# PROBABILISTIC METHODS IN MARKOV CHAINS

K. L. CHUNG
SYRACUSE UNIVERSITY

## 1. Introduction

To avoid constant repetition of qualifying phrases, we agree on the following notation, terminology, and conventions, unless otherwise specified.

$\mathbf{I}$ is a denumerable set of indices. The letters $i$, $j$, $k$, and $l$, with or without subscript, denote elements of $\mathbf{I}$.

$\bar{\mathbf{I}} = \mathbf{I} \cup \{\infty\}$ is the one-point compactification of $\mathbf{I}$ considered as an isolated set of real numbers; $\infty > i$.

$\mathbf{N}$ is the set of nonnegative integers used as ordinals. The letters $\nu$ and $n$ denote elements of $\mathbf{N}$.

$\mathbf{T} = [0, \infty)$; $\mathbf{T}^0 = (0, \infty)$. The letters $s$, $t$ and $u$, with or without subscript, denote elements of $\mathbf{T}^0$.

A statement or formula involving an unspecified element of $\mathbf{I}$ or $\mathbf{T}^0$ is meant to stand for every such element.

A sequence like $\{f_i\}$ is indexed by $\mathbf{I}$; a matrix like $(p_{ij})$ is indexed by $\mathbf{I} \times \mathbf{I}$; a sum like $\sum_j$ is over $\mathbf{I}$.

A function is real and finite valued. A function defined on $\mathbf{T}^0$ and having a right hand limit at zero is thereby extended to $\mathbf{T}$; if in addition it is continuous in $\mathbf{T}^0$ it is said to be continuous in $\mathbf{T}$.

A (*standard*) *transition matrix* is a matrix $(p_{ij})$ of functions on $\mathbf{T}^0$ satisfying the following conditions:

$$p_{ij}(t) \geqq 0, \tag{1.1}$$

$$\sum_j p_{ij}(t) p_{jk}(s) = p_{ik}(t + s), \tag{1.2}$$

$$\lim_{t \downarrow 0} p_{ii}(t) = 1, \tag{1.3}$$

$$\sum_j p_{ij}(t) = 1. \tag{1.4}$$

A (*temporally*) *homogeneous Markov chain*, or a *Markov chain with stationary transition probabilities*, associated with $\mathbf{I}$ and $(p_{ij})$, is a stochastic process $\{x_t\}$, $t \in \mathbf{T}$ or $t \in \mathbf{T}^0$, on the probability triple $(\Omega, \mathfrak{F}, \mathbf{P})$, with the generic sample point $\omega$, having the following properties:

**This paper was prepared with partial support of the Air Force Office of Scientific Research, under Contract AF 49(638)-265.**

(1.5) *For each $t$ in* $\mathbf{T}$ *or* $\mathbf{T}^0$ *respectively,* $x_t$ *is a discrete random variable, and the set of all possible values of all* $x_t$ *is* $\mathbf{I}$;

(1.6) *If* $t_1 < \cdots < t_n$, *then*

$$\mathbf{P}\{x(t_{\nu+1}, \omega) = i_{\nu+1}, 1 \leqq \nu \leqq n | x(t_1, \omega) = i_1\} = \prod_{\nu=1}^{n} p_{i_\nu i_{\nu+1}}(t_{\nu+1} - t_\nu).$$

An equivalent form of (1.6) is the *Markov property:*

(1.7) $$\mathbf{P}\{x(t_{n+1}, \omega) = i_{n+1} | x(t_\nu, \omega) = i_\nu, 1 \leqq \nu \leqq n\} = \mathbf{P}\{x(t_{n+1}, \omega) = i_{n+1} | x(t_n, \omega) = i_n\} = p_{i_n i_{n+1}}(t_{n+1} - t_n).$$

A *version* of the process will be chosen to have the following further properties:

(1.8) *For any denumerable set* $R$ *dense in* $\mathbf{T}$, *and every* $\omega \in \Omega$,

$$x(t, \omega) = \varliminf_{\substack{r \downarrow t \\ r \in R}} x(r, \omega)$$

*for all* $t$;

(1.9) *As a function of* $(t, \omega)$, $x(t, \omega)$ *is measurable with respect to the (uncompleted) product field* $\mathfrak{B} \times \mathfrak{F}$ *where* $\mathfrak{B}$ *is the usual Borel field on* $\mathbf{T}$.

The property (1.8) implies that the process is separable; the property (1.9) is called the Borel measurability of the process. Other properties of the process which follow from (1.5) to (1.9) for almost all $\omega$, may be supposed to hold for all $\omega$, so long as only denumerably many such properties are invoked.

