

TOPICS IN THE THEORY OF MARKOFF CHAINS

BY

J. L. DOOB

Let $P(t)$: $(p_{ij}(t))$ be a matrix (finite- or infinite-dimensional), depending on $t > 0$, whose elements satisfy the following conditions

$$(0.1) \quad p_{ij}(t) \geq 0, \quad \sum_j p_{ij}(t) = 1, \quad P(s)P(t) = P(t)P(s) = P(s+t).$$

Then $p_{ij}(t)$ can be considered a transition probability of a Markoff chain: A system is supposed which can assume various numbered states, and $p_{ij}(t)$ is the probability that the system is in the j th state at the end of a time interval of length t , if it was in the i th state at the beginning of the interval. The present paper will be divided into two parts. In the first, the regularity properties of $P(t)$, and its asymptotic properties as $t \rightarrow 0$, $t \rightarrow \infty$ are studied. These problems have been solved in the finite-dimensional case by Doeblin⁽¹⁾. In the infinite-dimensional case new situations can arise, and the results are somewhat different. The method of approach is new, depending on two theorems (Theorems 2 and 3) concerning matrices whose elements are non-negative, and which have row sums less than or equal to 1. The method of approach can also be applied to the study of the asymptotic properties of the powers of a matrix of non-negative elements, with row sums 1. In the second part of the paper, the actual transitions connected with Markoff chains are investigated: That is, the properties of the function $\xi(t)$, the number of the state which the given system assumes at time t , are investigated. The continuity properties of $\xi(t)$ are analyzed, and related to the regularity properties of the $p_{ij}(t)$.

LEMMA 1. *Suppose that the function $f(t)$, defined for all $t > 0$, satisfies the functional equation*

$$(1.1) \quad f(s+t) = \sum_n g_n(s)h_n(t) \quad (s, t > 0),$$

where $g_n(s)$, $h_n(s)$ are defined for $s > 0$, where $h_n(s)$ is measurable, and where for each fixed s , if $0 < a < b$, the series converges uniformly for $a \leq t \leq b$. Then $f(t)$ is continuous, for all $t > 0$.

Presented to the Society, December 31, 1941; received by the editors May 24, 1941.

⁽¹⁾ Bulletin des Sciences Mathématiques, (2), vol. 62 (1938), pp. 21–32, and vol. 63 (1939), pp. 35–37. In the following, these papers will be referred to as Doeblin (I). Fréchet has discussed the solutions of (0.1) in great detail, in the finite-dimensional case, with full references to earlier authors in his book *Traité du Calcul des Probabilités et de ses Applications*, vol. 1, Part 3, Book 2, *Méthode des fonctions arbitraires* . . . , Paris, 1938.

It will be sufficient to prove that if a value t_0 of t is given, and if $\{\delta_j\}$ is a sequence of numbers approaching 0, then $f(t_0 + \delta_{a_j}) \rightarrow f(t_0)$, for some subsequence $\{\delta_{a_j}\}$ of $\{\delta_j\}$. By a theorem of Auerbach⁽²⁾, there is, corresponding to each $h_n(t)$, a subsequence $\{\delta_{a_j}\}$ of $\{\delta_j\}$, such that

$$(1.2) \quad \lim_{j \rightarrow \infty} h_n(t + \delta_{a_j}) = h_n(t)$$

for almost all t in the interval $0 < t < t_0$. There is then, using the diagonal process, a subsequence $\{\delta_{a_j}\}$ of $\{\delta_j\}$ such that (1.2) is true for all n , $0 < t < t_0$, except possibly on a t -set of measure 0. If $0 < t < t_0$, and if j is large,

$$(1.3) \quad f(t_0 + \delta_{a_j}) = \sum_n g_n(t_0 - t) h_n(t + \delta_{a_j}),$$

and if t is not in the exceptional set, (1.3) implies, when $j \rightarrow \infty$,

$$(1.4) \quad f(t_0 + \delta_{a_j}) \rightarrow \sum_n g_n(t_0 - t) h_n(t) = f(t_0)$$

(because of the uniform convergence of the series in (1.3) with respect to j), as was to be proved.

THEOREM 1. *If the matrix function $P(t)$ satisfies (0.1), the measurability of the $p_{ij}(t)$ implies their continuity.*

This follows at once from Lemma 1. It has been shown by Doeblin (I) and it will be a corollary of results to be proved below, that the $p_{ij}(t)$ satisfying (0.1) are always continuous if the matrix $P(t)$ is finite-dimensional, even if measurability is not assumed. The following example shows that there are non-measurable solutions of (0.1).

In this example, the $p_{ij}(t)$ take on only the values 0, 1, and $P(t)$ is a permutation matrix. Hamel has shown that there is a function $f(x)$, defined for all real x , taking on only rational values, and satisfying the functional equation⁽³⁾ $f(x+y) = f(x) + f(y)$. Let $\{r_n\}$ be an enumeration of all the rational numbers, and let T_s be the transformation of these numbers taking r_j into $r_j + f(s)$. The transformation can be represented by a matrix $P(s) = (p_{ij}(s))$, where $p_{ij}(s) = 1$ if $T_s r_i = r_j$, and $p_{ij}(s) = 0$ otherwise. Then evidently $P(s+t) = P(s)P(t)$, and (0.1) is satisfied. The functions $p_{ij}(t)$ are not measurable, since they obviously are not continuous.

The following theorem describes completely the solutions of (0.1) which are independent of t . It will be useful to weaken (0.1) slightly. The theorem is essentially known, at least in an indirect form⁽⁴⁾.

⁽²⁾ *Fundamenta Mathematicae*, vol. 11 (1928), pp. 196–197.

⁽³⁾ *Mathematische Annalen*, vol. 60 (1905), pp. 459–462. To ensure that Hamel's $f(x)$ take on only rational values, we can set, using his notation, $f(a) = 1$, $f(b) = \dots = 0$.

⁽⁴⁾ Cf., for example, K. Yosida and S. Kakutani, *Japanese Journal of Mathematics*, vol. 16 (1939), pp. 47–55.

THEOREM 2. Let $U: (u_{ij})$ be a matrix of elements satisfying the following conditions:

$$(2.1) \quad u_{ij} \geq 0, \quad \sum_j u_{ij} \leq 1, \quad U^2 = U.$$

Then the subscripts can be divided into mutually exclusive classes⁽⁶⁾ F, G_1, G_2, \dots such that

(a) $u_{ij} = 0$, if $j \in F$;

(b) there are positive numbers $u_i, j \notin F$, such that⁽⁶⁾

$$u_{ij} = \delta_{IJ} u_i, \quad (\text{if } i \in I, j \in J),$$

$$\sum_{j \in J} u_{ij} = \sum_{j \in J} u_j = 1 \quad (i \in I = J);$$

(c) there are non-negative numbers $\{\rho_{iJ}\}$ such that

$$u_{ij} = \rho_{iJ} u_j \quad (\text{if } i \in F, j \in J).$$

Conversely, if the u_{ij} satisfy (a), (b), (c), then (2.1) is true.

Suppose (2.1) is true. Define F as the set of integers j with $p_{ij} = 0$ for all i . Then (a) is true. Unless U is the null matrix, there will be subscripts not in F . The extreme members of the inequality

$$(2.2) \quad \sum_k u_{ik} = \sum_{j,k} u_{ij} u_{jk} = \sum_j u_{ij} \left(\sum_k u_{jk} \right) \leq \sum_j u_{ij}$$

are equal; so if $\sum_k u_{jk} < 1, u_{ij} = 0: j \in F$. Let ξ_1, ξ_2, \dots be any numbers satisfying the conditions:

$$(2.3) \quad \sum_j |\xi_j| < \infty, \quad \sum_i \xi_i u_{ij} = \xi_j \quad (\text{all } j).$$

Then $\xi_j = 0$ if $j \in F$. If G is the set of integers j for which $\xi_j > 0$,

$$(2.4) \quad \sum_{j \in G} \xi_j = \sum_{j \in G} \sum_i \xi_i u_{ij} \leq \sum_{i,j \in G} \xi_i u_{ij} \leq \sum_{i \in G} \xi_i,$$

and there is an impossible inequality unless $u_{ij} = 0$ whenever $\xi_i < 0, \xi_j > 0$. Let i, j be any two distinct integers not in F . Unless the i th and j th columns of U are proportional (neglecting elements in the columns whose first subscripts are in F), there are integers $r, s (\notin F)$, such that

⁽⁶⁾ The class F may be absent, or the G_r may be absent. The latter case will arise when and only when U is the null matrix.

⁽⁶⁾ In the following, capital letters I, J, K will be used to denote the G_r , and a subscript i will always belong to the class I , and so on, unless the contrary is explicitly stated. The notation δ_{IJ} is the usual Kronecker δ .

$$(2.5) \quad \begin{vmatrix} u_{ri} & u_{rj} \\ u_{si} & u_{sj} \end{vmatrix} \neq 0.$$

Then λ, μ can be chosen so that $\lambda u_{ri} + \mu u_{si} < 0$, $\lambda u_{rj} + \mu u_{sj} > 0$. Since $\xi_k = \lambda u_{rk} + \mu u_{sk}$ provides a solution of (2.3) with $\xi_i < 0$, $\xi_j > 0$, it follows that $u_{ij} = 0$ if (2.5) is true. The subscripts not in F fall into classes, G_1, G_2, \dots , putting subscripts in the same class if the corresponding partial columns are proportional; if $i, j (\notin F)$ are not in the same class, $u_{ij} = 0$. The elements u_{ij} with i in a class G_r determine a matrix of rank 1. The rows of this matrix are therefore proportional, in fact identical, since the row sums are 1. We can thus write $u_{ij} = \delta_{IJ} u_j$ ($i \in I, j \in J$), and

$$(2.6) \quad \sum_{j \in J} u_j = \sum_i u_{i0i} = 1 \quad (j_0 \in J).$$

If $i \in F, j \in J$,

$$(2.7) \quad u_{ij} = \sum_k u_{ik} u_{kj} = \left(\sum_{k \in J} u_{ik} \right) u_j = \rho_{iJ} u_j$$

where ρ_{iJ} is defined by the sum in the parentheses. If $j \notin F$, u_j cannot vanish, since the elements of the j th column ($j \notin F$) cannot all vanish. We have now shown that (2.1) implies (a), (b), (c). Conversely, if (a), (b), (c) are true, (2.1) can be checked at once.

