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 Ω will be called a probability measure if the space Ω is itself in the field of definition and if the set function is defined as 1 on Ω.

1. **Probability measures defined on spaces of infinitely many dimensions.** Let Ω_0 (X) be any abstract space, consisting of elements ω_0 (x), and let F_{ω_0} (F_x) be a Borel field of subsets of $\Omega_0(X)$. We shall suppose that $\Omega_0 \epsilon F_{\omega_0}$, and $X \epsilon F_x$. If $f(\omega_0)$ is a function defined on Ω_0 which takes on values in X, and if the Ω_0-set defined by the condition $f(\omega_0) \epsilon E$ is in F_{ω_0} for every set E of F_x, then the function $f(\omega_0)$ will be called measurable on Ω_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 Λ_0 of F_{ω_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 Ω whose points are the aggregates ω: $\{x_n\}$, $n \epsilon \mathcal{A}$, $x_n \epsilon X$.§ If n runs through all real numbers t, Ω 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 $E_1-E_1 \cdot E_2$. A Borel field of sets is a field which contains with $E_1, E_2, \cdots$ their sum $\sum_1^\infty E_j$.

‡ Cf. Doob (I); Hopf (I); Khintchine (I); Kolmogoroff (II, pp. 24–30); Lévy (I and numerous papers); Łomnicki 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 Ω 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, Ω is the space of all sequences $(x_1, x_2, \cdots)$, $x_n \in X$; if $\mathcal{A}$ is the set of all integers $\cdots, -1, 0, 1, \cdots$, Ω 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 $\alpha_1, \cdots, \alpha_p$ be any finite set of subscripts, and let $E_1, \cdots, E_p$ be sets of F_x. The conditions

$$x_{\alpha_j} \in E_j, \qquad j = 1, \cdots, p, \tag{1.1}$$

determine a set of elements of Ω. The class of all Ω-sets defined in this way determines a Borel field F_ω of sets of Ω.* Evidently $\Omega \in F_\omega$. We shall define a probability measure $P(\Lambda)$ on the sets Λ of F_ω which will have as its value, on the set defined by (1.1), the P_0-measure of the Ω_0-set determined by the conditions

$$f_{\alpha_j}(\omega_0) \in E_j, \qquad j = 1, \cdots, p. \tag{1.2}$$

The P-measure on Ω is defined by means of a mapping of Ω_0 on Ω, which takes the sets of F_{ω_0} into sets of F_ω. Let ω_0 be a point of Ω_0. The map takes ω_0 into the point (ξ_n) of Ω defined by the equalities

$$\xi_n = f_n(\omega_0), \qquad n \in \mathcal{A}. \tag{1.3}$$

To the Ω-set determined by the conditions of (1.1) will then correspond the Ω_0-set determined by the conditions of (1.2). Then to any set Λ of F_ω will correspond a set Λ_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_{ω_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 $\{\mathbf{x}_n\}$) is reduced to the study of the properties of the space Ω. The earlier representation of the chance variable $\mathbf{x}_m$ by means of the function $f_m(\omega_0)$ defined on Ω_0 has been replaced by a new representation by means of the function $x_m(\omega)$, defined on Ω and taking on the value ξ_m at the point ω: (x_n) for which the mth coordinate is ξ_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 Ω correspond to similar relations between the chance variables $\{\mathbf{x}_n\}$; that is, between the original functions $f_n(\omega_0)$ in the sense that, if Λ is an Ω-set in the field F_ω, the corresponding Ω_0-set in the field F_{ω_0} is defined by conditions on the f's which, when imposed on the $\mathbf{x}$'s define Λ; and $P(\Lambda) = P_0(\Lambda_0)$. Due to the fact that the transformation from Ω_0 to Ω is not one-to-one, certain relations of the functions $\{f_n(\omega_0)\}$ may become distorted; thus an Ω_0-set in the field F_{ω_0} may not go into an Ω-set in the field F_ω. For example, if $\mathcal{A}$ is the set of real numbers t, so that Ω becomes the space of functions $x_t = x(t)$, and if X is the space of real numbers, it may be that $f_t(\omega_0)$ is a continuous function of t for all ω_0; on the other hand the set of elements $x(t)$ of Ω 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 $\mathbf{x}_1, \cdots, \mathbf{x}_N$ *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 $1, \cdots, N$. The space Ω becomes ordinary N-dimensional cartesian space. A probability measure is defined, in terms of the measure properties of $\mathbf{x}_j$ on the x_j-axis,† and the P-measure on Ω 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 $\mathbf{x}_j$ exists, or else that the expectation of $\mathbf{x}_1 \cdots \mathbf{x}_N$ 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 Ω_0; and a new representation was obtained in terms of the aggregate of functions $\{x_n(\omega)\}$ defined on Ω. 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 $\{\mathbf{x}_n\}$, distinguished by a subscript n (which is usually identified in some way with the time) varying in an aggregate $\mathcal{A}$. The chance variable $\mathbf{x}_n$, which takes on values in a space X, is considered defined by a physical process in the course of its development.‖ More specifically, it is supposed that there is a Borel field F_x of sets of points

* If a numerically-valued chance variable $\mathbf{x}$ 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 $\mathbf{x}$ 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 $0 \leqq x \leqq 1$, was published by Kac (I, pp. 47–50).

† The measure of the interval $a < x_j < b$ is defined as the probability that $a < \mathbf{x}_j < b$.

‡ 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 $a_j < x_j < b_j$, $j=1, \cdots, N$, 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 $\mathbf{x}_t$ 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 $\alpha_1, \cdots, \alpha_p$ are elements of $\mathcal{A}$, and if $E_1, \cdots, E_p$ are in F_x, a non-negative number is assigned to the p conditions

$$x_{\alpha_j} \in E_j, \qquad j = 1, \cdots, p. \tag{1.4}$$

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

$$x_{\alpha_j} \in E_j, \qquad j = 1, \cdots, p. \tag{1.4'}$$

It is supposed further that this (p-dimensional) measure function is additive for fixed subscripts $\alpha_1, \cdots, \alpha_p$, and that it can be defined on every set of the Borel field $F_{\alpha_1, \cdots, \alpha_p}$ (the field of Ω-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 Ω and field F_ω as described above. An Ω-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 $F_{\alpha_1, \cdots, \alpha_p}$ is the field of cylinder sets of F_ω over $x_{\alpha_1}, \cdots, x_{\alpha_p}$. What was just described was therefore the determination of a probability measure on the cylinder sets of F_ω over $x_{\alpha_1}, \cdots, x_{\alpha_p}$. Moreover the various measures, obtained by varying the coordinates involved, are coherent in the sense that if Λ is a cylinder set of F_ω over $x_{\alpha_1}, \cdots, x_{\alpha_p}$ and also a cylinder set over $x_{\beta_1}, \cdots, x_{\beta_q}$, then the probability measures, assigned to Λ 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_ω, not merely over the cylinder sets of F_ω 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_ω over a finite number of coordinates, which is a probability measure on the field of cylinder sets of F_ω over each such finite set of coordinates, can be so defined on the remaining sets of F_ω 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 Ω in the case where Ω is the set of natural numbers (with X, F_x defined as in Kolmogoroff's case) so that Ω is the space of sequences ω: $(x_1, x_2, \cdots)$. This latter case, for which in addition the chance variables concerned are independent, has been discussed by many other writers who map Ω on a linear interval and define the measure of an Ω-set by means of the Lebesgue measure of the corresponding set on the interval.*

Let F be the collection of Ω-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 $\Lambda_0, \Lambda_1, \cdots$ be sets of F. Then if $\Lambda_1, \Lambda_2, \cdots$ are disjunct, and if, in addition,

$$\Lambda_0 = \sum_1^\infty \Lambda_m, \tag{1.5}$$

we shall show that

$$P(\Lambda_0) = \sum_1^\infty P(\Lambda_m). \tag{1.6}$$

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 $E_1^{(\nu)}, E_2^{(\nu)}, \cdots$ of X which are involved in the restrictions on x_ν used to define $\Lambda_0, \Lambda_1, \cdots$.‡ We shall define a numerically-valued function $f_\nu(x)$, with domain X, so that each set $E_j^{(\nu)}$ is determined by simple inequalities imposed on $f_\nu(x)$. In order to define the function $f_\nu(x)$ we shall need the lemma which follows:

Lemma 1.1. *Let $\mathcal{E}_1, \mathcal{E}_2, \cdots$ 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 $\mathcal{E}_j$ 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 $\prod_1^\infty \mathcal{E}_{r_m}' = 0$ $(\sum_1^\infty \mathcal{E}_{r_m}' = \sum_1^\infty \mathcal{E}_j)$.*

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 $\mathcal{E}_j$. In the contrary case, there will be only a finite number of distinct sets $\mathcal{E}_r'$.

Using this lemma, we define $f_\nu(x)$ as follows: Identify the sets $\{E_j^{(\nu)}\}$ (ν fixed) with the sets $\{\mathcal{E}_j\}$ of the lemma, and set

$$f_\nu(x) = \limsup r, \qquad x \notin \mathcal{E}'_r \tag{1.7}$$

* 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 $\mathcal{A}$ 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 ν, there will be only one set $E_j^{(\nu)}$; namely X itself.

§ Annals of Mathematics, vol. 33 (1932), p. 602.

The set $\mathcal{E}_r'$ is characterized by the inequality

$$f_\nu(x) \leqq r. \tag{1.8}$$

Since every set $\mathcal{E}_j$ is in the field determined by the sets $\mathcal{E}_r'$, *every set $E_j^{(\nu)}$ is characterized by a finite number of inequalities imposed on $f_\nu(x)$.* Moreover, if $\tilde{E}$ is any Borel set of real numbers, the x-set E determined by the condition $f_\nu(x) \in \tilde{E}$ 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 $\tilde{\Omega}$ of points $\tilde{\omega}$: $\{\tilde{x}_n\}$, $n \in \mathcal{A}$, where $\tilde{x}_n$ is any real number.† Let $F_{\tilde{x}}$ be the field of Borel sets of the $\tilde{x}$-axis, and let $F_{\tilde{\omega}}$ be the Borel field of $\tilde{\Omega}$-sets determined by $F_{\tilde{x}}$ in the same way that F_ω is determined by F_x. We map Ω on $\tilde{\Omega}$, sending the point (x_n) of Ω into the point $(\tilde{x}_n)$ of $\tilde{\Omega}$ for which

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

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

$$\tilde{x}_{\alpha_j} \in \tilde{E}_j, \qquad j = 1, \cdots, p, \tag{1.10}$$

where $\tilde{E}_1, \cdots, \tilde{E}_p$ are Borel sets, the corresponding Ω-set Λ is determined by the conditions

$$f_{\alpha_j}(x_j) \in \tilde{E}_j, \qquad j = 1, \cdots, p. \tag{1.11}$$

It then follows from the definitions of F_ω and $F_{\tilde{\omega}}$ that the Ω-sets going into cylinder sets of $F_{\tilde{\omega}}$ over $\tilde{x}_{\alpha_1}, \cdots, \tilde{x}_{\alpha_p}$ are cylinder sets of F_ω over $x_{\alpha_1}, \cdots, x_{\alpha_p}$, and that the Ω-sets going into the sets of F are sets of F_ω. We now define a set function $\tilde{P}(\tilde{\Lambda})$, on the sets $\tilde{\Lambda}$ of $F_{\tilde{\omega}}$ which are cylinder sets over a finite number of coordinates, by setting $\tilde{P}(\tilde{\Lambda}) = P(\Lambda)$, where Λ is the Ω-set containing every element ω which is taken by the transformation into an element $\tilde{\omega}$ of $\tilde{\Lambda}$. The set function $\tilde{P}(\tilde{\Lambda})$ 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 $\tilde{\Lambda}_0, \tilde{\Lambda}_1, \cdots$ to which correspond the sets $\Lambda_0, \Lambda_1, \cdots$ of (1.5). The sets $\tilde{\Lambda}_1, \tilde{\Lambda}_2, \cdots$ are disjunct, and

$$\tilde{\Lambda}_0 = \sum_1^\infty \tilde{\Lambda}_m. \tag{1.5'}$$

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 $\{\mathcal{E}_r'\}$ can be written in the form $E_{r_0} + (E_{r_2} - E_{r_1}) + \cdots + (E_{r_N} - E_{r_{N-1}})$, with $r_0 < r_1 \leqq r_2 \leqq \cdots \leqq r_N$, or else in the same form without the first set E_{r_0}.