From now on a process $\{x_t\}$ having the properties (1.5) to (1.9) will be abbreviated as an "M.C." It is called an *open* M.C. iff the parameter set is $\mathbf{T}^0$. The set $\mathbf{I}$ is called its (*minimal*) *state space*, the matrix $(p_{ij})$ its *transition matrix*. The distribution of $x_0$, when defined, is called its initial distribution $\{p_i\}$, where $p_i = \mathbf{P}\{\Delta_i\}$ and $\Delta_i = \{\omega : x(0, \omega) = i\}$. When $p_i = 1$, the resulting $\mathbf{P}$ will be written as $\mathbf{P}_i$; for example,

(1.10) $$\mathbf{P}_i\{x(t, \omega) = j\} = p_{ij}(t) = \mathbf{P}\{x(s + t, \omega) = j | x(s, \omega) = i\}$$

whenever the last is defined.

The study of the theory of M.C.'s consists in:

(a) uncovering the properties of, and relations among, the functions $p_{ij}$;

(b) describing qualitatively and quantitatively the nature of the sample functions $x(\cdot, \omega)$, $\omega \in \Omega$; (less precisely, to analyze the evolution of the process in time).

Superficially at least, object (a) can be regarded as a purely "analytic" (as distinguished from "probabilistic" or "measure theoretic") program. We may simply wish to find as much information as possible about the set of functions satisfying (1.1) to (1.4). Or we may regard the matrices $\mathfrak{P}(t) = (p_{ij}(t))$ as forming a semigroup of operators and study the properties of the semigroup. A good number of papers have been written from such a standpoint eschewing probability itself "like the devil." For us however the most rewarding part of this study is the interplay between the "analytic" and "stochastic" aspects of the theory. It is the main purpose of this paper to show, by various illustrations from recent work, that the structure of the transition matrix on the one hand, and the behavior of the sample functions on the other, are so intimately connected that one can hardly strike a chord in the one without bringing out an echo from the other. The two sides of the theory of Markov chains induce, sustain, and complement each other.

## 2. Comments on the conditions (1.3) and (1.4)

It has long been observed that much of the analytic structure of a transition matrix $(p_{ij})$ remains unchanged if the condition (1.4) is replaced by the weaker one

$$\sum_j p_{ij}(t) \leqq 1. \tag{2.1}$$

A matrix $(p_{ij})$ satisfying (1.1), (1.2), (1.3), and (2.1) will be called a *substochastic transition matrix.* ( In distinction a transition matrix as defined in section 1 may be qualified as *stochastic.*) The above observation is easily justified by a simple reduction. Add a new index $\theta$ to $\mathbf{I}$ and define new elements as follows:

$$\begin{gathered} p_{i\theta}(t) = 1 - \sum_j p_{ij}(t), \\ p_{\theta\theta}(t) \equiv 1, \qquad p_{\theta j}(t) \equiv 0. \end{gathered} \tag{2.2}$$

The new matrix is stochastic and contains the old one. Probabilistically speaking, the new state $\theta$ is an absorbing state into which all the diminishing mass disappears. Thus $p_{i\theta}(t)$ is nondecreasing in $t$ and we have

$$p_{i\theta}(t + s) - p_{i\theta}(t) = \sum_j p_{ij}(t)p_{j\theta}(s). \tag{2.3}$$

This trivial equation will assume more interesting proportions as we proceed.

Not only can the condition (1.4) be weakened into (2.1), but it can be dropped completely for many analytic purposes. This is implicit in some known proofs, but it was first realized in its full import by W. B. Jurkat [5] when he dispensed with this condition in more difficult cases. This realization has an important analytic consequence, for the omission of the "row condition" (1.4) restores complete symmetry to the rows and columns of the matrices. They form then simply a semigroup of nonnegative matrices $\{\mathfrak{P}(t)\}$ converging to the identity matrix $I$ at $t = 0$. We shall not pursue the subject in this generality here since it has as yet no probabilistic interpretation.

Turning to the condition (1.3), let us first note that together with (1.1) and (1.2) it implies that every $p_{ij}$ is continuous in $\mathbf{T}$ (see after lemma 1 below). Indeed, if we regard the semigroup $\{\mathfrak{P}(t)\}$ as operating on absolutely convergent series, then the condition (1.3) is equivalent to the strong continuity of the semigroup (see [4], p. 636). Now in the terminology of semigroup theory there is an even stronger kind of continuity, namely that in the "uniform operator topology," which is equivalent here to the condition that the convergence in (1.3) be uniform with respect to all $i \in \mathbf{I}$. Using the notation to be introduced at the beginning of section 5 below, it can be shown (theorem II. 19.2 of [1]) that this condition is equivalent to the boundedness of the sequence $\{q_i\}$. In this case the matrix $Q = (q_{ij})$ is a bounded operator and we have (see [4], p. 635)