THEOREM 3. *Let \mathfrak{M} be a set of matrices (finite- or infinite-dimensional) with non-negative elements, and row sums less than or equal to 1. Suppose that the matrices in \mathfrak{M} form a group \mathfrak{M}' . There is then a $U \in \mathfrak{M}$ (the identity in \mathfrak{M}') with $U^2 = U$. If U is the null matrix, U is the only matrix in \mathfrak{M} , and \mathfrak{M}' consists only of the identity. If U is not the null matrix, we shall use the notation of Theorem 2 to describe its elements. The group \mathfrak{M}' is always isomorphic to a permutation group acting on the G_v . If (p_{ij}) is a matrix of \mathfrak{M} , and if the corresponding permutation takes I_1 into $I_2^{(7)}$, then*

$$(3.1) \quad \begin{aligned} p_{ij} &= \delta_{I_2 J} u_j && (i \in I_1, j \in J), \\ p_{ri} &= \rho_{r I_1} u_i && (r \in F, i \in I_2), \\ p_{ij} &= 0 && (j \in F). \end{aligned}$$

Evidently if U is the null matrix, it is the only matrix in \mathfrak{M} , and \mathfrak{M}' consists only of the identity. We shall assume from now on that U is not the null matrix, and use the notation of Theorem 2. Suppose that $P: (p_{ij}) \in \mathfrak{M}$. Then since $P = PU = UP$,

(7) As usual, letters I, J refer to the G_v .

$$(3.2) \quad p_{ik} = \sum_j p_{ij} u_{jk} = \sum_j u_{ij} p_{jk}.$$

If $k \in F$, (3.2) shows that $p_{ik} = 0$, the last equation of (3.1). If $i \in I$, $k \in K$, (3.2) becomes

$$(3.2') \quad p_{ik} = \left(\sum_{r \in K} p_{ir} \right) u_k,$$

$$(3.2'') \quad p_{ik} = \sum_{r \in K} u_r p_{rk}.$$

According to (3.2'), p_{ik}/u_k depends only on i , K , and according to (3.2''), p_{ik} depends only on I , k . Then p_{ik}/u_k depends only on I , K :

$$(3.3) \quad p_{ik} = \sigma_{IK} u_k \quad (i \in I, k \in K).$$

There is a $P' : (p'_{ij})$ in \mathfrak{M} which is the inverse of P in \mathfrak{M}' . If we write $p'_{ik} = \sigma'_{IK} u_k$, for $i, j \in F$, the equation $U = P'P$ implies

$$(3.4) \quad u_{ik} = \delta_{IK} u_k = \sum_j \sigma'_{IJ} \sigma_{JK} u_k.$$

The σ_{IJ} , σ'_{IJ} are non-negative and

$$(3.5) \quad \sum_j \sigma'_{IJ} = \sum_k p'_{ik} \leq 1.$$

If $I = K$ in (3.4), we obtain

$$(3.6) \quad 1 = \sum_j \sigma'_{IJ} \sigma_{JI} \leq \sum_j \sigma'_{IJ} \leq 1.$$

There must be equality throughout in (3.6); therefore if $\sigma_{JI} < 1$, it follows that $\sigma'_{IJ} = 0$. The matrices (σ_{IJ}) , (σ'_{IJ}) play symmetric roles; so if $\sigma'_{JI} < 1$, it follows that $\sigma_{IJ} = 0$. Then if $\sigma_{IJ} < 1$, $\sigma'_{JI} = 0 < 1$; so $\sigma_{IJ} = 0$. Each element in the matrix (σ_{IJ}) is either 1 or 0, and by (3.6) there is a 1 in each row of (σ'_{IJ}) and therefore in each row of (σ_{IJ}) . If $\sigma_{I_1 I_2} = 1$, the matrix (σ_{IJ}) defines the permutation of the G , taking I_1 into I_2 . The matrix (σ'_{IJ}) defines the inverse of this permutation. Equation (3.3) becomes the first equation of (3.1), equation (3.2) implies the second equation of (3.1), and the third equation of (3.1) has already been verified. The equations of (3.1) induce an isomorphism between the permutations defined by the (σ_{IJ}) permutation matrices and \mathfrak{M}' .

COROLLARY 1. *Suppose in Theorem 3 that \mathfrak{M} contains its limit matrices⁽⁸⁾. Then the corresponding permutation group on the G , has the property that each G_j*

⁽⁸⁾ The matrices $\{M^{(n)} : (m_{ij}^{(n)})\}$ will be said to converge to $M : (m_{ij})$, $M^{(n)} \rightarrow M$, if $m_{ij}^{(n)} \rightarrow m_{ij}$ for all i, j . The limit matrices of \mathfrak{M} are matrices which are limits of convergent sequences of matrices in \mathfrak{M} .

can go only into a finite number of the G_ν . If in addition it is supposed that corresponding to each $A \in \mathfrak{M}$ and positive integer n there is a $B \in \mathfrak{M}$ such that $B^n = A$, then \mathfrak{M} consists of only a single matrix, of the type described in Theorem 2.

Suppose that \mathfrak{M} contains its limiting matrices, and that some G_ν , say G_α , goes into infinitely many G_ν under the permutations of the group. Then there is a limiting matrix (p_{ij}) of \mathfrak{M} such that $p_{ij} = 0$ if $i \in G_\alpha$. But a matrix with these rows of zeros cannot be in \mathfrak{M} , so G_α cannot have the supposed property. The first part of the corollary is thus proved. Now suppose both hypotheses of the corollary are satisfied. It will be sufficient to prove that the group of permutations on the G_ν is the identity. Let G_α be any G_ν . We have already shown that G_α can go only into a finite number of G_ν , say G_{a_1}, \dots, G_{a_j} , under the permutations of the group. The permutations then permute G_{a_1}, \dots, G_{a_j} among themselves, and any element of the group of permutations on G_{a_1}, \dots, G_{a_j} has order a factor of $j!$. But any element in this group of permutations is by hypothesis the $j!$ th power of some other element; it must therefore be the identity. Then $j=1$, and G_α is transformed into itself by every permutation of the group, as was to be proved.

COROLLARY 2. Any matrix function $P(t) : (p_{ij}(t))$ with measurable elements $p_{ij}(t)$ satisfying (0.1) for all t (including 0 and negative values) is independent of t : $P(t) \equiv U$, where U has the properties described in Theorem 2.

We can assume that some $P(t)$ is not the null matrix, or there would be nothing to prove. The matrices $P(t)$ form a family \mathfrak{M} satisfying the conditions of Theorem 3. Moreover each $p_{ij}(t)$ is continuous, if $t > 0$, by Theorem 1, and so for all t , from (0.1). Using the notation of Theorem 3, if $i \notin F$, $p_{ij}(t) = u_j$ or $p_{ij}(t) = 0$. Then if $i \in F$, $p_{ij}(t)$ is independent of t . This means that \mathfrak{M}' consists only of the identity, so $P(t)$ is independent of t : $P(t) \equiv U$. The example above shows that the measurability of the $p_{ij}(t)$ is a necessary part of the hypotheses.

THEOREM 4. If the $p_{ij}(t)$ satisfying (0.1) are continuous, then $\lim_{t \rightarrow 0} P(t) = U$ exists. The matrix U is a non-null matrix of the type described in Theorem 2, and ⁽⁹⁾

$$(4.1) \quad UP(t) \leq P(t)U = P(t).$$

(In the following we shall use the notation of Theorem 2.) Moreover

$$(4.2) \quad p_{ij}(t) = 0 \quad (j \in F).$$

There are continuous functions $\Pi_{IJ}(t)$, satisfying (0.1) and

⁽⁹⁾ An inequality between two matrices is defined to mean the same inequality between their corresponding elements.

$$(4.3) \quad \lim_{t \rightarrow 0} \Pi_{IJ}(t) = \delta_{IJ}$$

such that

$$(4.4) \quad p_{ij}(t) = \Pi_{IJ}(t)u_j \quad (i \in I, j \in J).$$

There are continuous functions $\Pi_{iJ}(t)$ ($i \in F$) such that⁽¹⁰⁾

$$(4.5) \quad \begin{aligned} p_{ij}(t) &= \Pi_{iJ}(t)u_j & (i \in F, j \in F), \\ \Pi_{iK}(t) &\geq \sum_J \rho_{iJ} \Pi_{JK}(t). \end{aligned}$$

Conversely, if the $p_{ij}(t)$ satisfy (0.1) and if $\lim_{t \rightarrow 0} P(t)$ exists, the $p_{ij}(t)$ are continuous.

Neglecting subscripts in F , this theorem reduces the study of $P(t)$ to that of $(\Pi_{IJ}(t))$ in which case the limit matrix ($t \rightarrow 0$) is the identity.

Let $U: (u_{ij})$, $U': (u'_{ij})$ be limiting matrices of $P(t)$, $t \rightarrow 0$. Then (0.1) implies (4.1). The equal i th row sums in (4.1) are

$$(4.6) \quad \sum_j p_{ij}(t) \sum_k u_{jk} = \sum_k p_{ik}(t) = 1.$$

Since the row sums of U are less than or equal to 1, (4.6) implies that if $\sum_k u_{jk} < 1$, $p_{ij}(t) = 0$. Then in this case $u_{ij} = u'_{ij} = 0$ also. It follows from (4.1) that

$$(4.7) \quad \sum_j u'_{ij} u_{jk} \leq u'_{ik} \quad (U'U \leq U').$$

Summing over k , since $u'_{ij} = 0$ if $\sum_k u_{jk} < 1$, we see that both sides of (4.7) have sum $\sum_k u'_{ik}$; so there is equality in (4.7):

$$(4.7') \quad U'U = U'.$$

Replacing U by U' in the inequality $UP(t) \leq P(t)$, and letting t approach 0 in such a way that $P(t) \rightarrow U$, we obtain

$$(4.8) \quad U'U \leq U.$$

Then combining (4.7') and (4.8), we have $U' \leq U$, and by symmetry $U \leq U'$; so $U = U'$. There is thus only one limiting matrix $U: P(t) \rightarrow U$. Since equation (4.7') becomes $U^2 = U$, Theorem 2 is applicable. In the following, we shall use the notation of that theorem. If $k \in F$, $u_{jk} = 0$; therefore (using (4.1)) $p_{ik}(t) = 0$ also, for all i . Then U is not the null matrix. If $i, k \in F$, (4.1) implies

$$(4.9) \quad \left(\sum_{j \in K} p_{ij}(t) \right) u_k = p_{ik}(t) \quad (k \in K),$$

⁽¹⁰⁾ If the G_v contain only one subscript each, so that $p_{ij}(t) \rightarrow \delta_{ij}$ ($t \rightarrow 0$) if $i \in F$, then we can read $p_{ij}(t)$ for $\Pi_{IJ}(t)$, $p_{ij}(t)$ for $\Pi_{iJ}(t)$ throughout.

and

$$(4.10) \quad \sum_{j \in I} u_j p_{jk}(t) \leq p_{ik}(t) \quad (i \in I).$$

There is equality in (4.10) (and we shall refer to it as if it were so written), because summing over k gives 1 on both sides. Equations (4.9) and (4.10) imply that if $i, k \notin F$, $p_{ik}(t)/u_k$ depends only on I, K : $p_{ik}(t) = \Pi_{IK}(t)u_k$. Evidently the matrix function $(\Pi_{IJ}(t))$ satisfies (0.1), and (4.3) is true. If $i \in F$, and $k \notin F$, (4.1) implies

$$(4.11) \quad \sum_J \rho_{iJ} \Pi_{JK}(t) \leq \left(\sum_{j \in K} p_{ij}(t) \right) u_k = p_{ik}(t),$$

so that if $\Pi_{iK}(t)$ is defined as the parentheses in (4.11), (4.5) follows at once. Conversely, if $P(t) \rightarrow U$ ($t \rightarrow 0$),

$$(4.12) \quad \lim_{t \downarrow 0} P(s+t) = P(S)U.$$

The function $p_{ij}(t)$ having a right-hand limit for all t has at most denumerably many discontinuities, is therefore measurable, and continuous (Theorem 1).

THEOREM 5. *Let α be a given subscript. Then if $P(t)$ satisfies (0.1), $p_{\alpha j}(t)$ will be continuous and $\lim_{t \rightarrow 0} p_{\alpha j}(t)$ will exist, for all j , if $\sum_i p_{\alpha i}(t)$ converges uniformly in some interval $0 < t < t_0$.*

Doebelin (I) proved that if $P(t)$ satisfies (0.1) and is finite-dimensional, then the $p_{ij}(t)$ are continuous and have unique limits as $t \rightarrow 0$. This fact which evidently is a consequence of Theorem 5, can be proved directly as follows. Let \mathfrak{M} be the set of limiting matrices of $P(t)$, $t \rightarrow 0$. Then \mathfrak{M} satisfies the conditions of Theorem 3, Corollary 1, so \mathfrak{M} contains only a single matrix U . It follows that $P(t) \rightarrow U$, and the $p_{ij}(t)$ are then continuous, by Theorem 4.