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

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

$$\tilde{P}(\tilde{\Lambda}_0) = \sum_1^\infty \tilde{P}(\tilde{\Lambda}_m); \tag{1.6'}$$

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

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 $\dot{F}_\omega$), 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 Ω be the space whose elements ω are sequences $(\cdots, x_{-1}, x_0, x_1, \cdots)$ of points of X. Let F_x be a Borel field of sets of points of X, including X itself, and suppose that $E_1, \cdots, E_p$ are sets in F_x. The conditions

$$x_{\alpha_j} \in E_j, \qquad j = 1, \cdots, p, \tag{2.1}$$

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

Let $P(\Lambda)$ be a probability measure defined on the sets of F_ω. For a fixed set of coordinates $x_{\alpha_1}, \cdots, x_{\alpha_p}$, $P(\Lambda)$ becomes a probability measure defined essentially on the p-dimensional space of elements $(x_{\alpha_1}, \cdots, x_{\alpha_p})$, 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 Ω together with a probability measure defined on the sets of a field F_ω. 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 $x_j(\omega)$ taking on the value x_j at the point ω: $(\cdots, x_j, \cdots)$ is a measurable function on Ω‡ 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 $\cdots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \cdots$. Conversely,

* The field F was defined at the beginning of the proof.

† Cf. H. Hahn, Annali delle Università Toscane, Pisa, (2), vol. 2 (1933), pp. 433–436.

‡ 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 $\mathbf{x}_1, \mathbf{x}_2, \cdots$. To do this we need only restrict our attention to the cylinder sets of F_ω over $x_1, x_2, \cdots$.

We shall suppose throughout this paper that the probability measure $P(\Lambda)$ is so extended that it is defined on every set Λ_1 differing from a set Λ_0 of F_ω 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 P-measurable. The subsets of a set of P-measure 0 are measurable and of P-measure 0. The P-measure of a P-measurable set is the greatest lower bound of the P-measures of sets $M \supseteq \Lambda$ which are finite or denumerably infinite sums of sets of the field F defined above.† It follows from this fact that if Λ is P-measurable, to every positive number ϵ corresponds a set Λ_ϵ which is a cylinder set of F_ω over a finite number of coordinates, and which has the property that $P(\Lambda \cdot C\Lambda_\epsilon) + P(C\Lambda \cdot \Lambda_\epsilon) < \epsilon$.‡ If Λ is a P-measurable cylinder set over $x_{\alpha_1}, \cdots, x_{\alpha_p}$, it is determined by a condition of the form $(x_{\alpha_1}, \cdots, x_{\alpha_p}) \in E$, where E is a p-dimensional set of points $(x_{\alpha_1}, \cdots, x_{\alpha_p})$. The set E will be called an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure $P(\Lambda)$.

Let $f(\omega)$ be a numerically-valued function of ω. If for every number k the inequality $f(\omega) > k$ defines a set of F_ω (a P-measurable set), $f(\omega)$ will be called measurable with respect to F_ω (P-measurable). If $f(\omega)$ is measurable with respect to F_ω, then it is P-measurable; and conversely if $f(\omega)$ is P-measurable, there is a function $f_1(\omega)$, measurable with respect to F_ω and equal to $f(\omega)$ except possibly on a set of P-measure 0.§ This can be deduced from the following fact (which in turn follows readily from the approximation of P-measurable sets by means of cylinder sets of F_ω over a finite number of coordinates) that if $f(\omega)$ is any P-measurable function, to every positive number ϵ corresponds a function $f_\epsilon(\omega)$, measurable with respect to F_ω, depending on only a finite number of coordinates, and having the property that $|f(\omega) - f_\epsilon(\omega)| \leqq \epsilon$ except perhaps on an Ω-set of P-measure $\leqq \epsilon$. 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_ω, and if $\{n_j\}$ is any set of integers, $f(\omega)$ becomes a function of the coordinates $x_{n_1}, x_{n_2}, \cdots$ 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 Ω rather than Ω 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 Λ will be denoted by $C\Lambda$ throughout this paper.

§ The corresponding theorems for Borel and Lebesgue measurable functions are discussed by de la Vallée Poussin in his book *Intégrales de Lebesgue, Fonctions d'Ensemble, Classes de Baire*, 2d edition, Paris, 1934, pp. 34–40.

held fast, and this function of $x_{n_1}, x_{n_2}, \cdots$ will be also measurable with respect to F_ω.

In later sections we shall use the fact that if F_x is the Borel field determined by some denumerable collection of its sets, the same will be true of F_ω. Even without this hypothesis, it can be shown (by transfinite induction) that any given set Λ, in the field F_ω, is in the field F_ω' corresponding to some Borel field $F_x' \subseteq F_x$ such that F_x' 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_ω, a subfield F_x' of F_x 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_ω'.

The integral of the P-measurable function $f(\omega)$ over a P-measurable set Λ 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 Ω. If $f(\omega)$ depends only on a finite number of coordinates $x_{\alpha_1}, \cdots, x_{\alpha_p}$, and if Λ is a cylinder set over those coordinates, we shall use the notation

$$\int_\Lambda f(\omega)P(de_{\alpha_1,\cdots,\alpha_p})$$

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

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

$$x_j' = x_{j-1}, \qquad j = 0, \pm 1, \pm 2, \cdots.$$

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

3. **The conditional probability functions.*** Let Λ be a P-measurable set. The conditional probability function $P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ is defined as follows. If the set M is allowed to range through the P-measurable cylinder sets over $x_{\alpha_1}, \cdots, x_{\alpha_p}$, $P(\Lambda\mathrm{M})$ is a non-negative completely additive function of sets M which vanishes whenever $P(\mathrm{M})=0$. There is therefore a function $P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$, a non-negative P-measurable function depending only on the coordinates $x_{\alpha_1}, \cdots, x_{\alpha_p}$, 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.

$$\int_{M} P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)dP = P(\Lambda M) \tag{3.1}$$

for every set M.* The function $P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ is determined uniquely up to an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0. For a given set Λ, $P_{\alpha_1,\cdots,\alpha_p}(x^0_{\alpha_1}, \cdots, x^0_{\alpha_p}; \Lambda)$ is called the conditional probability of Λ if $x_{\alpha_j} = x^0_{\alpha_j}, j = 1, \cdots, p$. The subscripts determining the function are given by the subscripts in the argument, so there will be no danger of confusion if $P_{\alpha_1,\cdots,\alpha_p}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ is written simply as $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$. We shall need the following properties of $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ which are easily derived from the definition:†

(i) If $P(\Lambda) = 1$, then $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda) = 1$, except possibly on an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0. If $P(\Lambda) = 0$, then $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda) = 0$, except possibly on an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0.

(ii) If $\Lambda_1, \Lambda_2, \cdots$ are disjunct P-measurable sets, and if $\Lambda = \sum_1^\infty \Lambda_m$,

$$P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda) = \sum_1^\infty P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_m)$$

except possibly for an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0.

This implies the following fact:

(iii) If Λ', Λ'' are P-measurable, and if $\Lambda' \subseteq \Lambda''$, then

$$P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda') \leqq P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda'')$$

except possibly for an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0.

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

(iv) If $\Lambda_1, \Lambda_2, \cdots$ are P-measurable sets, and if

$$\Lambda_1 \supseteq \Lambda_2 \supseteq \cdots, \qquad \prod_1^\infty \Lambda_m = \Lambda,$$

then

$$\lim_{m\to\infty} P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_m) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda),$$

except possibly for an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0.

THEOREM 3.1. *If F_1 is any Borel field of P-measurable sets, including the set Ω, determined by a denumerable subcollection $\Lambda_1, \Lambda_2, \cdots$, and if $\alpha_1, \cdots, \alpha_p$ are any given subscripts, then $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ can be so defined that there is an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set E of P-measure 0 such that for $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ fixed, not in E, $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$, $(\Lambda \in F_1)$, 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 $\Lambda_1, \Lambda_2, \cdots$ with the sets $\mathcal{E}_1, \mathcal{E}_2, \cdots$ of Lemma 1.1, we obtain sets Λ_r' corresponding to the sets $\mathcal{E}_r'$ of that lemma. We then map Ω on the t-interval $0<t<1$ by the transformation which takes a point ω into t if

$$t = \underset{\omega \notin \Lambda_r'}{\text{L. U.B.}}\ r.$$

According to properties (i), (iii), (iv) of the conditional probability functions, if r, s are rational, with $r<s$, there is an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set E_{rs} of P-measure 0 such that if $(x_{\alpha_1}, \cdots, x_{\alpha_p}) \notin E_{rs}$,

$$(3.2) \qquad P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_{r'}') \leqq P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_s'),$$

and there is an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set E_r of P-measure 0 such that if r' approaches r from above (r, r' rational, $r \geqq 0$), and if $(x_{\alpha_1}, \cdots, x_{\alpha_p}) \notin E_r'$, then

$$(3.3) \qquad \lim_{r' \to r} P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda'_{r'}) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_r')$$

(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 $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ is fixed, not in E'. Then

$$F(r) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_r')$$

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 $0 \leqq t<1$ 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 $0 \leqq t<1$ 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 $0<t \leqq r$ is $F(r)$.* The Borel t-sets correspond to the sets of a certain Borel field $\tilde{F}$ of Ω-sets, under the transformation from ω to t, and a non-negative completely additive set function $\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ is thereby defined on these Ω-sets. The Borel sets of the interval (0, 1) are the sets of the Borel field determined by intervals of the type $0<t \leqq r$, 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 $0<t \leqq r$ (r rational) is the set Λ_r', so that the field $\tilde{F}$ includes every set Λ_r' and therefore every set Λ_r; $\tilde{F} \supseteq F_1$. By definition of $\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$, if r is rational,

* J. Radon, Sitzungsberichte der Akademie der Wissenschaften, Vienna, class IIa, vol. 122 (1913), pp. 1305–1313.

(3.4) $$\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_r') = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda_r').$$

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

(3.5) $$\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda),$$

except perhaps on an $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0; and to define $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ as required in the statement of the theorem, we need only re-define $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ as $\tilde{P}(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$.

If $\phi(x_{\alpha_1}, \cdots, x_{\alpha_p}, x_{\beta_1}, \cdots, x_{\beta_q}) = \phi((x_\alpha), (x_\beta))$ is a P-measurable and integrable function depending only on the indicated coordinates, we now define the function

(3.6) $$E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi) = \int \phi((x_\alpha), (x_\beta))P(x_{\alpha_1}, \cdots, x_{\alpha_p}; de_{\beta_1, \cdots, \beta_q})$$

not as an integral, but in such a way that the indicated integration actually gives $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi)$ when it can be carried out. Let M_α be a P-measurable cylinder set over $x_{\alpha_1}, \cdots, x_{\alpha_p}$. Then $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi)$ is defined as the P-measurable integrable function which satisfies

(3.7) $$\int_{\mathrm{M}_\alpha} E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi)dP = \int_{\mathrm{M}_\alpha} \phi((x_\alpha), (x_\beta))dP$$

for all sets M_α. The function $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi)$ is known as the conditional expectation of ϕ for given $(x_{\alpha_1}, \cdots, x_{\alpha_p})$.* Changing ϕ on a set of P-measure 0 does not affect $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi)$. If ϕ is the characteristic function of a P-measurable cylinder set over $x_{\beta_1}, \cdots, x_{\beta_q}$,

$$E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \phi) = P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda).$$

THEOREM 3.2. *Suppose that $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ can be defined so that for each $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ not in some $(x_{\alpha_1}, \cdots, x_{\alpha_p})$-set of P-measure 0, $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$ becomes a probability measure for Λ in the field of cylinder sets of F_ω over $(x_{\beta_1}, \cdots, x_{\beta_q})$. Then* (3.6) *can be interpreted as ordinary integration.*

This means, if $\phi((x_\alpha), (x_\beta))$ is P-measurable and integrable, that (a) whenever $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ is not in some exceptional set of P-measure 0, $\phi((x_\alpha), (x_\beta))$ for $(x_{\alpha_1}, \cdots, x_{\alpha_p}) = (x_{\alpha_1}^0, \cdots, x_{\alpha_p}^0)$ fixed is measurable in terms of the measure function $P(x_{\alpha_1}, \cdots, x_{\alpha_p}; de_{\beta_1, \cdots, \beta_q})$ (that is, that the cylinder set over $(x_{\beta_1}, \cdots, x_{\beta_q})$ defined by $\phi > k$ with $(x_{\alpha_1}, \cdots, x_{\alpha_p}) = (x_{\alpha_1}^0, \cdots, x_{\alpha_p}^0)$ is either in F_ω or differs from such a set by a subset of such a set on which $P(x_{\alpha_1}^0, \cdots,$

* This definition is due to Kolmogoroff (II, pp. 46–47).

$x^0_{\alpha_p}; \Lambda)$ vanishes); and that (b) the integral (3.6) exists and is $E(x_{\alpha_1}, \cdots, x_{\alpha_p}; \Lambda)$, 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 α and β. We shall suppose that $P(x_\alpha; \Lambda_\beta)$ 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 $\phi(x_\alpha, x_\beta)$ is P-measurable and integrable, then

$$(3.8) \quad E(x_\alpha; \phi) = \lim_{\lambda\to 0} \sum_{k=-\infty}^{+\infty} k\lambda P(x_\alpha; \lambda k \leqq \phi < \lambda(k+1)), \qquad \lambda > 0,\text{*}$$

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

$$\int \phi(x_\alpha, x_\beta) P(x_\alpha; de_\beta)$$

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_ω.