$$\mathfrak{P}(t) = e^{Qt}, \tag{2.4}$$

Hence this case, which includes the case of a finite set $\mathbf{I}$, may be regarded as "solved" analytically. Probabilistically, the uniform condition implies (but is not implied by) that almost every sample function of the M.C. is a *step function*, namely one whose only discontinuities are jumps. While this was the case first studied for continuous parameter Markov processes, the properties of a sample step function are not essentially different from those of a sample sequence arising from a "discrete skeleton" (see section 6) of the M.C. The study of continuous parameter M.C.'s would scarcely be any innovation if we were to confine ourselves to this "trivial" case and label any new phenomenon as "pathological."

## 3. Two analytical lemmas

The first lemma is theorem II. 2.3 of [1], from which a superfluous condition has been removed, even though that very mild condition is satisfied in all known instances of application. The added argument is due to D. G. Austin (oral communication).

LEMMA 1. *Let $(g_{ij})$ be a matrix of nonnegative functions on $\mathbf{T}^0$ satisfying the condition that for every $i$,*

$$\lim_{t \downarrow 0} g_{ii}(t) = 1. \tag{3.1}$$

*Let $\{f_j\}$ be nonnegative functions satisfying the following equations:*

$$\begin{aligned} f_j(s+t) &= \sum_i f_i(s) g_{ij}(t), \qquad j \in \mathbf{I}, \\ [\text{or} \quad f_i(s+t) &= \sum g_{ij}(s) f_j(t), \qquad i \in \mathbf{I}]. \end{aligned} \tag{3.2}$$

*Then each $f_j$ is continuous in $\mathbf{T}$.*

PROOF. It is proved in theorem II. 2.3 of [1] that each $f_j$ is left-continuous and has a finite right-hand limit $f_j(t+0) \geqq f_j(t)$ for every $t \in \mathbf{T}^0$, and that $f_j(0+)$ exists. Such a function has at most a denumerable set $D$ of discontinuities. If $D$ is not empty, let $t_0 \in D$ so that $f_j(t_0 + 0) > f_j(t_0)$. Then there exist $\epsilon$ and $\delta_0$ such that $f_j(t_0 + \delta) > f_j(t_0) + \epsilon$ if $0 < \delta < \delta_0$. There exists $s_0$ such that $g_{jj}(s) > 1/2$ if $0 < s < s_0$. Thus

$$f_j(t_0 + s + \delta) = \sum_i f_i(t_0 + \delta) g_{ij}(s) > [f_j(t_0) + \epsilon] g_{jj}(s) + \sum_{i \neq j} f_i(t_0 + \delta) g_{ij}(s). \tag{3.3}$$

Letting $\delta \downarrow 0$ and using Fatou's lemma, we have

$$f_j(t_0 + s + 0) > \frac{\epsilon}{2} + \sum_i f_i(t_0)g_{ij}(s) = \frac{\epsilon}{2} + f_j(t_0 + s). \tag{3.4}$$

Hence all points in $(t_0, t_0 + s_0)$ belong to $D$, a contradiction which proves the first part of the lemma. The second part is proved in the same way.

As a corollary we see that all $p_{ij}$ satisfying (1.1), (1.2) and (1.3) are continuous in $\mathbf{T}$, without recourse to the condition (1.4). We remark however that with (1.4) or (2.1) each $p_{ij}$ will be uniformly continuous in $\mathbf{T}$, which is not necessarily the case without it.

The second lemma is implicit in some previous work (see, for example, theorem II. 3.2 of [1]) but will be stated in a general form.

LEMMA 2. *Let* $(p_{ij})$ *be a matrix of functions satisfying* (1.1), (1.2), *and* (1.3). *Let* $\{F_j\}$ *be nonnegative, nondecreasing functions satisfying the equations*

$$\begin{aligned} &F_i(s + t) - F_i(s) = \sum_j p_{ij}(s)F_j(t), \qquad i \in \mathbf{I}, \\ &[\text{or}\quad F_j(s + t) - F_j(t) = \sum_i F_i(s)p_{ij}(t), \qquad j \in \mathbf{I}]. \end{aligned} \tag{3.5}$$

*Then each* $F_i$ *has a continuous derivative* $F'_i$ *satisfying*

$$\begin{aligned} &F'_i(s + t) = \sum_j p_{ij}(s)F'_j(t), \qquad t \in \mathbf{T}^0, \\ &[\text{or}\quad F'_j(s + t) = \sum_i F'_i(s)p_{ij}(t), \qquad s \in \mathbf{T}^0]. \end{aligned} \tag{3.6}$$