Proof of Theorem 5. Let G be the set of subscripts α with the property described in the theorem. The equation $P(s)P(t) = P(t)P(s)$ implies that if $A : (a_{ij})$ is a limiting matrix of $P(t)$, $t \rightarrow 0$, then

$$(5.1) \quad \sum_j a_{ij} p_{jk}(t) \leq \sum_i p_{ij}(t) a_{ik}.$$

If $i \in G$, then $\sum_j a_{ij} = 1$, and the sum over k on the left is 1, so that on the right is also 1. Then there is equality in (5.1):

$$(5.1') \quad \sum_j a_{ij} p_{jk}(t) = \sum_i p_{ij}(t) a_{ik} \quad (i \in G).$$

If $\sum_k a_{jk} < 1$, then $p_{ij}(t) = 0$, or the sum over k on the right in (5.1') would not be 1. If $j \notin G$, we can find an A with $\sum_k a_{jk} < 1$, whence it follows that

$p_{ij}(t) = 0$ if $i \in G, j \notin G$. Let $P'(t)$ be the matrix obtained by dropping all elements of $P(t)$ with a subscript not in G . Then $P'(t)$ satisfies (0.1) and has the property that any limiting matrix ($t \rightarrow \infty$) has row sums 1. The proof given above of Doebelin's result goes through word for word, applied to $P'(t)$. We have thus proved that $P_{ij}(t)$ is continuous, and $\lim_{t \rightarrow \infty} p_{ij}(t)$ exists, if $\alpha \in G$, and in addition that $p_{ij}(t) \equiv 0$ if $\alpha \in G, j \notin G$.

We now turn to an examination of the limiting matrices of $P(t)$, as $t \rightarrow \infty$.

THEOREM 6. Define the matrix $U: (u_{ij})$ by

$$(6.1) \quad \liminf_{t \rightarrow \infty} p_{ij}(t) = u_{ij}.$$

Then

(a) U is a limiting matrix of $P(t)$, as $t \rightarrow \infty$; U has the properties described in Theorem 2, and $P(t)U = UP(t) = U$;

(b) (6.1) can be sharpened to

$$(6.1') \quad \lim_{t \rightarrow \infty} p_{ij}(t) = u_{ij}$$

if i is a subscript such that $\sum_j u_{ij} = 1$ ⁽¹¹⁾.

(c) Using the notation of Theorem 2, and assuming that U is not the null matrix,

$$(6.2) \quad \begin{aligned} p_{ij}(t) &= 0 && (i \in I, j \notin I), \\ \sum_{r \in J} u_r p_{ri}(t) &= u_i && (j \in J), \\ \sum_{j \in K} p_{ij}(t) + \sum_{j \in F} p_{ij}(t) \rho_{jK} &= \rho_{iK} && (i \in F). \end{aligned}$$

If $i \in F$ or if $j \in F$ (6.1') is true. If $i \in F, p_{ij}(t)$ is continuous, and $\lim_{t \rightarrow \infty} p_{ij}(t)$ exists. Moreover

$$(6.3) \quad \lim_{t \rightarrow \infty} \sum_{r \in K} p_{ir}(t) = \rho_{iK}, \quad \lim_{t \rightarrow \infty} \sum_{j \in F} p_{ij}(t) \rho_{jK} = 0 \quad (i \in F).$$

The fact that if $P(t)$ is finite-dimensional (6.1') is always true, which follows from Theorem 6, can be proved directly as follows. The set of limiting matrices (in this case) of $P(t), t \rightarrow \infty$, is seen at once to have the properties required in Theorem 3, Corollary 1, so there is only one limiting matrix $U: P(t) \rightarrow U$. This argument breaks down in the infinite-dimensional case, in which a more detailed analysis is necessary.

Let \mathfrak{L} be the class of limiting matrices (a_{ij}) of $P(t)$, as $t \rightarrow \infty$. Then \mathfrak{L} includes all its limit matrices. This implies that \mathfrak{L} contains one or more matrices minimizing $\sum_{i,j} 2^{-i} a_{ij}$. Let \mathfrak{M} be the class of these minimal matrices. We shall show that \mathfrak{M} contains only one matrix, U , defined by (6.1). The proof will be carried through in several steps.

⁽¹¹⁾ It follows that if the matrices are finite-dimensional, (6.1') is true for all i, j , a fact due to Doebelin (I).

(i) If $A \in \mathfrak{L}$, then $P(t)A \in \mathfrak{L}$.

This can be deduced at once from (0.1).

(ii) If $A \in \mathfrak{M}$, then $AP(t) = P(t)A \in \mathfrak{M}$.

This follows, for (0.1) implies that if $A : (a_{ij}) \in \mathfrak{M}$, there is a $B : (b_{ij}) \in \mathfrak{L}$, depending on t and on A , such that

$$(6.4) \quad \sum_i b_{ij} p_{ik}(t) \leq \sum_j p_{ij}(t) b_{jk} = a_{ik}.$$

Summed over k , this means that

$$(6.5) \quad \sum_j b_{ij} \leq \sum_k a_{ik} \quad (i = 1, 2, \dots)$$

and only equality is possible, since $A \in \mathfrak{M}$. Then (6.4) becomes the equality $BP(t) = P(t)B = A$. Therefore,

$$AP(t) = P(t)BP(t) = P(t)A.$$

By (i), $P(t)A \in \mathfrak{L}$, and summing over the i th row of $P(t)A = AP(t)$ gives $\sum_j a_{ij}$, so $P(t)A \in \mathfrak{M}$, since $A \in \mathfrak{M}$.

(iii) If $A, B \in \mathfrak{M}$, then $AB = BA \in \mathfrak{M}$.

By (ii), if $A \in \mathfrak{M}$, it follows that $AP(t) \in \mathfrak{M}$. If $B \in \mathfrak{M}$, AB is a limiting matrix of $AP(t)$, $t \rightarrow \infty$, so $AB \in \mathfrak{M}$ since \mathfrak{M} is closed. Moreover by (ii), $AP(t) = P(t)A$, so $AB \geq BA$. By symmetry, the reverse inequality is also true; so $AB = BA$.

(iv) If $A \in \mathfrak{L}$ there is an $A' \in \mathfrak{M}$ with $A' \leq A$.

For if $A \in \mathfrak{L}$, there is, using (0.1), a $B \in \mathfrak{M}$ and a $C \in \mathfrak{L}$ such that $A \geq BC$, and since $BC \in \mathfrak{M}$ (as a limit of $BP(t) \in \mathfrak{M}$) this is the desired inequality.

(v) If $A, B \in \mathfrak{M}$, there is a $C \in \mathfrak{M}$ with $A = BC$.

We see this, for there is certainly, using (0.1), a $C \in \mathfrak{L}$ with $A \geq BC$. As we have seen, $BC \in \mathfrak{M}$, so there must be equality.

(vi) If $A \in \mathfrak{M}$, and if n is any positive integer, there is a $B \in \mathfrak{M}$ such that $A = B^n$.

For, since $P(t/n)^n = P(t)$, if $A \in \mathfrak{M}$, there is a $B_1 \in \mathfrak{L}$ such that $B_1^n \leq A$. By (iv) there is then a $B \in \mathfrak{M}$ with $B^n \leq B_1^n \leq A$. To show that there must be equality, we need only show that $B^n \in \mathfrak{L}$. Since $B \in \mathfrak{M}$, $BB = B^2 \in \mathfrak{M}$ by (iii). Then $BB^2 = B^3 \in \mathfrak{M}$, and so on.

Now (iii) and (v) imply that the matrices of \mathfrak{M} form a commutative group. The fact that \mathfrak{M} is closed and that (vi) is true shows that \mathfrak{M} has the properties required in Theorem 3, Corollary 1. There can therefore be only a single matrix $U : (u_{ij})$ in \mathfrak{M} , and U has the properties described in Theorem 2. Because of (iv), $\liminf_{t \rightarrow \infty} p_{ij}(t) = u_{ij}$. From now on we shall assume the notation of Theorem 2. The equality $P(t)U = UP(t) = U$ follows from (ii). If $\sum_j u_{aj} = 1$, no limiting value of $p_{aj}(t)$, $t \rightarrow \infty$, can be greater than u_{aj} , or there would be a limiting row having a sum greater than 1. Then if $\sum_j u_{aj} = 1$, $\lim_{t \rightarrow \infty} p_{aj}(t) = u_{aj}$, for all j . In particular, (6.1') is true if $i \notin F$. The equations of (6.2) are equiva-

lent to the equations $P(t)U = UP(t) = U$. If (a_{ij}) is a limiting matrix of $P(t)$ as $t \rightarrow 0$, (6.2) implies that

$$(6.6) \quad \sum_{r \in J} u_r a_{rj} = u_j \quad (j \in J).$$

Summing (6.6) over $j \in J$ we obtain

$$(6.7) \quad \sum_{r \in J} u_r \left(\sum_{j \in J} a_{rj} \right) = \sum_{j \in J} u_j = 1.$$

Then $\sum_j a_{rj} = 1$ if $r \in J$. This implies that $\sum_j p_{rj}(t)$ converges uniformly in some interval $0 < t < t_0$; so according to Theorem 5, $p_{rj}(t)$ is continuous, and has a unique limit as $t \rightarrow 0$, if $r \in J$. Then this is true for any subscript $r \notin F$. As $t \rightarrow \infty$ in the last equation of (6.2) the first sum has an inferior limit greater than or equal to ρ_{iK} . Then there must actually be convergence; the first equation of (6.3) is true. The second sum in the last equation of (6.2) must then approach 0; (6.3) is true. Equation (6.3) is impossible, since $\liminf_{t \rightarrow \infty} p_{ik}(t) = \rho_{iK} u_k$, ($i \in F$, $k \in K$) unless $p_{ik}(t) \rightarrow \rho_{iK} u_k$; so (6.1') is true if $j \notin F$. The proof of the theorem is now complete.

Regularity hypotheses imposed on the probability matrices can be used to simplify the above results. Thus suppose that there is a value t_0 of t such that $\sum_j p_{ij}(t_0)$ converges uniformly in i . It follows readily that $\sum_j p_{ij}(t)$ converges uniformly in i and $t \geq t_0$. This means that any limiting matrix of $P(t)$, $t \rightarrow \infty$, has row sums 1, so $P(t) \rightarrow U$, by Theorem 6. A less strong condition is that there be a value t_0 of t , a positive integer N and a positive ϵ such that $\sum_{j \leq N} p_{ij}(t_0) \geq \epsilon$ for all i . It follows readily that the same inequalities hold for $t \geq t_0$. Then

$$(6.8) \quad \sum_{k \leq N} u_{jk} \geq \epsilon;$$

so there can only be a finite number of G_v , and U cannot be the null matrix. Also if $j \in F$, (6.8) becomes

$$(6.8') \quad \sum_{k \leq N} \rho_{jK} u_k \geq \epsilon \quad (k \in K).$$

Then some $\rho_{jK} > 0$ for each $j \in F$, so by (6.3), $\lim_{t \rightarrow \infty} p_{ij}(t) = 0$, if $i \in F$, $j \in F$. Thus in this case also, $P(t) \rightarrow U$, as $t \rightarrow \infty$. The fact that $P(t) \rightarrow U$ under the above hypotheses can also be derived using general theorems of Doeblin⁽¹²⁾ or of Kryloff and Bogoliouboff⁽¹³⁾.

