STOCHASTIC PROCESSES WITH AN INTEGRAL-VALUED PARAMETER*

BY

J. L. DOOB

The purpose of this paper is to set up the measure relations of the most general stochastic process and to discuss the properties of the conditional probability functions of the processes depending on a parameter running through integral values. In particular, the study of temporally homogeneous processes of this type is shown to be essentially the study of measure preserving transformations. The well known results in the latter field are applied to develop and extend the theory of Markoff processes from a new point of view.

Throughout the paper, any non-negative completely additive function of point sets, defined on a Borel field of sets† of some abstract space \( \Omega \) will be called a probability measure if the space \( \Omega \) is itself in the field of definition and if the set function is defined as 1 on \( \Omega \).

1. Probability measures defined on spaces of infinitely many dimensions.

Let \( \Omega_0 (X) \) be any abstract space, consisting of elements \( \omega_0 (x) \), and let \( F_{\omega_0} (F_x) \) be a Borel field of subsets of \( \Omega_0 (X) \). We shall suppose that \( \Omega_0 \in F_{\omega_0} \) and \( X \in F_x \). If \( f(\omega_0) \) is a function defined on \( \Omega_0 \) which takes on values in \( X \), and if the \( \Omega_0 \)-set defined by the condition \( f(\omega_0) \in E \) is in \( F_{\omega_0} \) for every set \( E \) of \( F_x \), then the function \( f(\omega_0) \) will be called measurable on \( \Omega_0 \). Let \( \{ f_n(\omega_0) \} \) be any sequence of such measurable functions, where the subscript \( n \) ranges through any aggregate \( \mathcal{A} \), not necessarily denumerable. If a probability measure \( P_0(\Lambda_0) \) is defined on the sets \( \Lambda_0 \) of \( F_{\omega_0} \), the measurable function \( f_n(\omega_0) \), considered from the standpoint of probability, is a chance variable \( x_n \). The following method of analyzing the mutual relations of such a set of chance variables has been used, more or less explicitly, in recent years.‡ Consider the space \( \Omega \) whose points are the aggregates \( \omega: \{ x_n \}, n \in \mathcal{A}, x_n \in X. \)§ If \( n \) runs through all real numbers \( t \), \( \Omega \) consists of all functions of the real variable \( t \),

* Presented to the Society, April 9, 1937; received by the editors July 26, 1937.
† A field of sets is a collection of sets \( E \) containing, with \( E_1 \) and \( E_2 \), their sum \( E_1 + E_2 \) and difference \( E_1 - E_2 \). A Borel field of sets is a field which contains with \( E_1, E_2, \ldots \) their sum \( \sum E_i \).
‡ Cf. Doob (I); Hopf (I); Khintchine (I); Kolmogoroff (II, pp. 24–30); Lévy (I and numerous papers); Lomnicki and Ulam (I); Paley and Wiener (I, chaps. 9 and 10, and earlier papers by Wiener). The Roman numerals refer to the bibliography at the end of the paper.
§ Each subscript \( n \) determines a coordinate \( x_n \), and the space \( \Omega \) is thus a Cartesian space with a dimension corresponding to each element of \( \mathcal{A} \).
with range in $X$; if $\mathcal{A}$ is the set of natural numbers, $\Omega$ is the space of all sequences $(x_1, x_2, \ldots)$, $x_n \in X$; if $\mathcal{A}$ is the set of all integers $\cdots, -1, 0, 1, \ldots$, $\Omega$ is the space of all sequences $(\cdots, x_{-1}, x_0, x_1, \cdots)$, $x_n \in X$. The last example will be the one studied in later sections, but in the present section no restrictions will be made on $\mathcal{A}$. Let $\alpha_1, \ldots, \alpha_p$ be any finite set of subscripts, and let $E_1, \cdots, E_p$ be sets of $F_x$. The conditions

$$x_{\alpha_j} \in E_j, \quad j = 1, \cdots, p,$$

(1.1)

determine a set of elements of $\Omega$. The class of all $\Omega$-sets defined in this way determines a Borel field $F_\omega$ of sets of $\Omega$.\* Evidently $\Omega \in F_\omega$. We shall define a probability measure $P(\Lambda)$ on the sets $\Lambda$ of $F_\omega$ which will have as its value, on the set defined by (1.1), the $P_0$-measure of the $\Omega_0$-set determined by the conditions

$$f_{\alpha_j}(\omega_0) \in E_j, \quad j = 1, \cdots, p.$$

(1.2)

The $P$-measure on $\Omega$ is defined by means of a mapping of $\Omega_0$ on $\Omega$, which takes the sets of $F_{\omega_0}$ into sets of $F_\omega$. Let $\omega_0$ be a point of $\Omega_0$. The map takes $\omega_0$ into the point $(\xi_n)$ of $\Omega$ defined by the equalities

$$\xi_n = f_n(\omega_0), \quad n \in \mathcal{A}.$$

(1.3)

To the $\Omega$-set determined by the conditions of (1.1) will then correspond the $\Omega_0$-set determined by the conditions of (1.2). Then to any set $\Lambda$ of $F_\omega$ will correspond a set $\Lambda_0$ in the Borel field of sets determined by those sets which are defined by conditions of type (1.2). Since the latter sets are in $F_{\omega_0}$, $\Lambda_0 \in F_{\omega_0}$. We define $P(\Lambda)$ as $P_0(\Lambda_0)$. In this way the study of the mutual measure relations of the aggregate of functions $\{f_n(\omega_0)\}$ (the probability relations of the aggregate of chance variables $\{x_n\}$) is reduced to the study of the properties of the space $\Omega$. The earlier representation of the chance variable $x_m$ by means of the function $f_m(\omega_0)$ defined on $\Omega_0$ has been replaced by a new representation by means of the function $x_m(\omega)$, defined on $\Omega$ and taking on the value $\xi_m$ at the point $\omega$: $(x_n)$ for which the $m$th coordinate is $\xi_m$.\†

\* The (Borel) field determined by a given collection of point sets can be defined as the intersection of all the (Borel) fields of sets which include the sets of the given collection.

\† The measure relations of $\Omega$ correspond to similar relations between the chance variables $\{x_n\}$; that is, between the original functions $f_i(\omega_0)$ in the sense that, if $\Lambda$ is an $\Omega$-set in the field $F_\omega$, the corresponding $\Omega_0$-set in the field $F_{\omega_0}$ is defined by conditions on the $f_i$'s which, when imposed on the $x$'s define $\Lambda$; and $P(\Lambda) = P_0(\Lambda_0)$. Due to the fact that the transformation from $\Omega_0$ to $\Omega$ is not one-to-one, certain relations of the functions $\{f_n(\omega_0)\}$ may become distorted; thus an $\Omega_0$-set in the field $F_{\omega_0}$ may not go into an $\Omega$-set in the field $F_\omega$. For example, if $\mathcal{A}$ is the set of real numbers $t$, so that $\Omega$ becomes the space of functions $x(t)$, and if $X$ is the space of real numbers, it may be that $f_i(\omega_0)$ is a continuous function of $t$ for all $\omega_0$; on the other hand the set of elements $x(t)$ of $\Omega$ which are continuous functions of $t$ will never be measurable (in terms of $P$-measure). Cf. Doob (II) for a detailed treatment of this case.
As an example of the advantages of this procedure, we give a discussion of the following classical theorem:

**If** \( x_1, \ldots, x_N \) **is a set of mutually independent chance variables, the expectation of their product is the product of their expectations.**

In order to treat this theorem we take \( X \) as the space of real numbers, \( F_x \) as the field of Borel sets of \( X \) (or some more inclusive field), and \( \mathcal{A} \) as the set of integers \( 1, \ldots, N \). The space \( \Omega \) becomes ordinary \( N \)-dimensional cartesian space. A probability measure is defined, in terms of the measure properties of \( x_i \) on the \( x_i \)-axis, and the \( P \)-measure on \( \Omega \) is determined in the usual (multiplicative) way from these separate measures. The theorem in question is now an immediate consequence of Fubini’s theorem on the equality of a multiple integral and the corresponding iterated integrals. Incidentally Fubini’s theorem provides very sensitive sets of possible hypotheses: it is sufficient to suppose that the expectation of every \( x_i \) exists, or else that the expectation of \( x_1 \cdots x_N \) exists.

In the preceding discussion, the given aggregate of chance variables was considered in a given representation in terms of a corresponding aggregate of measurable functions \( \{ f_n(\omega_0) \} \), all defined on the same space \( \Omega_0 \); and a new representation was obtained in terms of the aggregate of functions \( \{ x_n(\omega) \} \) defined on \( \Omega \). If the chance variables are not given in some such representation, the problem becomes more difficult. A family of chance variables is generally considered as a family of entities \( \{ x_n \} \), distinguished by a subscript \( n \) (which is usually identified in some way with the time) varying in an aggregate \( \mathcal{A} \). The chance variable \( x_n \), which takes on values in a space \( X \), is considered defined by a physical process in the course of its development.

*If a numerically-valued chance variable \( x \) is represented by a measurable function \( f \) defined on a space on which some measure is defined, and if \( f \) is absolutely integrable, the expectation of \( x \) is defined as the integral of \( f \). In treating this theorem we shall assume that the \( N \) chance variables are represented by \( N \) numerically-valued functions defined and measurable on a space on which some measure is defined. The fact that such a representation is always possible will appear later in this section. A recent proof of the theorem in question, with somewhat more stringent hypotheses than those to be given, and with the chance variables represented by Lebesgue measurable functions defined on the interval \( 0 \leq x \leq 1 \), was published by Kac (I, pp. 47–50).*

† The measure of the interval \( a < x_j < b \) is defined as the probability that \( a < x_j < b \).

‡ Cf. Saks, *Théorie de l’Intégrale*, Warsaw, 1933, pp. 257–263, for the details for \( N = 2 \). The \( P \)-measure is determined by the fact that the \( P \)-measure of the \( N \)-dimensional interval \( a_j < x_j < b_j \), \( j = 1, \ldots, N \), is defined as the product of the (1-dimensional) measures of the sides \( a_j < x_j < b_j \) previously defined.

§ Saks, ibid., p. 262.

|| Thus the chance variable \( x_j \) may be the \( x \)-coordinate of the position of a particle (in a Brownian movement) at time \( t \).
of $X$ such that $X$ is in $F_x$, and that if $\alpha_1, \ldots, \alpha_p$ are elements of $\mathcal{A}$, and if $E_1, \ldots, E_p$ are in $F_x$, a non-negative number is assigned to the $p$ conditions

$$x_{\alpha_j} \in E_j, \quad j = 1, \ldots, p.$$ 

This number is called the probability that the conditions of (1.4) are satisfied. If $\alpha_1, \ldots, \alpha_p$ are kept fixed, these probability numbers assign measures to certain sets of the space of points $(x_{\alpha_1}, \ldots, x_{\alpha_p})$, the sets being those determined by conditions of the form

$$(1.4') \quad x_{\alpha_j} \in E_j, \quad j = 1, \ldots, p.$$ 

It is supposed further that this ($p$-dimensional) measure function is additive for fixed subscripts $\alpha_1, \ldots, \alpha_p$, and that it can be defined on every set of the Borel field $F_{\alpha_1, \ldots, \alpha_p}$ (the field of $\Omega$-sets determined by the sets, defined by (1.4'), on which the function is already defined), in such a way that the extended set function is a ($p$-dimensional) probability measure. Now consider the space $\Omega$ and field $F_\omega$ as described above. An $\Omega$-set, determined by conditions imposed on a certain set of coordinates, will be called a cylinder set over those coordinates. It is readily shown that the field $F_{\alpha_1, \ldots, \alpha_p}$ is the field of cylinder sets of $F_\omega$ over $x_{\alpha_1}, \ldots, x_{\alpha_p}$. What was just described was therefore the determination of a probability measure on the cylinder sets of $F_\omega$ over $x_{\alpha_1}, \ldots, x_{\alpha_p}$. Moreover the various measures, obtained by varying the coordinates involved, are coherent in the sense that if $\Lambda$ is a cylinder set of $F_\omega$ over $x_{\alpha_1}, \ldots, x_{\alpha_p}$ and also a cylinder set over $x_{\beta_1}, \ldots, x_{\beta_q}$, then the probability measures, assigned to $\Lambda$ in the two representations, will be the same.

To show that the present situation is no more general than that described above, when a probability measure was defined on all the sets of $F_\omega$, not merely over the cylinder sets of $F_\omega$ over a finite number of coordinates, it is necessary to prove the following theorem:

**Theorem 1.1.** A set function, defined on every cylinder set of $F_\omega$ over a finite number of coordinates, which is a probability measure on the field of cylinder sets of $F_\omega$ over each such finite set of coordinates, can be so defined on the remaining sets of $F_\omega$ that it becomes a probability measure on this field.

This theorem was proved by Kolmogoroff (I, pp. 27–30) under the additional hypothesis that $X$ is the set of real numbers, and that $F_x$ is the field of Borel sets of $X$. Daniell (I) discussed measures on $\Omega$ in the case where $\Omega$ is the set of natural numbers (with $X, F_x$ defined as in Kolmogoroff's case) so that $\Omega$ is the space of sequences $\omega$: $(x_1, x_2, \ldots)$. This latter case, for which in addition the chance variables concerned are independent, has been discussed by many other writers who map $\Omega$ on a linear interval and define the
measure of an \( \Omega \)-set by means of the Lebesgue measure of the corresponding set on the interval.*

Let \( F \) be the collection of \( \Omega \)-sets each of which is determined by conditions of the form (1.4') or is a finite sum of such sets. Then \( F \) is a field,† and the given set function \( P(\Lambda) \) is defined on the sets of \( F \). Let \( \Lambda_0, \Lambda_1, \ldots \) be sets of \( F \). Then if \( \Lambda_1, \Lambda_2, \ldots \) are disjunct, and if, in addition,

\[
\Lambda_0 = \sum_1^{\infty} \Lambda_m,
\]

we shall show that

\[
P(\Lambda_0) = \sum_1^{\infty} P(\Lambda_m).
\]

We prove (1.6) by reducing it to the corresponding result in the special case considered by Kolmogoroff. Fix a value of \( n, n = \nu \), and consider the sets \( E_1^{(\nu)}, E_2^{(\nu)}, \ldots \) of \( X \) which are involved in the restrictions on \( x \), used to define \( \Lambda_0, \Lambda_1, \ldots \).‡ We shall define a numerically-valued function \( f_r(x) \), with domain \( X \), so that each set \( E_i^{(\nu)} \) is determined by simple inequalities imposed on \( f_r(x) \). In order to define the function \( f_r(x) \) we shall need the lemma which follows:

**Lemma 1.1.** Let \( \mathcal{E}_1, \mathcal{E}_2, \ldots \) be any point sets of an abstract space. There is a collection of sets \( \{\mathcal{E}_r\} \) (where \( r \) is rational and \( 0 < r < 1 \)) with the following properties:

(i) each \( \mathcal{E}_i \) is in the field determined by the sets \( \mathcal{E}_r \) and conversely;
(ii) if \( r_1 < r_2 \), \( \mathcal{E}_{r_1} \subseteq \mathcal{E}_{r_2} \);
(iii) if \( r_m \to r \), where \( r_1 > r_2 > \cdots \), then \( \prod_1^\infty \mathcal{E}'_{r_m} = \mathcal{E}_r \);
(iv) if \( r_m \to 0 \) (\( r_m \to 1 \)), then \( \prod_1^\infty \mathcal{E}'_{r_m} = 0 \) (\( \sum_1^\infty \mathcal{E}'_{r_m} = \sum_1^\infty \mathcal{E}_i \)).

This lemma can be proved by a modification of the proof of a similar result due to von Neumann.§ We do not suppose that there are necessarily infinitely many distinct sets \( \mathcal{E}_i \). In the contrary case, there will be only a finite number of distinct sets \( \mathcal{E}_i \).

Using this lemma, we define \( f_r(x) \) as follows: Identify the sets \( \{E_i^{(\nu)}\} \) (\( \nu \) fixed) with the sets \( \{\mathcal{E}_r\} \) of the lemma, and set

\[
f_r(x) = \limsup r, \quad x \in \mathcal{E}_r
\]

* The details of such a map can be found in Paley and Wiener (I, pp. 145–146). Łomnicki and Ulam (I) treat the case for which \( \mathcal{F} \) is the set of natural numbers, and the chance variables concerned are independent (with no restriction on \( X \)); but their proof of the theorem stated above (Theorem 1.1) is defective. (The mistake is in the proof of Lemma 4, pp. 254–255.)
† Cf. Kolmogoroff (II, pp. 25–26).
‡ Except for a denumerable set of superscripts \( \nu \), there will be only one set \( E_i^{(\nu)} \); namely \( X \) itself.
The set $\mathcal{E}'$ is characterized by the inequality

$$f_\alpha(x) \leq r.$$  

Since every set $\mathcal{E}_i$ is in the field determined by the sets $\mathcal{E}_i'$, every set $E_i^{(r)}$ is characterized by a finite number of inequalities imposed on $f_\alpha(x)$. Moreover, if $\mathcal{E}$ is any Borel set of real numbers, the $x$-set $E$ determined by the condition $f_\alpha(x) \in E$ is in the field $F_x$. The latter fact is apparent if $E$ is an interval, and its truth then follows for $E$ any Borel set.

Now consider the space $\Omega$ of points $\tilde{\omega} = \{\tilde{x}_n\}, n \in \mathcal{A}$, where $\tilde{x}_n$ is any real number. Let $F_\mathcal{Z}$ be the field of Borel sets of the $\tilde{x}$-axis, and let $F_\Omega$ be the Borel field of $F_{\tilde{x}}$ sets determined by $F_\mathcal{Z}$ in the same way that $F_\mathcal{A}$ is determined by $F_\mathcal{X}$. We map $\Omega$ on $\tilde{\Omega}$, sending the point $(x_n)$ of $\Omega$ into the point $(\tilde{x}_n)$ of $\tilde{\Omega}$ for which

$$\tilde{x}_m = f_m(x_m), \quad m \in \mathcal{A}. \tag{1.9}$$

This mapping is a single-valued transformation of $\Omega$ into some subset of $\tilde{\Omega}$. If $\Lambda$ is the $\tilde{\Omega}$-set determined by the conditions

$$\tilde{x}_{a_j} \in \tilde{E}_j, \quad j = 1, \ldots, p, \tag{1.10}$$

where $\tilde{E}_1, \ldots, \tilde{E}_p$ are Borel sets, the corresponding $\Omega$-set $\Lambda$ is determined by the conditions

$$f_{a_j}(x_j) \in \tilde{E}_j, \quad j = 1, \ldots, p. \tag{1.11}$$

It then follows from the definitions of $F_\mathcal{A}$ and $F_\mathcal{Z}$ that the $\Omega$-sets going into cylinder sets of $F_\mathcal{A}$, $\tilde{x}_{a_1}, \ldots, \tilde{x}_{a_p}$ are cylinder sets of $F_\tilde{\mathcal{A}}$ over $x_{a_1}, \ldots, x_{a_p}$, and that the $\Omega$-sets going into the sets of $F_\omega$ are sets of $F_\tilde{\mathcal{A}}$. We now define a set function $\tilde{P}(\Lambda)$, on the sets $\Lambda$ of $F_\mathcal{A}$ which are cylinder sets over a finite number of coordinates, by setting $\tilde{P}(\Lambda) = P(\Lambda)$, where $\Lambda$ is the $\Omega$-set containing every element $\omega$ which is taken by the transformation into an element $\tilde{\omega}$ of $\tilde{\Lambda}$. The set function $\tilde{P}(\Lambda)$ is uniquely defined and is a probability measure on the field of cylinder sets over any fixed finite set of coordinates. According to the definition of $\tilde{\Omega}$, there are sets $\tilde{\Lambda}_0, \tilde{\Lambda}_1, \ldots$ to which correspond the sets $\Lambda_0, \Lambda_1, \ldots$ of (1.5). The sets $\tilde{\Lambda}_1, \tilde{\Lambda}_2, \ldots$ are disjunct, and

$$\tilde{\Lambda}_0 = \sum_{i=1}^{\infty} \tilde{\Lambda}_i. \tag{1.5'}$$

Now we are assuming Kolmogoroff's result that Theorem 1.1 is true if $X$ is

---

* One can readily determine these inequalities explicitly, using the fact that any set in the field determined by the sets $\{\mathcal{E}_i\}$ can be written in the form $E_{r_0} + (E_{r_1} - E_{r_0}) + \cdots + (E_{r_N} - E_{r_{N-1}})$, with $r_0 < r_1 \leq r_2 \leq \cdots \leq r_N$, or else in the same form without the first set $E_{r_0}$.

† The space $\Omega$ is the space of numerically-valued functions with domain $\mathcal{A}$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
the space of real numbers, and if $F_x$ is the field of Borel sets. Then $\tilde{P}(\tilde{\Lambda})$ must certainly be completely additive on its domain of definition, so that

\[(1.6') \quad \tilde{P}(\tilde{\Lambda}_0) = \sum_{i=1}^{\infty} \tilde{P}(\tilde{\Lambda}_m);\]

and this is equivalent to (1.6), since $P(\Lambda_m) = \tilde{P}(\tilde{\Lambda}_m)$, $m \geq 0$.

The proof, that the domain of definition of $P(\Lambda)$ can be extended as described in Theorem 1.1, is now immediate. By hypothesis, $P(\Lambda)$ is an additive function of sets on the field $F_*$, and the result just proved shows that $P(\Lambda)$ is completely additive on this field. It is then possible, according to a well known extension theorem, to extend the definition of $P(\Lambda)$ to all the sets of the Borel field determined by the sets of $F$ (in this case the Borel field will be $F_\omega$), in such a way that $P(\Lambda)$ becomes a completely additive function of sets on the larger field. The set function thus obtained is the probability measure described in Theorem 1.1.

2. Definition of a stochastic process. We can now state the definition of a stochastic process (of the type to be studied in this paper) suggested by the preceding section. Let $X$ be any abstract space of elements $x$, and let $\Omega$ be the space whose elements $\omega$ are sequences $(\ldots, x_{-1}, x_0, x_1, \ldots)$ of points of $X$. Let $F_x$ be a Borel field of sets of points of $X$, including $X$ itself, and suppose that $E_1, \ldots, E_p$ are sets in $F_x$. The conditions

\[(2.1) \quad x_{\alpha_j} \in E_j, \quad j = 1, \ldots, p,\]

($\alpha_1, \ldots, \alpha_p$ any $p$ distinct integers) determine a cylinder set over $x_{\alpha_1}, \ldots, x_{\alpha_p}$. The class of all such cylinder sets determines the Borel field $F_\omega$ of $\Omega$-sets.

Let $P(\Lambda)$ be a probability measure defined on the sets of $F_\omega$. For a fixed set of coordinates $x_{\alpha_1}, \ldots, x_{\alpha_p}$, $P(\Lambda)$ becomes a probability measure defined essentially on the $p$-dimensional space of elements $(x_{\alpha_1}, \ldots, x_{\alpha_p})$, and the converse (Theorem 1.1) is also true. A stochastic process depending on the parameter $n$ running through integral values is the combination of the space $\Omega$ together with a probability measure defined on the sets of a field $F_\omega$. More precisely, the process is the changing real entity of which the above is the mathematical abstraction. Examples of stochastic processes are given in §5.