REMARK. Taking the obvious differences, we see that the condition (3.5) is equivalent to the following: for any $t_1$ and $t_2$,

$$F_i(s + t_2) - F_i(s + t_1) = \sum_j p_{ij}(s)[F_j(t_2) - F_j(t_1)]. \tag{3.7}$$

PROOF. (A more elegant proof of this lemma has been given by Neveu [7].) By a theorem of Fubini on differentiation, we have for each $s$ and almost all $t$,

$$F'_i(s + t) = \sum_j p_{ij}(s)F'_j(t), \tag{3.8}$$

where $F'_j$ denotes an almost everywhere derivative. Hence by Fubini's theorem on product measures, (3.8) is also true if $t \notin Z$ and $s \notin Z(t)$ where $Z$ and $Z(t)$ are sets of Lebesgue measure zero. On the other hand we have by monotonicity and Fatou's lemma

$$F'_i(s + t) \geqq \sum_j p_{ij}(s)F'_j(t) \tag{3.9}$$

for every $s$ and $t$, if we agree now to take $F'_j$ as the right-hand lower derivate. Let $t_0 \notin Z$ and suppose for a certain $s_0$ we have

$$F'_i(s_0 + t_0) > \sum_j p_{ij}(s_0)F'_j(t_0). \tag{3.10}$$

Then it follows that if $s > s_0$, since $p_{ii}(t) > 0$ for all $t$,

$$\begin{aligned} F'_i(s + t_0) &\geqq \sum_j p_{ij}(s - s_0)F'_j(s_0 + t_0) > \sum_j p_{ij}(s - s_0) \sum_k p_{jk}(s_0)F'_k(t_0) \\ &= \sum_k p_{ik}(s)F'_k(t_0). \end{aligned} \tag{3.11}$$

This is impossible by the second sentence of the proof; hence (3.8) must hold for all $s$, if $t_0 \notin Z$. For an arbitrary $t > 0$, let $t = t_0 + t_1$ where $t_0 \notin Z$. It follows that

$$
\begin{aligned}
(3.12) \qquad F_i'(s+t) &= F_i'(s+t_1+t_0) = \sum_j p_{ij}(s+t_1)F_j'(t_0) \\
&= \sum_j \sum_k p_{ik}(s)p_{kj}(t_1)F_j'(t_0) = \sum_k p_{ik}(s)F_k'(t_1+t_0) \\
&= \sum_k p_{ik}(s)F_k'(t).
\end{aligned}
$$

Hence (3.8) holds for all $t > 0$, $s \geqq 0$. By the second part of lemma 1, each $F_j'$ is continuous and consequently $F_j$ has a continuous derivative. This proves the first part of lemma 2. The second part is proved in the same way.

## 4. Review of the strong Markov property

For a detailed discussion, see II. 8–9 of [1]. The reading of this section may be postponed until it becomes necessary.

Let $\{x_t\}$ be the M.C. defined in section 1. We denote by $\mathfrak{F}_t$ the augmented Borel field generated by $\{x_s, s \leqq t\}$. Let $\alpha$ be a nonnegative random variable with domain of definition $\Omega_\alpha$, where $\mathbf{P}(\Omega_\alpha) > 0$, which is "independent of the future," namely

$$(4.1) \qquad \{\omega : \alpha(\omega) < t\} \in \mathfrak{F}_t$$

for every $t \in \mathbf{T}^0$. Such a random variable will be called *optional.* The Borel field of sets $\Lambda$ (in $\mathfrak{F}$) such that for every $t$ we have

$$(4.2) \qquad \Lambda \cap \{\omega : \alpha(\omega) < t\} \in \mathfrak{F}_t$$

will be denoted by $\mathfrak{F}_\alpha$, the "past field relative to $\alpha$." Let

$$(4.3) \qquad y(t, \omega) = x[\alpha(\omega) + t, \omega], \qquad t \in \mathbf{T}^0.$$

It follows from (1.9) that $y_t = y(t, \cdot)$ with domain $\Omega_\alpha$ is a random variable. The process $\{y_t, t \in \mathbf{T}^0\}$ will be called the *post-$\alpha$ process* and the augmented Borel field it generates will be denoted by $\mathfrak{F}_\alpha'$, "the future field relative to $\alpha$." For any $\Lambda \in \mathfrak{F}_\alpha$ we put

$$(4.4) \qquad A(\Lambda; t) = \mathbf{P}\{\Lambda; \alpha(\omega) \leqq t\}.$$

The measure corresponding to this distribution function will be called the $A(\Lambda; \cdot)$ measure.