If there is a set of non-negative numbers p_1, p_2, \dots such that

$$(6.9) \quad \sum_i p_i p_{ij}(t) = p_j \quad (\text{all } j), \quad \sum_i p_i = 1,$$

⁽¹²⁾ Thesis, Paris, 1938, pp. 105-109.

⁽¹³⁾ Paris, Comptes Rendus de l'Académie des Sciences, vol. 204 (1937), pp. 1454-1456.

the set p_1, \dots will be called a set of (stationary) absolute probabilities. The number p_j can be considered as the probability of being in the j th state at time t . Any linear combination of absolute probabilities with non-negative coefficients is also a set of absolute probabilities, or proportional to a set. If U is defined as in Theorem 6, the second set of equations of (6.2) states that the i th row of U , if $i \notin F$, is a set of absolute probabilities. If $i \in F$, the i th row of U is a linear combination (coefficients ρ_{iK}) of the rows of elements with first subscripts not in F . Then every row of U is a set of absolute probabilities, or proportional to a set (if the row sum is less than 1). Moreover (6.9) implies that $\sum_i p_i u_{ij} = p_j$; so any set of absolute probabilities is a linear combination (non-negative coefficients) of rows of U . The states with subscripts in F then always have probability 0, regardless of the absolute probabilities. One simple consequence of these remarks is that if there is a solution to (6.9), U cannot be identically 0, and some row of U is also a solution of (6.9); there is a solution of (6.9) determined by the equations $p_j = \lim_{t \rightarrow \infty} p_{\alpha_j}(t)$, α fixed, not in F .

THEOREM 7. *Suppose that the $p_{ij}(t)$ satisfying (0.1) are continuous. Then if U is defined as in Theorem 6,*

$$(7.1) \quad \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T p_{ij}(t) dt = u_{ij}$$

for all i, j .

Let $Q(T)$ be the matrix with general element $q_{ij}(T)$:

$$q_{ij}(T) = \frac{1}{T} \int_0^T p_{ij}(t) dt.$$

Since

$$(7.2) \quad \sum_j q_{ij}(T) p_{jk}(t) = \sum_j p_{ij}(t) q_{jk}(T) = \frac{1}{T} \int_t^{T+t} p_{ik}(s) ds,$$

if U' is a limiting matrix of $Q(T)$, $T \rightarrow \infty$, it follows that

$$U'P(t) \leq P(t)U' = U'.$$

Since the row sums of $U'P(t)$ are the same as those of U' , there must be equality:

$$(7.3) \quad U'P(t) = P(t)U' = U'.$$

According to Theorem 6,

$$(7.4) \quad UP(t) = P(t)U = U.$$

It follows from (7.3) and (7.4) that

$$(7.5) \quad \begin{aligned} UU' &\leq U'U = U', \\ U'U &\leq UU' = U. \end{aligned}$$

Then $U = UU' \leq U' = U'U \leq U$, $U = U'$, as was to be proved.

The following is a simple example illustrating the fact that U in Theorems 6, 7 may be the null matrix. Let $p_{ij}(t) = 0$ if $j < i$, and otherwise define $p_{ij}(t)$ by

$$(7.6) \quad p_{ij}(t) = \frac{t^{j-i}}{(j-i)!} e^{-t}.$$

Evidently $p_{ij}(t) \rightarrow 0$, as $t \rightarrow \infty$. There can be no stationary absolute probabilities in this case.

In examining the successive transitions of the system, we shall assume that the system is initially in a state α , where α will be held fixed throughout the discussion. Let $\xi(t)$ be the number of the state assumed by the system at time t . Then $\xi(t)$, for each fixed value of t , is a chance variable: $\xi(0) \equiv \alpha$; $\xi(t) = j$ with probability $p_{\alpha j}(t)$ if $t > 0$. To discuss the continuity properties of $\xi(t)$ in t we shall assume a minimum of regularity properties of $P(t)$, to which we shall be led in a natural way. In order to discuss the probability measures under consideration, we must, as usual, find a space Ω^* of points ω , a measure defined on Ω^* , and a one-parameter family of measurable functions $x_t(\omega)$, $0 \leq t < \infty$, such that the probability relations of the chance variables $\{\xi(t)\}$ become measure relations of the functions $\{x_t(\omega)\}$. Let Ω^* be the space of all functions $x(t)$, $0 \leq t < \infty$, taking on the integral values used in the subscripts of $P(t)$. A probability measure on Ω^* is defined as follows. If $0 = t_0 < t_1 < \dots < t_n$, the conditions

$$(8.1) \quad x(t_j) = \nu_j \quad (j = 0, \dots, n)$$

determine a subset of Ω^* and the measure of this subset is defined by

$$(8.2) \quad \begin{aligned} P^* \{x(t_j) = \nu_j, j = 0, \dots, n\} \\ = \delta_{\alpha\nu_0} p_{\alpha\nu_1}(t_1) p_{\nu_1\nu_2}(t_2 - t_1) \dots p_{\nu_{n-1}\nu_n}(t_n - t_{n-1}). \end{aligned}$$

By a theorem of Kolmogoroff⁽¹⁴⁾, a completely additive measure function is determined on Ω^* by these conditions. Let $x_s(\omega)$ be the function of $\omega: x(t)$ which takes on the numerical value $f(s)$ if ω is the function $f(t)$. Then the probability relations of the chance variables $\{\xi(t)\}$ become measure relations of the measurable functions $\{x_t(\omega)\}$:

$$P^* \{x_t(\omega) = j\} = p_{\alpha j}(t)$$

and so forth. We shall sometimes write $x(t)$ instead of $x_t(\omega)$, so that " $x(t)$ " can mean, for example: (a) a point ω of Ω^* ; (b) a function $x_t(\omega)$ of ω ; (c) a num-

⁽¹⁴⁾ *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik, vol. 2, no. 4, pp. 24-30. The fact that our functions assume only integral values, whereas those of Kolmogoroff assume all values necessitates only trivial changes in the proof.

ber, the value of the function $x(t)$ at the point t . When there is any danger of confusion, the proper meaning will be explicitly stated. The function $x_i(\omega)$ is automatically defined on any subset of Ω^* , and it is usually desirable to restrict ω to be in a subset Ω of Ω^* , of outer measure 1, defining a P -measure on Ω by setting $P(\Lambda^* \cdot \Omega) = P^*(\Lambda^*)$ for any P^* -measurable set⁽¹⁵⁾ Λ^* . The probability relations of the chance variables $\{\xi(t)\}$ now become the measure relations of the functions $x_i(\omega)$, $\omega \in \Omega: P\{x_s(\omega) = j\} = p_{\alpha j}(s)$ and so on⁽¹⁶⁾. It has been shown⁽¹⁷⁾ that if any P^* -measure is given, there always corresponds an everywhere dense denumerable sequence of real numbers $R: \{r_j\}$ such that if I is any open interval, and if $s \in I$,

$$(8.3) \quad P^* \left\{ \text{G.L.B.}_{r_j \in I} x(r_j) \leq x(s) \leq \text{L.U.B.}_{r_j \in I} x(r_j) \right\} = 1^{(18)}.$$

It has been shown⁽¹⁹⁾ that Ω can be chosen to consist of all (possibly infinite-valued) functions $x(t)$ which satisfy the relation

$$(8.4) \quad \liminf_{r_j \rightarrow t} x(r_j) \leq x(t) \leq \limsup_{r_j \rightarrow t} x(r_j)$$

for all $t \in R$. Then if this is done,

$$(8.5) \quad \text{G.L.B.}_{r_j \in I} x(r_j) \leq x(s) \leq \text{L.U.B.}_{r_j \in I} x(r_j)$$

for all $s \in I$, $\omega: x(t)$ in Ω , sharpening (8.3). Such a space Ω is called quasi-separable, and the process: that is, the combination of Ω with its P -measure, is called a quasi-separable process.

A measure can be defined on the space $T \times \Omega$ of couples (t, ω) , as the product of Lebesgue measure on the t -axis and P -measure on ω . The process is called measurable if the function $x_i(\omega)$ is (t, ω) -measurable. The P^* -measure is then said to determine a measurable process. This hypothesis on the P^* -measure is certainly a minimum hypothesis. On the other hand, there are natural analytic restrictions on the $p_{ij}(t)$. Let G_α be the set of subscripts j such that $p_{\alpha j}(t) \neq 0$. Only the subscripts in G_α need be considered in analyzing the transitions of the system, supposed initially in state α . It follows readily from (0.1) that $p_{ij}(t) \equiv 0$ if $i \in G_\alpha, j \notin G_\alpha$. The matrix $P_\alpha(t): p_{ij}(t)$ with $i, j \in G_\alpha$ then satisfies (0.1), and it is this matrix $P_\alpha(t)$ which is essential to the discussion. The

⁽¹⁵⁾ Cf. Doob, these Transactions, vol. 42 (1937), pp. 108-110.

⁽¹⁶⁾ $P\{x_s(\omega) = j\}$ is to be interpreted as the Ω -measure of the set of all functions $x(t)$ in Ω for which $x(s) = j$.

⁽¹⁷⁾ Doob, these Transactions, vol. 47 (1940), p. 467.

⁽¹⁸⁾ This equality holds for each fixed s . The ω -set $\{x(s) \leq \phi, s \in I\}$, ϕ a P^* -measurable function, is not P^* -measurable, so each value of s must be considered separately in (8.3), or in probability relations of similar type. The subspace Ω is introduced in order to avoid this necessity.

⁽¹⁹⁾ Op. cit., pp. 468-469.

natural analytic hypotheses on $P_\alpha(t)$ would include the measurability of its elements. This, by Theorem 1, implies their continuity, and then (Theorem 4), $\lim_{t \rightarrow 0} P_\alpha(t) = U$ exists. The matrix U is the first determining factor of the regularity of the process. It is natural to suppose that it is the identity matrix. A glance at Theorem 4 shows that no other hypothesis can possibly be compatible with any sort of continuity in the transitions of the system.

These considerations lead to the following formulation of a natural hypothesis to be imposed on the matrix function $P_\alpha(t)$. We shall denote as hypothesis H_α the hypothesis that the system is initially in state α , and that if $i \in G_\alpha$, $\lim_{t \rightarrow 0} p_{ii}(t) = 1$. Then $\lim_{t \rightarrow 0} p_{ij}(t) = \delta_{ij}$ ($i \in G_\alpha$). Since $P_\alpha(t)$ satisfies (0.1), the $p_{ij}(t)$ ($i, j \in G_\alpha$) will be continuous (Theorem 4). Moreover $p_{ij}(t) \equiv 0$ if $i \in G_\alpha, j \notin G_\alpha$. Then if hypothesis H_α is true, and if $i \in G_\alpha, p_{ij}(t)$ is continuous for all j , and $\lim_{t \rightarrow 0} p_{ij}(t) = \delta_{ij}$. Hypothesis H_α implies the continuity of $p_{\alpha i}(t)$ for all j , even though α may not be in G_α . In fact the equation

$$p_{\alpha i}(s+h) = \sum_{i \in G_\alpha} p_{\alpha i}(s) p_{ij}(h)$$

shows that $p_{\alpha i}(t)$ is continuous for $t > s$, and therefore for all t . If $i \in G_\alpha$, then $p_{ii}(t) > 0$ for all t , and if $i = \alpha, j \in G_\alpha$ or if $i, j \in G_\alpha$ then $p_{ij}(t) = 0$ at most on a finite interval $0 < t \leq t_0$ (depending on i, j). The first fact follows from the inequality $p_{ii}(t) \geq p_{ii}(t/n)^n, n = 1, 2, \dots$, since $\lim_{n \rightarrow \infty} p_{ii}(t/n) = 1$. The second fact follows from the inequality $p_{ij}(t+h) \geq p_{ij}(t)p_{ij}(h)$ which implies that if $p_{ij}(t') > 0$, then $p_{ij}(t) > 0$, for $t > t'$.