(ii) Let $\phi_0(x_\alpha, x_\beta)$ be the characteristic function of a set Λ_0 in the field F_ω, of P-measure 0, and let $\Lambda(\xi)$, $(\xi \in X)$, be the cylinder set over x_β determined by the condition $\phi_0(\xi, x_\beta)=1$. Then, neglecting x_α-sets of P-measure 0 and using (i), we obtain

$$0 = E(x_\alpha, \phi_0) = \int \phi_0(x_\alpha, x_\beta) P(x_\alpha; de_\beta) = P(x_\alpha; \Lambda(x_\alpha));$$

that is, $P(x_\alpha; \Lambda(x_\alpha)) \equiv 0$. This same result will be true even if Λ_0 is only supposed P-measurable, since it can then be enclosed in a set Λ_0', in the field F_ω, which is of P-measure 0.† From this it follows that if $\phi(x_\alpha, x_\beta)$ is any function which vanishes except perhaps on an (x_α, x_β)-set of P-measure 0, then

* The function $P(x_\alpha; \lambda k \leqq \phi < \lambda(k+1))$ is the conditional probability, for given x_α, that $\lambda k \leqq \phi < \lambda(k+1)$. 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.

$$E(x_\alpha, \phi) = \int \phi(x_\alpha, x_\beta)P(x_\alpha; de_\beta) = 0,$$

if we neglect an x_α-set of P-measure 0.

(iii) Let $\phi(x_\alpha, x_\beta)$ be any P-measurable and integrable function depending only on x_α, x_β. 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_α, x_β)-set of P-measure 0, and where $\phi_1(x_\alpha, x_\beta)$ is measurable with respect to F_ω. Combining this fact with the results of (i) and (ii) we see that (neglecting x_α-sets of P-measure 0), for fixed $x_\alpha = x_\alpha{}^0$, $\phi(x_\alpha{}^0, x_\beta)$ is measurable in x_β with respect to the measure function $P(x_\alpha{}^0; de_\beta)$, and that

$$\int \phi(x_\alpha, x_\beta)P(x_\alpha, de_\beta) = \int \phi_1(x_\alpha, x_\beta)P(x_\alpha; de_\beta) = E(x_\alpha; \phi_1) = E(x_\alpha; \phi),$$

as was to be proved. Conversely it is readily seen that *if* (3.6) *exists as an integral, except possibly for an* (x_α)*-set of P-measure* 0, *and if the function of* $(x_{\alpha_1}, \cdots, x_{\alpha_p})$ *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_ω) is the Borel field determined by a denumerable collection of its sets. This will be true, for example, if X is euclidean space of $N \geqq 1$ dimensions, and if F_x is the field of Borel sets of X. A stronger statement can be made, however, since if ϕ_1 is measurable with respect to F_ω, there is always (cf. §2) a Borel field $F_x(\phi_1)$ of X-sets, (depending on ϕ_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 Ω-sets defined in terms of $F_x(\phi_1)$, as F_ω is defined in terms of F_x, then ϕ_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 ϕ under consideration.

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

THEOREM 3.3. *If* α, β, γ *are distinct integers, and if* $\phi(x_\alpha, x_\beta, x_\gamma)$ *is P-measurable, then*

$$E(x_\alpha; \phi) = E[x_\alpha; E(x_\alpha, x_\beta; \phi)]. \tag{3.9}$$

---

* The transition from the P-measurable function ϕ to the function ϕ_1, measurable with respect to F_ω, is made as in (ii) above.

In particular, if ϕ is the characteristic function of a P-measurable cylinder set Λ, over x_γ,

$$(3.10)\qquad P(x_\alpha;\Lambda) = E[x_\alpha; P(x_\alpha, x_\beta; \Lambda)] = \int P(x_\alpha, x_\beta; \Lambda)P(x_\alpha; de_\beta).^*$$

The following theorem will be useful:

THEOREM 3.4. *Let $\phi(x_{\alpha_1}, \cdots, x_{\alpha_p})$ be a P-measurable integrable function. Then*

$$(3.11)\qquad \int \phi dP = \int P(de_{\alpha_1}) \int P(x_{\alpha_1}; de_{\alpha_1}) \int \cdots \int \phi P(x_{\alpha_1}, \cdots, x_{\alpha_{p-1}}; de_{\alpha_p}).\dagger$$

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

$$\int \phi(x_{\alpha_1})dP = \int \phi(x_{\alpha_1})P(de_{\alpha_1}),$$

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 $E(x_{\alpha_1}, \cdots, x_{\alpha_{q-1}}; \phi)$. Since the theorem is supposed true for $p=q-1$, the right side of (3.11) then collapses to

$$\int E(x_{\alpha_1}, \cdots, x_{\alpha_{q-1}}; \phi)dP,$$

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

* The last expression is only symbolic for the second one unless the conditional expectation concerned can be expressed as an integral. Theorem 3.3 is true and will be used below in a somewhat more general form obtained by considering three groups of subscripts, an α-group, a β-group, and a γ-group, to replace α, β, γ.

† This expression is to be evaluated from right to left. If the expression can be considered as an iterated integral, the theorem becomes the generalization of Fubini's theorem (on the evaluation of a multiple integral by means of iterated integrals) to the most general measures on product spaces on which the field of measurable sets is defined by starting with the sets which are direct products of measurable sets in the component spaces. In the case considered in Fubini's theorem, $P(x_{\alpha_1}, \cdots, x_{\alpha_j}; de_{\alpha_{j+1}}) \equiv P(de_{\alpha_{j+1}}), j \geqq 1$. Lévy (I, p. 73) obtains the generalization (where the product space is n-dimensional euclidean space) by a change of variable which reduces the result to Fubini's theorem.

condition: If $\alpha<\beta$, and if Λ is a P-measurable cylinder set over $x_{\beta+1}, x_{\beta+2}, \cdots$, then, except perhaps for an $(x_\alpha, x_{\alpha+1}, \cdots, x_\beta)$-set of P-measure 0,

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

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

$$P(x_\alpha; \Lambda) = \int P(x_\beta; \Lambda) P(x_\alpha; de_\beta) \tag{3.12}$$

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

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

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

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

$$\begin{aligned} P(\Lambda) = \int Q(de_m) \int P(x_m; de_{m+1}) \int P(x_m, x_{m+1}; de_{m+2}) \int \\ \cdots \int \phi_\Lambda P(x_m, \cdots, x_{n-1}; de_n).\dagger \end{aligned} \tag{4.1}$$

* 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 Λ is also a cylinder set over $x_m, \cdots, x_{n'}$, $(n' \neq n)$, the expression (4.1) and the corresponding expression

$$P'(\Lambda) = \int Q(de_m) \int P(x_m; de_{m+1}) \int P(x_m, x_{m+1}; de_{m+2}) \int \cdots \int \phi_\Lambda P(x_m, \cdots, x_{n'-1}; de_{n'}) \tag{4.1'}$$

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

$$P'(\Lambda) = \int Q(de_m) \int \cdots \int P(x_m, \cdots, x_{n-1}; de_n) \int \cdots \int \phi_\Lambda P(x_m, \cdots, x_{n'-1}; de_{n'}),$$

and since ϕ_Λ can depend only on $x_m, \cdots, x_n$, the first integration gives $P(x_m, \cdots, x_{n'-1}; \Omega)$ (which is identically 1) multiplied by ϕ_Λ. Similarly the next integrations, up to the integration over x_n, give ϕ_Λ; hence $P'(\Lambda) = P(\Lambda)$. Evidently $P(\Lambda)$, as thus defined, is a probability measure on the field of cylinder sets of F_ω 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 $m, m+1, \cdots$, 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_ω over $x_m, x_{m+1}, \cdots$ in such a way that the extended set function is a probability measure. Since if Λ, M are respectively cylinder sets over x_{n+1} and $x_m, \cdots, x_n$, with characteristic functions ϕ_Λ, ϕ_{M},

$$\int_{\mathrm{M}} P(x_m, \cdots, x_n; \Lambda) P(de_{m,\cdots,n}) = \int P(x_m, \cdots, x_n; \Lambda)\phi_{\mathrm{M}} P(de_{m,\cdots,n})$$

$$= \int Q(de_m) P(x_m; de_{m+1}) \int \cdots \int \phi_{\mathrm{M}} P(x_m, \cdots, x_n; \Lambda) P(x_m, \cdots, x_{n-1}; de_n)$$

$$= \int Q(de_m) \int \cdots \int P(x_m, \cdots, x_{n-1}; de_n) \int \phi_\Lambda \phi_{\mathrm{M}} P(x_m, \cdots, x_n; de_{n+1})$$

$$= P(\Lambda \cdot \mathrm{M});$$

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 $x_m, x_{m+1}, \cdots$ 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 $P(x_m, \cdots, x_n; \Lambda_{n+1})$ 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_ω, 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*

$$\mathcal{E}_1 \supseteq \mathcal{E}_2 \supseteq \cdots, \quad \prod_1^\infty \mathcal{E}_\nu = 0 \tag{4.2}$$

implies that

$$\lim_{\nu \to \infty} Q_N(\mathcal{E}_\nu) = 0 \tag{4.3}$$

uniformly in N.

(i) *If*

$$\lim_{N \to \infty} Q_N(\mathcal{E}) = Q(\mathcal{E}) \tag{4.4}$$

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

$$\mathcal{E} = \sum_1^\infty \mathcal{E}_m, \qquad \mathcal{E}_m \cdot \mathcal{E}_n = 0, \qquad (m \neq n), \tag{4.5}$$

implies

$$Q(\mathcal{E}) = \sum_1^\infty Q(\mathcal{E}_m), \tag{4.6}$$

---

* Sitzungsberichte der Akademie der Wissenschaften, Vienna, class IIa, vol. 121 (1912), p. 286.

or (since $Q(\mathcal{E})$ is obviously additive) that

$$(4.7)\qquad Q(\mathcal{E})-\sum_{1}^{\nu}Q(\mathcal{E}_m)=Q\left(\sum_{\nu+1}^{\infty}\mathcal{E}_m\right)=\lim_{N\to\infty}Q_N\left(\sum_{\nu+1}^{\infty}\mathcal{E}_m\right)\to 0,\qquad \nu\to\infty.$$

Now

$$\sum_{1}^{\infty}\mathcal{E}_m\supseteq\sum_{2}^{\infty}\mathcal{E}_m\supset\cdots,\qquad \prod_{N=1}^{\infty}\sum_{N}^{\infty}\mathcal{E}_m=0,$$

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 $\{N_n\}$ such that $\lim_{n\to\infty} Q_{N_n}(\mathcal{E})$ exists for every set $\mathcal{E}$ of S which is in the field of sets determined by the sequence of sets $\{\mathcal{E}_n'\}$. 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 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 $m\leqq n$ and cylinder set Λ in F_ω 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 also*

$$(4.8)\qquad P(x_m,\cdots,x_n;\Lambda)\equiv P(x_n;\Lambda).$$

Suppose that for each fixed value of n, whenever $\Lambda_1, \Lambda_2, \cdots$ is a sequence of cylinder sets of F_ω over x_{n+1} satisfying

$$(4.9)\qquad \Lambda_1\supseteq\Lambda_2\supseteq\cdots,\qquad \prod_{1}^{\infty}\Lambda_\nu=0,$$

it is true that

$$(4.10)\qquad \lim_{\nu\to\infty}P(x_n;\Lambda_\nu)=0$$

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*

$$(4.11)\qquad P(x_n;\Lambda)=P(x_{n+1};T\Lambda),\dagger$$

---

* 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_ω over $x_N, x_{N+1}, \cdots, x_n$ with characteristic function ϕ_M, define $P_N(M)$ by

$$(4.12) \quad P_N(M) = \int Q(de_N) \int P(x_N; de_{N+1}) \int \cdots \int \phi_M P(x_{n-1}; de_n).$$

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 $x_m, x_{m+1}, \cdots, x_n$, we define $P(x_{m-1}; M)$ by

$$(4.13)\ P(x_{m-1}; M) = \int P(x_{m-1}; de_m) \int P(x_m; de_{m+1}) \int \cdots \int \phi_M P(x_{n-1}; de_n),$$

where ϕ_M is the characteristic function of M.