The following collection of assertions, valid for each optional $\alpha$, will be referred to as the *strong Markov property.*

(1) For every $\Lambda \in \mathfrak{F}_\alpha$ and $\mathrm{M} \in \mathfrak{F}_\alpha'$ we have

$$(4.5) \qquad \mathbf{P}\{\Lambda \mathrm{M} | y_0\} = \mathbf{P}\{\Lambda | y_0\}\mathbf{P}\{\mathrm{M} | y_0\}$$

almost everywhere on the set $\{\omega : y_0(\omega) \in \mathbf{I}\}$.

(2) The post-$\alpha$ process $\{y_t, t \in \mathbf{T}^0\}$ is an open M.C. which has the properties

corresponding to (1.8) and (1.9), and whose transition matrix is a part of $(p_{ij})$. In particular, $\{y_t, t \in \mathbf{T}\}$ is a M.C. on the set $\{\omega : y_0(\omega) \in \mathbf{I}\}$.

(3) For each $j \in \mathbf{I}$, $\Lambda \in \mathfrak{F}_\alpha$ and almost every $s \in (0, t)$ with respect to the $A(\Lambda; \cdot)$ measure, we have

(4.6) $$\mathbf{P}\{x(t, \omega) = j|\Lambda; \alpha(\omega) = s\} = \mathbf{P}\{y(t - s, \omega) = j|\Lambda; \alpha(\omega) = s\}.$$

One version of the conditional probability in (4.6), to be denoted by $r_j(s, t|\Lambda)$, is continuous in $t \in [s, \infty)$ for each $s \in \mathbf{T}$.

The following particular case of the strong Markov property, to be referred to as the *strongest Markov property*, will be applied in the sequel. The two fields $\mathfrak{F}_\alpha$ and $\mathfrak{F}'_\alpha$ are said to be independent iff for every $\Lambda \in \mathfrak{F}_\alpha$ and $\mathrm{M} \in \mathfrak{F}'_\alpha$ we have

(4.7) $$\mathbf{P}\{\Lambda\mathrm{M}|\Omega_\alpha\} = \mathbf{P}\{\Lambda|\Omega_\alpha\}\mathbf{P}\{\mathrm{M}|\Omega_\alpha\};$$

alternately, since $\Lambda \in \Omega_\alpha$,

(4.8) $$\mathbf{P}\{\Lambda\mathrm{M}\} = \mathbf{P}\{\Lambda\}\mathbf{P}\{\mathrm{M}|\Omega_\alpha\}.$$

(4) The fields $\mathfrak{F}_\alpha$ and $\mathfrak{F}'_\alpha$ are independent if and only if there exist functions $\{\rho_j\}$ on $\mathbf{T}^0$ such that for every $j \in \mathbf{I}$, $t \in \mathbf{T}^0$ and $\Lambda \in \mathfrak{F}_\alpha$ we have

(4.9) $$r_j(s, t|\Lambda) = \rho_j(t - s)$$

for almost all $s$ in $(0, t)$ with respect to the $A(\Lambda; \cdot)$ measure. We have then

(4.10) $$\rho_j(t) = \mathbf{P}\{y(t, \omega) = j|\Omega_\alpha\}$$

and $\rho_j$ is continuous in $\mathbf{T}$.

In particular, this is the case if for a fixed $j$ we have

(4.11) $$\mathbf{P}\{y(0, \omega) = j|\Omega_\alpha\} = 1.$$

## 5. Transition from and to a stable state

Let us introduce the following notation:

(5.1) $$-p'_{ii}(0) = \lim_{t \downarrow 0} \frac{1 - p_{ii}(t)}{t} = -q_{ii} = q_i \leqq \infty,$$

(5.2) $$p'_{ij}(0) = \lim_{t \downarrow 0} \frac{p_{ij}(t)}{t} = q_{ij} < \infty, \qquad i \neq j.$$

That these limits exist and have the indicated finiteness is well known (theorems II. 2.4 and II. 2.5 of [1]). Analytically, (5.1) follows from the subadditivity of $-\log p_{ii}(t)$ which is a consequence of (1.1), (1.2), and (1.3) without the intervention of (1.4). The corresponding basic property of sample functions is given in the formula

(5.3) $$\mathbf{P}_i\{x(s, \omega) \equiv i, 0 < s < t\} = e^{-q_i t},$$

where the right member stands for 0 if $q_i = \infty$ and $t > 0$. The state $i$ is called *stable* or *instantaneous* according as $q_i < \infty$ or $q_i = \infty$.