The following theorem shows the relations between various hypotheses it would be natural to assume.

THEOREM 8. *Suppose that $P^*\{x(0) = \alpha\} = 1$. Then the following three conditions on P^* -measure are equivalent.*

- (i) *The P^* -measure determines a measurable process.*
- (ii) *Hypothesis H_α is satisfied.*
- (iii) *For every $\tau > 0$,*

$$(8.6) \quad \lim_{t \rightarrow \tau} P^*\{x(t) = x(\tau)\} = 1.$$

In the usual language of measure theory, (8.6) states that $x_t(\omega) \rightarrow x_\tau(\omega)$ in measure. We shall prove a much stronger result below, Theorem 11. To prove Theorem 8, we prove that (i) implies (ii), (ii) implies (iii), and (iii) implies (i).

Proof that (i) implies (ii). Suppose that P^* -measure determines a measurable process. Then it follows⁽²⁰⁾ that for fixed $h > 0, \epsilon > 0, P^*\{|x(t+h) - x(t)| > \epsilon\}$ is Lebesgue measurable in t , and (as $h \rightarrow 0$) goes to 0 in measure on any finite t -interval. If $\epsilon < 1$,

$$(8.7) \quad P^*\{|x(t+h) - x(t)| > \epsilon\} = \sum_{i \in G_\alpha} p_{\alpha i}(t) [1 - p_{ii}(h)].$$

⁽²⁰⁾ Doob, these Transactions, vol. 42 (1937), p. 117.

Since the quantity in (8.7) goes to 0 as $h \rightarrow 0$ in measure on every finite t -interval, and since $p_{\alpha_j}(t) > 0$ if t is sufficiently large, $\lim_{h \rightarrow 0} p_{ij}(h) = 1$ if $j \in G_\alpha$. Then hypothesis H_α is satisfied.

Proof that (ii) implies (iii). If hypothesis H_α is true, we shall prove (8.6) by evaluating the probabilities involved. If $0 < \tau < t$,

$$(8.8) \quad \begin{aligned} P^*\{x(t) = x(\tau)\} &= \sum_j P^*\{x(t) = x(\tau) = j\} \\ &= \sum_{j \in G_\alpha} p_{\alpha_j}(\tau) p_{ij}(t - \tau) \end{aligned}$$

and if $0 < t < \tau$,

$$(8.8') \quad P^*\{x(t) = x(\tau)\} = \sum_{j \in G_\alpha} p_{\alpha_j}(t) p_{ij}(\tau - t).$$

Proof that (iii) implies (i). Condition (iii) is known (loc. cit. ⁽²⁰⁾) to imply that the P^* -measure determines a measurable process.

Now the series in (8.8) is majorized by $\sum_j p_{\alpha_j}(\tau)$, and that in (8.8') by $\sum_j p_{\alpha_j}(t)$. Then the series in (8.8) converges uniformly in t . The series $\sum_j p_{\alpha_j}(t)$ is a series of non-negative continuous functions, converging to the continuous function 1, so there is uniform convergence in a neighborhood of τ . Thus the series in (8.8) and (8.8') are uniformly convergent for t near τ , and when $t \rightarrow \tau$ both become $\sum_j p_{\alpha_j}(\tau) = 1$, as was to be proved.

THEOREM 9. *Suppose that hypothesis H_α is true. Then if $i \in G_\alpha$,*

$$(9.1) \quad \lim_{t \rightarrow 0} \frac{1 - p_{ii}(t)}{t} = q_i (\leq +\infty)$$

exists. If $q_i = 0$, $p_{ii}(t) \equiv 1$. If $q_i = \infty$, then if $j \neq i$, $j \in G_\alpha$,

$$(9.2) \quad \lim_{t \rightarrow 0} \frac{p_{ij}(t)}{1 - p_{ii}(t)} = \lim_{t \rightarrow 0} \frac{p_{ji}(t)}{1 - p_{ii}(t)} = 0.$$

If $q_i < \infty$, then for $j \neq i$, $j \in G_\alpha$ the limits

$$(9.3) \quad \lim_{t \rightarrow 0} \frac{p_{ij}(t)}{t} = q_{ij} (< +\infty), \quad \lim_{t \rightarrow 0} \frac{p_{ji}(t)}{t} = q_{ji} (< +\infty)$$

exist, and

$$(9.4) \quad \sum_i q_{ij} \leq q_i.$$

In the finite-dimensional case⁽²¹⁾, $q_i < \infty$ for all $i \in G_\alpha$, and there is equality in (9.4).

Let R be a denumerable everywhere dense t -set. Since when $t \rightarrow \tau$,

⁽²¹⁾ Doeblin (I) proved Theorem 9 in the finite-dimensional case.

$x(t) \rightarrow x(\tau)$ in measure (Theorem 8), it follows that

$$(9.5) \quad \liminf_{r \rightarrow \tau} x(r) \leq x(\tau) \leq \limsup_{r \rightarrow \tau} x(r) \quad (r \in R)$$

with probability 1 (that is, almost everywhere on Ω^*). Then (8.3) is satisfied. We shall also need the following fact: If I is any open t -interval, and if $t_1^{(n)} < \dots < t_n^{(n)}$ are points in I , with $\max_j (t_j^{(n)} - t_{j-1}^{(n)}) = \delta_n$, then if $\delta_n \rightarrow 0$,

$$(9.6) \quad \lim_{n \rightarrow \infty} \text{L.U.B.}_j x(t_j^{(n)}) = \text{L.U.B.}_{r \in R \cdot I} x(r)$$

with probability 1. This can be proved as follows. Because of the fact that when $t \rightarrow \tau$, $x(t) \rightarrow x(\tau)$ in measure, it surely is true that for each r in I ,

$$(9.7) \quad \liminf_{n \rightarrow \infty} \text{L.U.B.}_j x(t_j^{(n)}) \geq x(r)$$

with probability 1, and (9.7) implies (9.6), because of (8.3). In the same way we can prove

$$(9.8) \quad \lim_{n \rightarrow \infty} \text{G.L.B.}_j x(t_j^{(n)}) = \text{G.L.B.}_{r \in R \cdot I} x(r)$$

with probability 1.

Now let $i \in G_\alpha$, and choose τ so that $p_{\alpha i}(\tau) > 0$. Let $\phi_i(h)$ be the probability that if $x(\tau) = i$ then $x(r) = i$ for $\tau \leq r \leq \tau + h$ ($r \in R$). (If R is used to determine a quasi-separable process, $\phi_i(h)$ is the probability that if $x(\tau) = i$, then $x(t) = i$ for $\tau \leq t \leq \tau + h$.) According to (9.6) and (9.8), if $\tau = t_1^{(n)} < \dots \leq h$, and $\max_j (t_j^{(n)} - t_{j-1}^{(n)}) = \delta_n \rightarrow 0$, then

$$(9.9) \quad \begin{aligned} P^* \{ x(r) = i, \tau \leq r \leq \tau + h, (r \in R) \} &= \lim_{n \rightarrow \infty} P^* \{ x(t_j^{(n)}) = i, j \geq 1 \} \\ &= p_{\alpha i}(\tau) \lim_{n \rightarrow \infty} \prod_{j \geq 1} p_{ii}(t_{j+1}^{(n)} - t_j^{(n)}) = p_{\alpha i}(\tau) \phi_i(h). \end{aligned}$$

Let $\{\epsilon_n\}$ be any sequence of positive numbers converging to 0. To prove (9.1) it will be sufficient to let $t \rightarrow 0$ through the sequence $\{\epsilon_n\}$, and to show that there is a limit, which is independent of the sequence $\{\epsilon_n\}$. Choose the integers m_n so that $m_n \epsilon_n \uparrow h$. Then setting $t_{j+1}^{(n)} - \tau = j \epsilon_n$ in (9.9), $0 \leq j \leq m_n$,

$$(9.10) \quad \lim_{n \rightarrow \infty} p_{ii}(\epsilon_n)^{m_n} = \phi_i(h).$$

This implies that if $\phi_i(h) > 0$

$$(9.11) \quad \lim_{n \rightarrow \infty} m_n \log p_{ii}(\epsilon_n) = - \lim_{n \rightarrow \infty} h \frac{1 - p_{ii}(\epsilon_n)}{\epsilon_n} = \log \phi_i(h).$$

We have thus shown that unless $\phi_i(h) = 0$, (9.1) is true, and

$$(9.12) \quad \phi_i(h) = e^{-q_i h}.$$

On the other hand, if $\phi_i(h) = 0$, (9.1) is true with $q_i = \infty$, and then $\phi_i(h) \equiv 0$. Since $p_{ii}(t) \geq \phi_i(t)$, $q_i = 0$ implies that $p_{ii}(t) \equiv 1$. In proving (9.2) and (9.3) we can assume that $p_{ii}(t) < 1$ for all t , since otherwise (0.1) implies that $p_{ii}(t) \equiv 1$, so $q_i = 0$: in this case (9.2) is inapplicable; the first part of (9.3) is obvious, and the second is proved by a trivial modification of the proof below. To prove (9.2) we note that if $\eta > 0$, $j \neq i$, $j \in G_\alpha$,

$$(9.13) \quad p_{ij}(n\epsilon) \geq \sum_{k=0}^{n-1} p_{ii}(\epsilon)^k p_{ij}(\epsilon) \overline{p_{ij}(n-k-1)\epsilon} \geq (1-\eta) \frac{1-p_{ii}(\epsilon)^n}{1-p_{ii}(\epsilon)} p_{ij}(\epsilon),$$

if $n\epsilon$ is sufficiently small. Then if $q_i = \infty$, when $n \rightarrow \infty$ and $\epsilon \rightarrow 0$ so that $n\epsilon \rightarrow t$, (9.13) becomes

$$(9.14) \quad p_{ij}(t) \geq (1-\eta) \limsup_{\epsilon \rightarrow 0} \frac{p_{ij}(\epsilon)}{1-p_{ii}(\epsilon)},$$

for sufficiently small t . When $t \rightarrow 0$ this gives the first part of (9.2). Similarly if $i \neq j$, $j \in G_\alpha$

$$(9.13') \quad p_{ji}(n\epsilon) \geq \sum_{k=0}^{n-1} p_{ji}(n-k-1)\epsilon p_{ji}(\epsilon) p_{ii}(\epsilon)^k \geq (1-\eta) \frac{1-p_{ii}(\epsilon)^n}{1-p_{ii}(\epsilon)} p_{ji}(\epsilon)$$

is true for sufficiently small n , and then if $q_i = \infty$

$$(9.14') \quad p_{ji}(t) \geq (1-\eta) \limsup_{\epsilon \rightarrow 0} \frac{p_{ji}(\epsilon)}{1-p_{ii}(\epsilon)},$$

for sufficiently small t . When $t \rightarrow 0$ this gives the second part of (9.2). If $q_i < \infty$, (9.13) implies that

$$(9.15) \quad p_{ij}(t) \geq (1-\eta) \frac{1-e^{-q_i t}}{q_i} \limsup_{\epsilon \rightarrow 0} \frac{p_{ij}(\epsilon)}{\epsilon}.$$

Then

$$(9.16) \quad \liminf_{t \rightarrow 0} \frac{p_{ij}(t)}{t} \geq (1-\eta) \limsup_{\epsilon \rightarrow 0} \frac{p_{ij}(\epsilon)}{\epsilon}.$$

Since $\eta > 0$ is arbitrary, this implies that $\lim_{t \rightarrow 0} p_{ij}(t)/t$ exists, and the limit is finite, from (9.15). Similarly equation (9.13') implies that $\lim_{t \rightarrow 0} p_{ji}(t)/t$ exists and is finite. Moreover

$$(9.17) \quad \sum_{j \neq i} \frac{p_{ij}(t)}{t} = \frac{1-p_{ii}(t)}{t},$$

so that (9.4) is true. Equation (9.17) can also be written

$$(9.18) \quad \sum_{j \neq i} \frac{p_{ij}(t)}{1 - p_{ii}(t)} = 1.$$

Then in the finite-dimensional case $q_i = \infty$ is impossible (since each term of the sum goes to 0 with t if $q_i = \infty$), and (9.17) implies that there is equality in (9.4).