Now let m, n be any two integers with $m \leqq n$. The cylinder sets of F_ω over $x_m, \cdots, x_n$ constitute a Borel field $F_{m,n}$ determined by a denumerable subcollection.* Suppose that $\Lambda_1, \Lambda_2, \cdots$ are sets in the field $F_{m,n}$, and that

$$\Lambda_1 \supseteq \Lambda_2 \supseteq \cdots, \quad \prod_1^\infty \Lambda_\nu = 0.$$

Then

$$(4.14) \qquad \lim_{\nu \to \infty} P(x_{m-1}; \Lambda_\nu) = 0,$$

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

$$M_1 \supseteq M_2 \supseteq \cdots, \quad \prod_1^\infty M_\nu = 0.$$

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

$$(4.15) \quad P(x_{m-2}; \Lambda_\nu) = \int P(x_{m-1}; \Lambda_\nu) P(x_{m-2}; de_{m-1}) \leqq \epsilon + P(x_{m-2}; M_\nu).$$

The hypotheses of the theorem imply that

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

uniformly in x_{m-2}. If ν_0 is chosen so large that

* Cf. §2.

$$P(x_{m-2};\ \mathrm{M}_\nu) < \epsilon, \quad \nu > \nu_0$$

for all x_{m-2}, it follows from (4.15) that

$$P(x_{m-2};\ \Lambda_\nu) < 2\epsilon, \quad \nu > \nu_0$$

for all x_{m-2}. Then if $N < m-2$, and if $\nu > \nu_0$,

$$P_N(\Lambda_\nu) = \int Q(de_N) \int P(x_N;\ de_{N+1}) \int \cdots \int P(x_{m-2};\ \Lambda_\nu) P(x_{m-3};\ de_{m-2}) < 2\epsilon;$$

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 $\{P_{N_m}(\Lambda)\}$ of $\{P_N(\Lambda)\}$ converging to a limiting probability measure (defined on the field $F_{m,n}$). Since m, n are arbitrary except that $m \leqq n$, there is a further subsequence $\{P_{N_{a_m}}(\Lambda)\}$ such that

$$\lim_{m\to\infty} P_{N_{a_m}}(\Lambda) = P(\Lambda)$$

exists for every cylinder set Λ of F_ω 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_ω. If m, n are again any two integers with $m \leqq n$, and if Λ is a cylinder set of F_ω over x_{n+1}, it was shown in the general discussion preceding the statement of Theorem 4.1, that (if $N \leqq m$),

$$\int_{\mathrm{M}} P(x_n;\ \Lambda)dP_N = \int_{\mathrm{M}} P(x_N, \cdots, x_n;\ \Lambda)dP_N = P_N(\Lambda\mathrm{M}),$$

for every set M of $F_{m,n}$; which expresses the fact that the conditional probabilities at the Nth stage are the given ones. If N becomes negatively infinite only assuming values of the sequence $\{N_{a_m}\}$, this becomes

$$\int_{\mathrm{M}} P(x_n;\ \Lambda)dP = P(\Lambda\mathrm{M}),$$

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 Λ be a cylinder set of F_ω over a finite number of the coordinates $x_2, x_3, \cdots$. Consider the sequence of set functions $\{Q_N(\Lambda)\}$ where

$$Q_N(\Lambda) = \frac{1}{N}\sum_1^N \int P(x_1;\ T^m\Lambda)Q(de_1).$$

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 Λ. Moreover, if $P(\Lambda)$ is the limit, then

$$P(T\Lambda) = \lim_{m\to\infty} \frac{1}{N_m} \sum_{1}^{N_m} \int P(x_1;\, T^{j+1}\Lambda)Q(de_1) = \lim_{m\to\infty} \frac{1}{N_m} \sum_{2}^{N_m+1} \int P(x_1;\, T^{j}\Lambda)Q(de_1)$$

$$= \lim_{m\to\infty} \frac{1}{N_m} \sum_{1}^{N_m} \int P(x_1;\, T^{j}\Lambda)Q(de_1) = P(\Lambda),$$

so that if Λ is any cylinder set of F_ω over a finite number of coordinates, we can consistently define $P(\Lambda)$ as $P(T^m\Lambda)$, where m is chosen so large that $T^m\Lambda$ is a cylinder set over $x_2, x_3, \cdots$. 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_ω. Evidently $P(\Lambda)=P(T\Lambda)$ for every set of F_ω, 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* $N_1, N_2, \cdots$ *such that*

$$\lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{1}^{N_\nu} \int P(x_1;\, T^m\Lambda)Q(de_1) = P(\Lambda)$$

exists for all cylinder sets Λ *over a finite number of the coordinates* $x_2, x_3, \cdots$ *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 $E_1, \cdots, E_p$ are sets of F_x and $\alpha_1, \cdots, \alpha_p$ are distinct integers, the P-measure of the Ω-set determined by the m conditions $x_{\alpha_j} \in E_j$, $(j=1, \cdots, m)$, is the product of the P-measures of the m sets determined by the single conditions. This case is characterized by the fact that $P(x_{\alpha_1}, \cdots, x_{\alpha_p};\, \Lambda)$ does not depend on $x_{\alpha_1}, \cdots, x_{\alpha_p}$ if Λ is a P-measurable cylinder set over coordinates not including $x_{\alpha_1}, \cdots, x_{\alpha_p}$.*

II. Let X contain n elements, the numbers $1, \cdots, n$. We shall define a

* The corresponding P-measures on Ω 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

$$(5.1)\qquad \begin{aligned} p_{jk} &\geqq 0, &\qquad j, k &= 1, \cdots, n, \\ \sum_{k=1}^{n} p_{jk} &= 1, & j &= 1, \cdots, n. \end{aligned}$$

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

$$(5.2)\qquad \sum_{j=1}^{n} p_j p_{jk} = p_k, \qquad \sum_{1}^{n} p_j = 1, \qquad k = 1, \cdots, n.$$

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

$$(5.3)\qquad p_k = \lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} p_{jk}^{(m)}, \qquad (j \text{ fixed}),$$

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

$$(5.4)\qquad p_{jk}^{(1)} = p_{jk}, \qquad p_{jk}^{(m+1)} = \sum_{l=1}^{n} p_{jl}^{(m)} p_{lk}^{(1)}.$$

Evidently the matrix $(p_{jk}^{(m)})$ is the mth 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), $(x, y \epsilon X)$. 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 $x \epsilon E$, $y \epsilon F$ is the product of the X-measures of E and F.† Let $p(x, y)$ 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) $p(x, y)$ is non-negative;

(b) $p(x, y)$ is integrable in y for fixed x, and

$$\int p(x, y)dy = 1.\S \tag{5.5}$$

If Λ is a cylinder set over $x_{\nu+1}$ determined by the condition $x_{\nu+1} \epsilon E$, $(E \epsilon F_x)$, we define $P(x_\nu; \Lambda)$ by

$$P(x_\nu; \Lambda) = \int_E p(x, y)dy, \qquad \nu = 0, \pm 1, \cdots . \tag{5.6}$$

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

$$E_1 \supseteq E_2 \supseteq \cdots, \quad \prod_1^\infty E_m = 0 \tag{5.7}$$

implies

$$\lim_{m\to\infty} \int_{E_m} p(x, y)dy = 0 \tag{5.8}$$

uniformly in x. This can be interpreted as the uniform (in x) integrability of $p(x, y)$ 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 $p(x, y)$ is measurable with respect to the $X\times Y$ measure, as defined before its extension described in the preceding note, so that $p(x_0, x_1)$ is measurable with respect to F_ω, considered as a function defined on Ω. In any case $p(x, y)$ will be equal to such a function almost everywhere on $X\times Y$-space. It then follows (cf. Saks, ibid., p. 258) that $p(x, y)$ 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 $\int_E p(x, y)dy$. The condition will be satisfied if $p(x, y) \leqq \phi(y)$, 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 $\int 1dx < \infty$, the condition will be satisfied if $p(x, y)^2$ is integrable in y, and if there is a number K such that for every value of x,

$$\int p(x, y)^2 dy \leqq K.$$

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 $E \epsilon F_x$; and if Λ is determined by the condition $x_1 \epsilon E$, $Q(E) = P(\Lambda)$. Moreover (cf. equation (3.1)),

$$\int Q(dx) \int_E p(x, y)dy = Q(E). \tag{5.9}$$

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 $p(x)$ for which

$$Q(E) = \int_E p(x)dx,$$

for all sets $E \epsilon F_x$.* This function $p(x)$ satisfies the equation

$$\int p(x)dx \int_E p(x, y)dy = \int_E p(y)dy, \tag{5.9'}$$

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

$$\int_E dy \int p(x)p(x, y)dx = \int_E p(y)dy.$$

Then

$$\int \left\{ \int p(x)\left[p(x, y) - p(y)\right]dx \right\} dy = 0. \tag{5.10}$$

Since E is arbitrary, (5.10) implies

$$\int p(x)p(x, y)dx = p(y) \tag{5.11}$$

for almost all y. The function $p(y)$ 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" $p(y)$ can be obtained in the form

* Saks, *Théorie de l'Intégrale*, Warsaw, 1933, p. 257.

$$\int_E p(y)dy = \lim_{\nu\to\infty}\frac{1}{N_\nu}\sum_1^{N_\nu}\int_E p^{(m)}(x, y)dy, \qquad (x \text{ fixed}), \tag{5.12}$$

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

$$p^{(1)}(x, y) = p(x, y), \qquad p^{(m+1)}(x, y) = \int p^{(m)}(x, z)p(z, y)dz. \tag{5.13}$$

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

$$\begin{aligned}\int_{E_\nu} p^{(m)}(x, y)dy &= \int_{E_\nu} dy \int p^{(m-1)}(x, z)p(z, y)dz \\ &= \int p^{(m-1)}(x, z)dz \int_{E_\nu} p(z, y)dy \to 0,\end{aligned}$$

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 $\mu \geqq 1$, $p^{(\mu)}(x, y)$ 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_{jk}. 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_{jk} 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_{jk}) is given whose elements satisfy

$$p_{jk} \geqq 0, \qquad \sum_{k=1}^{\infty} p_{jk} = 1. \tag{5.14}$$

* 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):

$$\lim_{N\to\infty} \sum_{k=1}^{N} p_{jk} = 1 \tag{5.15}$$

uniformly in j. If the condition is satisfied, absolute probabilities $p_1, p_2, \cdots$ exist satisfying the conditions

$$p_j \geqq 0, \qquad j \geqq 1; \qquad \sum_{j=1}^{\infty} p_j p_{jk} = p_k, \qquad k \geqq 1; \qquad \sum_{1}^{\infty} p_j = 1. \tag{5.16}$$

Let $q_0, q_1, \cdots$ 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 $k\geqq j$. 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 Sx is defined on X which is one-to-one, takes X-measurable sets into X-measurable sets, and is X-measure preserving. If Λ is a cylinder set of F_ω over $x_{\nu+1}$, determined by the condition $x_{\nu+1} \epsilon E$, define $P(x_\nu; \Lambda)$ by

$$P(x_\nu; \Lambda) = 1 \quad \text{if} \quad Sx_\nu \epsilon E,$$
$$P(x_\nu; \Lambda) = 0 \quad \text{if} \quad Sx_\nu \notin E.$$

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 Ω-set determined by the conditions $x_{\alpha_j} \epsilon E_j$, $(j=1, \cdots, p)$, the P-measure of M can be defined as the X-measure of the set $(S^{-\alpha_1}E_1)(S^{-\alpha_2}E_2) \cdots (S^{-\alpha_p}E_p)$.

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 $\cdots, x_{-1}, x_0, x_1, \cdots$. The results of the preceding section show that the statement is not true if processes corresponding to a sequence of chance variables $\mathbf{x}_1, \mathbf{x}_2, \cdots$ 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* Λ *be any P-measurable set.*

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

$$\lim_{N\to\infty}\frac{1}{N}\sum_{1}^{N}\phi(T^m\omega) = \phi^*(\omega) \tag{6.1}$$

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

$$\lim_{N\to\infty}\frac{1}{N}\sum_{1}^{N}P(x_m;\, T^m\Lambda) = Q(\omega;\Lambda) \tag{6.2}$$

almost everywhere on Ω.

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

$$\lim_{N\to\infty}\frac{1}{N}\sum_{1}^{N}P(x_0;\, T^m\Lambda) = P^*(x_0;\Lambda) \tag{6.3}$$

exists almost everywhere on Ω; *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 Bogolioùboff.

† 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 Ω, 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 Λ is as described in (ii). Then if $m>-\nu$, it follows from equation (3.12) that

$$\int P(x_m;\, T^m\Lambda)P(x_0;\, de) = \int P(x_m;\, T^m\Lambda)P(x_0;\, de_m) = P(x_0;\, T^m\Lambda),$$

neglecting sets of P-measure 0, so that

$$\int \left\{\lim_{N\to\infty}\frac{1}{N}\sum_1^N P(x_m;\, T^m\Lambda)\right\}P(x_0;\, de) = \lim_{N\to\infty}\frac{1}{N}\sum_1^N\int P(x_m;\, T^m\Lambda)P(x_0;\, de)$$
$$= \lim_{N\to\infty}\frac{1}{N}\sum_{|\nu|+1}^N P(x_0;\, T^m\Lambda) = \lim_{N\to\infty}\frac{1}{N}\sum_1^N P(x_0;\, T^m\Lambda),$$

neglecting x_0-sets of P-measure 0.†

Corollary 1. *If the transformation T is metrically transitive,‡ then*

$$\phi^*(\omega) \equiv \int \phi(\omega)dP, \quad Q(\omega;\,\Lambda) = P^*(x_0;\,\Lambda) = P(\Lambda)$$

almost everywhere on Ω.