In the rest of this section let $i$ be fixed and $p_i = 1$ so that $\mathbf{P} = \mathbf{P}_i$. Define on $\Delta_i$ the *first exit time from* $i$:

$$\alpha(\omega) = \inf \{t : t > 0, x(t, \omega) \neq i\}. \tag{5.4}$$

Then (5.3) is equivalent to the assertion that $\alpha$ is a random variable with the distribution function

$$e_{q_i}(t) \overset{\text{def}}{=} 1 - e^{-q_i t}, \qquad t \in \mathbf{T}^0, \tag{5.5}$$

which reduces to the unit distribution $\epsilon$ if $q_i = \infty$. It is easy to see that $\alpha$ is optional. It may or may not be easy to see that $\mathfrak{F}_\alpha$ and $\mathfrak{F}'_\alpha$ are independent in the sense of (4) of section 4. For a tedious but rigorous proof of this fact, see theorem II. 15.2 of [1]; a partially analytic proof will be given later.

We have as a trivial identity valid for any $\alpha$:

$$\mathbf{P}\{x(t, \omega) = j\} = \mathbf{P}\{\alpha(\omega) \leqq t; x(t, \omega) = j\} + \mathbf{P}\{\alpha(\omega) > t; x(t, \omega) = j\}. \tag{5.6}$$

Now let $i$ be a stable state. The second term above is $\delta_{ij} \exp(-q_i t)$ by (5.3). The first term may be written as

$$\int_0^t \mathbf{P}\{x(t, \omega) = j | \alpha(\omega) = s\} \, d\mathbf{P} \{\alpha(\omega) \leqq s\} \tag{5.7}$$

by the definition of conditional probability. By (4) of section 4, and writing $r_{ij}$ for the $\rho_j$ then we see that (5.6) becomes

$$p_{ij}(t) = \int_0^t r_{ij}(t - s) q_i e^{-q_i s} \, ds + e^{-q_i t}\delta_{ij}. \tag{5.8}$$

Furthermore by (2) of section 4, we have

$$r_{ik}(t + s) = \sum_j r_{ij}(t) p_{jk}(s), \tag{5.9}$$

$$\sum_j r_{ij}(t) = 1. \qquad k \in \mathbf{I}; s, t \in \mathbf{T}^0. \tag{5.10}$$

The above formulas give an integral representation of $p_{ij}$ obtained by a precise analysis of the local behavior of a sample function at the exit from the stable state $i$. It is a clear example of the probabilistic method in reaching analytic conclusions.

For it follows from (5.9) and lemma 1 that $r_{ij}$ is continuous in $\mathbf{T}$. It is then an immediate consequence of (5.8) that $p_{ij}$ has a continuous derivative $p'_{ij}$ satisfying the following:

$$e^{-q_i t} \frac{d}{dt} [e^{q_i t} p_{ij}(t)] = p'_{ij}(t) + q_i p_{ij}(t) = q_i r_{ij}(t). \tag{5.11}$$

It follows furthermore from (1.2), (1.4), (5.9), and (5.10) that

$$\sum_j p'_{ij}(t) = 0, \tag{5.12}$$

$$\sum_j |p'_{ij}(t)| \leqq 2q_i, \qquad t \in \mathbf{T}^0, \tag{5.13}$$

$$\sum_k p'_{ik}(t) p_{kj}(s) = p'_{ij}(t + s), \tag{5.14}$$

namely that both the series in (1.2) and (1.4) can be differentiated term by term in $\mathbf{T}^0$ to yield absolutely convergent series—a by no means trivial analytical fact. Our proof shows that this is tied up with the fact that the post-$\alpha$ process is Markovian with the same transition matrix (curtailed). The critical case for $t = 0$ will be examined later in section 7.

The formula (5.8) has a dual which will be briefly discussed. Let $j$ be stable and $i$ arbitrary, then we have

$$p_{ij}(t) = \delta_{ij}e^{-q_it} + \int_0^t v_{ij}(s)e^{-q_j(t-s)}\,ds. \tag{5.15}$$

The function $v_{ij}$ represents a renewal density function; precisely $v_{ij}$ is the derivative of $V_{ij}$ where $V_{ij}(t)$ is the expected number of entrances into the state $j$ in the open interval $(0, t)$, under the hypothesis $p_i = 1$. Using the notation of section 6, we have in fact

$$V_{ij}(t) = \sum_{n=0}^{\infty} [F_{ij} * F_{jj}^{n*}](t), \tag{5.16}$$

where $*$ denotes the convolution of distribution functions, $F_{jj}^{0*} = \epsilon$, and $F_{jj}^{(n+1)*} = F_{jj}^{n*} * F_{jj}$; but this explicit formula will not be needed. From the probabilistic meaning we infer that

$$V_{ij}(s + t) - V_{ij}(s) = \sum_k p_{ik}(s)V_{kj}(t). \tag{5.17}$$

The existence of the continuous derivative $v_{ij}$ follows from (5.17) and lemma 2. (This is a better approach than that in section II. 16 of [1].) Furthermore it follows from (5.14) that

$$\frac{d}{dt}[p_{ij}(t)e^{q_jt}]e^{-q_jt} = p'_{ij}(t) + p_{ij}(t)q_j = v_{ij}(t); \tag{5.18}$$

$$\sum_k p_{ik}(s)p'_{kj}(t) = p'_{ij}(s + t), \tag{5.19}$$

where the series converges absolutely.