In discussing the continuity properties of $x(t)$ in t , it is usually convenient, because of measurability considerations, to choose a denumerable everywhere dense t -set R and then consider the functions $x(r)$ for $r \in R$. The continuity properties of $x(r)$ can be interpreted as continuity properties of $x(t)$, if the proper space Ω of the stochastic process is chosen, and this will sometimes be done below.

THEOREM 10. *Suppose that hypothesis H_α is true. Let τ be any positive number and let R be any denumerable everywhere dense set. Then $\lim_{r \rightarrow \tau} x(r) = x(\tau)$ ($r \in R$) with probability 1 if and only if whenever $p_{\alpha i}(\tau) > 0$, q_i is finite. If $q_i < \infty$ and if $p_{\alpha i}(\tau) > 0$, q_{ij}/q_i is the conditional probability that if $x(\tau) = i$, and if there is a discontinuity of $x(r)$ ($r \in R$) before $\tau + h$, then there is a first discontinuity before $\tau + h$, which is an isolated discontinuity where $x(r)$ jumps to j .*

The probability that $x(\tau) = i$ and that $x(r) = i$ for $r \in R$, $\tau - h < r < \tau + h$ is $p_{\alpha i}(\tau - h)\phi_i(2h)$. Then $\lim_{r \rightarrow \tau} x(r) = x(\tau)$ with probability 1 if and only if $\lim_{h \rightarrow 0} \phi_i(h) = 1$, that is, if and only if $q_i < \infty$, whenever $p_{\alpha i}(\tau) > 0$. This proves the first part of the theorem. The second part requires a more detailed analysis. Suppose that $q_i < \infty$ and that $p_{\alpha i}(\tau) > 0$. We shall evaluate the probability of the $x(t)$ -set Λ_η determined by the following conditions: $x(\tau) = i$; $x(r) = i$ for $r \in R$, $r > \tau$ on some interval of r -values; $x(r)$ then jumps to j , remaining equal to j on some interval of length at least η , the jump occurring before $\tau + h$. Let n be any positive integer, and define $\Lambda_{n,\eta}$ by

$$(10.1) \quad \Lambda_{n,\eta} = \sum_{m=1}^{n-2} \left\{ x(\tau) = i; x(r) \equiv i, \tau < r < \tau + \frac{m}{n} h; x\left(\tau + \frac{m+1}{n} h\right) = j; \right. \\ \left. x(r) \equiv j, \tau + \frac{m+1}{n} h < r < \tau + \frac{m+1}{n} h + \eta \right\}.$$

Then

$$(10.2) \quad P(\Lambda_{n,\eta}) = \sum_{m=1}^{n-2} p_{\alpha i}(\tau) e^{-mhq_i/n} p_{ij}(h/n) e^{-\eta q_j} \\ = p_{\alpha i}(\tau) \frac{e^{-hq_i/n} - e^{-hq_i(1-1/n)}}{1 - e^{-hq_i/n}} p_{ij}(h/n) e^{-\eta q_j} \\ \rightarrow p_{\alpha i}(\tau) \frac{1 - e^{-hq_i}}{q_i} q_{ij} e^{-\eta q_j}.$$

Now if $x(t) \in \Lambda_\eta$ it follows that $x(t) \in \Lambda_{n,\eta'}$ for sufficiently large n , whenever $\eta' < \eta$:

$$(10.3) \quad \Lambda_\eta \subset \liminf_{n \rightarrow \infty} \Lambda_{n,\eta'}.$$

If $x(t) \in \Lambda_{n,\eta}$ for infinitely many values of n , $x(t) \in \Lambda_\eta$, if no $\tau + (m/n)h \in R$:

$$(10.4) \quad \limsup_{n \rightarrow \infty} \Lambda_{n,\eta} \subset \Lambda_\eta.$$

Then if $\eta' < \eta$

$$(10.5) \quad \limsup_{n \rightarrow \infty} P(\Lambda_{n,\eta}) \leq P(\Lambda_\eta) \leq \liminf_{n \rightarrow \infty} P(\Lambda_{n,\eta'}).$$

The inferior and superior limits in (10.5) are actual limits, evaluated in (10.2). Since the limit function of η is continuous, we obtain⁽²²⁾, letting $\eta' \rightarrow \eta$,

$$(10.6) \quad P(\Lambda_\eta) = p_{\alpha i}(\tau)(1 - e^{-hqi}) \frac{q_{ij}}{q_i} e^{-\eta q_i}.$$

The probability that $x(\tau) = i$, that $x(r) = i$ for $r > \tau$ on some r -interval ($r \in R$), and that then $x(r)$ jumps to j where it remains for some r -interval, the jump occurring before $\tau + h$, is therefore

$$(10.7) \quad \lim_{\eta \rightarrow 0} P(\Lambda_\eta) = p_{\alpha i}(\tau)(1 - e^{-hqi})q_{ij}/q_i,$$

and this equality is equivalent to the statement of the theorem.

To make clear the meaning of Theorem 10, suppose that $\sum_i q_{ij} = q_i < \infty$ for all i in G_α . Then if $\tau > 0$, $\lim_{r \rightarrow \tau} x(r) = x(\tau)$ with probability 1. Excluding an $x(t)$ -set of Ω^* -measure 0, each $x(t)$ in the remainder Λ is then equal on R to $x(\tau)$ for r sufficiently near τ . According to the second part of the theorem, we can make the excluded Ω^* -set so large that if $x(t) \in \Lambda$ there will be a first discontinuity of $x(r)$ (if any) after τ , a jump. Now, applying the second part of the theorem, letting τ run through all rational numbers, we see that the excluded Ω^* -set can be made so large that if $x(t) \in \Lambda$, there will be a second discontinuity (if there is more than one), also a jump, a third, and so on. These discontinuities may cluster at a point, to give $x(r)$ a discontinuity which is no longer a jump.

We shall use a somewhat indirect method in examining in more detail the transitions of the system, that is, the discontinuities of $x(t)$. This method has the advantage of exhibiting analytically the relation between the regularity of the matrix function $(p_{ij}(t))$ and the discontinuities of $x(t)$.

⁽²²⁾ We have tacitly assumed the measurability of Λ_η . This is easily proved directly, or the above discussion can be modified, using inner and outer measures in (10.5), to furnish the proof that Λ_η is measurable, besides evaluating its measure. The restriction we have made on h is essential for (10.4), but evidently (10.5) and (10.6) are true without this restriction.

Let y_t (for each t in some point set) be a chance variable. The family of chance variables $\{y_t\}$ will be said to have the property \mathcal{E} if (for any natural number n), whenever $t_1 < \dots < t_{n+1}$,

$$(11.1) \quad E\{y_{t_1}, \dots, y_{t_n}; y_{t_{n+1}}\} = y_{t_n}^{(23)},$$

with probability 1⁽²⁴⁾. Suppose a family $\{y_t\}$ has the property \mathcal{E} , for t in some interval (a, b) . Then if $a < \tau < b$, $t_n \uparrow \tau$ implies that $\lim_{n \rightarrow \infty} y_{t_n} = y_{\tau-}$ exists with probability 1, and the limit $y_{\tau-}$ is independent (neglecting zero probabilities) of the particular sequence $\{t_n\}$. The chance variable $\{y_{\tau+}\}$ is defined similarly in terms of approach from above. Moreover, $y_{\tau-} = y_{\tau+} = y_{\tau}$ with probability 1, if τ is not in some set, which is at most denumerable. We shall call this set the set of fixed discontinuities. If R is any denumerable set, dense on (a, b) , y_r ($r \in R$), with probability 1, considered as a function of r alone is equal to a function defined on (a, b) , and continuous on the right at every point of (a, b) not a fixed discontinuity point. It will be useful below to say that a family of chance variables y_t has the property \mathcal{E}^* if the family y_{-t} has the property \mathcal{E} .

It is easily verified that if $T > 0$, and if y_t is defined by

$$(11.2) \quad y_t = p_{x(t)j}(T - t) \quad (0 < t < T),$$

then the family of chance variables has the property $\mathcal{E}^{(25)}$. We shall show that there are no fixed discontinuities if hypothesis H_α is true. To do this it will be sufficient to show that if t is given, and if $t_n \rightarrow t$, then some subsequence of $\{y_{t_n}\}$ converges to y_t with probability 1. Since according to Theorem 8, $x_{t_n}(\omega)$ converges to $x_t(\omega)$ in measure, some subsequence, $x_{\tau_n}(\omega)$, converges to $x_t(\omega)$ with probability 1. Then

$$(11.3) \quad y_{\tau_n} = p_{x(t)j}(T - \tau_n)$$

for large n , with probability 1, so that $y_{\tau_n} \rightarrow y_t$ with probability 1, because the $p_{ij}(t)$ are continuous if $i \in G_\alpha$ and for each t , $P^*\{x_t(\omega) \in G_\alpha\} = 1$. In a similar way it can be proved that if $T^* > 0$, and if y_t^* is defined by

$$(11.4) \quad y_t^* = \frac{p_{jx(t)}(t - T^*)}{p_{\alpha x(t)}(t)} \quad (t > T^*),$$

the family of chance variables $\{y_t^*\}$ has the property \mathcal{E}^* , and there are no fixed discontinuities, if hypothesis H_α is true. The chance variable $y_t^* \cdot p_{\alpha j}(T^*)$

⁽²³⁾ The notation $E\{y_{t_1}, \dots, y_{t_n}; y_{t_{n+1}}\}$ will be used to denote the conditional expectation of $y_{t_{n+1}}$ for given values of y_{t_1}, \dots, y_{t_n} , a function of the latter variables.

⁽²⁴⁾ The properties of such a family, summarized here, are proved in the author's paper in these Transactions, vol. 47 (1940), pp. 455-486. This paper will be referred to as " \mathcal{E} ."

⁽²⁵⁾ This fact is a result of the well known relations between conditional expectation functions, and is a special case of the fact that if $\{w_t\}$ is any family of chance variables, if z is a chance variable dependent on the w_t , and if $z_t = E\{w_s, s \leq t; z\}$, ($z_t =$ expectation of z for w_s given for $s \leq t$), then the family $\{z_t\}$ has the property \mathcal{E} .

bears the same relation to the inverse process (t decreasing) as y_t bears to the given process. For each t , the denominator in (11.4) vanishes only with probability 0 (hypothesis H_α). The following two regularity conditions on the $p_{ij}(t)$ will be useful.