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 Λ of measure not 0 or 1, which is invariant neglecting a set of P-measure 0, that is, if $T\Lambda=\Lambda+\Lambda_0-\Lambda_0'$, where $P(\Lambda_0)=P(\Lambda_0')=0$, then the P-measurable set $\sum_{-\infty}^{\infty}T^m\Lambda$ 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 Ω-set of positive P-measure, and that

$$\phi(T\omega) = c\phi(\omega), \qquad (|c| = 1,\, c\neq 1),$$

almost everywhere on Ω. If the transformation is metrically transitive, the invariance of $|\phi(\omega)|$ under T implies that $|\phi|\equiv$const. almost everywhere on Ω. (Cf. B. O. Koopman, Proceedings of the National Academy of Sciences, vol. 17 (1935), pp. 315–318.)

$$\lim_{m\to\infty} P(\mathrm{M}T^m\Lambda) \tag{6.4}$$

exists, when m is restricted to a certain increasing set of integers of measure 1,* *independent of the sets* Λ, M *which can be any P-measurable sets. If there is also metric transitivity, the limit in* (6.4) *is* $P(\Lambda)\cdot P(\mathrm{M})$.

Conversely, if the limit in (6.4) *exists, for all P-measurable sets* Λ, M *on some set of integers of measure* 1, *there are no angle variables*; *and if the limit is* $P(\Lambda)\cdot P(\mathrm{M})$, *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* $m_1, m_2, \cdots$ *be an increasing sequence of positive integers. Suppose that* $\{f(T^{m_j}\omega)\}$ *and* $\{f(T^{-m_j}\omega)\}$ *are sequences of functions convergent almost everywhere on* Ω. *Then if O is any open set of the complex plane, the* Ω*-sets defined by the conditions*

$$\lim_{j\to\infty} f(T^{m_j}\omega)\in O, \qquad \lim_{j\to\infty} f(T^{-m_j}\omega)\in O \tag{6.5}$$

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

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

$$\left| f(\omega) - f_\nu(\omega)\right| < 1/\nu \tag{6.6}$$

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

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

$$\lim_{m\to\infty}\frac{1}{m}\sum_{a_j\leq m} 1 = 1.$$

† 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_x 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 $\{m_j\}$ such that for each positive integer j, $f_j(T^{\mu_j}\omega)$ $(f_j(T^{-\mu_j}\omega))$ depends only on $x_{-1}, x_{-2}, \cdots$ $(x_1, x_2, \cdots)$. Since T is measure-preserving,

$$\begin{gathered} |f(T^{\mu_j}\omega) - f_j(T^{\mu_j}\omega)| < 1/j \\ (|f(T^{-\mu_j}\omega) - f_j(T^{-\mu_j}\omega)| < 1/j), \end{gathered} \tag{6.7}$$

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

$$\begin{gathered} |f(T^{\mu_\nu}\omega) - f_\nu(T^{\mu_\nu}\omega)| < 1/N, \qquad \nu = N, N+1, \cdots \\ (|f(T^{-\mu_\nu}\omega) - f_\nu(T^{-\mu_\nu}\omega)| < 1/N, \quad \nu = N, N+1, \cdots) \end{gathered} \tag{6.8}$$

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

$$\begin{gathered} \lim_{\nu\to\infty} f_\nu(T^{\mu_\nu}\omega) = \lim_{\nu\to\infty} f(T^{\mu_\nu}\omega) \\ \left(\lim_{\nu\to\infty} f_\nu(T^{-\mu_\nu}\omega) = \lim_{\nu\to\infty} f(T^{-\mu_\nu}\omega)\right) \end{gathered} \tag{6.9}$$

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

$$\lim_{\nu\to\infty} f_\nu(T^{\mu_\nu}\omega) \in O \qquad \lim_{\nu\to\infty} f_\nu(T^{-\mu_\nu}\omega) \in O$$

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

LEMMA 6.2. *The equality*

$$P(\cdots, x_{n-1}, x_n; \Lambda) = P(x_n; \Lambda)^*$$

holds almost everywhere on the space Ω of a Markoff process, for any P-measurable cylinder set Λ over $x_{n+1}, x_{n+2}, \cdots$.

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

$$\int_{\mathrm{M}} P(x_n; \Lambda)dP = P(\Lambda\mathrm{M}). \tag{6.10}$$

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 $x_m, x_{m+1}, \cdots, x_n$, (6.10) follows from the fact that $P(x_m, \cdots, x_n; \Lambda) \equiv P(x_n; \Lambda)$ almost everywhere on Ω.

* The conditional probability function $P(\cdots, x_{n-1}, x_n; \Lambda)$ is defined in the same way as the function $P(x_m, \cdots, x_n; \Lambda)$. Note that the finiteness of the set $\alpha_1, \cdots, \alpha_p$ 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 Λ is the Ω-set determined by the condition $x_1 \epsilon E$, $(0<P(\Lambda)<1)$, and*

(i) *Λ 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

$$(6.11)\qquad \begin{aligned} P(x_0;\Lambda) &= 1, & x_0 &\epsilon E \\ P(x_0;\Lambda) &= 0, & x_0 &\notin E. \end{aligned}$$

Proof of (i). If there is an invariant set Λ 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 $0<P(\mathrm{M})<1$. If $f(\omega)$ is the characteristic function of M, $f(\omega)=f(T\omega)$ on Ω so that

$$(6.12)\qquad \lim_{m\to\infty} f(T^m\omega) = \lim_{m\to\infty} f(T^{-m}\omega) = f(\omega)$$

on Ω. 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

$$0 \leqq P(\cdots, x_{-1}, x_0; \mathrm{M}_2) \leqq 1,$$

$$(6.13)\qquad \begin{aligned} \int_{\mathrm{M}_1} P(\cdots, x_{-1}, x_0; \mathrm{M}_2)dP &= P(\mathrm{M}_1\mathrm{M}_2) = P(\mathrm{M}_1), \\ \int_{C\mathrm{M}_1} P(\cdots, x_{-1}, x_0; \mathrm{M}_2)dP &= P(C\mathrm{M}_1\mathrm{M}_2) = 0 \end{aligned}$$

that

$$(6.14)\qquad \begin{aligned} P(\cdots, x_{-1}, x_0; \mathrm{M}_2) &= 1, & \omega &\epsilon \mathrm{M}_1, \\ P(\cdots, x_{-1}, x_0; \mathrm{M}_2) &= 0, & \omega &\notin \mathrm{M}_1, \end{aligned}$$

if we neglect sets of P-measure 0. Now, according to Lemma 6.2, $P(\cdots, x_{-1}, x_0; \mathrm{M}_2)=P(x_0; \mathrm{M}_2)$ almost everywhere on Ω. 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 $x_0 \epsilon E$, where E is the x_0-set on which $P(x_0; \mathrm{M}_2)=1$. Since we can suppose (cf. §2), altering $P(x_0; \mathrm{M}_2)$ on an x_0-set of P-measure 0 if necessary, that $P(x_0; \mathrm{M}_2)$ is measurable with respect to F_ω, we can suppose that E is in F_x. The Ω-set determined by the condition $x_0 \epsilon E$ 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

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

so that $T^{-1}\Lambda$, and therefore Λ, 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 Λ 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*

$$\phi(T\omega) = c\phi(\omega), \qquad |c| = 1,\ c \neq 1, \tag{6.15}$$

almost everywhere on Ω.

(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*

$$\psi(x_1) = c\psi(x_0), \tag{6.15'}$$

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,*

$$\psi(x_1) = \text{const.} = c\psi(x_0) \tag{6.16}$$

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 Λ_0, Λ_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,

$$\begin{aligned} P(x_0;\Lambda_1) &= 1, && x_0 \in E_0,\\ P(x_0;\Lambda_1) &= 0, && x_0 \notin E_0. \end{aligned} \tag{6.17}$$

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

$$\int \psi(x_1)P(x_0;de_1) = c\psi(x_0), \tag{6.18}$$

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 $n_1, n_2, \cdots$ such that

$$\lim_{\nu\to\infty} c^{n_\nu} = \lim_{\nu\to\infty} c^{-n_\nu} = 1.$$

Then

$$\lim_{\nu\to\infty} \phi(T^{n_\nu}\omega) = \lim_{\nu\to\infty} \phi(T^{-n_\nu}\omega) = \phi(\omega), \tag{6.19}$$

almost everywhere on Ω. If $\Lambda(O)$ is the Ω-set determined by the condition $\phi(\omega) \in O$ (O an open set of the complex ϕ-plane), the method used in the proof of the preceding theorem shows that $\Lambda(O)$ is a cylinder set over x_0, neglecting an Ω-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 Ω.

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

$$\int P(de_0) \int f_{\mathrm{M}} P(x_0; de_1) = \int P(x_0; \Lambda(x_0)) P(de_0) = 1,$$

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 Λ_0 of positive P-measure, $\psi(x_0)$ must take on $c^n\psi_0$ on a set Λ_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 $\psi_0, c\psi_0, \cdots$ are all distinct, so that the sets $\Lambda_0, \Lambda_1, \cdots$ are all disjunct. But this is impossible since then

$$1 \geqq P\left(\sum_0^{n-1} \Lambda_m\right) = nP(\Lambda_0), \qquad n = 1, 2, \cdots.$$

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 ψ-plane, so chosen that the P-measure of the Ω-set Λ_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 ψ is unimportant, since if ψ is an angle variable, the integrable function ψ_K, equal to ψ if $|\psi| \leqq K$ and otherwise equal to K, satisfies (6.15′) almost everywhere, so that ψ_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 Λ_1 is the Ω-set determined by the condition $c^{-1}\psi(x_1) \epsilon O$, 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 $\psi(x_1) = c\psi(x_0)$ almost everywhere on Ω, and if ψ is integrable, then

$$\int \psi(x_1)P(x_0;\, de_1) = c\int \psi(x_0)P(x_0;\, de_1) = c\psi(x_0),$$

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* $\cdots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \cdots$ *form an independent set (cf. §5, example* I) *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 Λ, as described in the statement of Theorem 6.2, is impossible, since in the case of independence $P(x_0; \Lambda) \equiv P(\Lambda)$; and the process is therefore metrically transitive. For the same reason, there can be no sets Λ_0, Λ_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* $\phi(x_0; x_1)$, *measurable with respect to* F_ω *and integrable over* Ω *in* x_1 *for fixed* x_0, *such that (except possibly for an* x_0*-set of P-measure* 0),

$$P(x_0;\, \Lambda) = \int_\Lambda \phi(x_0;\, x_1)P(de_1) \tag{6.20}$$

whenever Λ *is a cylinder set of* F_ω *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* Λ.

(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* Λ.

(iii) *If it is true that* $P(x_0; \Lambda)$ *(for* Λ *a cylinder set of* F_ω *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* Λ *if it vanishes at all, then a function* ϕ *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 ϕ 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 Λ is a cylinder set of F_ω over x_1, determined by the condition $x_1 \epsilon E$, then $P(x_0; T^m\Lambda)$ is 1 or 0 according as $S^m x_0$ is or is not in E. Then $\lim_{m\to\infty} P(x_0; T^m\Lambda)$ exists almost everywhere on Ω 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 Λ so that $P(\Lambda)\neq 0$, 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 Λ, M, such that

$$\lim_{\nu\to\infty} P(\mathrm{M}T^{a_\nu}\Lambda) = \lim_{\nu\to\infty}\int_{\mathrm{M}} P(x_1; T^{a_\nu}\Lambda)dP = Q(\mathrm{M}; \Lambda) \tag{6.21}$$

exists, where Λ, 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 Ω, then

$$\lim_{\nu\to\infty}\int P(x_1; T^{a_\nu}\Lambda)f(x_1)dP = \int f(x_1)Q(de_1; \Lambda).^* \tag{6.22}$$

In particular (cf. equation (3.12))

$$\begin{aligned}\lim_{\nu\to\infty}\int P(x_1; T^{a_\nu}\Lambda)\phi(x_0, x_1)P(de_1) &= \lim_{\nu\to\infty}\int P(x_1; T^{a_\nu}\Lambda)P(x_0; de_1)\\ &= \lim_{\nu\to\infty} P(x_0; T^{a_\nu}\Lambda)\end{aligned} \tag{6.23}$$

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

$$\int_{\mathrm{K}} Q(x_0; \Lambda)dP = Q(\mathrm{K}; \Lambda) \tag{6.24}$$

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

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

* The set function $Q(\mathrm{M}; \Lambda)$ is obviously additive in M for fixed Λ. Since $Q(\mathrm{M}; \Lambda)\leqq P(\mathrm{M})$, $Q(\mathrm{M}; \Lambda)$ for fixed Λ is a completely additive function of sets M; hence integration with respect to the differential element $Q(de_1; \Lambda)$ has a meaning.

except possibly on a set Λ_ϵ such that $P(\Lambda_\epsilon) \leqq \epsilon/6$. If M is a P-measurable cylinder set over x_1, and if $a_j < \nu$, then

$$(6.25) \qquad P(\mathrm{M}T^\nu\Lambda) = P(T^{a_j-\nu}\Lambda T^{a_j}\Lambda) = \int_{T^{a_j-\nu}\mathrm{M}} P(x_0;\, T^{a_j}\Lambda)dP,$$

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

$$\left| P(\mathrm{M}T^\nu\Lambda) - Q(T^{a_j-\nu}\mathrm{M};\, \Lambda)\right| = \left| \int_{T^{a_j-\nu}\mathrm{M}} [P(x_0;\, T^{a_j}\Lambda) - Q(x_0;\, \Lambda)]dP \right|$$
$$\leqq \epsilon/6 + 2P(\Lambda_\epsilon) \leqq \epsilon/2,$$

if $\nu > a_j > N$. Then since

$$Q(T^k\mathrm{M};\, \Lambda) = \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} P(T^k\mathrm{M}T^j\Lambda) = \lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N} P(\mathrm{M}T^{j-k}\Lambda)$$

is independent of k,

$$(6.26) \qquad \lim_{\nu\to\infty} P(\mathrm{M}T^\nu\Lambda) = Q(\mathrm{M};\Lambda),$$

so that (6.21) holds with $a_\nu = \nu$. The proof that (6.21) implies the existence of the limit in (6.23) can now be used to show that the existence of $\lim_{\nu\to\infty} P(\mathrm{M}T^\nu\Lambda)$ implies the existence of $\lim_{\nu\to\infty} P(x_0;\, T^\nu\Lambda)$. 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 Λ, 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 Λ is in the corresponding field F_ω', and such that $\phi(x_0, x_1)$ is measurable with respect to F_ω'.