The function $x_i(\omega)$ taking on the value $x_i$ at the point $\omega$: $(\ldots, x_i, \ldots)$ is a measurable function on $\Omega \dagger$ taking on values in $X$; and the sequence of functions $\cdots, x_{-1}(\omega), x_0(\omega), x_1(\omega), \cdots$ can then be considered as a representation of a sequence of chance variables $\cdots, x_{-1}, x_0, x_1, \cdots$. Conversely,

* The field $F$ was defined at the beginning of the proof.
‡ Cf. §1.
we have seen in §1 that any such sequence of chance variables can be represented in this way.* It is sometimes useful to consider the sequence of chance variables \(x_1, x_2, \ldots\). To do this we need only restrict our attention to the cylinder sets of \(F_\omega\) over \(x_1, x_2, \ldots\).

We shall suppose throughout this paper that the probability measure \(P(\Lambda)\) is so extended that it is defined on every set \(\Lambda_1\) differing from a set \(\Lambda_0\) of \(F_\omega\) by at most a subset of a set on which \(P(\Lambda)\) vanishes, if we set \(P(\Lambda_1) = P(\Lambda_0)\). The sets of this extended domain of definition will be called \(\mathcal{P}\)-measurable. The subsets of a set of \(\mathcal{P}\)-measure 0 are measurable and of \(\mathcal{P}\)-measure 0. The \(\mathcal{P}\)-measure of a \(\mathcal{P}\)-measurable set is the greatest lower bound of the \(\mathcal{P}\)-measures of sets \(M \supset A\) which are finite or denumerably infinite sums of sets of the field \(F\) defined above.† It follows from this fact that if \(\Lambda\) is \(\mathcal{P}\)-measurable, to every positive number \(\epsilon\) corresponds a set \(\Lambda_\epsilon\) which is a cylinder set of \(F_\omega\) over a finite number of coordinates, and which has the property that \(P(\Lambda \cdot \Lambda_\epsilon) + P(C\Lambda \cdot \Lambda_\epsilon) < \epsilon.\) If \(\Lambda\) is a \(\mathcal{P}\)-measurable cylinder set over \(x_{a_1}, \ldots, x_{a_p}\), it is determined by a condition of the form \((x_{a_1}, \ldots, x_{a_p}) \in E,\) where \(E\) is a \(p\)-dimensional set of points \((x_{a_1}, \ldots, x_{a_p})\). The set \(E\) will be called an \((x_{a_1}, \ldots, x_{a_p})\)-set of \(\mathcal{P}\)-measure \(P(\Lambda)\).

Let \(f(\omega)\) be a numerically-valued function of \(\omega\). If for every number \(k\) the inequality \(f(\omega) > k\) defines a set of \(F_\omega\) (a \(\mathcal{P}\)-measurable set), \(f(\omega)\) will be called measurable with respect to \(F_\omega\) (\(\mathcal{P}\)-measurable). If \(f(\omega)\) is measurable with respect to \(F_\omega\), then it is \(\mathcal{P}\)-measurable; and conversely if \(f(\omega)\) is \(\mathcal{P}\)-measurable, there is a function \(f_1(\omega)\), measurable with respect to \(F_\omega\) and equal to \(f(\omega)\) except possibly on a set of \(\mathcal{P}\)-measure 0.§ This can be deduced from the following fact (which in turn follows readily from the approximation of \(\mathcal{P}\)-measurable sets by means of cylinder sets of \(F_\omega\) over a finite number of coordinates) that if \(f(\omega)\) is any \(\mathcal{P}\)-measurable function, to every positive number \(\epsilon\) corresponds a function \(f_1(\omega)\), measurable with respect to \(F_\omega\), depending on only a finite number of coordinates, and having the property that \(|f(\omega) - f_1(\omega)| \leq \epsilon\) except perhaps on an \(\Omega\)-set of \(\mathcal{P}\)-measure \(\leq \epsilon\). Throughout the above if the given function depends only on some given set of coordinates, the approximating functions can be supposed to depend only on the same coordinates. If \(f(\omega)\) is measurable with respect to \(F_\omega\), and if \(\{n_i\}\) is any set of integers, \(f(\omega)\) becomes a function of the coordinates \(x_{n_1}, x_{n_2}, \ldots\) only, if the other coordinates are

* Cf., however, the note on p. 88 in accordance with which it may sometimes be necessary to define a stochastic process using a subspace of \(\Omega\) rather than \(\Omega\) itself, as in Doob (II).
† The extension theorem used in the proof of Theorem 1.1 defines \(P(\Lambda)\) in precisely this way.
‡ The complement of any set \(\Lambda\) will be denoted by \(C\Lambda\) throughout this paper.
held fast, and this function of \( x_{n_1}, x_{n_2}, \cdots \) will be also measurable with respect to \( F_\omega \).

In later sections we shall use the fact that if \( F_\omega \) is the Borel field determined by some denumerable collection of its sets, the same will be true of \( F_\omega \). Even without this hypothesis, it can be shown (by transfinite induction) that any given set \( \Lambda \), in the field \( F_\omega \), is in the field \( F_\omega' \) corresponding to some Borel field \( F_\omega' \subseteq F_\omega \) such that \( F_\omega' \) is the Borel field determined by some denumerable collection of its sets. It then follows that if \( f(\omega) \) is a function measurable with respect to \( F_\omega \), a subfield \( F_\omega' \) of \( F_\omega \) can be found which satisfies the denumerability condition just described and is such that \( f(\omega) \) is measurable with respect to the corresponding field \( F_\omega' \).

The integral of the \( P \)-measurable function \( f(\omega) \) over a \( P \)-measurable set \( \Lambda \) will be noted by

\[
\int_\Lambda f(\omega) dP;
\]

and if the domain of integration is not stated explicitly, it will be understood to be \( \Omega \). If \( f(\omega) \) depends only on a finite number of coordinates \( x_{a_1}, \cdots, x_{a_p} \), and if \( \Lambda \) is a cylinder set over those coordinates, we shall use the notation

\[
\int_\Lambda f(\omega) P(\omega_{a_1, \cdots, a_p})
\]

for the integral of \( f(\omega) \) over \( \Lambda \). Corresponding notation will be used for integration with respect to other probability measures.

Let \( T\omega \) be the transformation taking \( \omega: (\cdots, x_{-1}, x_0, x_1, \cdots) \) into \( \omega': (\cdots, x_{-2}, x_{-1}, x_0, \cdots) \), that is, the transformation defined by

\[
x_j' = x_{j-1}, \quad j = 0, \pm 1, \pm 2, \cdots
\]

If \( T\omega \) is measure preserving, the process is called temporally homogeneous.

3. The conditional probability functions.* Let \( \Lambda \) be a \( P \)-measurable set. The conditional probability function \( P_{a_1, \cdots, a_p}(x_{a_1}, \cdots, x_{a_p}; \Lambda) \) is defined as follows. If the set \( M \) is allowed to range through the \( P \)-measurable cylinder sets over \( x_{a_1}, \cdots, x_{a_p} \), \( P(\Lambda M) \) is a non-negative completely additive function of sets \( M \) which vanishes whenever \( P(M) = 0 \). There is therefore a function \( P_{a_1, \cdots, a_p}(x_{a_1}, \cdots, x_{a_p}; \Lambda) \), a non-negative \( P \)-measurable function depending only on the coordinates \( x_{a_1}, \cdots, x_{a_p} \), satisfying

* The ideas in this and the following section are not new, but there seems to be no systematic presentation of many of them in the literature, and some of the theorems have not been previously stated explicitly. (Known theorems and definitions are stated for later reference.) The importance and usefulness of the conditional probability and conditional expectation functions have been stressed most by P. Lévy.
(3.1) \[
\int_{M} P_{a_1, \ldots, a_p}(x_{a_1}, \ldots, x_{a_p}; \Lambda) dP = P(\Lambda M)
\]
for every set \(M\). The function \(P_{a_1, \ldots, a_p}(x_{a_1}, \ldots, x_{a_p}; \Lambda)\) is determined uniquely up to an \((x_{a_1}, \ldots, x_{a_p})\)-set of \(P\)-measure 0. For a given set \(\Lambda\), \(P_{a_1, \ldots, a_p}(x_0^{a_1}, \ldots, x_0^{a_p}; \Lambda)\) is called the conditional probability of \(\Lambda\) if \(x_{a_j} = x_0^{a_j}, j = 1, \ldots, p\). The subscripts determining the function are given by the subscripts in the argument, so there will be no danger of confusion if \(P_{a_1, \ldots, a_p}(x_{a_1}, \ldots, x_{a_p}; \Lambda)\) is written simply as \(P(x_{a_1}, \ldots, x_{a_p}; \Lambda)\). We shall need the following properties of \(P(x_{a_1}, \ldots, x_{a_p}; \Lambda)\) which are easily derived from the definition:

(i) If \(P(\Lambda) = 1\), then \(P(x_{a_1}, \ldots, x_{a_p}; \Lambda) = 1\), except possibly on an \((x_{a_1}, \ldots, x_{a_p})\)-set of \(P\)-measure 0. If \(P(\Lambda) = 0\), then \(P(x_{a_1}, \ldots, x_{a_p}; \Lambda) = 0\), except possibly on an \((x_{a_1}, \ldots, x_{a_p})\)-set of \(P\)-measure 0.

(ii) If \(A_1, A_2, \ldots\) are disjunct \(P\)-measurable sets, and if \(A = \bigcup_{m=1}^{\infty} A_m\), then
\[
P(x_{a_1}, \ldots, x_{a_p}; \Lambda) = \sum_{m=1}^{\infty} P(x_{a_1}, \ldots, x_{a_p}; \Lambda_m)
\]
except possibly for an \((x_{a_1}, \ldots, x_{a_p})\)-set of \(P\)-measure 0.

This implies the following fact:

(iii) If \(\Lambda', \Lambda''\) are \(P\)-measurable, and if \(\Lambda' \subseteq \Lambda''\), then
\[
P(x_{a_1}, \ldots, x_{a_p}; \Lambda') \leq P(x_{a_1}, \ldots, x_{a_p}; \Lambda'')
\]
except possibly for an \((x_{a_1}, \ldots, x_{a_p})\)-set of \(P\)-measure 0.

By taking complements in (ii) we obtain the following property:

(iv) If \(\Lambda_1, \Lambda_2, \ldots\) are \(P\)-measurable sets, and if
\[
\Lambda_1 \supseteq \Lambda_2 \supseteq \cdots, \quad \prod_{m=1}^{\infty} \Lambda_m = \Lambda,
\]
then
\[
\lim_{m \to \infty} P(x_{a_1}, \ldots, x_{a_p}; \Lambda_m) = P(x_{a_1}, \ldots, x_{a_p}; \Lambda),
\]
except possibly for an \((x_{a_1}, \ldots, x_{a_p})\)-set of \(P\)-measure 0.

**Theorem 3.1.** If \(F_1\) is any Borel field of \(P\)-measurable sets, including the set \(\Omega\), determined by a denumerable subcollection \(\Lambda_1, \Lambda_2, \ldots\), and if \(\alpha_1, \ldots, \alpha_p\) are any given subscripts, then \(P(x_{a_1}, \ldots, x_{a_p}; \Lambda)\) can be so defined that there is an \((x_{a_1}, \ldots, x_{a_p})\)-set \(E\) of \(P\)-measure 0 such that for \((x_{a_1}, \ldots, x_{a_p})\) fixed, not in \(E\), \(P(x_{a_1}, \ldots, x_{a_p}; \Lambda)\), \((\Lambda \in F_1)\), is a probability measure.

* This definition is due to Kolmogoroff (I, pp. 41–44).
† Cf. Kolmogoroff (II, pp. 43–44).
If we identify the sets $A_1, A_2, \ldots$ with the sets $E_1, E_2, \ldots$ of Lemma 1.1, we obtain sets $A'_1$ corresponding to the sets $E'_1$ of that lemma. We then map $\Omega$ on the $t$-interval $0 < t < 1$ by the transformation which takes a point $\omega$ into $t$ if

$$t = L. \ U. \ B. \ \omega.$$  

According to properties (i), (iii), (iv) of the conditional probability functions, if $r, s$ are rational, with $r < s$, there is an $(x_{a_1}, \ldots, x_{a_p})$-set $E_{rs}$ of $P$-measure 0 such that if $(x_{a_1}, \ldots, x_{a_p}) \in E_{rs},$

$$P(x_{a_1}, \ldots, x_{a_p}; A'_r) \leq P(x_{a_1}, \ldots, x_{a_p}; A'_s),$$  

and there is an $(x_{a_1}, \ldots, x_{a_p})$-set $E_r$ of $P$-measure 0 such that if $r'$ approaches $r$ from above ($r, r'$ rational, $r \geq 0$), and if $(x_{a_1}, \ldots, x_{a_p}) \in E'_r,$ then

$$\lim_{r' \to r} P(x_{a_1}, \ldots, x_{a_p}; A'_r) = P(x_{a_1}, \ldots, x_{a_p}; A'_r)$$  

(where if $r = 0$ the right side is replaced by 0). Let $E'$ be defined by

$$E' = \sum_{r,s} E_{rs} + \sum_r E_r,$$

and suppose that $(x_{a_1}, \ldots, x_{a_p})$ is fixed, not in $E'$. Then

$$F(r) = P(x_{a_1}, \ldots, x_{a_p}; A'_r)$$

is a monotone non-decreasing function of $r$, defined for rational values of $r$ between 0 and 1 and continuous on the right at these values. Define $F(t)$ for every value of $t$ in the interval $0 \leq t < 1$ as $\lim_{r \to t} F(r)$ ($r$ rational, $r > t$). This is consistent with the previous definition if $t$ is rational, and $F(r)$ thus becomes a monotone non-decreasing function $F(t)$ defined for $0 \leq t < 1$ and continuous on the right. There is a non-negative completely additive function of Borel sets on the $t$-interval $(0, 1)$ determined by the condition that its value on the interval $0 < t \leq r$ is $F(r)$.* The Borel $t$-sets correspond to the sets of a certain Borel field $\tilde{F}$ of $\Omega$-sets, under the transformation from $\omega$ to $t$, and a non-negative completely additive set function $\tilde{P}(x_{a_1}, \ldots, x_{a_p}; \Lambda)$ is thereby defined on these $\Omega$-sets. The Borel sets of the interval $(0, 1)$ are the sets of the Borel field determined by intervals of the type $0 < t \leq r$, for $r$ rational, so that the sets of $\tilde{F}$ are the sets of the Borel field determined by the images of such intervals. The image of the interval $0 < t \leq r$ (r rational) is the set $\Lambda'_r$, so that the field $\tilde{F}$ includes every set $\Lambda'_r$ and therefore every set $\Lambda_r$; $\tilde{F} \supseteq F_1$. By definition of $\tilde{P}(x_{a_1}, \ldots, x_{a_p}; \Lambda)$, if $r$ is rational,

(3.4) \[ \tilde{P}(x_{a_1}, \ldots, x_{a_p}; \Lambda') = P(x_{a_1}, \ldots, x_{a_p}; \Lambda'). \]

We deduce from this, using the properties of the conditional probability functions listed above, that if \( \Lambda \) is a set in the field \( F_1 \),

(3.5) \[ \tilde{P}(x_{a_1}, \ldots, x_{a_p}; \Lambda) = P(x_{a_1}, \ldots, x_{a_p}; \Lambda), \]

except perhaps on an \((x_{a_1}, \ldots, x_{a_p})\)-set of \( P \)-measure 0; and to define \( P(x_{a_1}, \ldots, x_{a_p}; \Lambda) \) as required in the statement of the theorem, we need only re-define \( \tilde{P}(x_{a_1}, \ldots, x_{a_p}; \Lambda) \) as \( \tilde{P}(x_{a_1}, \ldots, x_{a_p}; \Lambda) \).

If \( \phi(x_{a_1}, \ldots, x_{a_p}, x_{\beta_1}, \ldots, x_{\beta_q}) = \phi((x_a), (x_\beta)) \) is a \( P \)-measurable and integrable function depending only on the indicated coordinates, we now define the function

(3.6) \[ E(x_{a_1}, \ldots, x_{a_p}; \phi) = \int \phi((x_a), (x_\beta))P(x_{a_1}, \ldots, x_{a_p}; d\beta_1, \ldots, \beta_q) \]

not as an integral, but in such a way that the indicated integration actually gives \( E(x_{a_1}, \ldots, x_{a_p}; \phi) \) when it can be carried out. Let \( M_a \) be a \( P \)-measurable cylinder set over \( x_{a_1}, \ldots, x_{a_p} \). Then \( E(x_{a_1}, \ldots, x_{a_p}; \phi) \) is defined as the \( P \)-measurable integrable function which satisfies

(3.7) \[ \int_{M_a} E(x_{a_1}, \ldots, x_{a_p}; \phi) dP = \int_{M_a} \phi((x_a), (x_\beta)) dP \]

for all sets \( M_a \). The function \( E(x_{a_1}, \ldots, x_{a_p}; \phi) \) is known as the conditional expectation of \( \phi \) for given \((x_{a_1}, \ldots, x_{a_p})\). Changing \( \phi \) on a set of \( P \)-measure 0 does not affect \( E(x_{a_1}, \ldots, x_{a_p}; \phi) \). If \( \phi \) is the characteristic function of a \( P \)-measurable cylinder set over \( x_{\beta_1}, \ldots, x_{\beta_q} \),

\[ E(x_{a_1}, \ldots, x_{a_p}; \phi) = P(x_{a_1}, \ldots, x_{a_p}; \Lambda). \]

**Theorem 3.2.** Suppose that \( P(x_{a_1}, \ldots, x_{a_p}; \Lambda) \) can be defined so that for each \((x_{a_1}, \ldots, x_{a_p})\) not in some \((x_{a_1}, \ldots, x_{a_p})\)-set of \( P \)-measure 0, \( P(x_{a_1}, \ldots, x_{a_p}; \Lambda) \) becomes a probability measure for \( \Lambda \) in the field of cylinder sets of \( F_\omega \) over \( (x_{\beta_1}, \ldots, x_{\beta_q}) \). Then (3.6) can be interpreted as ordinary integration.

This means, if \( \phi((x_a), (x_\beta)) \) is \( P \)-measurable and integrable, that (a) whenever \((x_{a_1}, \ldots, x_{a_p})\) is not in some exceptional set of \( P \)-measure 0, \( \phi((x_a), (x_\beta)) \) for \((x_{a_1}, \ldots, x_{a_p}) = (x^0_{a_1}, \ldots, x^0_{a_p}) \) is measurable in terms of the measure function \( P(x_{a_1}, \ldots, x_{a_p}; d\beta_1, \ldots, \beta_q) \) (that is, that the cylinder set over \((x_{\beta_1}, \ldots, x_{\beta_q})\) defined by \( \phi > \kappa \) with \( (x_{a_1}, \ldots, x_{a_p}) = (x^0_{a_1}, \ldots, x^0_{a_p}) \) is either in \( F_\omega \) or differs from such a set by a subset of such a set on which \( P(x^0_{a_1}, \ldots, x^0_{a_p}) \).
to \( x^0_{a_0}; \Lambda \) vanishes); and that (b) the integral (3.6) exists and is \( E(x_{a_1}, \ldots, x_{a_p}; \Lambda) \), if we neglect sets of \( P \)-measure 0.

We shall suppose that \( p = q = 1 \) to simplify the notation, and we can then drop the subscript 1 from \( \alpha \) and \( \beta \). We shall suppose that \( P(x_{a}; \Lambda_{x}) \) is already defined to satisfy the conditions of the theorem. In order to avoid confusion we shall reserve the integral sign throughout this proof for actual integration.

(i) According to a theorem of Kolmogoroff (II, pp. 48–49), if \( \phi(x_{a}, x_{\beta}) \) is \( P \)-measurable and integrable, then

\[
\left( 3.8 \right) \quad E(x_{a}; \phi) = \lim_{\lambda \to 0} \sum_{k=-\infty}^{+\infty} k\lambda P(x_{a}; \lambda k \leq \phi < \lambda(k + 1)), \quad \lambda > 0, *
\]

(if \( \lambda \) approaches 0 taking on only a denumerable set of values), except perhaps for an \( x_{a} \)-set of \( P \)-measure 0. The series in (3.8) converges absolutely for each value of \( \lambda \), except perhaps for an \( x_{a} \)-set of \( P \)-measure 0. In particular, suppose that \( \phi(x_{a}, x_{\beta}) \) is measurable with respect to \( F_{\omega} \). Then for fixed \( x_{a} \), \( \phi(x_{a}, x_{\beta}) \) becomes a function of \( x_{\beta} \) which is measurable with respect to \( F_{\omega} \) (cf. §2). The existence of the limit on the right, for a fixed value of \( x_{a} \), is exactly a condition that the integral

\[
\int \phi(x_{a}, x_{\beta})P(x_{a}; d\varepsilon_{\beta})
\]

exist; and in fact the limit is equal to this integral. Theorem 3.2 thus follows from Kolmogoroff’s result for a function which is measurable with respect to \( F_{\omega} \).

(ii) Let \( \phi_{0}(x_{a}, x_{\beta}) \) be the characteristic function of a set \( \Lambda_{0} \) in the field \( F_{\omega} \), of \( P \)-measure 0, and let \( \Lambda(\xi), (\xi \epsilon X) \), be the cylinder set over \( x_{\beta} \) determined by the condition \( \phi_{0}(\xi, x_{\beta}) = 1 \). Then, neglecting \( x_{a} \)-sets of \( P \)-measure 0 and using (i), we obtain

\[
0 = E(x_{a}, \phi_{0}) = \int \phi_{0}(x_{a}, x_{\beta})P(x_{a}; d\varepsilon_{\beta}) = P(x_{a}; \Lambda(x_{a}));
\]

that is, \( P(x_{a}; \Lambda(x_{a})) = 0 \). This same result will be true even if \( \Lambda_{0} \) is only supposed \( P \)-measurable, since it can then be enclosed in a set \( \Lambda_{0}' \), in the field \( F_{\omega} \), which is of \( P \)-measure 0.† From this it follows that if \( \phi(x_{a}, x_{\beta}) \) is any function which vanishes except perhaps on an \( (x_{a}, x_{\beta}) \)-set of \( P \)-measure 0, then

* The function \( P(x_{a}; \lambda k \leq \phi < \lambda(k + 1)) \) is the conditional probability, for given \( x_{a} \), that \( \lambda k \leq \phi < \lambda(k + 1) \). Kolmogoroff’s result does not require the hypotheses of Theorem 3.2.

† Cf. §2. The result is a generalization of the fact that if \( E \) is a Lebesgue measurable set of measure 0, in two-dimensional \((x, y)\) space, the intersection of \( E \) with the line \( x = c \) will be of (one-dimensional) Lebesgue measure 0 for almost all values of \( c \).
\[ E(x_\alpha, \phi) = \int \phi(x_\alpha, x_\beta) P(x_\alpha; d\theta) = 0, \]

if we neglect an \( x_\alpha \)-set of \( P \)-measure 0.