CONDITION $C(\beta)$. Let β be in G_α . Then there are numbers η, δ such that for all $i \neq \beta$ in G_α , and all $s < \delta$

$$(11.5) \quad p_{i\beta}(s) < 1 - \eta.$$

CONDITION $C^*(\beta, \tau)$. Let β be in G_α and let τ be a positive number with $p_{\alpha\beta}(\tau) > 0$. There are positive numbers η, δ such that if $0 < s < \delta, i \notin \beta, p_{\alpha i}(\tau) > 0$, then

$$(11.5^*) \quad p_{\beta i}(s) < \frac{p_{\alpha i}(\tau + s)}{p_{\alpha\beta}(\tau)} (1 - \eta).$$

Under hypothesis H_α , if i is fixed and $\delta \rightarrow 0$ in (11.5), the inequality becomes $0 \leq 1 - \eta$, and under the same circumstances, (11.5*) becomes

$$(11.6) \quad 0 \leq \frac{p_{\alpha i}(\tau)}{p_{\alpha\beta}(\tau)} (1 - \eta).$$

Then conditions $C(\beta)$ and $C^*(\beta, \tau)$ are certainly always satisfied in the finite-dimensional case, under hypothesis H_α , for all possible β and pairs β, τ ($\beta \in G_\alpha$), respectively.

Condition $C(\beta)$ can be put in an interesting alternate form. If condition $C(\beta)$ is not satisfied, there is a sequence of distinct integers $\{i_\nu\}$ in G_α , and a sequence $\{s_\nu\}$, $s_\nu \rightarrow 0$, such that $p_{i_\nu\beta}(s_\nu) \rightarrow 1$. Now if $t > 0$, and if ν is so large that $s_\nu < t$,

$$(11.7) \quad p_{i_\nu j}(t) = p_{i_\nu\beta}(s_\nu) p_{\beta j}(t - s_\nu) + \sum_{k \neq \beta} p_{i_\nu k}(s_\nu) p_{kj}(t - s_\nu).$$

If $j \notin G_\alpha$, $p_{i_\nu j}(t) \equiv 0$. If $j \in G_\alpha$, the sum on the right is at most

$$\sum_{k \neq \beta} p_{i_\nu k}(s_\nu) = 1 - p_{i_\nu\beta}(s_\nu) \rightarrow 0.$$

Then (11.7) implies

$$(11.8) \quad \lim_{\nu \rightarrow \infty} p_{i_\nu j}(t) = p_{\beta j}(t),$$

for all j, t . Thus (under hypothesis H_α) condition $C(\beta)$ is satisfied if (and, as is easily seen, only if) no sequence of distinct rows (whose elements have first subscripts in G_α) converges, element by element to the β th row, for all t . An analogous but less elegant form of $C^*(\beta, \tau)$ can be obtained.

The following theorem makes Theorem 10 more precise.

THEOREM 11. *Suppose that hypothesis H_α is true. Let τ be any positive number, and suppose that R is any denumerable set having τ as a limit point. Then*

$$(11.9) \quad \lim_{r \rightarrow \tau} \frac{x(r) - x(\tau)}{1 + x(r)^2} = 0 \quad (r \in R)$$

with probability 1. If $p_{\alpha\beta}(\tau) > 0$, then $\lim_{r \rightarrow \tau} x(r) = \beta$ whenever $x(\tau) = \beta$, (with probability 1) if and only if $q_\beta < \infty$. If $C(\beta)$ or $C^*(\beta, \tau)$ is satisfied, then $q_\beta < \infty$.

Since, as we have seen in Theorem 8, if $r \rightarrow \tau$, $x_r(\omega) \rightarrow x_\tau(\omega)$ in measure, that is, $x(r) \rightarrow x(\tau)$ in measure, it is impossible that $|x(r)| \rightarrow \infty$ with positive probability along any sequence of r -values approaching τ . Therefore, neglecting 0 probabilities (Ω^* -sets of measure 0), (11.9) implies that $x(r)$ always has $x(\tau)$ as a limiting value when $r \rightarrow \tau$, (the only finite limiting value) but $x(r)$ may also have $\pm \infty$ as a limiting value. In the course of the proof of Theorem 10, we have already proved that if $p_{\alpha\beta}(\tau) > 0$, $\lim_{r \rightarrow \tau} x(r) = \beta$ whenever $x(\tau) = \beta$ with probability 1, if and only if $q_\beta < \infty$: in fact this statement follows directly from our evaluation of $\phi_\beta(h)$. To prove (11.9) we shall use the families of chance variables $\{y_t\}$, $\{y_t^*\}$ introduced above. Since these families have no fixed discontinuities,

$$(11.10) \quad \lim_{r \rightarrow \tau} p_{x(r)j}(T - r) = p_{x(\tau)j}(T - \tau),$$

with probability 1. Let Λ be an ω -set of measure 1, such that the following conditions are satisfied, if $x(t) \in \Lambda$:

- (a) (11.10) is true for all j and rational $T > \tau$;
- (b) if $p_{\alpha j}(r) = 0$, then $x(r) \neq j$; if $p_{\alpha j}(\tau) = 0$, then $x(\tau) \neq j$.

Now suppose that $x_0(t) \in \Lambda$, and suppose that $x_0(\tau) = \beta$. Suppose that $\liminf_{r \rightarrow \tau} |x_0(r)| < +\infty$. Then there is an integer γ such that $x_0(r) = \gamma$ for infinitely many values of r , as $r \rightarrow \tau$. Condition (a) implies

$$(11.11) \quad \lim_{r \rightarrow \tau} p_{\gamma j}(T - r) = p_{\gamma j}(T - \tau) = p_{\beta j}(T - \tau),$$

for all j and rational $T > \tau$. (We are using the fact that because of condition (b), $p_{\alpha\gamma}(t) \neq 0$, so, in accordance with hypothesis H_α , $p_{\alpha\gamma}(t)$ is continuous in t .) Because of hypothesis H_α , $\lim_{t \rightarrow 0} p_{\gamma j}(t) = \delta_{\gamma j}$, $\lim_{t \rightarrow 0} p_{\alpha j}(t) = \delta_{\alpha j}$ for all j . Thus if we let $T \rightarrow \tau$, (11.11) implies that $\delta_{\gamma j} = \delta_{\beta j}$ for all j : $\gamma = \beta$. We have now proved that the only possible limiting values of $x_0(r)$ as $r \rightarrow \tau$ are $x_0(\tau)$, $\pm \infty$; (11.9) is true for $x(t) \in \Lambda$, and hence is true with probability 1. If the matrices are finite-dimensional, there can be only finite limiting values of $x_0(r)$; so $x(r) \rightarrow x(\tau)$ ($r \in R$) with probability 1, as we have already proved in Theorem 10. Now suppose that $(p_{ij}(t))$ is infinite-dimensional, and suppose that for some $x_0(t)$ in Λ , $x_0(r)$ does not approach $x_0(\tau) = \beta$ ($r \in R$). Then there must be a sequence of integers $\{i_\nu\}$ and a sequence $\{r_\nu\}$ such that $x_0(r_\nu) = i_\nu \rightarrow \pm \infty$, ($r_\nu \rightarrow \tau$). Then (11.10) becomes

$$(11.12) \quad \lim_{\nu \rightarrow \infty} p_{i_\nu j}(T - r_\nu) = p_{\beta j}(T - \tau)$$

for all j and rational $T > \tau$. It follows readily from (11.12) with $j = \beta$ that there is a subsequence $\{j_\nu\} = \{i_{a_\nu}\}$ of $\{i_\nu\}$ and a sequence of values $\{T_\nu\}$ of T , $T_\nu \downarrow \tau$, such that

$$(11.13) \quad \lim_{\nu \rightarrow \infty} p_{i_\nu \beta}(T_\nu - r_{a_\nu}) = \lim_{\nu \rightarrow \infty} p_{\beta \beta}(T_\nu - \tau) = 1.$$

This evidently contradicts condition $C(\beta)$. Thus if $C(\beta)$ is satisfied, $q_\beta < \infty$. The family of chance variables

$$\left\{ \frac{p_{ix(r)}(r - \tau)}{p_{\alpha x(r)}(r)} \right\} \quad (r > \tau)$$

has, as we have seen, the property \mathcal{E}^* . Since these chance variables are non-negative, there is convergence when $r \downarrow \tau$, with probability 1⁽²⁶⁾. Since $x(r) \rightarrow x(\tau)$ in measure,

$$(11.14) \quad \lim_{r \downarrow \tau} \frac{p_{\beta x(r)}(r - \tau)}{p_{\alpha x(r)}(r)} = \frac{1}{p_{\alpha \beta}(\tau)}$$

almost everywhere where $x(\tau) = \beta$. We can suppose Λ has been chosen so that (11.14) is true if $x(t) \in \Lambda$. Unless $x(r) \rightarrow \beta$ whenever $x(t) \in \Lambda$, $r \downarrow \tau$, and $x(\tau) = \beta$, there is an $x_0(t)$ in Λ with $x_0(\tau) = \beta$, a sequence of integers $\{i_\nu\}$, and a sequence $\{r_\nu\}$ such that $x_0(r_\nu) = i_\nu \rightarrow \pm \infty$, $r_\nu \downarrow \tau$. Then, using (11.14),

$$(11.15) \quad \frac{p_{\beta i_\nu}(r_\nu - \tau)}{p_{\alpha i_\nu}(r_\nu)} \rightarrow \frac{1}{p_{\alpha \beta}(\tau)}$$

which contradicts $C^*(\beta, \tau)$. Thus if $C^*(\beta, \tau)$ is satisfied, $q_\beta < \infty$. (We are using here the fact which is implicit in the discussion of $\phi_\beta(h)$ above that if $x(\tau) = \beta$, $\lim_{r, \tau} x(r) = \beta$ with probability 1 if and only if $q_\beta < \infty$.)

We shall need a somewhat stronger condition than $C^*(\beta, \tau)$ below. We shall say that condition $C^{**}(\beta, \tau)$ is satisfied if $p_{\alpha \beta}(\tau) > 0$ and if there are positive numbers η, δ such that if $0 < s_1 \leq s_2 < \delta$, $i \neq \beta$, $p_{\alpha i}(\tau + s_1) > 0$, then

$$(11.16) \quad p_{\beta i}(s_2) < \frac{p_{\alpha i}(\tau + s_1)}{p_{\alpha \beta}(\tau)} (1 - \eta).$$

Condition $C^{**}(\beta, \tau)$ is always satisfied in the finite-dimensional case, under hypothesis H_α , if $\beta \in G_\alpha$ since (11.16) becomes (11.6) when s_1 and s_2 approach 0.

⁽²⁶⁾ This is true if $r \downarrow \tau$ along any sequence of values (Doob, these Transactions, vol. 47 (1940), p. 460, Theorem 1.3) and this means the truth of the statement when $r \downarrow \tau$, $r \in R$ (Doob, these Transactions, vol. 42 (1937), p. 111, Theorem 1.3, or, in another formulation, Duke Mathematical Journal, vol. 4 (1938), pp. 758-759, Lemma 2).

THEOREM 12. *Suppose that hypothesis H_α is true. Let R be any denumerable everywhere dense t -set. Then there is a set Λ of functions $x(t)$, of probability 1, such that if $x(t) \in \Lambda$ the following statements are true.*

(a) *If the matrices are finite-dimensional, $x(r)$ ($r \in R$) has only isolated jumps as discontinuities.*

(b) *Either $\lim_{r \downarrow \tau} |x(r)| = \infty$ ($r \in R$), or there is an integer β , depending on the function $x(t)$ and on τ , such that*

$$(12.1) \quad \lim_{r \downarrow \tau} \frac{x(r) - \beta}{1 + |x(r)|^2} = 0 \quad (r \in R).$$

If condition $C(\beta)$ is satisfied, then either $\lim_{r \downarrow \tau} |x(r)| = \infty$ or there is an integer β depending on the function $x(t)$ and on τ , such that $\lim_{r \downarrow \tau} x(r) = \beta$ ($r \in R$).