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

$$(6.27) \qquad \lim_{\nu\to\infty} P(\mathrm{M}T^\nu\Lambda) = \lim_{\nu\to\infty}\int_{\mathrm{M}} P(x_1;\, T^\nu\Lambda)dP$$

exists. Thus

$$(6.28) \qquad \lim_{\nu\to\infty}\int f(x_1)\overline{g(x_\nu)}dP^*$$

exists, if $f(x_1)$, $g(x_1)$ are characteristic functions of cylinder sets of F_ω 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 ξ 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

$$\lim_{\nu\to\infty} \int \psi(x_1)\overline{\psi(x_\nu)}dP = \lim_{\nu\to\infty} c^{-\nu+1} \int |\psi(x_1)|^2 dP, \qquad |c| = 1, c \neq 1,$$

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, $\lim_{\nu\to\infty} P(x_0; T^\nu\Lambda)$, which we know exists for Λ 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 Λ is a P-measurable cylinder set over x_1, $\lim_{\nu\to\infty} P(x_0;\, T^\nu\Lambda) = P(\Lambda)$ 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 Λ over x_1, $(0 < P(\Lambda) < 1)$, which is invariant under T (if we neglect a set of P-measure 0). Then

$$P(x_0;\, \Lambda) = P(x_0;\, T^\nu\Lambda) \to P(\Lambda), \qquad \nu \to \infty,$$

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 Λ is a cylinder set of F_ω 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

$$P(\Lambda) = \int P(x_0;\, \Lambda) P(de_0).$$

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

$$\tilde{P}(\mathrm{M}) = \int P(de_0) \int f_{\mathrm{M}}(x_0,\, x_1) P(de_1), \tag{6.29}$$

where $f_{\mathrm{M}}(x_0, x_1)$ 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.

$$P(\mathrm{M}) = \int P(de_0) \int f_{\mathrm{M}}(x_0, x_1) P(x_0; de_1), \tag{6.30}$$

and the integration need not be taken symbolically. Let $\mathrm{M}(x_0)$ be the cylinder set over x_1 defined by the equation $f_{\mathrm{M}}(x_0, x_1) = 1$. If we integrate in (6.30), then if $P(\mathrm{M}) = 0$, $P(x_0; \mathrm{M}(x_0)) = 0$, 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 $P(\xi; \mathrm{M}(\xi)) = 0$, $P(\mathrm{M}(\xi)) = 0$. Then if $P(\mathrm{M}) = 0$,

$$\tilde{P}(\mathrm{M}) = \int P(\mathrm{M}(x_0)) P(de_0) = 0;$$

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

$$\tilde{P}(\mathrm{M}) = \int_{\mathrm{M}} \phi(x_0, x_1) dP. \tag{6.31}$$

In particular if M is the intersection of Λ_0 (a cylinder set of F_ω determined by the condition $x_0 \in E_0$) and Λ (a cylinder set of F_ω determined by the condition $x_1 \in E_1$), (6.31) becomes

$$P(\Lambda_0) P(\Lambda) = \int_{E_0} P(de_0) \int_{E_1} P(x_0; de_1),$$

so that

$$\int_{E_0} P(de_0) \left\{ P(\Lambda) - \int_{E_1} \phi(x_0, x_1) P(x_0; de_1) \right\} = 0.$$

This equation is to hold for every set E_0 in the field F_x, 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*

$$\begin{gathered} P(x_0; \Lambda) \geqq \lambda_0 P(\Lambda) \\ P(x_{-\nu}, \cdots, x_0; \Lambda) \geqq \lambda_\nu P(x_{-\nu+1}, \cdots, x_0; \Lambda), \quad \nu = 1, 2, \cdots, \end{gathered} \tag{6.32}$$

for every P-measurable cylinder set Λ over x_1,† where $0 < \lambda_\nu \leqq 1$, and if

* Saks, ibid., p. 257.

† The inequalities are to hold with probability 1 for each set Λ.

$$\prod_{1}^{\infty} \lambda_m^m = \lambda > 0, \tag{6.33}$$

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

If the process is a Markoff process, we can take $\lambda_\nu = 1$, 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 Λ_2 be a P-measurable cylinder set over $x_1, \cdots, x_n$, $n \geqq 1$. Then if f is the characteristic function of Λ_2, and if $m \geqq 1$ (cf. equation (3.10)), then

$$\begin{aligned} P(x_{-m}, \cdots, x_0; \Lambda_2) &= \int P(x_{-m}, \cdots, x_0; de_1) \int P(x_{-m}, \cdots, x_1; de_2) \int \\ &\qquad \cdots \int f P(x_{-m}, \cdots, x_{n-1}; de_n) \\ &= 1 - \int P(x_{-m}, \cdots, x_0; de_1) \int \\ &\qquad \cdots \int (1 - f) P(x_{-m}, \cdots, x_{n-1}; de_n) \\ &\leqq 1 - \int \lambda_m \lambda_{m-1} \cdots \lambda_0 P(de_1) \int \lambda_{m+1} \cdots \lambda_1 P(x_1; de_2) \int \\ &\qquad \cdots \int \lambda_{m+n-1} \cdots \lambda_{n-1} (1 - f) P(x_1, \cdots, x_{n-1}; de_n) \\ &\leqq 1 - \lambda \int P(de_1) \int P(x_1; de_2) \int \\ &\qquad \cdots \int (1 - f) P(x_1, \cdots, x_{n-1}; de_n) \\ &= 1 - \lambda [1 - P(\Lambda_2)]. \end{aligned} \tag{6.34}$$

If Λ_1 is a P-measurable cylinder set over $x_{-m}, \cdots, x_0$, $(m \geqq 0)$, then

$$P(\Lambda_1 \Lambda_2) = \int_{\Lambda_1} P(x_{-m}, \cdots, x_0; \Lambda_2) dP \leqq P(\Lambda_1) \{ 1 - \lambda [1 - P(\Lambda_2)] \}. \tag{6.35}$$

Since this inequality is true for any sets Λ_1, Λ_2 as described, it is true for any P-measurable cylinder sets Λ_1, Λ_2 over $x_0, x_{-1}, \cdots; x_1, x_2, \cdots$ respectively. Now suppose there is a function $\phi(\omega)$, a complex-valued P-measurable function which does not vanish almost everywhere on Ω, and such that, for some constant c of modulus 1,

$$\phi(T\omega) = c\phi(\omega) \tag{6.36}$$

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

$$\phi(T^\nu\omega) = c^\nu\phi(\omega), \qquad \nu = 1, 2, \cdots,$$

almost everywhere on Ω, if the integers $n_1, n_2, n_3, \cdots$ are chosen so that $\lim_{\nu\to\infty} c^{n_\nu} = 1$, it follows that

$$\lim_{\nu\to\infty} \phi(T^{n_\nu}\omega) = \lim_{\nu\to\infty} \phi(T^{-n_\nu}\omega) = \phi(\omega) \tag{6.37}$$

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

$$P(\Lambda) \leqq P(\Lambda)\{1 - \lambda[1 - P(\Lambda)]\}, \tag{6.38}$$

which implies that $P(\Lambda) = 0$, or that $P(\Lambda) = 1$. Since O is arbitrary, this means that there is a constant ϕ_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).† Suppose that F_x is the Borel field determined by a denumerable collection of its sets,‡ 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 λ, $(0 < \lambda \leqq 1)$, such that whenever Λ is a cylinder set of F_ω over x_1,

$$P(x_0; \Lambda) \geqq \lambda P(x_0'; \Lambda) \tag{6.39}$$

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

$$P(\Lambda) = \int P(x_0; \Lambda)P(de_0) \leqq \frac{1}{\lambda} P(x_0'; \Lambda) \tag{6.40}$$

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

* The point ϕ_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 Ω-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 hypotheses of Theorem 6.5 (iii) are satisfied, so that, according to part (ii) of that theorem, $P(x_0; T^m\Lambda)\to P(\Lambda)$ 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

$$\int P(x_1; T^m\Lambda)P(x_0; de_1) = P(x_0; T^m\Lambda) \tag{6.41}$$

we have just seen that the integrand converges to $P(\Lambda)$ except possibly on an x_1-set of P-measure 0. It follows readily from (6.39) that if $P(\Lambda)=0$, then $P(x_0; \Lambda)\equiv 0$,* so that the exceptional set is of $P(x_0; de_1)$-measure 0 for each value of x_0. Then term by term integration in (6.41) gives Kolmogoroff's result, that $P(x_0; T^m\Lambda)\to P(\Lambda)$ for all x_0.† The assumption made above, that F_x is the Borel field determined by a denumerable collection of its sets, is unnecessary, since in any case, if Λ is preassigned, F_x can be chosen to satisfy the denumerability condition and the condition that Λ lies in F_ω.

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 variables, the conditional probabilities become absolute probabilities. If the process 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 $p_1, \cdots, p_n$ always exist in case II, and can be obtained in the form

$$p_k = \lim_{\nu\to\infty} \frac{1}{N_\nu}\sum_{m=1}^{N_\nu} p_{jk}^{(m)}, \qquad (j \text{ fixed}). \tag{7.1}$$

THEOREM 7.1. (i) *Except possibly on an Ω-set of P-measure* 0,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{m=1}^{N} p_{x_m k}. \tag{7.2}$$

exists.§

* In fact $P(\Lambda)\geqq\lambda P(x_0'; \Lambda)$ 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-measure. In (7.2), $p_{x_m k}$ is a chance variable, a function of ω: $(\cdots, x_{-1}, x_0, \cdots)$ taking on the value p_{rk} at ω 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 Ω in an obvious way) is considered.

(ii) *If (j, k) is any pair of subscripts, there exists*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} p_{jk}^{(m)} = q_{jk}. \tag{7.3}$$

The existence of the limit in (7.3) was proved by Fréchet (I, p. 151) by means of an explicit determination of $p_{jk}^{(m)}$, 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 $p_1, \cdots, p_n$ 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 $p_1, \cdots, p_n$ 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

$$p_k = \frac{1}{n} \lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} \left(\sum_{j=1}^{n} p_{jk}^{(m)} \right). \tag{7.4}$$

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 $p_{jk}^{(m+\nu)}$ in the following form:

$$p_{jk}^{(m+\nu)} = \sum_{l=1}^{n} p_{jl}^{(\nu)} p_{lk}^{(m)}, \tag{7.5}$$

and if we set

$$\Pi_{jk}^{(N)} = \frac{1}{N} \sum_{m=1}^{N} p_{jk}^{(m)},$$

we obtain the relation

$$\frac{\nu+\mu}{\mu} \Pi_{jk}^{(\nu+\mu)} - \frac{\nu}{\mu} \Pi_{jk}^{(\nu)} = \sum_{l=1}^{n} p_{jl}^{(\nu)} \Pi_{lk}^{(\mu)}. \tag{7.6}$$

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

$$\lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} \left[\sum_{j=1}^{n} \left(\sum_{l \in J} p_{jl}^{(m)} \right) \right] = 0. \tag{7.7}$$

This implies that

$$\liminf_{m\to\infty} \sum_{j=1}^{n} \sum_{l \in J} p_{jl}^{(m)} = 0. \tag{7.8}$$

Moreover, it has already been shown that

$$(7.9) \qquad \lim_{\mu\to\infty} \Pi_{lk}^{(\mu)} = q_{lk}, \qquad l \notin J,$$

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

$$(7.10) \quad \limsup_{N\to\infty} \Pi_{jk}^{(N)} = \limsup_{\mu\to\infty} \sum_{l\in J} p_{jl}^{(\nu)} \Pi_{lk}^{(\mu)} + \sum_{l\notin J} p_{jl}^{(\nu)} q_{lk} \leqq \sum_{l\in J} p_{jl}^{(\nu)} + \sum_{l\notin J} p_{jl}^{(\nu)} q_{lk}$$

and

$$(7.11) \qquad \liminf_{N\to\infty} \Pi_{jk}^{(N)} \geqq \sum_{l\notin J} p_{jl}^{(\nu)} q_{lk},$$

so that

$$(7.12) \qquad \limsup_{N\to\infty} \Pi_{jk}^{(N)} - \liminf_{N\to\infty} \Pi_{jk}^{(N)} \leqq \sum_{l\in J} p_{jl}^{(\nu)}.$$