(iii) Let \( \phi(x_\alpha, x_\beta) \) be any \( P \)-measurable and integrable function depending only on \( x_\alpha, x_\beta \). Then it can be expressed in the form

\[ \phi(x_\alpha, x_\beta) = \phi_0(x_\alpha, x_\beta) + \phi_1(x_\alpha, x_\beta) \]

(cf. §2), where \( \phi_0(x_\alpha, x_\beta) \) vanishes except perhaps on an \((x_\alpha, x_\beta)\)-set of \( P \)-measure 0, and where \( \phi_1(x_\alpha, x_\beta) \) is measurable with respect to \( F_\omega \). Combining this fact with the results of (i) and (ii) we see that (neglecting \( x_\alpha \)-sets of \( P \)-measure 0), for fixed \( x_\alpha = x_\alpha^0 \), \( \phi(x_\alpha^0, x_\beta) \) is measurable in \( x_\beta \) with respect to the measure function \( P(x_\alpha^0; d\theta) \), and that

\[ \int \phi(x_\alpha, x_\beta) P(x_\alpha; d\theta) = \int \phi_1(x_\alpha, x_\beta) P(x_\alpha; d\theta) = E(x_\alpha; \phi_1) = E(x_\alpha; \phi), \]

as was to be proved. Conversely it is readily seen that if (3.6) exists as an integral, except possibly for an \((x_\alpha)\)-set of \( P \)-measure 0, and if the function of \((x_\alpha, \cdots, x_\alpha)\) thus obtained is integrable, then \( \phi((x_\alpha), (x_\beta)) \) is itself integrable, and the original definition of conditional expectation is applicable.

According to Theorem 3.1, the hypotheses of Theorem 3.2 are always satisfied if \( F_x \) (and therefore \( F_\omega \)) is the Borel field determined by a denumerable collection of its sets. This will be true, for example, if \( X \) is euclidean space of \( N \geq 1 \) dimensions, and if \( F_x \) is the field of Borel sets of \( X \). A stronger statement can be made, however, since if \( \phi_1 \) is measurable with respect to \( F_\omega \), there is always (cf. §2) a Borel field \( F_x(\phi_1) \) of \( X \)-sets, (depending on \( \phi_1 \)), which is the Borel field determined by a denumerable collection of its sets, such that if \( F_\omega(\phi_1) \) is the Borel field of \( \Omega \)-sets defined in terms of \( F_x(\phi_1) \), as \( F_\omega \) is defined in terms of \( F_x \), then \( \phi_1 \) is measurable with respect to \( F_\omega(\phi_1) \). Theorem 3.1 then shows, since only sets of \( F_x(\phi_1) \) (or \( F_\omega(\phi_1) \)) are involved, that (3.6) can always be interpreted as integration,* if we define \( P(x_\alpha; \Lambda_\beta) \) in a way depending on the function \( \phi \) under consideration.

The following theorem, proved by Kolmogoroff (II, pp. 47–48), is stated for future reference:

**Theorem 3.3.** If \( \alpha, \beta, \gamma \) are distinct integers, and if \( \phi(x_\alpha, x_\beta, x_\gamma) \) is \( P \)-measurable, then

\[ (3.9) \quad E(x_\alpha; \phi) = E[x_\alpha; E(x_\alpha, x_\beta; \phi)]. \]

*The transition from the \( P \)-measurable function \( \phi \) to the function \( \phi_1 \), measurable with respect to \( F_\omega \), is made as in (ii) above.
In particular, if $\phi$ is the characteristic function of a $P$-measurable cylinder set $\Lambda$, over $x_\gamma$,

\begin{equation}
(3.10) \quad P(x_\alpha; \Lambda) = E[x_\alpha; P(x_\alpha, x_\beta; \Lambda)] = \int P(x_\alpha, x_\beta; \Lambda)P(x_\alpha; \, de_\beta).\]
\end{equation}

The following theorem will be useful:

**Theorem 3.4.** Let $\phi(x_{a_1}, \cdots, x_{a_p})$ be a $P$-measurable integrable function. Then

\begin{equation}
(3.11) \quad \int \phi dP = \int P(de_{a_1}) \int P(x_{a_1}; de_{a_1}) \int \cdots \int \phi P(x_{a_1}, \cdots, x_{a_{q-1}}; de_{a_{q-1}}).\]
\end{equation}

This theorem can be considered as a corollary to the preceding one, but a direct proof will be given by induction. If $p=1$, (3.11) becomes

\[ \int \phi(x_{a_1})dP = \int \phi(x_{a_1})P(de_{a_1}), \]

which is certainly true. Suppose that $q>1$ and that the theorem is true for $p<q$. In (3.11) (with $p=q$), the first (symbolic integration) gives $E(x_{a_1}, \cdots, x_{a_{q-1}}; \phi)$. Since the theorem is supposed true for $p=q-1$, the right side of (3.11) then collapses to

\[ \int E(x_{a_1}, \cdots, x_{a_{q-1}}; \phi)dP, \]

and this is equal to the left side of (3.11) by the definition of conditional expectation.

Most stochastic processes which have been discussed in the mathematical literature are Markoff processes, that is, processes which satisfy the following
condition: If $\alpha < \beta$, and if $\Lambda$ is a $P$-measurable cylinder set over $x_\beta$, $x_{\beta+2}$, $\ldots$, then, except perhaps for an $(x_\alpha, x_{\alpha+1}, \ldots, x_\beta)$-set of $P$-measure 0,

$$P(x_\alpha, \ldots, x_\beta; \Lambda) = P(x_\beta; \Lambda).$$

It follows at once that if $\alpha_1 < \alpha_2 < \ldots < \alpha_p$, and if $\Lambda$ is a $P$-measurable cylinder set over $x_{\alpha_p+1}$, $x_{\alpha_p+2}$, $\ldots$, then if we neglect an $(x_{\alpha_1}, \ldots, x_{\alpha_p})$-set of $P$-measure 0, $P(x_{\alpha_1}, \ldots, x_{\alpha_p}; \Lambda) = P(x_{\alpha_p}; \Lambda)$. If $\alpha < \beta$, and if $\Lambda$ is a cylinder set over $x_\beta$, $x_{\beta+1}$, $\ldots$, (3.10) implies that

$$P(x_\alpha; \Lambda) = \int P(x_\beta; \Lambda)P(x_\alpha; d\beta)$$

for a Markoff process. Markoff processes are sometimes carelessly discussed in the literature as if they were the general case, as if (3.12) followed from the definition of probability.

4. Probability measures in terms of the conditional probability functions. In §3, the conditional probability functions were derived from the measure relations of a stochastic process. In this section the converse problem will be discussed. It will be seen that more is supposed below to be true of the conditional probabilities than is true in the general case, but the hypotheses are wide enough to cover the applications to be made.

Let $F_x$, $F_\omega$ be fields as described in §2. Suppose that for every pair of integers $m$, $n$, with $m \leq n$ and cylinder set $\Lambda_{n+1}$ in the field $F_\omega$ over $x_{n+1}$, a function $P(x_m, \ldots, x_n; \Lambda_{n+1})$ is defined and has the following properties:

(i) For fixed $x_m, \ldots, x_n$, $P(x_m, \ldots, x_n; \Lambda_{n+1})$ is a probability measure on the field of sets $\Lambda_{n+1}$.

(ii) For fixed $\Lambda_{n+1}$, $P(x_m, \ldots, x_n; \Lambda_{n+1})$ is measurable with respect to $F_\omega$.

Let $Q(\Lambda)$ be a probability measure defined on the field of cylinder sets of $F_\omega$ over $x_m$. There is then, as we shall now show, a uniquely determined probability measure defined on the cylinder sets of $F_\omega$ over $x_m, x_{m+1}, \ldots$, having the given functions $P(x_m, \ldots, x_n; \Lambda_{n+1})$ as its conditional probability functions and equal to $Q(\Lambda)$ if $\Lambda$ is a cylinder set of $F_\omega$ over $x_m$. If $\phi$ is the characteristic function of any cylinder set of $F_\omega$ over $x_m, \ldots, x_n$, $P(\Lambda)$ is defined by an iterated integral

$$P(\Lambda) = \int Q(d\epsilon_m) \int P(x_m; d\epsilon_{m+1}) \int P(x_m, x_{m+1}; d\epsilon_{m+2}) \int \ldots \int \phi_{\Lambda} P(x_m, \ldots, x_{n-1}; d\epsilon_n).$$

* A detailed discussion of the physical meaning of a Markoff process is given in Kolmogoroff (I).

† Cf. P. Lévy (I, pp. 121–123).
To show that this defines $P(\Lambda)$ uniquely it is necessary to show that if $\Lambda$ is also a cylinder set over $x_m, \ldots, x_{n'}$, $(n' \neq n)$, the expression (4.1) and the corresponding expression

$$P'(\Lambda) = \int Q(dem) \int P(x_m; dem+1) \int P(x_m, x_{m+1}; dem+2) \int$$

$$\cdots \int \phi_\Lambda P(x_m, \ldots, x_{n'-1}; dem')$$

(4.1')

are equal. We can suppose without restricting generality that $n' > n$. Then

$$P'(\Lambda) = \int Q(dem) \int \cdots \int P(x_m, \ldots, x_{n-1}; dem) \int$$

$$\cdots \int \phi_\Lambda P(x_m, \ldots, x_{n'-1}; dem'),$$

and since $\phi_\Lambda$ can depend only on $x_m, \ldots, x_n$, the first integration gives $P(x_m, \ldots, x_{n-1}; \Omega)$ (which is identically 1) multiplied by $\phi_\Lambda$. Similarly the next integrations, up to the integration over $x_n$, give $\phi_\Lambda$; hence $P'(\Lambda) = P(\Lambda)$. Evidently $P(\Lambda)$, as thus defined, is a probability measure on the field of cylinder sets of $F_\omega$ over any finite set of coordinates with subscripts at least equal to $m$. It then follows from Theorem 1.1, as applied to the case where $\mathcal{A}$ is the set of integers $m, m+1, \ldots$, and where $X, F_x$ are as here given, that the domain of definition of $P(\Lambda)$ can be extended to include all the cylinder sets of $F_\omega$ over $x_m, x_{m+1}, \ldots$ in such a way that the extended set function is a probability measure. Since if $\Lambda, M$ are respectively cylinder sets over $x_{n+1}$ and $x_m, \ldots, x_n$, with characteristic functions $\phi_\Lambda, \phi_M$,

$$\int_M P(x_m, \ldots, x_n; \Lambda)P(dem, \ldots, n) = \int P(x_m, \ldots, x_n; \Lambda)\phi_M P(dem, \ldots, n)$$

$$= \int Q(dem)P(x_m; dem+1) \int \cdots \int \phi_M P(x_m, \ldots, x_n; \Lambda)P(x_m, \ldots, x_{n-1}; dem)$$

$$= \int Q(dem) \int \cdots \int P(x_m, \ldots, x_{n-1}; dem) \int \phi_\Lambda \phi_M P(x_m, \ldots, x_n; dem+n)$$

$$= P(\Lambda \cdot M);$$

the conditional probability functions determined by $P(\Lambda)$ are actually the given ones. The set function $P(\Lambda)$ thus exists and it is uniquely determined by $Q(E)$ and the given conditional probability functions, since (4.1) holds for $P(\Lambda)$ either as a definition or as a theorem, because of Theorem 3.4.
What made this problem simple was the fact that the given conditional probability functions were entirely independent of each other; that is, there were no necessary relations between the given functions. This was possible because only cylinder sets over \( x_m, x_{m+1}, \ldots \) were being considered, for \( m \) fixed. If \( m \) is not to be kept fixed, the set of conditional probability functions can no longer be chosen independently of each other. We shall only consider the problem in detail for conditional probability functions corresponding to Markoff processes. In the treatment just given, if we had supposed that \( P(x_m, \ldots, x_n; \Lambda_{n+1}) \) depended only on \( x_n \), the resulting process would have been a Markoff process. To extend the results to the consideration of all the sets of \( F_n \), we shall need the following lemma:

**Lemma 4.1.** Let \( \{Q_N(\mathcal{E})\} \) be a sequence of probability measures defined on the sets of some Borel field \( S \) of sets of an abstract space. Suppose that

\[
(4.2) \quad \mathcal{E}_1 \supseteq \mathcal{E}_2 \supseteq \cdots, \prod_{i=1}^{\infty} \mathcal{E}_r = 0
\]

implies that

\[
(4.3) \quad \lim_{N \to \infty} Q_N(\mathcal{E}_r) = 0
\]

uniformly in \( N \).

(i) If

\[
(4.4) \quad \lim_{N \to \infty} Q_N(\mathcal{E}) = Q(\mathcal{E})
\]

exists for every \( \mathcal{E} \) in \( S \), the set function \( Q(\mathcal{E}) \) is a probability measure.

(ii) If \( S \) is the Borel field determined by a denumerable collection of its sets, there is a subsequence \( \{Q_{N_m}(\mathcal{E})\} \) of \( \{Q_N(\mathcal{E})\} \) converging to a limiting probability measure.

This lemma can be considered as a generalization of Helly's theorem.* Its proof will only be sketched.

**Proof of (i).** We need only prove that

\[
(4.5) \quad \mathcal{E} = \sum_{m=1}^{\infty} \mathcal{E}_m, \quad \mathcal{E}_m \cdot \mathcal{E}_n = 0, \quad (m \neq n),
\]

implies

\[
(4.6) \quad Q(\mathcal{E}) = \sum_{m=1}^{\infty} Q(\mathcal{E}_m),
\]

* Sitzungsberichte der Akademie der Wissenschaften, Vienna, class IIa, vol. 121 (1912), p. 286.
or (since \( Q(\mathcal{E}) \) is obviously additive) that

\[
Q(\mathcal{E}) - \sum_{1}^{\nu} Q(\mathcal{E}_m) = Q\left( \sum_{r+1}^{\infty} \mathcal{E}_m \right) = \lim_{N \to \infty} Q_N\left( \sum_{r+1}^{\infty} \mathcal{E}_m \right) \to 0, \quad \nu \to \infty.
\]

Now

\[
\sum_{1}^{\infty} \mathcal{E}_m \geq \sum_{2}^{\infty} \mathcal{E}_m \geq \cdots, \quad \prod_{N=1}^{\infty} \sum_{N}^{\infty} \mathcal{E}_m = 0,
\]

so that the hypotheses of the lemma imply (4.7).

**Proof of (ii).** Let \( \mathcal{E}_1', \mathcal{E}_2', \cdots \) be a denumerable collection of sets of \( S \), such that the Borel system of sets determined by the sequence \( \{ \mathcal{E}_m' \} \) is \( S \). By a familiar procedure, we can find a sequence of integers \( \{ N_m \} \) such that

\[
\lim_{n \to \infty} Q_{N_n}(\mathcal{E}) \text{ exists for every set } \mathcal{E} \text{ of } S \text{ which is in the field of sets determined by the sequence of sets } \{ \mathcal{E}_m' \}. \]

The hypotheses then imply that \( \lim_{n \to \infty} Q_{N_n}(\mathcal{E}) \) exists for every set \( \mathcal{E} \in S \),* and the remainder of part (ii) then follows from (i).

**Theorem 4.1.** Suppose that \( \mathcal{F}_x \) is the Borel field of sets determined by a denumerable subcollection of its sets, and that, for every pair of integers \( m, n \), with \( m \leq n \) and cylinder set \( \Lambda \) in \( \mathcal{F}_x \) over \( x_{n+1} \), a (conditional probability) function

\[
P(x_m, \cdots, x_n; \Lambda)
\]

is defined which has properties (i), (ii) given above, and for which

\[
P(x_m, \cdots, x_n; \Lambda) \equiv P(x_n; \Lambda).
\]

**Suppose that for each fixed value of \( n \), whenever \( \Lambda_1, \Lambda_2, \cdots \) is a sequence of cylinder sets of \( \mathcal{F}_x \) over \( x_{n+1} \) satisfying

\[
\Lambda_1 \supseteq \Lambda_2 \supseteq \cdots, \quad \prod_{1}^{\infty} \Lambda_r = 0,
\]

it is true that

\[
\lim_{r \to \infty} P(x_n; \Lambda_r) = 0
\]

uniformly in \( x_n \). Then

(i) there is a Markoff process with the given conditional probability functions, and

(ii) if \( x_n = x_{n+1} \) implies

\[
P(x_n; \Lambda) = P(x_{n+1}; TA),\]

* This can be proved by transfinite induction.

† The transformation \( T \) was defined at the end of §2.
there is a Markoff temporally homogeneous process with the given conditional probability functions.

Proof of (i). Let $Q(E)$ be any probability measure defined on the sets of $F_x$. Then if $M$ is any cylinder set of $F_\omega$ over $x_N, x_{N+1}, \ldots, x_n$ with characteristic function $\phi_M$, define $P_N(M)$ by

$$
P_N(M) = \int Q(de_N) \int P(x_N; de_{N+1}) \int \cdots \int \phi_M P(x_{n-1}; de_n).
$$

It was shown above that this determines $P_N(M)$ uniquely. In the following, if $M$ is any cylinder set over a finite number of coordinates $x_m, x_{m+1}, \ldots, x_n$, we define $P(x_{m-1}; M)$ by

$$
P(x_{m-1}; M) = \int P(x_{m-1}; de_m) \int P(x_m; de_{m+1}) \int \cdots \int \phi_M P(x_{n-1}; de_n),
$$

where $\phi_M$ is the characteristic function of $M$.

Now let $m, n$ be any two integers with $m \leq n$. The cylinder sets of $F_\omega$ over $x_m, x_{m+1}, \ldots, x_n$ constitute a Borel field $F_{m,n}$ determined by a denumerable sub-collection.* Suppose that $A_1, A_2, \ldots$ are sets in the field $F_{m,n}$, and that $0 = A_0 = A_1 = \cdots$, $\lim_{n \to \infty} A_n = 0$.

Then

$$
\lim_{n \to \infty} P(x_{m-1}; A_n) = 0,
$$

for all $x_{m-1}$. If $\epsilon > 0$, and if $M_\mu$ is the cylinder set over $x_{m-1}$ on which there is a value of $\mu \geq \nu$ such that $P(x_{m-1}; A_\mu) \geq \epsilon$, then it follows from (4.14) that

$$
M_1 \geq M_2 \geq \cdots, \quad \lim_{n \to \infty} M_n = 0.
$$

It follows from the definition of $M_\mu$ and from the fact that the conditional probability functions are less than or equal to 1, that

$$
P(x_{m-2}; A_\mu) = \int P(x_{m-1}; A_\mu) P(x_{m-2}; de_{m-1}) \leq \epsilon + P(x_{m-2}; M_\mu).
$$

The hypotheses of the theorem imply that

$$
\lim_{n \to \infty} P(x_{m-2}; M_\mu) = 0
$$

uniformly in $x_{m-2}$. If $\nu_0$ is chosen so large that

* Cf. §2.
for all \( x_{m-2} \), it follows from (4.15) that
\[
P(x_{m-2}; \Lambda_r) < 2\epsilon, \quad \nu > \nu_0
\]
for all \( x_{m-2} \). Then if \( N < m-2 \), and if \( \nu > \nu_0 \),
\[
P_N(\Lambda_r) = \int Q(d\epsilon_N) \int P(x_N; d\epsilon_{N+1}) \int \cdots \int P(x_{m-2}; \Lambda_r) P(x_{m-3}; d\epsilon_{m-2}) < 2\epsilon;
\]
so that if the field \( F_{m,n} \) is identified with the field \( S \) of Lemma 4.1, the hypotheses of the lemma are satisfied. There is therefore a subsequence \( \{P_N(\Lambda)\} \) of \( \{P_N(\Lambda)\} \) converging to a limiting probability measure (defined on the field \( F_{m,n} \)). Since \( m, n \) are arbitrary except that \( m \leq n \), there is a further subsequence \( \{P_{N_m}(\Lambda)\} \) such that
\[
\lim_{m \to \infty} P_{N_m}(\Lambda) = P(\Lambda)
\]
exists for every cylinder set \( \Lambda \) of \( F_{\omega} \) over a finite number of coordinates. Since \( P(\Lambda) \) satisfies the hypotheses of Theorem 1.1, its domain of definition can be extended to include all the sets of \( F_\omega \). If \( m, n \) are again any two integers with \( m \leq n \), and if \( \Lambda \) is a cylinder set of \( F_\omega \) over \( x_{n+1} \), it was shown in the general discussion preceding the statement of Theorem 4.1, that (if \( N \leq m \)),
\[
\int_M P(x_n; \Lambda) dP_N = \int_M P(x_N, \cdots, x_n; \Lambda) dP_N = P_N(\Lambda M),
\]
for every set \( M \) of \( F_{m,n} \); which expresses the fact that the conditional probabilities at the \( N \)th stage are the given ones. If \( N \) becomes negatively infinite only assuming values of the sequence \( \{N_m\} \), this becomes
\[
\int_M P(x_n; \Lambda) dP = P(\Lambda M),
\]
so that the conditional probability functions of the new \( P \)-measure are the given ones.

**Proof of (ii).** Suppose that (4.11) is satisfied, and let \( Q(E) \) be as in the proof of (i). Let \( \Lambda \) be a cylinder set of \( F_\omega \) over a finite number of the coordinates \( x_2, x_3, \cdots \). Consider the sequence of set functions \( \{Q_N(\Lambda)\} \) where
\[
Q_N(\Lambda) = \frac{1}{N} \sum_{1}^{N} \int P(x_1; T^m\Lambda)Q(d\epsilon_1).
\]
A slight modification of the argument just used shows that some subsequence
\{Q_{N_m}(\Lambda)\} of \{Q_N(\Lambda)\} converges for every such set \(\Lambda\). Moreover, if \(P(\Lambda)\) is the limit, then
\[
\begin{align*}
P(T\Lambda) &= \lim_{m \to \infty} \frac{1}{N_m} \sum_{1}^{N_m} \int P(x_1; T^{i+1}\Lambda)Q(\text{dei}) \\
&= \lim_{m \to \infty} \frac{1}{N_m} \sum_{1}^{N_m} \int P(x_1; T^{i}\Lambda)Q(\text{dei}) = P(\Lambda),
\end{align*}
\]
so that if \(\Lambda\) is any cylinder set of \(F_\omega\) over a finite number of coordinates, we can consistently define \(P(\Lambda)\) as \(P(P^{mA})\), where \(m\) is chosen so large that \(T^m\Lambda\) is a cylinder set over \(x_2, x_3, \ldots\). The set function thus defined satisfies the conditions of Theorem 1.1, hence it can be extended to become a probability measure defined on all the sets of \(F_\omega\). Evidently \(P(\Lambda) = P(T\Lambda)\) for every set of \(F_\omega\), and, as in the proof of part (i), the conditional probability functions are the given ones. We have proved incidentally the following corollary:

**Corollary.** In part (ii) of Theorem 4.1, if \(Q(E)\) is any probability measure defined on the field \(F_x\), there is an increasing sequence of positive integers \(N_1, N_2, \ldots\) such that
\[
\lim_{r \to \infty} \frac{1}{N_r} \sum_{1}^{N_r} \int P(x_1; T^r\Lambda)Q(\text{dei}) = P(\Lambda)
\]
exists for all cylinder sets \(\Lambda\) over a finite number of the coordinates \(x_2, x_3, \ldots\) and determines a possible choice of the probability measure \(P(\Lambda)\).