Having deduced the preceding results by probabilistic methods, we are now ready for an analytic short cut based on hindsight. The fact that $(\exp q_it)p_{ij}(t)$ is nondecreasing in $t$, as shown in (5.11), can be proved directly as follows. Since $p_{ii}(h) \geqq \exp(-q_ih)$ by the subadditivity mentioned in connection with (5.1) [or probabilistically as a consequence of (5.3)], we have

$$e^{q_i(t+h)}p_{ij}(t + h) \geqq e^{q_ih}p_{ii}(h)e^{q_it}p_{ij}(t) \geqq e^{q_it}p_{ij}(t). \tag{5.20}$$

Let $P_{ij}(t) = \int_0^t p_{ij}(s)\,ds$. Then we have by partial integration,

$$p_{ij}(t) - \delta_{ij} + q_iP_{ij}(t) = \int_0^t e^{-q_is}D[e^{q_is}p_{ij}(s)]\,ds, \tag{5.21}$$

where $D$ denotes an almost everywhere derivative. Since this derivative is nonnegative, the left member of (5.21) is a nondecreasing function of $t$. Now a trivial calculation based on (1.2) yields

$$(5.22)\qquad \sum_j [p_{ij}(t) - \delta_{ij} + q_i P_{ij}(t)] p_{jk}(s) = p_{ik}(t+s) + q_i P_{ik}(t+s) - p_{ik}(s) - q_i P_{ik}(s).$$

Thus the conditions for the second part of lemma 2 are satisfied if we take $F_j(t)$ to be the left member of (5.21). It follows that $p_{ij}$ has a continuous derivative satisfying (5.14). In an exactly dual way (5.19) can be proved. We remark also that neither proof utilizes (1.4).

As far as the analytic part is concerned, the above approach is the simplest. We can now retrace our steps to *define* $r_{ij}$ by means of the second equation in (5.11), verify (5.8) as a consequence, and using (4) of section 4, conclude that the two fields $\mathfrak{F}_\alpha$ and $\mathfrak{F}'_\alpha$ are independent.

We add the following remarks before turning to another illustration of this kind. The rather complete success of the methods developed in this section depends on the primary fact that the *set of constancy*,

$$(5.23)\qquad S_i(\omega) = \{t : x(t, \omega) = i\}$$

for a fixed stable $i$, consists of a sequence of disjoint intervals without clustering in the finite (theorem II. 5.7 of [1]). Thus the endpoints of these intervals form natural relay points in the analysis of the sample functions, with the length of an interval (sojourn time) corresponding analytically to the smoothing exponential factor exp $(\pm q_i t)$. It is not known whether suitable substitutes for (5.11) and (5.18), or (5.8) and (5.15), exist in the general case where both $i$ and $j$ are arbitrary. On the other hand, it has been proved by D. Ornstein [8] (see also Jurkat [5] and the appendix in [1]) that the equations (5.12), (5.14), and (5.19) remain valid in the general case. This can be proved by the development in the next section.

## 6. First entrance and last exit

Let $i \neq j$ and let $\Delta_{ij}$ be the subset of $\Delta_i$ where the following infimum is finite:

$$(6.1)\qquad \alpha_{ij}(\omega) = \inf\{t : t > 0, x(t, \omega) = j\}.$$

It is verified that $\alpha_{ij}$ is an optional random variable, and in view of the last sentence in section 4, the strongest Markov property applies with $y(0, \omega) = j$ on $\Delta_{ij}$, and the $\rho_j$ in (4.10) reducing to $p_{jj}$ (in general $\rho_k = p_{jk}$). Now if $\alpha = \alpha_{ij}$ in (5.6), the second term vanishes by definition and we obtain, by what has just been said,

$$(6.2)\qquad p_{ij}(t) = \int_0^t p_{jj}(t - s)\, dF_{ij}(s),$$

where

$$(6.3)\qquad F_{ij}(t) = \mathbf{P}_i\{\alpha_{ij}(\omega) \leqq t\}.$$

It is easy to see that $F_{ij}$ is continuous in $\mathbf{T}$ but more will be shown presently. The formula (6.2) is the *first entrance formula from $i$ to $j$*. The definitions (6.1) and (6.3) may be extended to the case $i = j$, yielding $F_{ii}(t) \equiv 1$. The last definition, as well as (6.4) below, differs from that given in section II.11 of [1] but the latter agrees with that in the appendix there.