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

$$(7.13) \qquad \limsup_{N\to\infty} \Pi_{jk}^{(N)} = \liminf_{N\to\infty} \Pi_{jk}^{(N)}$$

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_{jk} depends only on k; or*

(iii) *the equations*

$$(7.14) \qquad \sum_{j=1}^{n} x_j p_{jk} = x_k, \qquad k = 1, \cdots, n,$$

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.*

$$\begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ \cdot & \cdot & R_3 \end{pmatrix}$$

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 ν boxes along the main diagonal, if ν 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 $p_1, \cdots, p_n$ be a set of absolute probabilities corresponding to it. Then according to Theorem 6.1, Corollary 1,

$$q_{jk} = p_k, \qquad k = 1, \cdots, n,$$

if $p_j > 0$. If $p_1', \cdots, p_n'$ is a second set of absolute probabilities corresponding to the given matrix, then

$$\tfrac{1}{2}(p_1 + p_1'), \cdots, \tfrac{1}{2}(p_n + p_n')$$

is also a set of absolute probabilities corresponding to the given matrix, so that if $p_j+p_j'>0$,

$$q_{jk}=\tfrac{1}{2}(p_k+p_k'), \qquad k=1,\cdots,n.$$

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

$$q_{j_0k}=p_k=\tfrac{1}{2}(p_k+p_k'), \qquad k=1,\cdots,n,$$

that is,

$$p_k=p_k', \qquad k=1,\cdots,n.$$

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, $p_1,\cdots,p_n$. It can be verified directly that for each value of j, $q_{j1},\cdots,q_{jn}$ is a set of absolute probabilities corresponding to the given matrix,* so that $q_{jk}=p_k$ for all j, k. If the matrix is not metrically transitive, the process determined by the matrix of conditional probabilities (p_{jk}) 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

$$0<p_k,\quad k\in K, \qquad \sum_{k\in K}p_k<1, \tag{7.15}$$

$$\sum_{l\in K}p_{kl}^{(m)}=1, \qquad k\in K, \qquad m=1,2,\cdots.\dagger \tag{7.16}$$

Then

$$\sum_{l\in K}\Pi_{kl}^{(m)}=1, \qquad k\in K, \qquad m=1,2,\cdots,$$

so that, using the fact that $\Pi_{kl}^{(m)}\to q_{kl}=p_l$, and (7.16), we obtain

$$\sum_{l\in K}p_l=\lim_{m\to\infty}\sum_{l\in K}\Pi_{kl}^{(m)}=1, \qquad k\in K,$$

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

---

* In general, it can be shown that if $q_1,\cdots,q_n$ 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 $q_1,\cdots,q_n$ 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 Ω-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 $p_{jk}^{(m)}$.

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 $p_1, \cdots, p_n$ 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 $(x_1, \cdots, x_n)$ be a solution of (7.14). Then it can be verified directly that

$$\sum_{j=1}^{n} x_j p_{jk}^{(m)} = x_k, \qquad k = 1, \cdots, n; \; m = 1, 2, \cdots, \tag{7.17}$$

so that

$$\sum_{j=1}^{n} x_j \Pi_{jk}^{(N)} = x_k, \qquad k = 1, \cdots, n; \; N = 1, 2, \cdots. \tag{7.18}$$

Letting m become infinite in (7.18) we obtain

$$\sum_{j=1}^{n} x_j q_k = q_k \sum_{j=1}^{n} x_j = x_k, \qquad k = 1, \cdots, n. \tag{7.19}$$

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

$$q_k \sum_{j=1}^{n} p_j = q_k = p_k, \qquad k = 1, \cdots, n.$$

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 $p_1, \cdots, p_n$, and we have seen that $q_{jk}=p_k=q_k$, $(j, k=1, \cdots, n)$. Then if $(x_1, \cdots, x_n)$ is any solution of (7.14), it is linearly dependent on $(p_1, \cdots, p_n)$ 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

$$(7.20)\qquad \begin{aligned} x_1 + \cdots + x_n &= 0, \\ \sum_{j=1}^{n} x_j p_{jk} &= x_k, \qquad k = 2, \cdots, n, \end{aligned}$$

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

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

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 $(q_{jk}) = (p_k a_j)$ (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:†

$$(7.22)\qquad \begin{aligned} p_j > 0, \qquad j \leqq \alpha, \qquad 0 < \alpha \leqq n, \\ p_{\alpha+1} = p_{\alpha+2} = \cdots = p_n = 0. \end{aligned}$$

Since

$$\sum_{j=1}^{n} p_j p_{jk} = p_k, \qquad k = 1, \cdots, n,$$

* 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

$$p_{jk} = 0, \quad \text{if} \quad j \leqq \alpha, \qquad k > \alpha, \tag{7.23}$$

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 $p_j > 0$ if $j \in K$, that $\sum_{j \in K} p_j < 1$, and that

$$p_{jk} = 0, \qquad \begin{cases} j \in K, & k \notin K \\ j \notin K, & k \in K, \quad j \leqq \alpha.* \end{cases} \tag{7.24}$$

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 $\sum_{j \in K} p_j < 1$, R_2 cannot be absent. If no p_j 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,

$$\lim_{m \to \infty} p_{jk}^{(m)} = q_{jk} \tag{7.25}$$

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* ν *groups* $J_1, \cdots, J_\nu$, *of consecutive subscripts, such that* $p_{jk} = 0$, *unless* p_{jk}, *is in some one of the square matrices* $S_1, \cdots, S_\nu$ *for which* $j \in J_r$, $k \in J_{r+1}$, $(r < \nu)$, or $j \in j_\nu$, $k \in J_1$.†

$$\begin{array}{c} \\ J_1 \\ J_2 \\ J_3 \\ J_4 \end{array} \begin{array}{c} \begin{array}{cccc} J_1 & J_2 & J_3 & J_4 \end{array} \\ \left(\begin{array}{cccc} 0 & S_1 & 0 & 0 \\ 0 & 0 & S_2 & 0 \\ 0 & 0 & 0 & S_3 \\ S_4 & 0 & 0 & 0 \end{array}\right) \end{array}, \qquad \nu = 4,$$

Fig. 2

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

* Cf. Theorem 6.2. The Ω-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 $P(x_0; C\Lambda) = 0$, $(x_0 \in E)$, 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.

$p_1, \cdots, p_n$ 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

$$p_{rk} = \phi_{rk} p_k, \qquad k = 1, \cdots, n.$$

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

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

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 $q_{j1}, \cdots, q_{jn}$ is a set of absolute probabilities corresponding to the given matrix, that the absolute probabilities $p_1, \cdots, p_n$ are given by

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

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

$$p_{jk}^{(\mu+\nu)} = \sum_{l\in J} p_{jl}^{(\mu)} p_{lk}^{(\nu)} + \sum_{l\notin J} p_{jl}^{(\mu)} p_{lk}^{(\nu)}. \tag{7.27}$$

Then letting ν become infinite in (7.27), we obtain

$$\begin{aligned} \limsup_{m\to\infty} p_{jk}^{(m)} &\leqq \sum_{l\in J} p_{jl}^{(\mu)} + \sum_{l\notin J} p_{jl}^{(\mu)} q_{lk} \\ \liminf_{m\to\infty} p_{jk}^{(m)} &\geqq \sum_{l\notin J} p_{jl}^{(\mu)} q_{lk}, \end{aligned} \tag{7.28}$$

so that

$$\limsup_{m\to\infty} p_{jk}^{(m)} - \liminf_{m\to\infty} p_{jk}^{(m)} \leqq \sum_{l\in J} p_{jl}^{(\mu)} \qquad \mu = 1, 2, \cdots. \tag{7.29}$$

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

$$\limsup_{m\to\infty} p_{jk}^{(m)} = \liminf_{m\to\infty} p_{jk}^{(m)}$$

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 $x_1, \cdots, x_n$, not all 0, such that

$$\sum_{k=1}^{n} p_{jk}x_k = cx_j, \qquad j = 1, \cdots, n. \tag{7.30}$$

Then

$$\sum_{k=1}^{n} p_{jk}^{(m)} x_k = c^m x_j, \quad j = 1, \cdots, n; m = 1, 2, \cdots . \tag{7.31}$$

When m becomes infinite, the left side converges, to $\sum_{k=1} q_{jk}x_k$, 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 $p_1, \cdots, p_n$ be any set of absolute probabilities corresponding to the given matrix. We shall prove that the temporally homogeneous process defined in terms of $p_1, \cdots, p_n$ 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 $\psi(j) = \psi_j$, (6.18) becomes

$$\sum_{k=1}^{n} p_{jk}\psi_k = c\psi_j, \qquad |c| = 1, c \neq 1, \tag{7.32}$$

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 $j \notin J$, we have seen above that $p_{jk} = 0$ for $k \in J$, so that changing ψ_k for $k \in J$ does not affect (7.32). We shall attempt to re-define ψ_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 $\psi_j, j \in J$:

$$\sum_{k=1}^{n} p_{jk}\psi_k = c\psi_j, \qquad j \in J; \tag{7.33}$$

or, if we set

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

the system

$$\sum_{k\in J} p_{jk}\psi_k = c\psi_j - \alpha_j, \qquad j \in J, \tag{7.34}$$

that is,

$$\sum_{k\in J} (p_{jk} - c\delta_{jk})\psi_k = -\alpha_j, \qquad j \in J. \tag{7.35}$$

If these equations have a solution, this solution, when combined with the ψ_j's for $j \notin J$, 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

$$\sum_{k\in J} p_{jk}\psi_k = c\gamma_j, \qquad j \in J. \tag{7.36}$$

But then if γ_j is defined as 0 for $j \notin J$, the set $\gamma_1, \cdots, \gamma_n$ 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 $(p_{jk}^{(\nu+1)})$, $(p^{(2\nu+1)})$, $\cdots$ are of the same form, whereas the matrices $(p_{jk}^{(2)})$, $(p_{jk}^{(\nu+2)})$, $(p_{jk}^{(2\nu+2)})$, $\cdots$ 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), $\cdots$ instead of (J_1J_2), (J_2J_3), $\cdots$. 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 $p_1, \cdots, p_n$ of the absolute probabilities, so that (Theorem 6.3 (i)) there is a function $\psi(x_0)$ such that

$$\psi(x_1) = c\psi(x_0), \qquad |c| = 1, c \neq 1, \tag{7.37}$$

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 $a_1, a_2, \cdots$ be those non-zero values in the set $\psi_1, \cdots, \psi_n$

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 $\nu>1$, such that if $a_1, a_2, \cdots, a_\nu$ are chosen successively as described, so that $a_2=ca_1, \cdots, a_\nu=c^{\nu-1}a_1$, the next application of the algorithm will give $ca_\nu=a_1$, and hence $c^\nu=1$. Then c is a νth root of unity, and $1<\nu\leqq n$. If $n_{r-1}<x_0\leqq n_r$* (so that $\psi(x_0)=a_r$), then if $r<\nu$, it follows that $\psi(x_1)=a_{r+1}$ necessarily (if $r=\nu$, $\psi(x_1)=a_1$ necessarily), if we neglect x_1-sets of P-measure 0, that is, subscripts k for which $p_k=0$; and

$$p_{jk} = 0 \quad \text{if} \quad \begin{cases} j \in J_r, & k \notin J_{r+1}, \quad p_k > 0, \qquad r = 1, \cdots, \nu - 1, \\ j \in J_\nu, & k \notin J_1, \quad p_k > 0. \end{cases}$$

These equations describe the $(J_1+\cdots+J_\nu)^2$ matrix R_1. The fact that the Ω-set determined by the condition $x_1 \in J_1+\cdots+J_\nu$ 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 Ω-set, determined by the condition $x_0 \in J_1+\cdots+J_\nu$, 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, $\lim_{m\to\infty} p_{jk}^{(rm)}=p_k$, so that, according to (i) the matrix $(p_{jk}^{(r)})$ 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 $x_1, \cdots, x_n$ (not all 0) such that

$$\sum_{j=1}^{n} x_j p_{jk} = c x_k, \qquad k = 1, \cdots, n.$$

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 ν is chosen so that $c^\nu = 1$, $(\nu \leqq n)$, then $\sum_{j=1}^{n} x_j p_{jk}^{(\nu)} = c^\nu x_k = x_k$. Now the (uniquely defined) absolute probabilities $p_1, \cdots, p_n$ satisfy the equations

$$\sum_{j=1}^{n} p_j p_{jk}^{(\nu)} = p_k, \qquad k = 1, \cdots, n.$$

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

$$\sum_{j=1}^{n} p_j = 1, \qquad \sum_{j=1}^{n} x_j = 0.$$

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

$$(7.38) \qquad \int p(x)dx = 1, \qquad \int p(x)p(x, y)dx = p(y)$$

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* $m \geqq 1$, *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

$$\int_E p(x)dx = \lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} \int Q(dx) \int_E p^{(m)}(x, y)dy. \tag{7.39}$$

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 Bogolioùboff (I)).