The proof becomes particularly simple if a value \(x_1^{(0)}\) of \(x_1\) is chosen, and if \(Q(E)\) is defined as 1 or 0 according as \(E\) does or does not contain \(x_1^{(0)}\). The integral then becomes \(P(x_1^{(0)}; T^m\Lambda)\).

5. **Examples.** The examples discussed in this section are simple illustrative examples, all of Markoff processes, which will be studied in detail in §7.

I. The type of stochastic process most frequently studied is that in which the chance variables form an independent set, that is, in which if \(E_1, \ldots, E_p\) are sets of \(F_x\) and \(\alpha_1, \ldots, \alpha_p\) are distinct integers, the \(P\)-measure of the \(\Omega\)-set determined by the \(m\) conditions \(x_{\alpha_j} \in E_{j}\) \((j = 1, \ldots, m)\), is the product of the \(P\)-measures of the \(m\) sets determined by the single conditions. This case is characterized by the fact that \(P(x_{\alpha_1}, \ldots, x_{\alpha_p}; \Lambda)\) does not depend on \(x_{\alpha_1}, \ldots, x_{\alpha_p}\) if \(\Lambda\) is a \(P\)-measurable cylinder set over coordinates not including \(x_{\alpha_1}, \ldots, x_{\alpha_p}\).

II. Let \(X\) contain \(n\) elements, the numbers 1, \ldots, \(n\). We shall define a

* The corresponding \(P\)-measures on \(\Omega\) have been examined by many writers, referred to in §1 and §2.
Markoff temporally homogeneous process. Let \((p_{jk})\) be an \(n^2\) matrix of elements which satisfy the conditions

\[
\begin{align*}
& p_{jk} \geq 0, \quad j, k = 1, \ldots, n, \\
& \sum_{k=1}^{n} p_{jk} = 1, \quad j = 1, \ldots, n.
\end{align*}
\]

The element \(p_{jk}\) is identified with the conditional probability that \(x_{v+1} = k\) if \(x_v = j, \; v = 0, \pm 1, \ldots\). The \(P\)-measure will be completely determined if the probability \(p_k\) that \(x_v = k\) (which is to be independent of \(v\)) is assigned. The hypotheses of Theorem 4.1 (ii) are satisfied, so the existence of the “absolute probabilities” \(p_1, \ldots, p_n\) is assured. These satisfy (cf. equation (3.1))

\[
\sum_{j=1}^{n} p_{i} p_{jk} = p_k, \quad \sum_{j=1}^{n} p_{i} = 1, \quad k = 1, \ldots, n.
\]

According to the corollary to Theorem 4.1, the absolute probabilities can be obtained in the form

\[
p_k = \lim_{v \to \infty} \frac{1}{N_v} \sum_{m=1}^{N_v} p_{i}^{(m)}, \quad (j \text{ fixed}),
\]

where the set \(N_1, N_2, \ldots\) is an increasing set of positive integers, and \(p_{i}^{(m)}\) is the conditional probability that \(x_{v+m} = k\) if \(x_v = j\). The element \(p_{i}^{(m)}\) is determined (cf. equation (3.12)) by

\[
(5.4) \quad p_{i}^{(m+1)} = \sum_{l=1}^{n} p_{i}^{(m)} p_{lk}.
\]

Evidently the matrix \((p_{i}^{(m)})\) is the \(m\)th power of the matrix \((p_{jk})\), and its elements satisfy (5.1).

III. Let \(X\) be arbitrary, but suppose that \(F_x\) is the Borel field of sets determined by some denumerable subcollection. A non-negative completely additive set function (not necessarily always finite-valued) is supposed defined on \(X\),† and the integral, with respect to this measure, of an \(X\)-measurable

* This classical Markoff process, the one originally studied by Markoff, is discussed by Hostinsky (II), who gives an extensive bibliography. References to more recent work will be given in §7. Fréchet has announced a new book on Markoff processes in which he will probably study this case in detail.

† It is supposed that there is a monotone increasing sequence of sets, each of finite \(X\)-measure, whose sum is \(X\), and that the \(X\)-measure of any \(X\)-measurable set is the limit of the \(X\)-measure of its intersection with the sets of the sequence. We shall suppose that this set function is extended as usual so that it is defined (and 0) on the subsets of sets of \(F_x\) on which it vanishes. The sets for which the extended set function is defined will be called \(X\)-measurable, and \(X\)-measurable functions are then defined in the usual way.
function \( f(x) \) over an \( X \)-measurable set \( E \) will be denoted by \( \int_E f(x) \, dx \).*

Let \( X \times Y \) be the product space of pairs \( (x, y), (x, y \in X) \). A measure can then be defined on \( X \times Y \) by the condition that if \( E, F \) are \( X \)-sets in \( F_x \), the \( X \times Y \)-measure of the set determined by \( x \in E, y \in F \) is the product of the \( X \)-measures of \( E \) and \( F \).† Let \( \rho(x, y) \) be a function defined on \( X \times Y \)-space which is measurable with respect to the measure just defined,‡ and which satisfies the following conditions:

(a) \( \rho(x, y) \) is non-negative;
(b) \( \rho(x, y) \) is integrable in \( y \) for fixed \( x \), and

\[
\int \rho(x, y) \, dy = 1. \quad \text{§}
\]

If \( \Lambda \) is a cylinder set over \( x_{r+1} \) determined by the condition \( x_{r+1} \in E_i (E_i \in F_x) \), we define \( P(x_{r}; \Lambda) \) by

\[
P(x_{r}; \Lambda) = \int_E \rho(x, y) \, dy, \quad \nu = 0, \pm 1, \ldots .
\]

By Theorem 4.1 (ii) these conditional probability functions are those of a temporally homogeneous Markov process if, whenever \( E_1, E_2, \ldots \) are sets in the field,

\[
E_1 \supseteq E_2 \supseteq \cdots , \quad \prod_{i=1}^{\infty} E_m = 0
\]

implies

\[
\lim_{m \to \infty} \int_{E_m} \rho(x, y) \, dy = 0
\]

uniformly in \( x \). This can be interpreted as the uniform (in \( x \)) integrability of \( \rho(x, y) \) with respect to \( y \), that is, the uniform (in \( x \)) absolute continuity (in \( E \))

---

* It will be supposed, as usual, that integrability means absolute integrability, and that a non-negative function is integrable if and only if its integral on the sequence of sets in the preceding note is bounded. In many applications, \( X \) is supposed to be a Borel set \( \mathcal{E} \) of \( n \)-dimensional euclidean space and \( F_x \) the field of Borel subsets of \( \mathcal{E} \); and the set function is supposed to be Borel measure.

† Saks, *Théorie de l'Intégrale*, Warsaw, 1933, pp. 257–263. As usual we suppose that the measure is further extended so that subsets of sets of measure 0 are measurable and of measure 0.

‡ We shall suppose further that \( \rho(x, y) \) is measurable with respect to the \( X \times Y \) measure, as defined before its extension described in the preceding note, so that \( \rho(x_n, x_i) \) is measurable with respect to \( F_{x_n} \), considered as a function defined on \( \Omega \). In any case \( \rho(x, y) \) will be equal to such a function almost everywhere on \( X \times Y \)-space. It then follows (cf. Saks, ibid., p. 258) that \( \rho(x, y) \) is \( X \)-measurable in \( x (y) \) for each fixed value of \( y (x) \).

§ As in the previous sections, when no region of integration is explicitly prescribed integration will be over the whole space.
of the set function $\int_B \rho(x, y) dy$. The condition will be satisfied if $\rho(x, y) \leq \phi(y)$, for all $x, y$, where $\phi(x)$ is $X$-measurable and integrable over $X$. If the $X$-measure of $X$ is finite, that is, if $\int dx < \infty$, the condition will be satisfied if $\rho(x, y)^2$ is integrable in $y$, and if there is a number $K$ such that for every value of $x$,

$$\int \rho(x, y)^2 dy \leq K.$$ 

If (5.7) implies (5.8), the measure function $P(\Lambda)$ given by Theorem 4.1 becomes, on the cylinder sets over $x_1$, a function $Q(E)$ of sets $E \in F_x$; and if $\Lambda$ is determined by the condition $x_1 \in E$, $Q(E) = P(\Lambda)$. Moreover (cf. equation (3.1)),

$$\int Q(dx) \int_B \rho(x, y) dy = Q(E).$$

If the $X$-measure of $E$ vanishes, (5.9) shows that $Q(E) = 0$, that is, $Q(E)$ is absolutely continuous. There is then an $X$-measurable function $\rho(x)$ for which

$$Q(E) = \int_B \rho(x) dx,$$

for all sets $E \in F_x$. This function $\rho(x)$ satisfies the equation

$$\int \rho(x) dx \int_B \rho(x, y) dy = \int_B \rho(y) dy,$$

so that, if the order of integration is interchanged ($\rho(x, y)$, $\rho(x)$ are non-negative),

$$\int_B dy \int_B \rho(x) \rho(x, y) dx = \int_B \rho(y) dy.$$

Then

$$\int \left\{ \int_B \rho(x) [\rho(x, y) - \rho(y)] dx \right\} dy = 0.$$

Since $E$ is arbitrary, (5.10) implies

$$\int \rho(x) \rho(x, y) dx = \rho(y)$$

for almost all $y$. The function $\rho(y)$ can now be changed on a set of $X$-measure 0 to make (5.11) true for all $y$. According to the corollary to Theorem 4.1, the "absolute probability density" $\rho(y)$ can be obtained in the form

(5.12) \[ \int_E p(y)dy = \lim_{r \to \infty} \frac{1}{N_r} \sum_{i=1}^{N_r} \int_E p^{(m)}(x, y)dy, \quad (x \text{ fixed}), \]

where \( N_1, N_2, \ldots \) is an increasing sequence of positive integers, and \( \int_E p^{(m)}(x, y)dy \) is the conditional probability that \( x_{r+m} \in E \) if \( x_r = x \). The function \( p^{(m)}(x, y) \) is determined (cf. equation (3.12)) by

(5.13) \[ p^{(1)}(x, y) = p(x, y), \quad p^{(m+1)}(x, y) = \int p^{(m)}(x, z)p(z, y)dz. \]

Evidently the function \( p^{(m)}(x, y) \) satisfies the conditions (a), (b) imposed on \( p(x, y) \). Moreover if the sequence of sets satisfies (5.7), and if (5.7) implies (5.8), then

\[ \int_{E_{x'}} p^{(m)}(x, y)dy = \int_{E_{x'}} dy \int p^{(m-1)}(x, z)p(z, y)dz \]

\[ = \int p^{(m-1)}(x, z)dz \int_{E_{x'}} p(z, y)dy \to 0, \]

uniformly in \( x \), so that \( p^{(m)}(x, y) \) satisfies the condition of uniform integrability if \( p(x, y) \) does. A slight modification of the proof shows that if, for some integer \( \mu \geq 1 \), \( p^{(\mu)}(x, y) \) satisfies the condition of uniform integrability, the function \( p^{(m)}(x, y) \) for \( m > \mu \) will also satisfy the condition; and then a suitable modification of the proof of Theorem 4.1 (ii) will show that it is sufficient to assure the existence of an absolute probability density (given, for example, by (5.12)) to suppose that for some \( m, p^{(m)}(x, y) \) satisfies the uniform integrability condition.

The Markoff process considered here is very general.* Example II is a special case. To show this we need only define \( X \)-measure suitably and define the function \( p(x, y) \) in terms of \( p_{ik} \). The space \( X \) has points denoted by the numbers 1, \( \cdots, n \). If \( E \) is any set of \( X \) containing \( r \) elements, define the \( X \)-measure of \( E \) as \( r \). The function \( p(x, y) \) is defined as \( p_{ik} \) for \( x = j, y = k \). More generally we can consider a space \( X \) whose points are the numbers 1, 2, \( \cdots \). The field \( F_x \) is to be the field of all subsets of \( X \), and the \( X \)-measure of a set is the number of points in it. A matrix \( (p_{ik}) \) is given whose elements satisfy

(5.14) \[ p_{ik} \geq 0, \quad \sum_{k=1}^{\infty} p_{ik} = 1. \]

* Hostinsky discusses this type in (I) and (II) and gives an extensive bibliography in (II). Cf. also the forthcoming book by Fréchet. Further references will be given in §7.
The condition of uniform integrability becomes here the condition of uniform convergence in (5.14):

\[
\lim_{N \to \infty} \sum_{k=1}^{N} p_{jk} = 1
\]

uniformly in \( j \). If the condition is satisfied, absolute probabilities \( p_1, p_2, \ldots \) exist satisfying the conditions

\[
\begin{align*}
p_j &\geq 0, \quad j \geq 1; \\
\sum_{j=1}^{\infty} p_j p_{jk} &= p_k, \quad k \geq 1; \\
\sum_{j=1}^{\infty} p_j &= 1.
\end{align*}
\]

Let \( q_0, q_1, \ldots \) be a sequence of non-negative numbers whose sum is 1. Suppose that \( p_{jk} = 0 \) if \( k < j \), and \( p_{jk} = q_{k-j} \) if \( k \geq j \). If \( q_0 < 1 \), it is readily seen that no process exists, temporally homogeneous or not, having the given conditional probabilities.* A particular case in which this is obvious is obtained by setting \( q_1 = 1 \).

IV. The following example is again that of a temporally homogeneous Markoff process. The space \( X \) is arbitrary, but we suppose that a probability measure is defined on the field \( F_x \), and that a transformation \( S_x \) is defined on \( X \) which is one-to-one, takes \( X \)-measurable sets into \( X \)-measurable sets, and is \( X \)-measure preserving. If \( \Lambda \) is a cylinder set of \( F_x \) over \( x_{r+1} \), determined by the condition \( x_{r+1} \in E \), define \( P(x_r; \Lambda) \) by

\[
P(x_r; \Lambda) = 1 \quad \text{if} \quad Sx_r \in E,
\]

\[
P(x_r; \Lambda) = 0 \quad \text{if} \quad Sx_r \notin E.
\]

The condition of Theorem 4.1 is not satisfied, but there is nevertheless a temporally homogeneous Markoff process with these conditional probability functions, for if \( M \) is an \( \Omega \)-set determined by the conditions \( x_{aj} \in E_j \), \( (j = 1, \ldots, p) \), the \( P \)-measure of \( M \) can be defined as the \( X \)-measure of the set \((S^{-a_1}E_1)(S^{-a_2}E_2) \cdots (S^{-a_p}E_p)\).

6. Temporally homogeneous processes. A stochastic process suggests the transformation idea in its very phraseology; for example, “the conditional probability that \( x_1 \) belong to a set \( E \) if \( x_0 = x_0^{(0)} \)”; and in §1 an explicit transformation \( T \) was defined to exploit this suggestion. In this section we shall consider only temporally homogeneous processes (for which \( T \) is measure-preserving). The theory of temporally homogeneous processes uses to a large extent the terminology of the theory of measure-preserving transformations, and in this section we shall see that this has a complete justification, in every

* This statement refers only to a process corresponding to a sequence of chance variables \( \ldots, x_{-1}, x_0, x_1, \ldots \). The results of the preceding section show that the statement is not true if processes corresponding to a sequence of chance variables \( x_1, x_2, \ldots \) are being considered.
detail, through the mediation of the transformation $T$. The present section will then essentially be an independent study of the measure-preserving transformation $T$, with particular stress on the case where the $P$-measure satisfies the conditions imposed on the $P$-measure corresponding to a Markoff process. We shall apply the theory of measure-preserving transformations, as developed by Birkhoff, Koopman, and von Neumann.

Suppose that a given process is temporally homogeneous. The ergodic theorem gives the following result.†

**Theorem 6.1.** Let the given process be temporally homogeneous, and let $\Lambda$ be any $P$-measurable set.

(i) If $\phi(\omega)$ is $P$-measurable and integrable, there is a $P$-measurable function $\phi^*(\omega)$ such that

$$
\lim_{N \to \infty} \frac{1}{N} \sum_{1}^{N} \phi(T^m \omega) = \phi^*(\omega)
$$

almost everywhere on $\Omega$. In particular there is a $P$-measurable function $Q(\omega; \Lambda)$ such that

$$
\lim_{N \to \infty} \frac{1}{N} \sum_{1}^{N} P(x_m; T^m \Lambda) = Q(\omega; \Lambda)
$$

almost everywhere on $\Omega$.

(ii) If the process is a Markoff process, and if $\Lambda$ is any $P$-measurable cylinder set over $x_v, x_{v+1}, \cdots$ for some integer $v$,

$$
\lim_{N \to \infty} \frac{1}{N} \sum_{1}^{N} P(x_0; T^m \Lambda) = P^*(x_0; \Lambda)
$$

exists almost everywhere on $\Omega$; that is, except possibly on an $x_0$-set of $P$-measure 0.

The fact that the limit exists in (6.2) is apparently new. Results closely related to (ii), with more restrictive hypotheses on the conditional probability functions,§ have been proved by Fréchet and (jointly) by Kryloff and Bogoliouboff.

† Conversely, as was seen in example IV, a measure-preserving transformation gives rise to a certain (Markoff) temporally homogeneous process, which is necessarily of a very special type.

‡ The form of the ergodic theorem used here (due to Birkhoff) is the following: If $T \omega$ is a measure-preserving transformation of an abstract space $\Omega$, then part (i) of the following theorem (Theorem 6.1) holds. A simple proof was given by Khintchine, Mathematische Annalen, vol. 107 (1933), pp. 485–488. The function $\phi(x, N)/N$ of Khintchine’s proof corresponds to the average in (6.1). For a complete treatment of the ergodic and related theorems see E. Hopf, *Ergodentheorie*, Ergebnisse der Mathematik, vol. 5, no. 2, which appeared so late that detailed reference to it could not be made in this paper.

§ The only restriction on the conditional probability functions made in Theorem 6.1 is that there should actually exist a corresponding temporally homogeneous process; that is, that there should exist “absolute probabilities.” Exact references will be given in §7.
Proof of (i). The first part of (i) is a restatement of the ergodic theorem. The second part of (i) is an application of the first part, with \( \phi(\omega) = P(x_0; \Lambda) \).

Proof of (ii). Suppose that \( \Lambda \) is as described in (ii). Then if \( m > -\nu \), it follows from equation (3.12) that
\[
\int P(x_m; T^m\Lambda) P(x_0; \text{de}) = \int P(x_m; T^m\Lambda) P(x_0; \text{de}_m) = P(x_0; T^m\Lambda),
\]
neglecting sets of \( P \)-measure 0, so that
\[
\int \left\{ \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} P(x_m; T^n\Lambda) \right\} P(x_0; \text{de}) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int P(x_m; T^n\Lambda) P(x_0; \text{de})
\]
\[
= \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} P(x_0; T^n\Lambda) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} P(x_0; T^n\Lambda),
\]
neglecting \( x_0 \)-sets of \( P \)-measure 0.†

Corollary 1. If the transformation \( T \) is metrically transitive,‡ then
\[
\phi^*(\omega) = \int \phi(\omega) dP, \quad Q(\omega; \Lambda) = P^*(x_0; \Lambda) = P(\Lambda)
\]
almost everywhere on \( \Omega \).

This corollary is merely a rephrasing, pertinent to the case being considered, of the ergodic theorem for metrically transitive systems.§

Corollary 2. If \( F_x \) is the Borel field determined by a denumerable collection of its sets, and if there are no angle variables,‖ then

† If the first two of the above expressions are considered as actual integrals, the admissibility of the transition from the first to the second follows from Lebesgue's theorem on the admissibility of term by term integration of a uniformly bounded convergent sequence of measurable functions. However, even if the integrals are considered merely as symbols for conditional expectations, the proof of Lebesgue's theorem can be extended to this case.

‡ Metric transitivity means here that no \( P \)-measurable set of measure not 0 or 1 is invariant under \( T \). If there is a \( P \)-measurable set \( \Lambda \) of measure not 0 or 1, which is invariant neglecting a set of \( P \)-measure 0, that is, if \( T\Lambda = \Lambda + \Delta_0 - \Delta_0' \), where \( P(\Delta_0) = P(\Delta_0') = 0 \), then the \( P \)-measurable set \( \sum \sigma^\infty T^n\Lambda \) has measure \( P(\Lambda) \neq 0, 1 \) and is invariant under \( T \), so there cannot be metric transitivity. Hence the content of the definition is not changed if invariance up to a set of \( P \)-measure 0 is substituted for actual invariance.

§ Cf. Khintchine, ibid., p. 488.

‖ An angle variable is a complex-valued \( P \)-measurable function \( \phi(\omega) \) such that \( |\phi| > 0 \) on an \( \Omega \)-set of positive \( P \)-measure, and that
\[
\phi(T\omega) = c\phi(\omega), \quad (|c| = 1, c \neq 1),
\]
almost everywhere on \( \Omega \). If the transformation is metrically transitive, the invariance of \( |\phi(\omega)| \) under \( T \) implies that \( |\phi(\omega) = \text{const. almost everywhere on } \Omega \). (Cf. B. O. Koopman, Proceedings of the National Academy of Sciences, vol. 17 (1935), pp. 315–318.)
exists, when \( m \) is restricted to a certain increasing set of integers of measure 1,* independent of the sets \( \Lambda, M \) which can be any \( P \)-measurable sets. If there is also metric transitivity, the limit in (6.4) is \( P(\Lambda) \cdot P(M) \).

Conversely, if the limit in (6.4) exists, for all \( P \)-measurable sets \( \Lambda, M \) on some set of integers of measure 1, there are no angle variables; and if the limit is \( P(\Lambda) \cdot P(M) \), there is metric transitivity.

This theorem was proved by Koopman and von Neumann in the metrically transitive case, for a one-parameter family of transformations \( \{ T_t \} \), \(-\infty < t < \infty \).† Their proof is applicable, with insignificant modifications, to the family, considered here, of transformations \( T_n = T^n \).

**Lemma 6.1.** Let \( f(\omega) \) be any complex-valued \( P \)-measurable function, and let \( m_1, m_2, \ldots \) be an increasing sequence of positive integers. Suppose that \( \{ f(T^{n\omega}) \} \) and \( \{ f(T^{-m\omega}) \} \) are sequences of functions convergent almost everywhere on \( \Omega \). Then if \( O \) is any open set of the complex plane, the \( \Omega \)-sets defined by the conditions

\[
(6.5) \quad \lim_{j \to \infty} f(T^{j\omega}) \in O, \quad \lim_{j \to -\infty} f(T^{-j\omega}) \in O
\]

are respectively cylinder sets over \( x_{-1}, x_{-2}, \ldots \) and \( x_1, x_2, \ldots \) (if we neglect sets of \( P \)-measure 0).

To any positive integer \( \nu \) corresponds (cf. §2) a \( P \)-measurable function \( f_\nu(\omega) \) depending on only a finite number of coordinates, and having the property that

\[
(6.6) \quad | f(\omega) - f_\nu(\omega) | < 1/\nu
\]

except perhaps on an \( \Omega \)-set of \( P \)-measure at most \( 2^{-\nu} \). There is a subsequence

* A set of integers \( a_1, a_2, \ldots \), (\( a_1 < a_2 < \cdots \)), is said to have measure 1 if

\[
\lim_{m \to \infty} \frac{1}{m} \sum_{a_i \leq m} 1 = 1.
\]

† Proceedings of the National Academy of Sciences, vol. 17 (1935), pp. 315–318. To extend their proof to the non-metrically transitive case, it is only necessary to allow a wider interpretation of their projection operator \( E_0 \). The hypotheses of topological character they impose on their space are unnecessary in this application.

