imp. Acad. Tokyo vol. 16 (1940) pp. 43-48.
9. N. Wiener, *The ergodic theorem*, Duke Math. J. vol. 5 (1939) pp. 1-18.
10. (Added in proof.) A. M. Yaglom, *The ergodic principle for Markov processes with stationary distributions*, C. R. (Doklady) Acad. Sci. URSS. N.S. vol. 54 (1947) pp. 347-349. The author supposes that  $P^{(n)}(x, A)$  is absolutely continuous with respect to a given self-reproducing distribution  $\Phi(A)$ , with a positive continuous density, and proves that then  $\lim_{n \rightarrow \infty} P^{(n)}(x, A) = \Phi(A)$ . This is a special case of Theorem 5.

UNIVERSITY OF ILLINOIS,  
URBANA, ILL.