(b*) *The statement of (b) remains true with $r \uparrow \tau$, instead of $r \downarrow \tau$, replacing condition $C(\beta)$ by $C^{**}(\beta, \tau)$.*

(c) *For each τ , there will be an integer β (and $\beta = x(\tau)$), as described in (b), (b*), with probability 1.*

Theorem 12 is closely related to work of Doeblin and Feller, with which it will be compared below.

The families $\{y_i\}$, $\{y_i^*\}$ have the properties \mathcal{E} , \mathcal{E}^* , respectively. There is therefore an ω -set Λ of probability 1, such that if $x(t) \in \Lambda$, the corresponding y_i , (y_i^*) coincide on R with functions everywhere continuous on the right (left), for all j , rational T , T^* . (It has been proved above that there are no points of fixed discontinuity.) We can also suppose that $x(r) \neq j$ unless $p_{\alpha j}(r) > 0$, if $x(t) \in \Lambda$. Then if $x_0(t) \in \Lambda$, and if $T > \tau > T^* > 0$, T , T^* rational, the following limits exist:

$$(12.2) \quad \lim_{r \downarrow \tau} p_{x_0(r)j}(T - r),$$

$$(12.2^*) \quad \lim_{r \uparrow \tau} \frac{p_{jx_0(r)}(r - T^*)}{p_{\alpha x_0(r)}(r)}.$$

If $x_0(r)$ takes on a subscript β for values of r approaching τ from above, we can evaluate the limit in (12.2):

$$(12.3) \quad \lim_{r \downarrow \tau} p_{x_0(r)j}(T - r) = \lim_{r \downarrow \tau} p_{\beta j}(T - r) = p_{\beta j}(T - \tau).$$

Then β is uniquely determined, for if γ had the same defining property, we should have $p_{\beta j}(T - \tau) = p_{\gamma j}(T - \tau)$ for all rational $T > \tau$, where β, γ are both in G_α . When $T \downarrow \tau$ this means (using hypothesis H_α) that $\delta_{\beta j} = \delta_{\gamma j}$ for all j , impossible unless $\beta = \gamma$. Thus (12.1) is true. Then in the finite-dimensional case, $\lim_{r \downarrow \tau} x_0(r) = \beta$, that is, $x_0(r) = \beta$ for r sufficiently near τ ($r > \tau$). In the infinite-dimensional case, unless $\lim_{r \downarrow \tau} x_0(r) = \beta$, we have proved that $\lim \sup_{r \downarrow \tau} |x_0(r)| = \infty$, and the method of proof of Theorem 11 can be car-

ried through to find a contradiction to condition $C(\beta)$. We have now proved that Theorem 12(b) is true, supposing however that there is a subscript β as described. (In the finite-dimensional case there is always such a β .) In the infinite-dimensional case, if there is no such β , $\lim_{r \downarrow \tau} |x(r)| = \infty$. This finishes the proof of (b). The discussion when $r \uparrow \tau$ is carried on in the same way, using the existence of the limit in (12.2*). Theorem 12(a) is now obviously true. For a given τ , $x(r) \rightarrow x(\tau)$ in measure, when $r \rightarrow \tau$, so there will be an integer β as described above, and $\beta = x(\tau)$, with probability 1.

As usual in this sort of discussion, instead of saying that $x(r)$ ($r \in R$) has the above described properties with probability 1, we could say that if a space Ω of a stochastic process is chosen properly, all the $x(t)$ in Ω will have the above properties, where t ranges through all values.

Doebelin has considered a general Markoff process in which the transition probability of going from state i at time t to state j at time t' is not supposed necessarily to be a function of $t' - t$, and in which it is not supposed that the number of possible states is denumerably infinite. His hypotheses, when translated into our notation, and simplified because of the more special process being considered here, become

$$(12.4) \quad \lim_{t \rightarrow 0} p_{ii}(t) = 1$$

uniformly in i . This hypothesis, combined with the hypothesis that the process is initially in state α is considerably stronger than hypothesis H_α (except in the finite-dimensional case, when, assuming hypothesis H_α , Doebelin's condition is always applicable) and evidently also implies condition $C(\beta)$ for all $\beta \in G_\alpha$. Doebelin showed that under his hypotheses, and assuming some given initial state, neglecting an ω -set of measure 0, $x(r)$ ($r \in R$) has only isolated jumps as discontinuities⁽²⁷⁾.

Conversely, suppose that the process is initially in state α , and that the ω -measure has the property that $x(r)$ ($r \in R$, a denumerable everywhere dense t -set) has only isolated jumps as discontinuities, with probability 1. Theorem 11 shows that in this case there must be continuity at each fixed τ , with probability 1, that is, $q_i < \infty$ if $i \in G_\alpha$. Also, by Theorem 10, $\sum_j q_{ij} = q_i$. Let $P_{ij}^{(n)}(t)$ be the probability⁽²⁸⁾ that if $x(\tau) = i$ then $x(\tau + t) = j$ and $x(r)$ has n jumps in going from i to j , between τ and $\tau + t$. It is easily verified that if $i \in G_\alpha$

$$(12.5) \quad \begin{aligned} P_{ik}^{(0)}(t) &= \delta_{ik} e^{-q_i t} \\ P_{ik}^{(n+1)}(t) &= \sum_j \int_0^t P_{ij}^{(n)}(s) q_{jk} e^{-q_k(t-s)} ds \end{aligned} \quad (n \geq 0),$$

⁽²⁷⁾ Skandinavisk Aktuarietidskrift, vol. 22 (1939), pp. 211-222.

⁽²⁸⁾ For this conditional probability to have a meaning we must suppose that $p_{\alpha i}(\tau) > 0$; τ can always be so chosen, if $i \in G_\alpha$.

and obviously

$$(12.6) \quad p_{ij}(t) = \sum_{n \geq 0} P_{ij}^{(n)}(t).$$

Moreover if we suppose only hypothesis H_α to be true, and that $\sum_j q_{ij} = q_i < \infty$, ($i \in G_\alpha$), then if $i \in G_\alpha$, considerations analogous to those used in the proof of Theorem 10 show that $P_{ij}^{(n)}(t)$ as defined in (12.5) will have the probability meaning described above. On the other hand, (12.6) is now true if and only if the only discontinuities of $x(\tau)$ are isolated jumps, with probability 1. Feller has found necessary and sufficient conditions on the q_i, q_{ij} that (12.6) be true⁽²⁹⁾. The above remarks give a complete justification for Feller's probability interpretation (ibid., p. 498) of the $P_{ij}^{(n)}(t)$. (He did not need this interpretation in his proofs.) The details given in Theorem 12 on the character of $x(t)$ at a discontinuity which is not a jump round out Feller's description, arising from an entirely different background (ibid., pp. 512-513).

Suppose again that hypothesis H_α is satisfied and that $\sum_j q_{ij} = q_i < \infty$ for all $i \in G_\alpha$. The differential equations

$$(13.1) \quad p'_{ik}(t) = -q_i p_{ik}(t) + \sum_{j \neq i} q_{ij} p_{jk}(t) \quad (i \in G_\alpha)$$

satisfied by the $p_{ij}(t)$, due to Kolmogoroff⁽³⁰⁾ are well known, but their intimate relation to the continuity properties of the $x(t)$ seems less well known. In fact, under the above hypotheses, we have shown (Theorem 10) that (with probability 1) if $x(\tau) = i \in G$, there is a first discontinuity of $x(\tau)$ ($\tau \in \mathcal{R}$, an everywhere dense denumerable t -set) after τ , an isolated jump. Then the probability of going from i to k is the sum of the probabilities of going from i to j on the first jump, and then to k (summed over j). Considerations analogous to those used in the proof of Theorem 10 now show that, evaluating the above probabilities,

$$(13.2) \quad p_{ik}(t) = \sum_{j \neq i} \int_0^t e^{-q_i(t-s)} q_{ij} p_{jk}(s) ds + \delta_{ik} e^{-q_i t}.$$

The equations of (13.1) are obtained by differentiating those of (13.2), in which the series can obviously be differentiated term by term. The second set of differential equations obtained by Kolmogoroff

$$(13.3) \quad p'_{ik}(t) = -q_k p_{ik}(t) + \sum_{j \neq k} p_{ij}(t) q_{jk} \quad (i \in G_\alpha)$$

⁽²⁹⁾ These Transactions, vol. 48 (1940), pp. 506-507. Feller's results are applicable to considerably more general stochastic processes than those considered here.

⁽³⁰⁾ Mathematische Annalen, vol. 104 (1931), p. 429. Kolmogoroff imposed further restrictions on the $p_{ij}(t)$. Feller (op. cit., p. 495) obtained (13.1) with substantially our hypotheses given above.

does not seem always to be true without further hypotheses. We shall show that the truth of (13.3) is equivalent to the imposition of certain regularity properties on the $x(t)$. The probability

$$p_{ik}(t_2) - p_{ik}(t_1)e^{-q_k(t_2-t_1)} \quad (i \neq k, t_2 > t_1)$$

is at least equal to the probability that if $x(\tau-t_2)=i$, then $x(\tau)$ goes to j at some point between $\tau-t_1$ and τ , when $x(\tau)$ jumps to k , remaining at k until $t=\tau$, summed over $j \neq k$. Thus

$$(13.4) \quad p_{ik}(t_2) - p_{ik}(t_1)e^{-q_k(t_2-t_1)} \geq \sum_{j \neq k} \int_{t_1}^{t_2} p_{ij}(s)q_{jk}e^{q_k(t_2-s)} ds.$$

Moreover there is equality if and only if when $x(\tau)=i$, there is a last discontinuity of $x(\tau)$ before τ , which is a jump, with probability 1. Dividing (13.4) by t_2-t_1 and letting $t_2-t_1 \rightarrow 0$ we obtain

$$(13.5) \quad p'_{ik}(t) \geq -q_k p_{ik}(t) + \sum_j p_{ij}(t)q_{jk} \quad (i \in G_\alpha).$$

It is easily verified that (13.5) also follows directly from (0.1). Since there is equality in (13.5) if and only if there is equality in (13.4), we have obtained the following theorem.

THEOREM 13. *Suppose that hypothesis H_α is satisfied and that $\sum_j q_{ij} = q_i < \infty$ for all $i \in G_\alpha$. Then (13.1) is always true; (13.3) is true (for all t) if and only if when $x(\tau)=i$, there is, with probability 1, a last discontinuity of $x(\tau)$ before τ , which is an isolated discontinuity (a jump).*

It is interesting to note that if $\bar{p}_{ij}(t)$ is the probability that if $x(\tau)=i$ then $x(\tau+t)=j$ and the transition from i to j ⁽³¹⁾ is accomplished in a finite number of isolated jumps, then $\bar{p}_{ij}(t)$ evidently satisfies (0.1) except that $\sum_j \bar{p}_{ij}(t)$ may be less than 1. Moreover (13.1) is also true for the $\bar{p}_{ij}(t)$ since the derivation for the $p_{ij}(t)$ applies equally well to $\bar{p}_{ij}(t)$. And the derivation we have given of (13.5), when applied to the $\bar{p}_{ij}(t)$ actually gives equality: (13.3) is true of the $\bar{p}_{ij}(t)$ in all cases. The latter fact was also proved by Feller.

⁽³¹⁾ Strictly speaking, we should restrict i, j to lie in G_α .