‡ We use the fact that the \( P \)-measurable complex-valued function whose absolute values squared are integrable, form a unitary space \( H \) when distance and inner product are defined in the usual way (cf. Stone, Linear Transformations in Hilbert Space, American Mathematical Society Colloquium Publications, vol. 15, pp. 23–29). The hypothesis that \( F_\nu \) is the Borel field determined by a denumerable collection of its sets means that \( H \) is separable and thus is either finite-dimensional or a Hilbert space. The theorem (and proof) of Koopman and von Neumann is valid in the finite case also.
\{\mu_j\} of \{m_i\} such that for each positive integer \(j\), \(f_j(T^{t_i}\omega)\) (\(f_j(T^{-t_i}\omega)\)) depends only on \(x_{-1}, x_{-2}, \ldots, (x_1, x_2, \ldots)\). Since \(T\) is measure-preserving,

\[
\frac{\left| f(T^{t_i}\omega) - f_j(T^{t_i}\omega) \right|}{1/j} < 1/j, \\
\frac{\left| f(T^{-t_i}\omega) - f_j(T^{-t_i}\omega) \right|}{1/j} < 1/j,
\]

\(j\) fixed, except perhaps on an \(\Omega\)-set of \(P\)-measure at most \(2^{-i}\). Then

\[
\left| f(T^{u\omega}) - f(T^{u\omega}) \right| < 1/N, \quad v = N, N + 1, \ldots \\
\left| f(T^{-u\omega}) - f(T^{-u\omega}) \right| < 1/N, \quad v = N, N + 1, \ldots
\]

except perhaps on a set of \(P\)-measure at most \(2^{-N+1}\); so that

\[
\lim_{v \to \infty} f(T^{t_i\omega}) = \lim_{v \to \infty} f(T^{-t_i\omega})
\]

almost everywhere on \(\Omega\). Then the sets defined by (6.5) are the same as the sets defined by the conditions

\[
\lim_{v \to \infty} f(T^{t_i\omega}) \in O \quad \lim_{v \to \infty} f(T^{-t_i\omega}) \in O
\]

respectively, if we neglect sets of \(P\)-measure 0. This fact implies the truth of the lemma.

**Lemma 6.2.** The equality

\[
P(\ldots, x_{n-1}, x_n; \Lambda) = P(x_n; \Lambda)
\]

holds almost everywhere on the space \(\Omega\) of a Markoff process, for any \(P\)-measurable cylinder set \(\Lambda\) over \(x_{n+1}, x_{n+2}, \ldots\).

We need only show that, if \(M\) is a \(P\)-measurable cylinder set over \(\ldots, x_{n-1}, x_n\), then

\[
\int_M P(x_n; \Lambda) dP = P'(\Lambda M).
\]

It is evidently sufficient to prove (6.10) for sets \(M\) which are cylinder sets over a finite number of coordinates. If \(M\) is such a cylinder set, over \(x_m, x_{m+1}, \ldots, x_n\), (6.10) follows from the fact that \(P(x_m, \ldots, x_n; \Lambda) = P(x_n; \Lambda)\) almost everywhere on \(\Omega\).

*The conditional probability function \(P(\ldots, x_{n-1}, x_n; \Lambda)\) is defined in the same way as the function \(P(x_m, \ldots, x_n; \Lambda)\). Note that the finiteness of the set \(\alpha_1, \ldots, \alpha_p\) was not used in the definition of the latter function.*
Theorem 6.2. A temporally homogeneous Markoff process is metrically transitive* if and only if there is no set $E$ of the field $F_x$ such that if $\Lambda$ is the $\Omega$-set determined by the condition $x_1 \in E$, ($0 < P(\Lambda) < 1$), and

(i) $\Lambda$ is invariant under $T$, if we neglect a set of $P$ measure 0, or
(ii) if we neglect $x_0$-sets of $P$-measure 0,

then

$$
P(x_0; \Lambda) = 1, \quad x_0 \in E
$$
$$
P(x_0; \Lambda) = 0, \quad x_0 \not\in E.
$$

Proof of (i). If there is an invariant set $\Lambda$ of the type described in the theorem, the process cannot be metrically transitive, by the definition of metric transitivity. Conversely, if the process is not metrically transitive, there is a $P$-measurable set $M$ invariant under $T$, and $0 < P(M) < 1$. If $f(\omega)$ is the characteristic function of $M$, $f(\omega) = f(T\omega)$ on $\Omega$ so that

$$
\lim_{m \to \infty} f(T^m\omega) = \lim_{m \to -\infty} f(T^{-m}\omega) = f(\omega)
$$
on $\Omega$. Then by Lemma 6.1, $M$ can be considered either as a cylinder set over $x_{-1}, x_{-2}, \cdots$ (when we denote it by $M_1$), or as a cylinder set over $x_1, x_2, \cdots$ (when we denote it by $M_2$), neglecting sets of $P$-measure 0. It follows from

$$
0 \leq P(\cdots, x_{-1}, x_0; M_2) \leq 1,
$$

$$
\int_{M_1} P(\cdots, x_{-1}, x_0; M_2) dP = P(M_1M_2) = P(M_1),
$$

$$
\int_{C M_1} P(\cdots, x_{-1}, x_0; M_2) dP = P(CM_1M_2) = 0
$$

that

$$
P(\cdots, x_{-1}, x_0; M_2) = 1, \quad \omega \in M_1,
$$

$$
P(\cdots, x_{-1}, x_0; M_2) = 0, \quad \omega \not\in M_1,
$$

if we neglect sets of $P$-measure 0. Now, according to Lemma 6.2, $P(\cdots, x_{-1}, x_0; M_2) = P(x_0; M_2)$ almost everywhere on $\Omega$. Then if a set of $P$-measure 0 is neglected, $M$ must be a cylinder set over $x_0$; this set is determined by the condition $x_0 \in E$, where $E$ is the $x_0$-set on which $P(x_0; M_2) = 1$. Since we can suppose (cf. §2), altering $P(x_0; M_2)$ on an $x_0$-set of $P$-measure 0 if necessary, that $P(x_0; M_2)$ is measurable with respect to $F_\omega$, we can suppose that $E$ is in $F_\omega$. The $\Omega$-set determined by the condition $x_0 \in E$ is invariant (up

* If the transformation $T$ is metrically transitive, or has angle variables, the same will be said to be true of the corresponding stochastic process.
to a set of $P$-measure 0); so that it is the same as any set determined by a condition $x_n \in E$, up to a set of $P$-measure 0.

**Proof of (ii).** If there is a set $E$, as described in the theorem, for which the hypotheses of (ii) are true, then

$$P(T^{-1}A; \Lambda) = \int_{T^{-1}A} P(x_0; \Lambda) dP = P(\Lambda),$$

so that $T^{-1}A$, and therefore $A$, is invariant under $T$ up to a set of $P$-measure 0; and the process cannot be metrically transitive. Conversely, if the process is not metrically transitive, an invariant set $A$ of the type described in part (i) exists, and (6.14) becomes precisely (6.11).

**Theorem 6.3.** Suppose that $\phi(\omega)$ is an angle variable of a temporally homogeneous Markoff process, so that

(6.15) \[ \phi(T\omega) = c\phi(\omega), \quad |c| = 1, \quad c \neq 1, \]

almost everywhere on $\Omega$.

(i) The function $\phi(\omega)$ can be considered as a function of $x_0$ alone, namely, $\phi(\omega) = \psi(x_0)$, so that (6.15) becomes

(6.15') \[ \psi(x_1) = c\psi(x_0), \]

and the possible exceptional set is an $(x_0, x_1)$-set of $P$-measure 0.

(ii) If the hypotheses of Theorem 3.2 are satisfied, and if the conditional probability functions are supposed defined as described in the statement of that theorem, then for each value of $x_0$,

(6.16) \[ \psi(x_1) = \text{const.} = c\psi(x_0) \]

on a cylinder set $\Lambda(x_0)$ over $x_1$ such that $P(x_0; \Lambda(x_0)) = 1$ except possibly on an $x_0$-set of $P$-measure 0.

(iii) If $\psi(x_0)$ takes on any non-zero value on a set of positive $P$-measure, $c$ is a root of unity.

(iv) There exist $P$-measurable cylinder sets $\Lambda_0, \Lambda_1$ over $x_0, x_1$ respectively, determined by the conditions $x_0 \in E_0, x_1 \in E_1$, such that $0 < P(\Lambda_i) < 1$, and if we neglect $x$-sets of $P$-measure 0,

(6.17) \[ P(x_0; \Lambda_1) = 1, \quad x_0 \in E_0, \]
\[ P(x_0; \Lambda_1) = 0, \quad x_0 \notin E_0. \]

(v) The function $\psi(x_0)$, if it is integrable, satisfies the integral equation

(6.18) \[ \int \psi(x_1) P(x_0; dx_1) = c\psi(x_0), \]
except possibly for an \( x_0 \)-set of \( P \)-measure 0.*

**Proof of (i).** Suppose that (6.15) is satisfied. There is an increasing sequence of positive integers \( n_1, n_2, \ldots \) such that

\[
\lim_{r \to \infty} c^{n_r} = \lim_{r \to \infty} c^{-n_r} = 1.
\]

Then

\[
(6.19) \quad \lim_{r \to \infty} \phi(T^{n_r} \omega) = \lim_{r \to \infty} \phi(T^{-n_r} \omega) = \phi(\omega),
\]

almost everywhere on \( \Omega \). If \( \Lambda(0) \) is the \( \Omega \)-set determined by the condition \( \phi(\omega) \in O \) (\( O \) an open set of the complex \( \phi \)-plane), the method used in the proof of the preceding theorem shows that \( \Lambda(0) \) is a cylinder set over \( x_0 \), neglecting an \( \Omega \)-set of \( P \)-measure 0. It follows readily from this that there is a \( P \)-measurable function \( \psi(x_0) \), depending only on \( x_0 \) and such that \( \phi(\omega) = \psi(x_0) \) almost everywhere on \( \Omega \).

**Proof of (ii).** Let \( \Lambda(x_0) \) be the cylinder set over \( x_1 \), determined by the condition \( x_1 \in E(x_0) \), on which \( \psi(x_1) = c\psi(x_0) \). The \( (x_0, x_1) \)-set \( M \), determined by the condition that \( x_1 \in E(x_0) \) for each value of \( x_0 \), is of \( P \)-measure 1, and its measure can be expressed, according to Theorem 3.4, as

\[
\int P(d\omega_0) \int f_M P(x_0; d\omega_1) = \int P(x_0; \Lambda(x_0)) P(d\omega_0) = 1,
\]

where \( f_M \) is the characteristic function of \( M \). Then \( P(x_0; \Lambda(x_0)) = 1 \) except possibly on an \( x_0 \)-set of \( P \)-measure 0.

**Proof of (iii).** If \( \psi(x_0) \) takes on a value \( \psi_0 \neq 0 \) on a set \( \Lambda_0 \) of positive \( P \)-measure, \( \psi(x_0) \) must take on \( c^n\psi_0 \) on a set \( \Lambda_n \) of the same \( P \)-measure (since \( \psi(x_n) = c^n\psi(x_0) \)). The number \( c \) must then be a root of unity; for if not, the numbers \( \psi_0, c\psi_0, \ldots \) are all distinct, so that the sets \( \Lambda_0, \Lambda_1, \ldots \) are all disjoint. But this is impossible since then

\[
1 \geq P \left( \sum_{0}^{n-1} \Lambda_m \right) = nP(\Lambda_0), \quad n = 1, 2, \ldots .
\]

Evidently the fact that the process is a Markoff process was not needed in this proof of (iv).

**Proof of (iv).** If \( O \) is an open set of the complex \( \psi \)-plane, so chosen that the \( P \)-measure of the \( \Omega \)-set \( \Lambda_0 \), determined by the condition \( \psi(x_0) \in O \), is posi-

* The left side of (6.18) is the conditional expectation \( E(x_0; \psi) \), and the conditions under which it can be considered an integral were considered in §3. The integrability condition imposed on \( \psi \) is unimportant, since if \( \psi \) is an angle variable, the integrable function \( \psi_K \), equal to \( \psi \) if \( |\psi| \leq K \) and otherwise equal to \( K \), satisfies (6.15') almost everywhere, so that \( \psi_K \) is also an angle variable if \( K \) is chosen so large that \( |\psi_K| > 0 \) on a set of positive \( P \)-measure.
tive and less than 1; and if \( \Lambda_1 \) is the \( \Omega \)-set determined by the condition \( c^{-1}\phi(x_1) \in O \), then \( \Lambda_0 = \Lambda_1 \) (if we neglect sets of \( P \)-measure 0). Equation (6.17) then follows at once. (Cf. the proof of Theorem 6.2 (ii).)

**Proof of (v).** If \( \phi(x_1) = \alpha \phi(x_0) \) almost everywhere on \( \Omega \), and if \( \phi \) is integrable, then

\[
\int \phi(x_1)P(x_0; \, d\xi) = c \int \phi(x_0)P(x_0; \, d\xi) = \alpha \phi(x_0),
\]

except possibly on an \( x_0 \)-set of \( P \)-measure 0, as was to be proved.

**Theorem 6.4.** A temporally homogeneous process for which the corresponding sequence of chance variables \( \cdots, x_{-1}, x_0, x_1, \cdots \) form an independent set (cf. §5, example 1) is metrically transitive and has no angle variables.*

A process of this type is a very special case of a Markoff process, so Theorems 6.2 and 6.3 are applicable. A set \( \Lambda \), as described in the statement of Theorem 6.2, is impossible, since in the case of independence \( P(x_0; \Lambda) = P(\Lambda) \); and the process is therefore metrically transitive. For the same reason, there can be no sets \( \Lambda_0, \Lambda_1 \), as described in Theorem 6.2 (ii); and the process therefore has no angle variables.

**Theorem 6.5.** Suppose the measure relations of a temporally homogeneous Markoff process have the following property: There is a function \( \phi(x_0; \xi) \), measurable with respect to \( F_\xi \) and integrable over \( \Omega \) in \( \xi_1 \) for fixed \( x_0 \), such that (except possibly for an \( x_0 \)-set of \( P \)-measure 0),

\[
(6.20) \quad P(x_0; \Lambda) = \int_{\Lambda} \phi(x_0; \xi)P(\, d\xi)
\]

whenever \( \Lambda \) is a cylinder set of \( F_\xi \) over \( x_1 \).

(i) The process has no angle variables if and only if \( \lim_{m \to \infty} P(x_0; T^m \Lambda) \) exists (except possibly on an \( x_0 \)-set of \( P \)-measure 0) for every such set \( \Lambda \).

(ii) The process has no angle variables and is metrically transitive if and only if \( \lim_{m \to \infty} P(x_0; T^m \Lambda) = P(\Lambda) \) (except possibly on an \( x_0 \)-set of \( P \)-measure 0) for every such set \( \Lambda \).

(iii) If it is true that \( P(x_0; \Lambda) \) (for \( \Lambda \) a cylinder set of \( F_\xi \) over \( x_1 \)) can be defined to be a probability measure, for each fixed value of \( x_0 \), which vanishes identically in \( x_0 \) for a given set \( \Lambda \) if it vanishes at all, then a function \( \phi \) exists and satisfies the hypotheses of the theorem.

Before proving the theorem, we shall give an example of a temporally homogeneous stochastic process which has no angle variables, and for which

---

* This result was proved by Doob (I, pp. 761–763), and Hopf (I, p. 95).
the limit described in (i) does not exist. This example will therefore show that some condition, such as the existence of the function $\phi$ as described, is necessary in the theorem. The example is a particular case of example IV of §5. In example IV suppose that the transformation $S$ is metrically transitive and has no angle variable. Then it is readily seen that the transformation $T$ is metrically transitive and has no angle variable. On the other hand, if $A$ is a cylinder set of $F_\omega$ over $x_1$, determined by the condition $x_1 \in E$, then $P(x_0; T^mA)$ is 1 or 0 according as $S^m x_0$ is or is not in $E$. Then $\lim_{m \to \infty} P(x_0; T^mA)$ exists almost everywhere on $\Omega$ only if $S^m x_0$, for large $m$, is finally always in $E$ or never in $E$ for almost all $x_0$. This implies that $E$ is invariant under $S$ (up to a set of $X$-measure 0), which is impossible, since $S$ is metrically transitive, if we choose $A$ so that $P(A) \neq 0, 1$.

Proof of (i). Suppose first that there are no angle variables and that $F_x$ is the Borel field determined by a denumerable collection of its sets. According to Theorem 6.1, Corollary 2, there is an increasing sequence of integers, independent of $A, M$, such that

$$\lim_{r \to \infty} P(M T^r A) = \lim_{r \to \infty} \int_M P(x_1; T^r A) dP = Q(M; \Lambda)$$

exists, where $A, M$ are $P$-measurable cylinder sets over $x_1$. From this it follows readily that if $f(x_1)$ is $P$-measurable and integrable over $\Omega$, then

$$\lim_{r \to \infty} \int P(x_1; T^r A) f(x_1) dP = \int f(x_1) Q(dx_1; A).$$

In particular (cf. equation (3.12))

$$\lim_{r \to \infty} \int P(x_1; T^r A) \phi(x_0, x_1) P(dx_1) = \lim_{r \to \infty} \int P(x_1; T^r A) P(x_0; dx_1) = \lim_{r \to \infty} P(x_0; T^r A)$$

exists. We shall denote this limit by $Q(x_0; \Lambda)$. Evidently

$$\int_K Q(x_0; \Lambda) dP = Q(K; \Lambda)$$

if $K$ is a $P$-measurable cylinder set over the coordinates $x_m, x_{m+1}, \ldots, x_0$ ($m \leq 0$). If $\epsilon > 0$, there is an integer $N = N(\epsilon)$ so large that if $a_j > N$,

$$|P(x_0; T^{a_j} \Lambda) - Q(x_0; \Lambda)| \leq \epsilon/6$$

*The set function $Q(M; \Lambda)$ is obviously additive in $M$ for fixed $\Lambda$. Since $Q(M; \Lambda) \leq P(M)$, $Q(M; \Lambda)$ for fixed $\Lambda$ is a completely additive function of sets $M$; hence integration with respect to the differential element $Q(dx_1; \Lambda)$ has a meaning.
except possibly on a set $\Lambda_\epsilon$ such that $P(\Lambda_\epsilon) \leq \epsilon/6$. If $M$ is a $P$-measurable cylinder set over $x_1$, and if $a_i < \nu$, then

\[(6.25) \quad P(MT^*\Lambda) = P(T^{a_i^*}\Lambda T^{a_i}\Lambda) = \int_{a_i^* - r_M} P(x_0; T^{a_i}\Lambda) dP, \]

so that, if use is made of (6.24) and (6.25),

\[ |P(MT^*\Lambda) - Q(T^{a_i^*}M; \Lambda)| = \left| \int_{a_i^* - r_M} [P(x_0; T^{a_i}\Lambda) - Q(x_0; \Lambda)] dP \right| \leq \epsilon/6 + 2P(\Lambda_\epsilon) \leq \epsilon/2, \]

if $\nu > a_i^* > N$. Then since

\[Q(T^k M; \Lambda) = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^N P(T^k M T^i \Lambda) = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^N P(M T^{i-k} \Lambda)\]

is independent of $k$,

\[(6.26) \quad \lim_{j \to \infty} P(MT^*\Lambda) = Q(M; \Lambda), \]

so that (6.21) holds with $a_* = \nu$. The proof that (6.21) implies the existence of the limit in (6.23) can now be used to show that the existence of $\lim_{v \to \infty} P(MT^\Lambda)$ implies the existence of $\lim_{v \to \infty} P(x_0; T^\Lambda)$. The hypothesis that $F_x$ is the Borel field determined by a denumerable collection of its sets can now be removed; since if this is not true, we can preassign the set $\Lambda$, and then replace the field $F_x$ by a smaller field $F_x'$ for which the denumerability hypothesis is true, such that (cf. §2) the set $\Lambda$ is in the corresponding field $F_x'$, and such that $\phi(x_0, x_1)$ is measurable with respect to $F_x'$.

Conversely suppose that $\lim_{v \to \infty} P(x_0; T^\Lambda)$ exists for every set $\Lambda$, as described in the theorem. Then if $M$ is a $P$-measurable cylinder set over $x_1$,

\[(6.27) \quad \lim_{v \to \infty} P(MT^\Lambda) = \lim_{v \to \infty} \int_M P(x_1; T^\Lambda) dP \]

exists. Thus

\[(6.28) \quad \lim_{v \to \infty} \int f(x_1) g(x_0) dP^* \]

exists, if $f(x_1), g(x_1)$ are characteristic functions of cylinder sets of $F_\omega$ over $x_1$. The limit can then be shown to exist (using a familiar method of approximation) if $f(x_1), g(x_1)$ are any bounded complex-valued $P$-measurable functions

---

* If $\xi$ is a complex number, the notation $\bar{\xi}$ will be used, as is customary, to denote its conjugate complex number.
depending only on $x_1$. Now if there is an angle variable, there is, as was seen above, a bounded angle variable. If $\psi(x_1)$ is a bounded angle variable, we set $f=g=\psi$ in (6.28) and find that the limit

$$\lim_{r \to \infty} \int \psi(x_1)\overline{\psi(x_r)}dP = \lim_{r \to \infty} \int |\psi(x_1)|^2dP,$$

must exist. This is absurd; hence there can be no angle variable.

Proof of (ii). If the process has no angle variables, and if it is also metrically transitive, $\lim_{r \to \infty} P(x_0; T^r\Lambda)$, which we know exists for $\Lambda$ a $P$-measurable cylinder set over $x_1$ by part (i), must be $P(\Lambda)$ since (Theorem 6.1, Corollary 1)

$$\lim_{N \to \infty} \frac{1}{N} \sum_{1}^{N} P(x_0; T^m\Lambda) = P(\Lambda),$$

if we neglect $x_0$-sets of $P$-measure 0 throughout. Conversely if, whenever $\Lambda$ is a $P$-measurable cylinder set over $x_1$, $\lim_{r \to \infty} P(x_0; T^r\Lambda) = P(\Lambda)$ except possibly on an $x_0$-set of $P$-measure 0, the process can have no angle variables, according to (i). If the process is not metrically transitive, there is a $P$-measurable cylinder set $\Lambda$ over $x_1$, ($0 < P(\Lambda) < 1$), which is invariant under $T$ (if we neglect a set of $P$-measure 0). Then

$$P(x_0; \Lambda) = P(x_0; T^r\Lambda) \to P(\Lambda), \quad r \to \infty,$$

that is, $P(x_0; \Lambda) = P(\Lambda)$ except possibly on an $x_0$-set of $P$-measure 0. This is incompatible with (6.11). The process therefore has no angle variables and is metrically transitive, as was to be proved.