THEOREM 7.5. (i) *Except possibly on an* Ω-*set of* P-*measure* 0,

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

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

(ii) *The limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^{N} \int_E p^{(m)}(x, y)dy = q(x, E) \tag{7.41}$$

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,

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

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 Bogolioùboff (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 Ω; 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 Ω.

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

$$\phi(x) > 0, \qquad \int \phi(x)dx = 1.*$$

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

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

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

$$(7.43) \qquad \frac{\nu + \mu}{\mu} \Pi^{(\nu+\mu)}(x, y) - \frac{\nu}{\mu} \Pi^{(\nu)}(x, y) = \int p^{(\nu)}(x, z)\Pi^{(\mu)}(z, y)dz.$$

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

$$(7.44) \qquad \lim_{N\to\infty} \int_E \Pi^{(N)}(x, y)dy = q(x, E), \qquad x \in CE_0,$$

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

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

* If $\int 1dx = \lambda < \infty$, we can take $\phi(x) \equiv 1/\lambda$. 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 $\sum_1^\infty E_j = X$ and that $\int_{E_j} 1dx = \lambda_j < \infty$. There is a sequence of positive numbers $\{\epsilon_n\}$ such that $\epsilon_0 + \sum_1^\infty \epsilon_n(\lambda_{n+1} - \lambda_n) = 1$, and we define $\phi(x)$ as ϵ_0 on E_1 and ϵ_n on $E_{n+1} - E_n$, for $n > 1$.

This implies that

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

Equation (7.45) means that some subsequence $\{\phi(x)\int_{E_0} p^{(a_j)}(x, y)\}$ of $\{\phi(x)\int_{E_0} p^{(m)}(x, y)dy\}$, when integrated over X, converges to 0. This implies that $\phi(x)\int_{E_0} p^{(a_j)}(x, y)dy$ 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

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

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

$$\begin{aligned} \liminf_{N\to\infty} \int_E \Pi^{(N)}(x, y)dy &= \liminf_{\mu\to\infty} \int p^{(\nu)}(x, z)dz \int_E \Pi^{(\mu)}(z, y)dy \\ &\geqq \liminf_{\mu\to\infty} \int_{CE_0} p^{(\nu)}(x, z)dz \int_E \Pi^{(\mu)}(z, y)dy. \end{aligned} \tag{7.47}$$

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

$$\liminf_{N\to\infty} \int_E \Pi^{(N)}(x, y)dy \geqq \int_{CE_0} p^{(\nu)}(x, z)q(z, E)dz, \qquad \nu = 1, 2, \cdots.$$

---

* Convergence in measure was defined and discussed by F. Riesz, Comptes Rendus de l'Académie des Sciences, Paris, vol. 148 (1909), pp. 1303–1305.

On the other hand,

$$
(7.48)\quad
\begin{aligned}
\limsup_{N\to\infty}\int_E \Pi^{(N)}(x,\,y)ay &= \limsup_{\mu\to\infty}\left\{\int_{E_0} p^{(\nu)}(x,\,z)dz\int_E \Pi^{(\mu)}(z,\,y)dy\right.\\
&\qquad\left. + \int_{CE_0} p^{(\nu)}(x,\,z)dz\int_E \Pi^{(\mu)}(z,\,y)dy\right\}\\
&\leqq \int_{E_0} p^{(\nu)}(x,\,z)dz + \int_{CE_0} p^{(\nu)}(x,\,z)q(z,\,E)dz,\\
&\qquad\qquad \nu = 1,\,2,\,\cdots.
\end{aligned}
$$

Then

$$
(7.49)\quad \limsup_{N\to\infty}\int_E \Pi^{(N)}(x,\,y)dy - \liminf_{N\to\infty}\int_E \Pi^{(N)}(x,\,y)dy \leqq \int_{E_0} p^{(\nu)}(x,\,z)dz,\qquad \nu = 1,\,2,\,\cdots.
$$

If ν is allowed to increase without limit through the sequence $\{b_j+1\}$ 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 μ, $p^{(\mu)}(x,\,y)$ 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 $f(z) = p^{(\mu)}(z,\,y)$, then

$$
\begin{aligned}
\lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}\int p^{(m)}(x,\,z)p^{(\mu)}(z,\,y)dz &= \lim_{N\to\infty}\frac{1}{N}\sum_{m=\mu+1}^{N+\mu} p^{(m)}(x,\,y)\\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N} p^{(m)}(x,\,y)
\end{aligned}
$$

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*

$$\int \psi(x)p(x, y)dx = \psi(y), \tag{7.50}$$

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*

$$p(x, y) = 0, \qquad x \in F_j, \qquad y \notin F_j, \qquad (j = 1, 2), \tag{7.51}$$

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

$$q(x, E) = \int_E p(y)dy. \tag{7.52}$$

(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

$$\int \psi(x)dx \int_E \Pi^{(N)}(x, y)dy = \int_E \psi(y)dy, \qquad N = 1, 2, \cdots;$$

and if $N \to \infty$, this becomes†

$$\int \psi(x)dx \int_E p(y)dy = \int_E \psi(y)dy. \tag{7.53}$$

If α is defined as $\int \psi(x)dx$, (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, $\psi(y) \equiv \alpha p(y)$, 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), (ii), (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 Bogolioùboff (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_{i1}, F_{i2}, of positive measure, such that $p(x, y)=0$ if $x \epsilon F_{ij}$, $y \notin F_{ij}$, if we neglect (x, y)-sets of (x, y)-measure 0. In the contrary case the function $p(x, y)$, considered only for $x \epsilon F_i$, $y \epsilon F_i$, is metrically transitive. Now it is readily seen that the uniform integrability condition prevents the existence of infinitely many X-measurable sets $F_1, F_2, \cdots$, of positive X-measure, such that if $x \epsilon F_i$, $y \notin F_i$, then $p(x, y)=0$ neglecting (x, y)-sets of (x, y)-measure 0. Hence there is at most a finite number μ of such sets, and $p(x, y)$ is metrically transitive when considered defined only for $x, y \epsilon F_i$. Let $p_i(x)$ be the corresponding uniquely defined absolute probability density, and define $p_i(x)=0$ if $x \notin F_i$. Then if $\rho_1, \cdots, \rho_\mu$ are non-negative numbers with sum 1, $p(x)=\sum_{i=1}^{\mu} \rho_i p_i(x)$ 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 $\sum_{i=1}^{n} \rho_i \int_{E \cdot E_i} p_i(x)dx$.

THEOREM 7.7. *The function $p(x, y)$ has no angle variables if and only if*

(i) *whenever $E \epsilon F_x$,*

$$\lim_{m\to\infty} \int_E p^{(m)}(x, y)dy = q(x, E) \tag{7.54}$$

exists for all $x \epsilon X$ (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 $E_1, \cdots, E_\nu$, $\nu>1$, of positive X-measure, such that (if we neglect an (x, y)-set of (x, y)-measure* 0),

$$p(x, y) = 0, \qquad \begin{cases} x \epsilon E_r, & y \notin E_{r+1}, \qquad r = 1, \cdots, \nu - 1, \\ x \epsilon E_\nu, & y \notin E_1. \end{cases} \tag{7.55}$$

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

$$\int \psi(y)p(x, y)dy = c\psi(x).$$

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

* Cf. Kryloff und Bogolioùboff (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 $E \epsilon F_x$ and for all x (except perhaps values in a set on which the integral of $p(x)$ vanishes),

$$\int_E p(x, y)dy = \int_E \phi(x, y)p(y)dy. \tag{7.56}$$

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

$$\int p(x)dx \int_{E_0} p(x, y)dy = \int_{E_0} p(y)dy = 0, \tag{7.57}$$

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

$$\begin{aligned} \phi(x, y) &= p(x, y)/p(y), & p(y) &> 0 \\ \phi(x, y) &= 0, & p(y) &= 0, \end{aligned} \tag{7.58}$$

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 $E_1, \cdots, E_\nu$ 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

$$\psi(x_1) = c\psi(x_0), \qquad |c| = 1, c \neq 1, \tag{7.59}$$

if we neglect an (x_0, x_1)-set of P-measure 0. Then

$$\int [\psi(x_1) - c\psi(x_0)]P(de_{01}) = \int\int p(x)p(x, y)[\psi(y) - c\psi(x)]dx\, dy = 0,$$

so that

$$p(x)p(x, y)[\psi(y) - c\psi(x)] = 0 \tag{7.60}$$

for almost all (x, y), and

$$p(x, y)[\psi(y) - c\psi(x)] = 0, \qquad x \notin E_0, \tag{7.61}$$

if we neglect (x, y)-sets of zero measure. Let ξ be a point not in E_0, such that

$$\int_{E_0} p(\xi, y)dy = p(\xi, y)[\psi(y) - c\psi(\xi)] = 0$$

(where the second is to hold for almost all values of y). This may exclude (cf. (7.5)) a ξ-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 $c\psi(\xi)$ on a set of positive P-measure. Now let a be any value, not equal 0, assumed by ψ on a set of positive P-measure. According to Theorem 6.3 (iii), c must be a root of unity. Let ν be the smallest exponent r for which $c^r = 1$. The function $\psi(x)$ takes on values $a, ca, \cdots, c^{\nu-1}a$ on subsets $E_1, \cdots, E_\nu$ 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 $x_0 \in E + \cdots + E_\nu$ is invariant up to a set of P-measure 0, implies the conditions of (7.55).

The set $E_1 + \cdots + E_\nu$ is one of the sets F (corresponding to invariant Ω-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 $E \in F_x$, $\lim_{m\to\infty} \int_E p^{(m)}(x, y)dy$ 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 $\lim_{m\to\infty} p^{(m)}(x, y)$ 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. BOGOLIOÙBOFF.
Cf. N. Kryloff and N. Bogolioùboff below.

P. J. DANIELL.
(I) *Integrals in an infinite number of dimensions*, Annals of Mathematics, (2), vol. 20 (1918–1919), pp. 281–288.

W. DOEBLIN and R. FORTET.
(I) *Sur deux notes de Mm. Kryloff et Bogolioùboff*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1937), pp. 1699–1701.

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
(I) *Compléments à la théorie des probabilités discontinues "en chaîne,"* Annali delle Università Toscane, Pisa, (2), vol. 2 (1933), pp. 131–164.
(II) *Sur l'allure asymptotique des densités itérées dans le problème des probabilités "en chaîne,"* Bulletin de la Société Mathématique de France, vol. 62 (1934), pp. 68–83.

M. FRÉCHET and J. HADAMARD
(I) Zeitschrift für angewandte Mathematik und Mechanik, vol. 13 (1933), pp. 92–97.

J. HADAMARD.
Cf. M. Fréchet and J. Hadamard.

E. HOPF.
(I) *On causality, statistics and probability*, Journal of Mathematics and Physics of the Massachusetts Institute of Technology, vol. 13 (1934), pp. 51-102.

B. HOSTINSKY.
(I) *Application du calcul des probabilités à la théorie du mouvement Brownien*, Annales de l'Institute Henri Poincaré, vol. 3 (1932), pp. 1–74.
(II) *Méthodes générales du calcul des probabilités*, Mémorial des Sciences Mathématiques, no. 52.

M. KAC.
(I) *Sur les fonctions indépendantes* I, Studia Mathematica, vol. 6 (1936), pp. 46–58.

A. KHINTCHINE.
(I) *Zur mathematischen Begründung der statistischen Mechanik*, Zeitschrift für angewandte Mathematik und Mechanik, vol. 13 (1933), pp. 101–103.

A. KOLMOGOROFF.
(I) *Über die analytische Methoden in der Wahrscheinlichkeitsrechnung*, Mathematische Annalen, vol. 104 (1931), pp. 415–458.
(II) *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 3.

N. KRYLOFF and N. BOGOLIOÙBOFF.
(I) *Sur les probabilités en chaîne*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1933), pp. 1386–1388.
(II) *Les propriétés ergodiques des suites des probabilitiés en chaîne*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1937), pp. 1454–1456.

P. LÉVY.
(I) *Théorie de l'Addition des Variables Aléatoires*, Paris, 1937.

Z. Łomnicki and S. Ulam.

(I) *Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités*, I *Variables indépendantes*, Fundamenta Mathematicae, vol. 23 (1934), pp. 237–278.

R. von Mises.

(I) *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Leipzig and Vienna, 1931.

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

(I) *Fourier Transforms in the Complex Domain*, American Mathematical Society Colloquium Publications, vol. 19, 1934.

V. Romanovsky.

(I) *Recherches sur les chaînes de Markoff*, Acta Mathematica, vol. 66 (1936), pp. 147–251.

S. Ulam.

Cf. Z. Łomnicki and S. Ulam.

N. Wiener.

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

University of Illinois,
Urbana, Ill.