Proof of (iii). We shall use the hypotheses of (iii) only to derive the fact that if $\Lambda$ is a cylinder set of $F_0$ over $x_1$, and if $P(x_0; \Lambda) = 0$ for some value of $x_0$, then $P(\Lambda) = 0$. This fact is obvious from the equation

$$P(\Lambda) = \int P(x_0; \Lambda)P(dx_0).$$

Now consider the field of cylinder sets of $F_0$ over $x_0, x_1$. One probability measure is already defined on this field, namely $P$-measure. We define a second probability measure $\tilde{P}(M)$ for $M$ in this field by

$$(6.29) \quad \tilde{P}(M) = \int P(dx_0) \int f_M(x_0, x_1)P(dx_1),$$

where $f_M(x_0, x_1)$ is the characteristic function of $M$. According to Theorem 3.4,* this new measure is essentially a measure in two-dimensional ($x_0, x_1$)-space, obtained in the usual (multiplicative) way (cf. Saks, Théorie de l'Intégrale, Warsaw, 1933, pp. 257–263) from a given measure ($P$-measure) on the $x_0$-axis and a given measure ($P$-measure) on the $x_1$-axis.

* This new measure is essentially a measure in two-dimensional ($x_0, x_1$)-space, obtained in the usual (multiplicative) way (cf. Saks, Théorie de l'Intégrale, Warsaw, 1933, pp. 257–263) from a given measure ($P$-measure) on the $x_0$-axis and a given measure ($P$-measure) on the $x_1$-axis.
\[(6.30) \quad P(M) = \int P(\mu_0) \int f_M(x_0, x_1)P(x_0; \mu_1),\]

and the integration need not be taken symbolically. Let \(M(x_0)\) be the cylinder set over \(x_1\) defined by the equation \(f_M(x_0, x_1) = 1\). If we integrate in \(6.30\), then if \(P(M) = 0\), \(P(x_0; M(x_0)) = 0\), except possibly on an \(x_0\)-set of \(P\)-measure 0. It has already been shown that for each value of \(x_0 = \xi\) such that \(P(\xi; M(\xi)) = 0\), \(P(M(\xi)) = 0\). Then if \(P(M) = 0\),

\[
\tilde{P}(M) = \int P(M(x_0))P(\mu_0) = 0;
\]

hence the set function \(\tilde{P}(M)\) is absolutely continuous with respect to \(P(M)\). There is therefore* a function \(\phi(x_0, x_1)\), measurable with respect to \(F_\omega\), such that if \(M\) is a cylinder set of \(F_\omega\) over \(x_0, x_1\),

\[(6.31) \quad \tilde{P}(M) = \int M \phi(x_0, x_1)dP.\]

In particular if \(M\) is the intersection of \(\Lambda_0\) (a cylinder set of \(F_\omega\) determined by the condition \(x_0 \in E_0\)) and \(\Lambda\) (a cylinder set of \(F_\omega\) determined by the condition \(x_1 \in E_1\)), \(6.31\) becomes

\[
P(\Lambda_0)P(\Lambda) = \int_{E_0}P(\mu_0)\int_{E_1}P(x_0; \mu_1),
\]

so that

\[
\int_{E_0}P(\mu_0)\left\{P(\Lambda) - \int_{E_1}\phi(x_0, x_1)P(x_0; \mu_1)\right\} = 0.
\]

This equation is to hold for every set \(E_0\) in the field \(F_z\), so that the quantity in the brace must vanish, except possibly on an \(x_0\)-set of \(P\)-measure 0, as was to be proved.

**Theorem 6.6.** If the conditional probability functions of a temporally homogeneous stochastic process satisfy the conditions

\[(6.32) \quad P(x_0; \Lambda) \geq \lambda P(\Lambda)\]

\[
P(x_{-\nu}, \cdots, x_0; \Lambda) \geq \lambda \nu P(x_{-\nu+1}, \cdots, x_0; \Lambda), \quad \nu = 1, 2, \cdots,
\]

for every \(P\)-measurable cylinder set \(\Lambda\) over \(x_1\),† where \(0 < \lambda, \leq 1\), and if

---

* Saks, ibid., p. 257.
† The inequalities are to hold with probability 1 for each set \(\Lambda\).
(6.33) \[ \prod_{m=1}^{n} \lambda_m^m = \lambda > 0, \]

then the process is metrically transitive and has no angle variables.

If the process is a Markoff process, we can take \( \nu = 1 \), whenever \( \nu > 0 \), leaving only the first inequality of (6.32) as an actual condition. Theorem 6.2 gave a much more sensitive condition.

Let \( \Lambda_2 \) be a \( P \)-measurable cylinder set over \( x_1, \ldots, x_n, n \geq 1 \). Then if \( f \) is the characteristic function of \( \Lambda_2 \), and if \( m \geq 1 \) (cf. equation (3.10)), then

\[
P(x_m, \ldots, x_0; \Lambda_2) = \int P(x_m, \ldots, x_0; \Lambda_1) \int \cdots \int fP(x_m, \ldots, x_{m-1}; \Lambda_1) \int \cdots \int \frac{1}{\lambda_m} \cdots \cdots \cdots \int \frac{1}{\lambda_n} P(x_m, \ldots, x_{m-1}; \Lambda_1) \int \cdots \int \frac{1}{\lambda_n} P(x_m, \ldots, x_{m-1}; \Lambda_1)
\]

(6.34)

\[
\leq 1 - \int P(x_m, \ldots, x_0; \Lambda_1) \int \cdots \int (1 - f)P(x_m, \ldots, x_{m-1}; \Lambda_1)
\]

If \( \Lambda_1 \) is a \( P \)-measurable cylinder set over \( x_m, \ldots, x_0, (m \geq 0) \), then

(6.35) \( P(\Lambda_1 \Lambda_2) = \int P(x_m, \ldots, x_0; \Lambda_2) \, dP \leq P(\Lambda_1) \{ 1 - \lambda [1 - P(\Lambda_2)] \}. \)

Since this inequality is true for any sets \( \Lambda_1, \Lambda_2 \) as described, it is true for any \( P \)-measurable cylinder sets \( \Lambda_1, \Lambda_2 \) over \( x_0, x_{-1}, \ldots; x_1, x_2, \ldots \) respectively. Now suppose there is a function \( \phi(\omega) \), a complex-valued \( P \)-measurable function which does not vanish almost everywhere on \( \Omega \), and such that, for some constant \( c \) of modulus 1,
(6.36) \( \phi(T\omega) = c\phi(\omega) \)

almost everywhere on \( \Omega \). To prove the theorem, it is sufficient to show that \( \phi(\omega) \) is identically a constant almost everywhere on \( \Omega \). Since

\[
\phi(T^n\omega) = c^n\phi(\omega), \quad n = 1, 2, \ldots,
\]

almost everywhere on \( \Omega \), if the integers \( n_1, n_2, n_3, \ldots \) are chosen so that \( \lim_{n \to \infty} e^{\pi in} = 1 \), it follows that

\[
(6.37) \lim_{n \to \infty} \phi(T^n\omega) = \lim_{n \to \infty} \phi(T^{-n}\omega) = \phi(\omega)
\]

almost everywhere on \( \Omega \). Let \( \Lambda = \Lambda(\omega) \) be the \( \Omega \)-set defined by \( \phi(\omega) \in O \) where \( O \) is an open set of the complex \( \phi \)-plane. According to Lemma 6.1, (6.37) implies that \( \Lambda(\omega) \) can be considered (neglecting sets of \( P \)-measure 0) as a cylinder set over both \( x_1, x_2, \ldots \), and \( x_{-1}, x_{-2}, \ldots \). Then in (6.35) we can take \( \Lambda_1 = \Lambda_2 = \Lambda \), obtaining

\[
(6.38) P(\Lambda) \leq P(\Lambda) \{1 - \lambda [1 - P(\Lambda)]\},
\]

which implies that \( P(\Lambda) = 0 \), or that \( P(\Lambda) = 1 \). Since \( O \) is arbitrary, this means that there is a constant \( \phi_0 \) such that \( \phi(\omega) = \phi_0 \) almost everywhere,\(^*\) as was to be proved.

As an application of the theorems of this section, we shall show how to derive a theorem of Kolmogoroff (I, p. 425).\(^t\) Suppose that \( F_x \) is the Borel field determined by a denumerable collection of its sets,\(^t\) and that conditional probability functions are given, as in Theorem 4.1 (ii), except that instead of supposing that (4.9) implies (4.10), we suppose, with Kolmogoroff, the validity of the stronger condition that there is a number \( \lambda \), \( 0 < \lambda \leq 1 \), such that whenever \( A \) is a cylinder set of \( F_o \) over \( x_1, \ldots, x_\lambda \),

\[
(6.39) P(x_0; A) = \lambda P(x_0'; A)
\]

for all \( x_0, x_0' \). There is then, according to Theorem 4.1 (ii) a temporally homogeneous Markoff process with the given conditional probability functions. From (6.36) (interchanging \( x_0 \), \( x_0' \)) we find that

\[
(6.40) P(\Lambda) = \int P(x_0; \Lambda)P(dx_0) \leq \frac{1}{\lambda} P(x_0'; \Lambda)
\]

for all \( x_0' \). Then according to Theorem 6.6, with \( \lambda_0 = \lambda, \lambda_1 = \lambda_2 = \cdots = 0 \), the

\* The point \( \phi_0 \) of the complex plane is the intersection of the interiors of all the circles of rational radii with centers at points whose coordinates are rational and for which the corresponding \( \Omega \)-sets are of \( P \)-measure 1.

† The proof to be given cannot compare in simplicity or elegance with that of Kolmogoroff. It is only given to show the significance of Kolmogoroff's hypothesis, (6.39) below, and the place of such a theorem in this development.

‡ This hypothesis will be eliminated below.
process is metrically transitive and has no angle variables. Moreover the hypo-
theses of Theorem 6.5 (iii) are satisfied, so that, according to part (ii) of
that theorem, \( P(x_0; T^mA) \rightarrow P(A) \) except possibly on an \( x_0 \)-set of \( P \)-measure 0.
We shall show that this exceptional set is actually empty. In the integral following

\[
\int P(x_1; T^mA)P(x_0; \, de) = P(x_0; T^mA)
\]

we have just seen that the integrand converges to \( P(A) \) except possibly on an
\( x_1 \)-set of \( P \)-measure 0. It follows readily from (6.39) that if \( P(A) = 0 \), then
\( P(x_0; \Lambda) = 0 \),* so that the exceptional set is of \( P(x_0; de) \)-measure 0 for each
value of \( x_0 \). Then term by term integration in (6.41) gives Kolmogoroff's re-
sult, that \( P(x_0; T^mA) \rightarrow P(A) \) for all \( x_0 \).† The assumption made above, that
\( F_\xi \) is the Borel field determined by a denumerable collection of its sets, is
unnecessary, since in any case, if \( \Lambda \) is preassigned, \( F_\xi \) can be chosen to satisfy
the denumerability condition and the condition that \( \Lambda \) lies in \( F_\omega \).

7. Application to the examples of §5. In this section we apply the results
of §6 to a detailed study of the examples of §5.

I. In this case, that of a sequence of mutually independent chance var-
iables, the conditional probabilities become absolute probabilities. If the proc-
ess is temporally homogeneous, it is always metrically transitive and has no
angle variables (Theorem 6.4). The ergodic theorem, as applied in Theorem
6.1, gives the strong law of large numbers.‡

II. We have seen above (§5) that absolute probabilities \( p_1, \ldots, p_n \) al-
ways exist in case II, and can be obtained in the form

\[
(7.1) \quad p_k = \lim_{\nu \to \infty} \frac{1}{N} \sum_{m=1}^{\nu} p^{(m)}_{jk}, \quad (j \text{ fixed}).
\]

\textbf{Theorem 7.1.} (i) Except possibly on an \( \Omega \)-set of \( P \)-measure 0,

\[
(7.2) \quad \lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} p_{x_m k}.
\]

exists.§

* In fact \( P(\Lambda) \geq \Lambda P(x_0'; \, \Lambda) \) for all \( x_0' \).

† Kolmogoroff actually obtains more, since he obtains an estimate of the speed of convergence.

‡ Cf. Doob (I, pp. 764–765); Hopf (I, p. 83); Khintchine (I).

§ Part (i) supposes that some set of absolute probabilities is accepted, thus determining \( P \)-mea-
sure. In (7.2), \( p_{x_m k} \) is a chance variable, a function of \( \omega: (\cdots, x_{-1}, x_0, \cdots) \) taking on the value \( p_{\omega k} \) at \( \omega \) if \( x_m = r \). Since only cylinder sets over \( x_1, x_2, \cdots \) are involved in the theorem, the result holds when
only the space of points \( (x_1, x_2, \cdots) \) (on which \( P \)-measure is defined in terms of that on \( \Omega \) in an ob-
vious way) is considered.
(ii) If \((j, k)\) is any pair of subscripts, there exists

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} \hat{p}_{jk}^{(m)} = q_{jk}.
\]

The existence of the limit in (7.3) was proved by Fréchet (I, p. 151) by means of an explicit determination of \(\hat{p}_{jk}^{(m)}\), as a function of \(j, k, m\), derived from the theory of linear difference equations. The proof given here will hold in the more general case III. It will be remembered that (7.3) is an integrated form of (7.2).

Let \(p_1, \ldots, p_n\) be some set of absolute probabilities corresponding to the given matrix. The first part of the theorem is simply the first part of Theorem 6.1 (cf. equation (6.2)) in this special case. The proof of (ii) requires more care. We shall first choose a particular set of absolute probabilities \(p_1, \ldots, p_n\) obtained by applying the corollary of Theorem 4.1 (ii), where we define the set function \(Q(E)\) to be equal to the number of points in \(E\) divided by \(n\). This convention gives absolute probabilities defined by

\[
\hat{p}_k = \frac{1}{n} \lim_{r \to \infty} \frac{1}{N_r} \sum_{m=1}^{N_r} \left( \sum_{j=1}^{n} \hat{p}_{jk}^{(m)} \right).
\]

According to Theorem 6.1, the limit in (7.3) exists for all pairs of subscripts \(j, k\) for which \(p_j > 0\). Let \(J\) be the set of subscripts \(j\) for which \(p_j = 0\). Then the limit in (7.3) exists if \(j \notin J\). We can write \(\hat{p}_{jk}^{(m+r)}\) in the following form:

\[
\hat{p}_{jk}^{(m+r)} = \sum_{l=1}^{n} \hat{p}_{jl}^{(r)} \hat{p}_{lk}^{(m)} ,
\]

and if we set

\[
\Pi_{jk}^{(N)} = \frac{1}{N} \sum_{m=1}^{N} \hat{p}_{jk}^{(m)} ,
\]

we obtain the relation

\[
\frac{\nu + \mu}{\mu} \Pi_{jk}^{(n+\mu)} - \frac{\nu}{\mu} \Pi_{jk}^{(n)} = \sum_{l=1}^{n} \hat{p}_{jl}^{(r)} \Pi_{lk}^{(m)} .
\]

Now according to (7.4), since \(p_l = 0\) if \(l \in J\),

\[
\lim_{r \to \infty} \frac{1}{N_r} \sum_{m=1}^{N_r} \left[ \sum_{j=1}^{n} \left( \sum_{l \in J} \hat{p}_{jl}^{(m)} \right) \right] = 0.
\]

This implies that

\[
\lim \inf_{m \to \infty} \sum_{j=1}^{n} \sum_{l \in J} \hat{p}_{jl}^{(m)} = 0.
\]
Moreover, it has already been shown that

\[(7.9) \lim_{\mu \to \infty} \Pi_{ik}^{(u)} = q_{ik}, \quad l \not\in J,\]

exists. Then letting \(\mu\) become infinite in (7.6), and using (7.9), we obtain

\[(7.10) \limsup_{N \to \infty} \Pi^{(N)}_{jk} = \limsup_{\mu \to \infty} \sum_{l \in J} \mathcal{P}_{jl}^{(u)} \Pi_{lk}^{(u)} + \sum_{l \not\in J} \mathcal{P}_{jl}^{(u)} q_{lk} \leq \sum_{l \in J} \mathcal{P}_{jl}^{(v)} + \sum_{l \not\in J} \mathcal{P}_{jl}^{(v)} q_{lk}\]

and

\[(7.11) \liminf_{N \to \infty} \Pi^{(N)}_{jk} \geq \sum_{l \not\in J} \mathcal{P}_{jl}^{(v)} q_{lk},\]

so that

\[(7.12) \limsup_{N \to \infty} \Pi^{(N)}_{jk} - \liminf_{N \to \infty} \Pi^{(N)}_{jk} \leq \sum_{l \in J} \mathcal{P}_{jl}^{(v)}.

This inequality is true for \(v = 1, 2, \cdots\), so that, using (7.8), we obtain

\[(7.13) \limsup_{N \to \infty} \Pi^{(N)}_{jk} = \liminf_{N \to \infty} \Pi^{(N)}_{jk}\]

as was to be proved.

Since, in general, there is not a unique set of absolute probabilities \(p_1, \cdots, p_n\), a given matrix may correspond to several temporally homogeneous processes. If all these processes are metrically transitive, the matrix \((p_{jk})\) will be called metrically transitive. If none of these processes has angle variables, the matrix will be said to have no angle variables. Otherwise the matrix will be said to be not metrically transitive, or to have angle variables, as the case may be.

**Theorem 7.2.** The matrix \((p_{jk})\) is metrically transitive if and only if

\[(i)\] there is a single set of absolute probabilities \((p_1, \cdots, p_n)\); or

\[(ii)\] the limit \(q_{ik}\) depends only on \(k\); or

\[(iii)\] the equations

\[(7.14) \sum_{j=1}^{n} x_j p_{jk} = x_k, \quad k = 1, \cdots, n,\]

have only a single linearly independent solution in \((x_1, \cdots, x_n)^*\), that is, the matrix \((p_{jk} - \delta_{jk})\) has rank \(n - 1\); or

\[\*\] This condition is not the same as that of (i) since the absolute probabilities are restricted to be non-negative.
(iv) the characteristic equation of the matrix \((p_{jk})\) has 1 as a simple root; or
(v) the matrix \((p_{jk})\) cannot be put in the form of Fig. 1 (where \(R_1, R_2, R_3\) are square matrices and the 0's represent blocks consisting entirely of 0-elements, in which \(R_3\), but not \(R_1\) or \(R_2\), may be absent, by means of some permutation applied to both rows and columns.

\[
\begin{pmatrix}
R_1 & 0 & 0 \\
0 & R_2 & 0 \\
\vdots & \vdots & R_3
\end{pmatrix}
\]

Fig. 1

It would be very difficult to give complete references to previous work on the various parts of this and the following theorems, and such references are perhaps made unnecessary by Fréchet's forthcoming book. Since the time of Markoff, various writers have rediscovered and extended his results, independently of Markoff and of each other. It is hoped that this paper will provide a certain unity to these results, and it is claimed that the terminology used to describe the various cases is of more general validity and less ad hoc than that previously used. The methods, and some of the results, are new. Fréchet and Hadamard (I) have given a historical discussion of some of them. The equivalence of (i)–(iv) was shown by Fréchet (I) in the most detailed treatment of case II which has as yet appeared. The equivalence of (ii) and (v) is somewhat related to more specialized results of von Mises (I, pp. 533–549). The equivalence of (iv) and (v) (in a somewhat different form, with the additional hypothesis that no column of \((p_{jk})\) contains only 0 elements) was obtained by Romanovsky (I, pp. 154–155) by applying theorems of Frobenius. The matrix can be further decomposed if 1 is a root of multiplicity >2. As Romanovsky proves, and as follows readily here also, \(R_1, R_2\) can be replaced by \(\nu\) boxes along the main diagonal, if \(\nu\) is the multiplicity of 1 as a root of the characteristic equation. A complete proof of each part of Theorem 7.2 will be given, since the method will be available for the treatment of case III, and the details of the latter case will then be omitted.

**Proof of** (i). Suppose that the given matrix is metrically transitive, and let \(p_1, \ldots, p_n\) be a set of absolute probabilities corresponding to it. Then according to Theorem 6.1, Corollary 1,

\[ q_{jk} = p_k, \quad k = 1, \ldots, n, \]

if \(p_1 > 0\). If \(p_1', \ldots, p_n'\) is a second set of absolute probabilities corresponding to the given matrix, then

\[ \frac{1}{2}(p_1 + p_1'), \ldots, \frac{1}{2}(p_n + p_n') \]
is also a set of absolute probabilities corresponding to the given matrix, so that if \( p_i + p_j > 0 \),

\[ q_{ik} = \frac{1}{2}(p_k + p_k'), \quad k = 1, \ldots, n. \]

Combining these two results, if we choose \( j_0 \) so that \( p_{j_0} > 0 \),

\[ q_{ik} = p_k = \frac{1}{2}(p_k + p_k'), \quad k = 1, \ldots, n, \]

that is,

\[ p_k = p_k', \quad k = 1, \ldots, n. \]

Thus metric transitivity implies that there is only a single set of absolute probabilities.

Conversely, suppose that there is only a single set of absolute probabilities, \( p_1, \ldots, p_n \). It can be verified directly that for each value of \( j, q_1, \ldots, q_n \) is a set of absolute probabilities corresponding to the given matrix,* so that \( q_{ik} = p_k \) for all \( j, k \). If the matrix is not metrically transitive, the process determined by the matrix of conditional probabilities \( (p_{ij}) \) and the absolute probabilities is not metrically transitive (that is, the corresponding transformation \( T \) is not metrically transitive), so that there is, according to Theorem 6.2, a set of subscripts \( K \), such that

\begin{equation}
0 < p_k, \quad k \in K, \quad \sum_{k \in K} p_k < 1, \tag{7.15}
\end{equation}

\begin{equation}
\sum_{i \in K} \Pi_{k}^{(m)} = 1, \quad k \in K, \quad m = 1, 2, \ldots. \tag{7.16}
\end{equation}

Then

\[ \sum_{i \in K} \Pi_{k}^{(m)} = 1, \quad k \in K, \quad m = 1, 2, \ldots, \]

so that, using the fact that \( \Pi_{k}^{(m)} \rightarrow q_{k} = p_{l} \), and (7.16), we obtain

\[ \sum_{i \in K} p_{l} = \lim_{m \rightarrow \infty} \sum_{i \in K} \Pi_{k}^{(m)} = 1, \quad k \in K, \]

contradicting (7.15). The matrix is thus metrically transitive.

* In general, it can be shown that if \( q_1, \ldots, q_n \) is a linear combination of columns of the matrix \( (q_{ij}) \), where the coefficients of the combination are non-negative and have sum 1, then \( q_1, \ldots, q_n \) is a set of absolute probabilities corresponding to the given matrix, and conversely every set of absolute probabilities corresponding to the given matrix can be obtained in this way. Cf. the discussion of case III below.

† The \( \Omega \)-set determined by the condition \( x_0 \in K \) is invariant under \( T \) up to an \( x_0 \)-set of \( P \)-measure 0. Equation (7.16) for \( m = 1 \) is then the first equation of (6.11), and it follows for \( m > 1 \) by direct verification in view of the definition of \( p_{ij}^{(m)} \).
Proof of (ii). If the matrix is metrically transitive, it was shown in the proof of (i) that \( q_{jk} = p_k \) for all \( j, k \) (where \( p_1, \ldots, p_n \) is the uniquely determined set of absolute probabilities) and \( q_{jk} \) therefore depends only on \( k \). Conversely if \( q_{jk} = q_k \) is independent of \( j \), we shall show that the absolute probabilities are uniquely determined by investigating the solutions of (7.14). Let \( (x_1, \ldots, x_n) \) be a solution of (7.14). Then it can be verified directly that

\[
\sum_{j=1}^n x_j p_{jk}^{(m)} = x_k, \quad k = 1, \ldots, n; \quad m = 1, 2, \ldots,
\]

so that

\[
\sum_{j=1}^n x_j \Pi_{jk}^{(N)} = x_k, \quad k = 1, \ldots, n; \quad N = 1, 2, \ldots.
\]

Letting \( m \) become infinite in (7.18) we obtain

\[
\sum_{j=1}^n x_j q_k = q_k \sum_{i=1}^n x_i = x_k, \quad k = 1, \ldots, n.
\]

Then if \( (p_1, \ldots, p_n) \) is a set of absolute probabilities, since \( p_1, \ldots, p_n \) is a solution of (7.14) and since \( p_1 + \cdots + p_n = 1 \),

\[
q_k \sum_{i=1}^n p_i = q_k = p_k, \quad k = 1, \ldots, n.
\]

Thus the absolute probabilities are uniquely determined; which fact implies, according to (i) that the matrix \( (p_{jk}) \) is metrically transitive.

Proof of (iii). If the system (7.14) has only a single linearly independent solution, the absolute probabilities (which constitute a particular solution) are surely uniquely determined, so the matrix is metrically transitive (according to (i)). Conversely, according to (i), if the matrix is metrically transitive, there is a unique set of absolute probabilities \( p_1, \ldots, p_n \), and we have seen that \( q_{jk} = p_k = q_k, (j, k = 1, \ldots, n) \). Then if \( (x_1, \ldots, x_n) \) is any solution of (7.14), it is linearly dependent on \( (p_1, \ldots, p_n) \) by (7.19).

Proof of (iv). Since a set of absolute probabilities is a solution of (7.14), 1 is always a root of the characteristic equation of the matrix \( (p_{jk}) \). If the matrix is metrically transitive, there is only a single linearly independent solution of (7.14), according to (iii), and we shall show that 1 is a simple root of the characteristic equation of the matrix. Suppose the contrary. If all the columns of the matrix \( (p_{jk} - \lambda \delta_{jk}) \) are added to the first, every element of the first column becomes \( 1 - \lambda \). If \( 1 - \lambda \) is factored from the determinant, the determinant still vanishes for \( \lambda = 1 \), by hypothesis. Then the equations
\[ x_1 + \cdots + x_n = 0, \]

(7.20) \[ \sum_{j=1}^{n} x_j p_{jk} = x_k, \quad k = 2, \ldots, n, \]

have a non-trivial solution (expressing the fact that the rows of the determinant are linearly dependent). If the last \( n - 1 \) equations are subtracted from the first, the first becomes

(7.21) \[ \sum_{j=1}^{n} x_j p_{j1} = -x_2 - \cdots - x_n = x_1. \]

Thus \((x_1, \cdots, x_n)\) is a solution of (7.14), and since \(x_1 + \cdots + x_n = 0\), it is not linearly dependent on \((p_1, \cdots, p_n)\); which contradicts the fact that there is only a single linearly independent solution of (7.14).

Conversely, if 1 is a simple root of the characteristic equation of the matrix \((p_{jk})\), the system (7.14) has only a single linearly independent solution.\(^*\)

**Proof of (v).** If the given matrix is metrically transitive, we shall prove, using the fact that \(q_{jk} = p_k\) for all \(j, k\) (where \(p_1, \cdots, p_n\) is the uniquely determined set of absolute probabilities), that the matrix \((p_{jk})\) cannot be put in the form of Fig. 1, with \(R_1\) and \(R_2\) both present, by a transformation of the type described. (Such a transformation corresponds to a relabeling of the points of \(X\).) If, on the contrary, the matrix can be put in this form, it is no restriction to assume that it is already in this form. It can then be verified directly that the iterated matrix \((p_{jk}^{(m)})\), and therefore \(\Pi_{jk}^{(N)}\), will also be in this form with the same blocks \(R_1, R_2\). But then each column of the limiting matrix \((q_{jk}) = (p_{jk}a_{ij})\) (with \(a_1 = \cdots = a_n = 1\)) contains zeros, so that \(p_1 = \cdots = p_n = 0\), contrary to fact.

To show the converse we shall assume that the matrix is not metrically transitive and put it in the form described. Let \(p_1, \cdots, p_n\) be a set of absolute probabilities for which the corresponding process is not metrically transitive. If any \(p\)'s vanish, we can assume they are the last ones:\(^†\)

\[ (7.22) \quad p_j > 0, \quad j \leq \alpha, \quad 0 < \alpha \leq n, \]

\[ p_{\alpha+1} = p_{\alpha+2} = \cdots = p_n = 0. \]

Since

\[ \sum_{j=1}^{n} p_j p_{jk} = p_k, \quad k = 1, \cdots, n, \]

\(^*\) This follows from elementary matrix theory and does not depend upon the particular properties of our matrix \((p_{jk})\).

\(^†\) This assumption implies the possibility of a matrix transformation of the type described in the theorem.
it follows that

\[(7.23) \quad p_{jk} = 0, \quad \text{if} \quad j \leq \alpha, \quad k > \alpha,\]

where the inequalities on \(j, k\) are to hold simultaneously. Because of the fact that there is not metric transitivity, there is a set of subscripts \(K\) such that \(p_i > 0\) if \(j \in K\), that \(\sum_{i \in K} p_i < 1\), and that

\[(7.24) \quad p_{jk} = 0, \quad \begin{cases} j \in K, & k \in K \\ j \in K, & k \in K, \quad j \leq \alpha. \end{cases}\]

We can assume that \(K\) consists of the first \(a\) subscripts; then equations (7.23) and (7.24) describe the form of Fig. 1. Since \(K\) is not empty, \(R_1\) cannot be absent; since \(\sum_{i \in K} p_i < 1\), \(R_2\) cannot be absent. If no \(p_i\) vanishes, \(R_3\) is absent.

**Theorem 7.3.** The matrix \((p_{jk})\) has no angle variables if and only if

(i) for every pair of subscripts \(j, k,\)

\[(7.25) \quad \lim_{m \to \infty} p_{jk}^{(m)} = q_{jk}\]

exists; or

(ii) 1 is the only root of modulus 1 of the characteristic equation of \(p_{jk}\); or

(iii) the matrix cannot be put in the form of Fig. 1 by a transformation of the type described in Theorem 7.2 (v), where \(R_2, R_3\) may not be present, and where \(R_1\) is itself in the form of Fig. 2, obtained by dividing the subscripts into \(v\) groups \(J_1, \ldots, J_v\), of consecutive subscripts, such that \(p_{jk} = 0\), unless \(p_{jk}\), is in some one of the square matrices \(S_1, \ldots, S_v\), for which \(j \in J_r, k \in J_{r+1}, (r < v)\), or \(j \in J_v, k \in J_1.\)

\[\begin{array}{cccc}
J_1 & J_2 & J_3 & J_4 \\
J_1 & 0 & S_1 & 0 & 0 \\
J_2 & 0 & 0 & S_2 & 0 \\
J_3 & 0 & 0 & 0 & S_3 \\
J_4 & S_4 & 0 & 0 & 0 \\
\end{array} \]

Fig. 2

**Proof of (i).** Suppose that the matrix \((p_{jk})\) has no angle variables. Let

* Cf. Theorem 6.2. The \(\Omega\)-set determined by the condition \(x_0 \in K\) is invariant under \(T\), if we neglect a set of \(P\)-measure 0. The first set of equations in (7.24) is obtained from the first equation of (6.11), which implies that \(P(x_0; C\Lambda) = 0, (x_0 \in \mathcal{E})\), except perhaps for an \(x_0\)-set of \(P\)-measure 0, and the second set of equations in (7.24) is obtained directly from the second equation of (6.11).

† The equivalence of (i), (ii) was shown by Fréchet (I, q.v. for earlier references). Romanovsky (I) has discussed matrices like that of Fig. 2. Cf. also Doeblin and Fortet (I, p. 1700). The latter authors omit mention of exceptional points, implying here the possible existence of \(R_3\).
$p_1, \ldots, p_n$ be a set of absolute probabilities corresponding to this matrix. The condition of Theorem 6.5 will be satisfied if, whenever $r$ is a subscript such that $p_r > 0$, $p_{rk}$ can be put in the form

$$p_{rk} = \phi_{rk} p_k, \quad k = 1, \ldots, n.$$  

This will be possible if, for $r$ fixed, $p_{rk} = 0$ whenever $p_k = 0$; but this is true, since

$$\sum_{j=1}^{n} p_{ij} p_{jk} = p_k.$$  

Thus the condition of Theorem 6.5 is satisfied so that the limit in (7.25) exists for all $j, k$ for which $p_j > 0$. Then if $J$ is the set of subscripts for which the absolute probabilities vanish, the limit in (7.25) exists if $j \notin J$. We shall suppose, using the results of Theorem 7.1 and the fact that for any subscript $j$, $q_{ji}, \ldots, q_{jn}$ is a set of absolute probabilities corresponding to the given matrix, that the absolute probabilities $p_1, \ldots, p_n$ are given by

$$(7.26) \quad p_k = \frac{1}{n} \sum_{j=1}^{n} q_{jk} = \frac{1}{n} \lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{n} \sum_{j=1}^{N} p_{jk}^{(m)}.$$  

We can write $p_{jk}^{(\mu + \nu)}$ in the form

$$(7.27) \quad p_{jk}^{(\mu + \nu)} = \sum_{i \in J} p_{ji}^{(\mu)} p_{ik}^{(\nu)} + \sum_{i \notin J} p_{ji}^{(\mu)} p_{ik}^{(\nu)}.$$  

Then letting $\nu$ become infinite in (7.27), we obtain

$$\lim_{m \to \infty} \sup_{\nu} p_{jk}^{(m)} \leq \sum_{i \in J} p_{ji}^{(\mu)} + \sum_{i \notin J} p_{ji}^{(\mu)} q_{ik},$$  

so that

$$\lim_{m \to \infty} \inf_{\nu} p_{jk}^{(m)} \geq \sum_{i \in J} p_{ji}^{(\mu)} q_{ik},$$  

so that

$$(7.29) \quad \lim_{m \to \infty} \sup_{\nu} p_{jk}^{(m)} - \lim_{m \to \infty} \inf_{\nu} p_{jk}^{(m)} \leq \sum_{i \in J} p_{ji}^{(\mu)} \quad \mu = 1, 2, \ldots.$$  

Now as in the proof of Theorem 7.1 we know that (7.8) is true, and combining this with (7.29) we obtain

$$\lim_{m \to \infty} \sup_{\nu} p_{jk}^{(m)} = \lim_{m \to \infty} \inf_{\nu} p_{jk}^{(m)}$$  

as was to be proved.
Conversely, if the limit in (7.25) exists for every pair of subscripts \( j, k \), Theorem 6.5 shows that there can be no angle variables in any process corresponding to the matrix \((p_{jk})\); thus the matrix \((p_{jk})\) has no angle variables.

**Proof of (ii).** If there are no angle variables corresponding to the given matrix, there can be no root of the characteristic equation of the matrix \((p_{jk})\) of modulus 1, other than 1; for if \( c \) is such a root, there is a set of constants \( x_1, \ldots, x_n \), not all 0, such that

\[
\sum_{k=1}^{n} p_{jk} x_k = c x_j, \quad j = 1, \ldots, n.
\]

Then

\[
\sum_{k=1}^{n} p^{(m)}_{jk} x_k = c^m x_j, \quad j = 1, \ldots, n; \quad m = 1, 2, \ldots.
\]

When \( m \) becomes infinite, the left side converges, to \( \sum_{k=1}^{n} q_{jk} x_k \), whereas the right side, if \( j \) is chosen so that \( x_j \neq 0 \), does not converge. Then the characteristic root \( c \) is impossible.

Conversely, suppose that there is no root of the characteristic equation of modulus 1 other than 1. Let \( p_1, \ldots, p_n \) be any set of absolute probabilities corresponding to the given matrix. We shall prove that the temporally homogeneous process defined in terms of \( p_1, \ldots, p_n \) and \((p_{jk})\) can have no angle variable, by showing that the existence of an angle variable implies the existence of a root (not equal to 1) of the characteristic equation, of modulus 1. If \( \psi(x_0) \) is an angle variable, and if \( \psi(j) = \psi_i \), (6.18) becomes

\[
\sum_{k=1}^{n} p_{jk} \psi_k = \alpha \psi_j, \quad |c| = 1, c \neq 1,
\]

for those values of \( j \) for which \( p_j > 0 \). Let these make up the set of subscripts not belonging to the set \( J \). If \( j \in J \), we have seen above that \( p_{jk} = 0 \) for \( k \in J \), so that changing \( \psi_k \) for \( k \in J \) does not affect (7.32). We shall attempt to re-define \( \psi_j \) for \( j \in J \) to make (7.32) valid for all \( j \). To do this, we must solve the following system of equations for \( \psi_j, j \in J \):

\[
\sum_{k=1}^{n} p_{jk} \psi_k = \alpha \psi_j, \quad j \in J;
\]

or, if we set

\[
\sum_{k \notin J} p_{jk} \psi_k = \alpha_j,
\]

the system
(7.34) \[ \sum_{k \in J} p_{jk} \psi_k = c \psi_j - \alpha_j, \quad j \in J, \]
that is,

(7.35) \[ \sum_{k \in J} (p_{jk} - c \delta_{jk}) \psi_k = - \alpha_j, \quad j \in J. \]

If these equations have a solution, this solution, when combined with the \( \psi'_j \)'s for \( j \in J \), satisfies (7.32) for all \( j \), so that the matrix \( (p_{jk}) \) has the number \( c \) as a root of its characteristic equation. On the other hand, if the system (7.35) has no solution, the matrix \( (p_{jk}) \), with \( j, k \) restricted to \( J \), has \( c \) as a root of its characteristic equation, so that there is a set of numbers \( \{\gamma_j\}, j \in J \), not all 0, satisfying

(7.36) \[ \sum_{k \in J} p_{jk} \psi_k = c \gamma_j, \quad j \in J. \]

But then if \( \gamma_j \) is defined as 0 for \( j \notin J \), the set \( \gamma_1, \ldots, \gamma_n \) provides a non-trivial solution of (7.32) for all \( j \), so that again the matrix \( (p_{jk}) \) has the number \( c \) as a root of its characteristic equation. In any case then, the hypothesis that there is an angle variable implies the existence of a root \( c \neq 1, |c| = 1 \), (which is the characteristic value corresponding to the angle variable) of the characteristic equation of the matrix.

**Proof of (iii).** If there are no angle variables, the matrix \( (p_{jk}) \) cannot be put in the form described; for it is readily verified that if \( (p_{jk}) \) is in this form, the matrices \( (p_{jk}^{(r+1)}), (p_{jk}^{(2r+1)}), \ldots \) are of the same form, whereas the matrices \( (p_{jk}^{(2)}), (p_{jk}^{(r+2)}), (p_{jk}^{(2r+2)}), \ldots \) are of the same form except that the non-zero blocks of \( R_1 \) are the matrices determined by \( (J_1J_3), (J_2J_4), \ldots \) instead of \( (J_1J_2), (J_2J_3), \ldots \). Then if \( p_{jk}^{(m)} \rightarrow q_{jk} \), the submatrix \( R_1 \) of \( (q_{jk}) \) must have only 0 elements; but this contradicts the fact that the sum of the elements in each row of \( R_1 \) is 1. (It would also have been possible to prove this part by giving an explicit definition of an angle variable corresponding to the given matrix.)

Conversely, suppose that there is an angle variable corresponding to some choice \( p_1, \ldots, p_n \) of the absolute probabilities, so that (Theorem 6.3 (i)) there is a function \( \psi(x_0) \) such that

(7.37) \[ \psi(x_1) = c \psi(x_0), \quad |c| = 1, c \neq 1, \]
except on an \( (x_0, x_1) \)-set of \( P \)-measure 0. Since \( \psi(x_0) \) (which takes on at most \( n \) values) necessarily takes on some value on a set of positive \( P \)-measure, \( c \) must be a root of unity (Theorem 6.3 (ii)). This fact will appear again below. Let \( \psi(j) = \psi_j \). Let \( a_1, a_2, \ldots \) be those non-zero values in the set \( \psi_1, \ldots, \psi_n \).
for which the corresponding probability \( p_j \) is positive. Define \( J_k \) as the set of subscripts \( j \) for which \( p_j > 0 \), and \( \psi_j = a_k \). Let \( n_k \) be the number of subscripts in \( J_k \). We can assume (transforming the matrix as described above if necessary) that \( J_1 \) consists of the first \( n_1 \) subscripts, \( J_2 \) of the next \( n_2 \) subscripts, and so on. According to (7.37), some \( a_j \) will necessarily be \( ca_1 \), and we can suppose it to be \( a_2 \). In the same way, some \( a_j \) will necessarily be \( ca_2 \), and we can assume it to be \( a_3 \) (unless it is \( a_1 \)), \( \cdots \). Continuing this, we will necessarily find a first integer \( v > 1 \), such that if \( a_1, a_2, \cdots, a_v \) are chosen successively as described, so that \( a_2 = ca_1, \cdots, a_v = c^{v-1}a_1 \), the next application of the algorithm will give \( ca_v = a_1 \), and hence \( c^v = 1 \). Then \( c \) is a \( v \)th root of unity, and \( 1 < v \leq n \). If \( n_{r-1} < x_0 \leq n_r \) (so that \( \psi(x_0) = a_r \)), then if \( r < v \), it follows that \( \psi(x_1) = a_{r+1} \) necessarily (if \( r = v \), \( \psi(x_1) = a_1 \) necessarily), if we neglect \( x_1 \)-sets of \( P \)-measure 0, that is, subscripts \( k \) for which \( p_k = 0 \); and

\[
\begin{align*}
    p_{jk} = 0 & \quad \text{if} \quad \begin{cases} 
        j \in J_r, & k \in J_{r+1}, \quad p_k > 0, \\
        j \in J_r, & k \in J_1, \quad p_k > 0.
    \end{cases} \quad r = 1, \cdots, v - 1.
\end{align*}
\]

These equations describe the \((J_1 + \cdots + J_v)^2\) matrix \( R_1 \). The fact that the \( \Omega \)-set determined by the condition \( x_1 \in J_1 + \cdots + J_v \) is invariant under the transformation \( T \) up to a set of \( P \)-measure 0 means that the matrix \((p_{jk})\) can be put in the form of Fig. 1, as was shown in the proof of the preceding theorem, except that in this case the matrices \( R_2, R_3 \) may be absent. The \( R_3 \) is absent if every \( p_j \) is positive; \( R_2 \) is absent if the \( \Omega \)-set, determined by the condition \( x_0 \in J_1 + \cdots + J_v \), has \( P \)-measure 1.

**Theorem 7.4.** The matrix \((p_{jk})\) is metrically transitive and has no angle variables if and only if

(i) for every pair of subscripts \( j, k, \lim_{m \to \infty} p_{jk}^{(m)} \) exists and is independent of \( j \); or

(ii) the root 1 is the only root of the characteristic equation of modulus 1 and is itself a simple root; or

(iii) the matrix cannot be put in the form of Fig. 1 if either both \( R_1 \) and \( R_2 \) are present, or if \( R_1 \) is present and has the form of Fig. 2; or

(iv) each matrix \((p_{jk}^{(1)}), \cdots, (p_{jk}^{(n)})\) is metrically transitive.

Only the last part requires any comment. Suppose that the matrix \((p_{jk})\) is metrically transitive and has no angle variables. Then \( p_{jk}^{(m)} \to p_k \) (where \( p_1, \cdots, p_n \) is the uniquely determined set of absolute probabilities). In particular, \( \lim_{m \to \infty} p_{jk}^{(m)} = p_k \), so that, according to (i) the matrix \((p_{jk}^{(2)})\) is metrically transitive. Conversely, suppose that the matrices \((p_{jk}^{(1)}), \cdots, (p_{jk}^{(n)})\) are

* Take \( n_0 = 0 \).
metrically transitive. We shall prove that the matrix \((p_{jk})\) can have no angle variable. If there is an angle variable, its characteristic value \(c, (|c| = 1, c \neq 1)\), is a root of the characteristic equation of the matrix \((p_{jk})\) (cf. the proof of Theorem 7.3 (iii)), and we have seen that this number \(c\) must be a root of unity of order less than or equal to \(n\). Then there is a set of numbers \(x_1, \ldots, x_n\) (not all 0) such that

\[
\sum_{j=1}^{n} x_j p_{jk} = c x_k, \quad k = 1, \ldots, n.
\]

Moreover, if we sum over \(k\),

\[
\sum_{j=1}^{n} x_j = c \sum_{k=1}^{n} x_k,
\]

which implies that \(\sum_{j=1}^{n} x_j = 0\). If \(\nu\) is chosen so that \(c^\nu = 1, (\nu \leq n)\), then

\[
\sum_{j=1}^{n} x_j p_{jk}^{(\nu)} = c^\nu x_k = x_k.
\]

Now the (uniquely defined) absolute probabilities \(p_1, \ldots, p_n\) satisfy the equations

\[
\sum_{j=1}^{n} p_j p_{jk}^{(\nu)} = p_k, \quad k = 1, \ldots, n.
\]

If the matrix \((p_{jk}^{(\nu)})\) is to be metrically transitive, the sets \(p_1, \ldots, p_n, x_1, \ldots, x_n\) must be linearly dependent (Theorem 7.2 (iii)), but this is impossible, since

\[
\sum_{j=1}^{n} p_j = 1, \quad \sum_{j=1}^{n} x_j = 0.
\]

The hypothesis that the matrix \((p_{jk})\) has an angle variable has thus led to a contradiction.

III. In this example, special conditions must be imposed on the conditional probability density \(p(x, y)\) to insure the existence of an absolute probability function. If there is an absolute probability function, it was shown in §5 that it is determined by an \(X\)-integrable density function \(p(x)\) which can be supposed to satisfy

\[
(7.38) \quad \int p(x)dx = 1, \quad \int p(x)p(x, y)dx = p(y)
\]

for all \(y \in X\). In the following we shall mean by an absolute probability density \(p(x)\) an \(X\)-integrable function, which is non-negative and satisfies (7.38) for all \(y\). It follows that two absolute probability densities which are equal almost everywhere on \(X\) are identical. We shall assume throughout that \(p^{(m)}(x, y)\), for
some integer \( m \geq 1 \), satisfies the condition of uniform integrability discussed in §5 insuring the existence of at least one absolute probability density function. In fact, we have seen in the corollary to Theorem 4.1 that if \( Q(E) \) is any probability measure defined on the sets of \( F_x \), we can obtain a probability density \( p(x) \) in the form

\[
(7.39) \quad \int_B p(x) \, dx = \lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} \int Q(dx) \int_B p^{(m)}(x, y) \, dy.
\]

The denumerability hypothesis on the field \( F_x \) and the uniform integrability hypothesis were employed in order to obtain a certain compactness in an aggregate of set functions (cf. §4) through which (7.39) was derived. More abstract formulations are possible (cf. the hypotheses of Kryloff and Bogoliouboff (I)).

**Theorem 7.5.** (i) Except possibly on an \( \Omega \)-set of \( P \)-measure 0,

\[
(7.40) \quad \lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} p(x_m, y) = q(\omega; y)^* 
\]

exists for each value of \( y \) for which \( p(y) \) is finite-valued.

(ii) The limit

\[
(7.41) \quad \lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} \int_B p^{(m)}(x, y) \, dy = q(x, E)
\]

exists for all \( x \in X \) and every set \( E \in F_x \). If there is a value of \( m \) for which \( p^{(m)}(x, y) \) is a bounded function of \( x \) for each value of \( y \),

\[
(7.42) \quad \lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} p^{(m)}(x, y) = q(x, y)
\]

exists for all \( x, y \).

Part (i) of the theorem is new. Part (ii), which is an integrated form of (i), generalizes Theorem 7.1 (ii). The existence of the limit in (7.42) was proved by Fréchet (II, p. 81) who supposes that \( X \) is the closed cover of a bounded domain of euclidean space, and that there is a value of \( m \) such that \( p^{(m)}(x, y) \) is a bounded function. Fréchet’s results were generalized by Kryloff and Bogoliouboff (I) to a form which is substantially identical to Theorem 7.5 (ii) (cf. the note above on the hypotheses of the present discussion) but less general than Theorem 6.5 (ii).

**Proof of (i).** Let \( p(x) \) be an absolute probability density corresponding to the given conditional probability density. For each fixed value of \( y \) for which

* Cf. the note to Theorem 7.1 (i).
\( p(y) < \infty, p(x_0, y) \) is a \( P \)-measurable function, depending only on \( x_0 \), which is integrable on \( \Omega \); namely

\[
\int p(x_0, y) dP = \int p(x, y) p(x) dx = p(y).
\]

Then according to Theorem 6.1 (if \( \phi(\omega) = p(x_0, y) \)) the limit in (7.40) exists almost everywhere on \( \Omega \).

**Proof of (ii).** This proof follows the outline of the special case considered in Theorem 7.1 (ii). There is an \( X \)-measurable function \( \phi(x) \), such that

\[
\phi(x) > 0, \quad \int \phi(x) dx = 1. \ast
\]

If in (7.39) we define \( Q(E) = \int_E \phi(x) dx \), an absolute probability density \( p(x) \) is obtained in the form

\[
(7.39') \int_E p(x) dx = \lim_{\nu \to \infty} \frac{1}{N_{\nu}} \sum_{m=1}^{N_{\nu}} \int p(x) dx \int_E p^{(m)}(x, y) dy,
\]

where the sequence \( \{N_{\nu}\} \) is independent of \( E \). We shall suppose that \( p(x) \) is defined by (7.39'). If we define \( \Pi^{(N)}(x, y) \) by

\[
\Pi^{(N)}(x, y) = \frac{1}{N} \sum_{m=1}^{N} p^{(m)}(x, y),
\]

we find (cf. equation (7.6)) that

\[
(7.43) \quad \frac{\nu + \mu}{\mu} \Pi^{(\nu+\mu)}(x, y) - \frac{\nu}{\mu} \Pi^{(\nu)}(x, y) = \int \Pi^{(\nu)}(x, z) \Pi^{(\mu)}(z, y) dz.
\]

Let \( E_0 \) be the \( X \)-set on which \( p(x) = 0 \). Then if \( E \subset F \), the limit in (7.41) exists, according to Theorem 6.1 (ii), for \( x \) almost everywhere in the complement of \( E_0 \), and

\[
(7.44) \quad \lim_{\nu \to \infty} \int_E \Pi^{(\nu)}(x, y) dy = q(x, E), \quad x \in CE_0,
\]

except possibly for an \( X \)-set of measure 0. According to equation (7.41),

\[
\lim_{\nu \to \infty} \int_{E_0} \phi(x) dx \int_{E_0} \Pi^{(\nu)}(x, y) dy = \lim_{\nu \to \infty} \frac{1}{N_{\nu}} \sum_{m=1}^{N_{\nu}} \int_{E_0} \phi(x) dx \int_{E_0} p^{(m)}(x, y) dy = 0.
\]

\( \ast \) If \( f1dx = \lambda < \infty \), we can take \( \phi(x) = 1/\lambda \). Otherwise we use the fact (cf. §5) that there is an increasing sequence \( E_1 \subset E_2 \subset \cdots \) of \( X \)-measurable sets such that \( \sum_{n=1}^{\infty} E_n = X \) and that \( \int_{E_1} 1dx = \lambda_1 < \infty \). There is a sequence of positive numbers \( \{e_n\} \) such that \( e_0 + \sum_{n=1}^{\infty} e_n(\lambda_{n+1} - \lambda_n) = 1 \), and we define \( \phi(x) \) as \( \phi_0 \) on \( E_1 \) and \( \phi_n \) on \( E_{n+1} - E_n \), for \( n > 1 \).
This implies that

\[(7.45) \liminf_{m \to \infty} \int \phi(x) dx \int p^{(m)}(x, y) dy = 0.\]

Equation (7.45) means that some subsequence \( \{\phi(x) \int_{E_0} p^{(a_i)}(x, y) \} \) of \( \{\phi(x) \int_{E_0} p^{(m)}(x, y) dy \} \), when integrated over \( X \), converges to 0. This implies that \( \phi(x) \int_{E_0} p^{(a_i)}(x, y) dy \) converges in measure to 0* which in turn implies that a further subsequence \( \{\phi(x) \int p^{(b_j)}(x, y) dy\} \) converges to 0 for almost all \( x \). Since \( \phi(x) > 0 \),

\[\lim_{j \to \infty} \int_{E_0} p^{(b_j)}(x, y) dy = 0\]

for almost all \( x \). Now

\[\int_{E_0} p^{(b_j+1)}(x, y) dy = \int p(x, z) dz \int_{E_0} p^{(b_j)}(z, y) dy.\]

The \( z \)-integrand (for \( x \) fixed), \( p(x, z) \int p^{(b_j)}(z, y) dy \), converges to 0, for almost all \( z \), according to what has been just shown, and is less than or equal to the \( z \)-integrable function \( p(x, z) \). Then, by a well known integration theorem, we can go to the limit under the integral sign, so that

\[(7.46) \lim_{j \to \infty} \int_{E_0} p^{(b_j+1)}(x, y) dy = 0, \quad x \in X.\]

If both sides of (7.43) are integrated with respect to \( y \) over a set \( E \) in the field, then if \( \mu \to \infty \), we obtain

\[\liminf_{N \to \infty} \int_E \Pi^{(N)}(x, y) dy = \liminf_{\mu \to \infty} \int p^{(v)}(x, z) dz \int_E \Pi^{(v)}(z, y) dy\]

\[\geq \liminf_{\mu \to \infty} \int_{CE_0} p^{(v)}(x, z) dz \int_E \Pi^{(v)}(z, y) dy.\]

Now the \( z \)-integrand \( p^{(v)}(x, z) \int_E \Pi^{(v)}(z, y) dy \) converges, as \( \mu \to \infty \), to \( p^{(v)}(x, z) q(z, E) \) for almost all \( z \) in \( CE_0 \), according to (7.44). Moreover this integrand is less than or equal to the \( z \)-integrable function \( p^{(v)}(x, z) \). Then we can go to the limit under the integral sign, and obtain

\[\liminf_{N \to \infty} \int_E \Pi^{(N)}(x, y) dy \geq \int_{CE_0} p^{(v)}(x, z) q(z, E) dz, \quad v = 1, 2, \ldots .\]

On the other hand,

\[
\limsup_{N \to \infty} \int_E \Pi^{(N)}(x, y) \, dy = \limsup_{n \to \infty} \left\{ \int_{E_n} p^{(v)}(x, z) \, dz \int_E \Pi^{(u)}(z, y) \, dy \right\}
\]

\[
+ \int_{C_{E_0}} p^{(v)}(x, z) \, dz \int_E \Pi^{(u)}(z, y) \, dy
\]

\[
\leq \int_{E_0} p^{(v)}(x, z) \, dz + \int_{C_{E_0}} p^{(v)}(x, z) q(z, E) \, dz,
\]

\[
\nu = 1, 2, \ldots.
\]

Then

\[
\limsup_{N \to \infty} \int_E \Pi^{(N)}(x, y) \, dy - \liminf_{N \to \infty} \int_E \Pi^{(N)}(x, y) \, dy \leq \int_{E_0} p^{(v)}(x, z) \, dz,
\]

\[
\nu = 1, 2, \ldots.
\]

If \( \nu \) is allowed to increase without limit through the sequence \( \{b_i + 1\} \) of (7.46), it follows that the limit in (7.41) exists for all \( x \), as was to be proved.

Now in addition to the other hypotheses, suppose that for some integer \( \mu \), \( p^{(\mu)}(x, y) \) is bounded in \( x \) for each \( y \). The existence of the limit in (7.41) is readily seen to imply that if \( f(x) \) is a bounded \( X \)-measurable function,

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} \int p^{(m)}(x, z) f(z) \, dz
\]

exists for all \( x \). If we take \( f(z) = p^{(v)}(z, y) \), then

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} \int p^{(m)}(x, z) p^{(v)}(z, y) \, dz = \lim_{N \to \infty} \frac{1}{N} \sum_{m=\mu+1}^{N+\mu} p^{(m)}(x, y)
\]

\[
= \lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} p^{(m)}(x, y)
\]

exists for all \( x, y \), as was to be proved.

A given function \( p(x, y) \) may correspond to several temporally homogeneous processes. If all these processes are metrically transitive, the function \( p(x, y) \) will be called metrically transitive. If none of these processes has angle variables, \( p(x, y) \) will be said to have no angle variables. Otherwise \( p(x, y) \) will be said to be not metrically transitive, or to have angle variables, as the case may be.

**Theorem 7.6.** The function \( p(x, y) \) is metrically transitive if and only if

(i) there is only a single absolute probability density \( p(x) \); or

(ii) \( q(E, x) \) (or \( q(x, y) \) if the hypotheses of the second part of Theorem 7.5 (ii) are satisfied) is independent of \( x \); or
(iii) the integral equation

\[(7.50) \int \psi(x)p(x, y)dx = \psi(y),\]

in the $X$-measurable integrable function $\psi(x)$, has only a single linearly independent solution; or

(iv) there are disjunct $X$-sets $F_1, F_2$ in $F_x$, of positive $X$-measure, such that

\[(7.51) p(x, y) = 0, \quad x \in F_j, \quad y \notin F_j, \quad (j = 1, 2),\]

if we neglect $(x, y)$-sets of $(x, y)$-measure 0.*

The various parts of this theorem are proved by exactly the methods of the proof of Theorem 7.2. In the case of metric transitivity, the limit $q(x, E)$ can be expressed simply by

\[(7.52) q(x, E) = \int_E p(y)dy.\]

(This equation corresponds to the equation $q_{jk} = p_k$ for all $j, k$ in case II, when the given matrix is metrically transitive.) As an example of the proofs used, we prove (iii).

Proof of (iii). If the function $p(x, y)$ is metrically transitive, let $p(x)$ be the uniquely determined probability density. Then (7.52) is true. If $\psi(x)$ is $X$-measurable and integrable, and satisfies (7.50), it follows that

\[\int E \psi(x)dx \int E \Pi^{(N)}(x, y)dy = \int E \psi(y)dy, \quad N = 1, 2, \ldots ;\]

and if $N \to \infty$, this becomes†

\[(7.53) \int \psi(x)dx \int E p(y)dy = \int E \psi(y)dy.\]

If $\alpha$ is defined as $\int E \psi(x)dx$, (7.53) implies, since $E$ is arbitrary, that $\psi(y) = \alpha p(y)$ for almost all $y$. Since the functions $p(y)$, and $\psi(y)$ satisfy their integral equation identically, $\psi(y) = \alpha p(y)$, as was to be proved. Conversely, if there is a solution of (7.52), uniquely determined (up to a constant factor), then the absolute probability density, which is a solution, is uniquely determined; hence there is metric transitivity, according to part (i).

* The equivalence of (i), (iii), (iii) was proved by Fréchet (I), in the case (described above) which he considered. The equivalence of (ii) and (iv) was announced by Kryloff and Bogoliùboff (II), whose hypotheses apparently exclude the possibility of exceptional values in (7.51).

† The $x$-integrand $\psi(x)\int E \Pi^{(N)}(x, y)dy$ converges for all $x$ to $\psi(x)\int E p(y)dy$ and is uniformly less than or equal to $\psi(x)$; so we can go to the limit under the $x$-integral sign.
There is some interest in developing this theorem further.* Suppose a function \( p(x, y) \) is not metrically transitive. Then \( X \) sets \( F_1, F_2 \) exist as described in (iv). Now it may be that \( F_i \) itself contains \( X \)-measurable sets \( F_{a1}, F_{a2} \) of positive measure, such that \( p(x, y) = 0 \) if \( x \in F_{ai}, y \notin F_{ai} \), if we neglect \( (x, y) \)-sets of \((x, y)\)-measure 0. In the contrary case the function \( p(x, y) \), considered only for \( x \in F_{ai}, y \in F_i \), is metrically transitive. Now it is readily seen that the uniform integrability condition prevents the existence of infinitely many \( X \)-measurable sets \( F_1, F_2, \ldots \), of positive \( X \)-measure, such that if \( x \in F_i, y \notin F_i \), then \( p(x, y) = 0 \) neglecting \((x, y)\)-sets of \((x, y)\)-measure 0. Hence there is at most a finite number \( \mu \) of such sets, and \( p(x, y) \) is metrically transitive when considered defined only for \( x, y \in F_i \). Let \( \rho_i(x) \) be the corresponding uniquely defined absolute probability density, and define \( \rho_i(x) = 0 \) if \( x \notin F_i \). Then if \( \rho_1, \ldots, \rho_\mu \) are non-negative numbers with sum 1, \( p(x) = \sum_{i=1}^\mu \rho_i \rho_i(x) \) is an absolute probability density for \( p(x, y) \), and conversely, any absolute probability density for \( p(x, y) \) is such a linear combination. The limit \( q(x, E) \) of Theorem 7.6 (ii) must be \( \sum_{i=1}^\mu \rho_i \int_{E \times E_i} \rho_i(x) \, dx \).

**Theorem 7.7.** The function \( p(x, y) \) has no angle variables if and only if
(i) whenever \( E \in F_x \),

\[
\lim_{m \to \infty} \int_E p^{(m)}(x, y) \, dy = q(x, E)
\]

exists for all \( x \in X \) (or, in case the hypotheses of the second part of Theorem 7.5 (ii) are satisfied, whenever \( \lim_{m \to \infty} p^{(m)}(x, y) \) exists for all \( x, y \)); or

(ii) it is impossible to find disjunct sets \( E_1, \ldots, E_\nu, \nu > 1 \), of positive \( X \)-measure, such that (if we neglect an \((x, y)\)-set of \((x, y)\)-measure 0),

\[
p(x, y) = 0, \quad \begin{cases} x \in E_r, & y \notin E_{r+1}, \\ x \notin E_r, & y \in E_1. \end{cases} \quad r = 1, \ldots, \nu - 1,
\]

In case II, there seems to be no essential difference between the existence of angle variables and the existence of solutions (not equal to 1, of modulus 1) of the characteristic equation of the given matrix. However, in the present case it seems possible to obtain more general results by considering angle variables rather than solutions of the integral equation

\[
\int \psi(y) p(x, y) \, dy = \alpha \psi(x).
\]

The greater adaptability of angle variables is shown, for example, by the fact

* Cf. Kryloff und Bogoliouboff (I, II), Doeblin and Fortet (I).
that the existence of an angle variable implies the existence of a bounded angle variable (as we have seen above). Fréchet was able to extend the usual Fredholm theory of integral equations to his kernels \( p(x, y) \), and so could obtain the complete analogue of Theorem 7.3; and in the present treatment also, if the Fredholm theory is available, the proof of Theorem 7.3 goes right through in case III.

**Proof of (i).** Suppose that \( p(x, y) \) has no angle variables. Let \( p(x) \) be an absolute probability density corresponding to \( p(x, y) \). We show first that the hypotheses of Theorem 6.5 are satisfied, so that there is a function \( \phi(x, y) \) such that for every set \( E \in F_x \) and for all \( x \) (except perhaps values in a set on which the integral of \( p(x) \) vanishes),

\[
\int_E p(x, y)dy = \int_E \phi(x, y)p(y)dy.
\]

Let \( E_0 \) be the \( x \)-set on which \( p(x) = 0 \). Since

\[
\int_{E_0} p(x)dx \int_{E_0} p(x, y)dy = \int_{E_0} p(y)dy = 0,
\]

\( \int_{E_0} p(x, y)dy = 0 \), except possibly on an \( x \)-set on which the integral of \( p(x) \) vanishes. Then if \( \phi(x, y) \) is defined by

\[
\phi(x, y) = \frac{p(x, y)}{p(y)}, \quad \phi(y) > 0
\]

\[
\phi(x, y) = 0, \quad p(y) = 0,
\]

it is readily verified that (7.56) holds, except possibly for values of \( x \) for which the integral of \( p(x) \) vanishes. The hypotheses of Theorem 6.5 are therefore satisfied, and as the process has, by hypothesis, no angle variables, whatever absolute probability density is chosen, the limit in (7.56) must exist almost everywhere on \( CE_0 \). A suitable generalization of the proof of the corresponding part of Theorem 7.3 then completes the proof. Conversely, if the limit in (7.54) exists for all \( x \), Theorem 6.5 states that the function \( p(x, y) \) has no angle variables. The transition from (7.54) to the unintegrated form (\( \lim_{m \to \infty} p^{(m)}(x, y) \)) is easily made as in the proof of Theorem 7.5

**Proof of (ii).** If there are sets \( E_1, \cdots, E_n \) as described in (ii), an angle variable can easily be explicitly defined, or the proof of the corresponding part of Theorem 7.3 can be generalized to show that \( p(x, y) \) must have angle variables. Conversely suppose there is an angle variable, so that (cf. Theorem 6.3) there is an \( X \)-measurable function \( \psi(x) \) such that

\[
\psi(x_1) = c\psi(x_0), \quad |c| = 1, c \neq 1,
\]

if we neglect an \( (x_0, x_1) \)-set of \( P \)-measure 0. Then
\[
\int [\psi(x_1) - \alpha\psi(x_0)]P(dx_0) = \int\int p(x)p(x, y)[\psi(y) - \alpha\psi(x)]dx\,dy = 0,
\]
so that
\[
(7.60) \quad p(x)p(x, y)[\psi(y) - \alpha\psi(x)] = 0
\]
for almost all \((x, y)\), and
\[
(7.61) \quad p(x, y)[\psi(y) - \alpha\psi(x)] = 0, \quad x \notin E_0,
\]
if we neglect \((x, y)\)-sets of zero measure. Let \(\xi\) be a point not in \(E_0\), such that
\[
\int_{E_0} p(\xi, y)dy = p(\xi, y)[\psi(y) - \alpha\psi(\xi)] - 0
\]
(where the second is to hold for almost all values of \(y\)). This may exclude (cf. (7.5)) a \(\xi\)-set of measure 0, besides \(E_0\). Then, since \(p(\xi, y) > 0\) on \(CE_0\) on a set of positive \(y\)-measure (its integral over \(CE_0\) is 1), \(\psi(x_1)\) takes on the value \(\alpha\psi(\xi)\) on a set of positive \(P\)-measure. Now let \(a\) be any value, not equal 0, assumed by \(\psi\) on a set of positive \(P\)-measure. According to Theorem 6.3 (iii), \(c\) must be a root of unity. Let \(\nu\) be the smallest exponent \(r\) for which \(c^r = 1\).

The function \(\psi(x)\) takes on values \(a, ca, \cdots, c^{r-1}a\) on subsets \(E_1, \cdots, E_r\) of \(CE_0\). The fact that if \(x_0 \in E_r\), then \(x_1 \in E_{r+1}\), \((r = 1, \cdots, \nu - 1)\), (if \(x_0 \in E_\nu\), then \(x_1 \in E_1\)) necessarily, if we neglect sets of \(P\)-measure 0, and that the set determined by the condition \(x_0 \in E_1 + \cdots + E_r\) is invariant up to a set of \(P\)-measure 0, implies the conditions of (7.55).

The set \(E_1 + \cdots + E_r\) is one of the sets \(F\) (corresponding to invariant \(\Omega\)-sets) analyzed above. In general there will be then a finite number of \(F\)-sets, and if the function \(p(x, y)\) has any angle variables, one or more of these \(F\)-sets will be divided into a finite number of \(E\)-sets.*

Combining the previous theorems we obtain finally the theorem:

**Theorem 7.8.** The function \(p(x, y)\) is metrically transitive and has no angle variables if and only if

(i) whenever \(E \in F_\nu\), \(\lim_{m \to \infty} \int_E \rho^{(m)}(x, y)dy\) exists for all \(x\) and is independent of \(x\) (or, in case the conditions of the second part of Theorem 7.5 (ii) are satisfied, if \(\lim_{m \to \infty} \rho^{(m)}(x, y)\) exists for all \(x, y\) and is independent of \(x\)); or

(ii) it is impossible to find sets as described in Theorems 7.6 (iv) or 7.7 (ii);

or

(iii) every function \(p(x, y), p^{(2)}(x, y), \cdots\) is metrically transitive.

* This decomposition of \(X\) was announced by Doeblin and Fortet (I). The hypothesis, made here, that absolute and conditional probabilities are given by density functions, is unnecessary, as we only need enough hypotheses to assure the fact that an angle variable will assume one of its values on a set of positive \(P\)-measure.
IV. The case requires little comment. The properties of the transformation $S$ correspond to similar properties of $T$; for example, if one has angle variables, so has the other.

Bibliography

N. Bogoliûboff.
Cf. N. Kryloff and N. Bogoliûboff below.

P. J. Daniell.

W. Doeblin and R. Fortet.

J. L. Doob.
(I) Probability and statistics, these Transactions, vol. 36 (1934), pp. 759–775.
(II) Stochastic processes depending on a continuous parameter, these Transactions, vol. 42 (1937), pp. 107–140.

R. Fortet.
Cf. W. Doeblin and R. Fortet.

M. Fréchet

M. Fréchet and J. Hadamard

J. Hadamard.

E. Hopf.

B. Hostinsky.
(II) Méthodes générales du calcul des probabilités, Mémorial des Sciences Mathématiques, no. 52.

M. Kac.

A. Khintchine.

A. Kolmogoroff.
(II) Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergebnisse der Mathematik, vol. 2, no. 3.

N. Kryloff and N. Bogoliûboff.

P. Lévy.
Z. Łomnicki and S. Ulam.


R. von Mises.


R. E. A. C. Paley and N. Wiener.


V. Romanovsky.


S. Ulam.

Cf. Z. Łomnicki and S. Ulam.

N. Wiener.


University of Illinois,

Urbana, Ill.