To proceed further we must introduce the *taboo probability functions*

$$(6.4)\qquad {}_jp_{ik}(t) = \mathbf{P}_i\{x(t, \omega) = k;\, x(s, \omega) \neq j,\, 0 < s < t\}.$$

It follows from the stochastic continuity of the M.C. [equivalent to condition (1.3)] that ${}_jp_{ik}(t) \equiv 0$ if $i = j$ or $k = j$. These probabilities are well defined on account of the separability of the process. We observe that

$$(6.5)\qquad F_{ij}(t) = 1 - \sum_k {}_jp_{ik}(t), \qquad i \neq j.$$

For fixed $j$, the matrix $({}_jp_{ik})$ with $i$ and $k$ in $\mathbf{I} - \{j\}$, is a substochastic transition matrix:

$$(6.6)\qquad \sum_k {}_jp_{ik}(t)\,{}_jp_{kl}(s) = {}_jp_{il}(t + s).$$

It is unnecessary to exclude $j$ from the summation since the corresponding term vanishes. For this substochastic transition matrix, $F_{ij}$ plays the role of $p_{i\theta}$ in section 2. It follows at once [compare (2.3)] that

$$(6.7)\qquad F_{ij}(s + t) - F_{ij}(s) = \sum_k {}_jp_{ik}(s)F_{kj}(t), \qquad i \neq j.$$

Hence an application of lemma 2 shows that each $F_{ij}$ has a continuous derivative $f_{ij}$ satisfying

$$(6.8)\qquad f_{ij}(s + t) = \sum_k {}_jp_{ik}(s)f_{kj}(t),$$

and consequently (6.2) can be improved into

$$(6.9)\qquad p_{ij}(t) = \int_0 f_{ij}(s)p_{jj}(t - s)\, ds, \qquad i \neq j.$$

It turns out that the formula (6.9) has a dual which has been proved in general only recently (the case where $i$ is stable being previously known). To motivate this dualization it is best to consider the discrete parameter analogues.

For each $h \in \mathbf{T}^0$ the stochastic process $\{x_{nh}, n \in \mathbf{N}\}$ is called the *discrete skeleton* of $\{x_t, t \in \mathbf{T}\}$ at the scale $h$. It is a discrete parameter homogeneous Markov chain with the *n-step transition matrix* $(p_{ij}^{(n)})$. Let

$$(6.10)\qquad {}_jp_{ik}^{(n)}(h) = \mathbf{P}_i\{x(nh, \omega) = k,\, x(\nu h, \omega) \neq j,\, 1 \leqq \nu \leqq n - 1\}$$

be the corresponding taboo probabilities. The analogue of (6.9) is then

$$(6.11)\qquad p_{ij}^{(n)}(h) = \sum_{\nu=1}^{n} {}_jp_{ij}^{(\nu)}(h)p_{jj}^{(n-\nu)}(h), \qquad n \geqq 1,$$

where ${}_jp_{ij}^{(\nu)}(h)$ may be denoted by $f_{ij}^{(\nu)}(h)$ for comparison with (6.9) but is preferably written as shown with a view to dualization. This is a very old formula and is basic in the so-called theory of "recurrent events" (see section I. 8 of [1]). Now in the discrete parameter case the reasoning leading to (6.11) can be immediately dualized by interchanging "$i$" and "$j$," "first" and "last," "entrance" and "exit," to yield the dual:

$$p_{ij}^{(n)}(h) = \sum_{\nu=0}^{n-1} p_{ii}^{(\nu)}(h)\,{}_ip_{ij}^{(n-\nu)}(h), \qquad n \geqq 1. \tag{6.12}$$

These two formulas (6.11) and (6.12), valid also for $i = j$, are particular cases of theorem I. 9.1 of [1]. Since the taboo probabilities can be defined algebraically, they appear as simple algebraic consequences of the operation of matrix multiplication, apart from questions of convergence. Now if (1.4) or the weaker (2.1) holds, then

$$\sum_{n=1}^{\infty} {}_jp_{ij}^{(n)}(h) \leqq 1, \tag{6.13}$$

which greatly facilitates the passage to limit in (6.11) as $h \downarrow 0$. The same however cannot be said of the series $\sum_{n=1}^{\infty} {}_ip_{ij}^{(n)}(h)$. Thus it is desirable to execute the limit operation without the advantage of (1.4), but making defter use of (1.3). The main idea is to consider a *sequence* of $h \downarrow 0$ such that

$$\sum_{nh \leqq t} {}_jp_{ij}^{(n)}(h) \quad \text{and} \quad \sum_{nh \leqq t} {}_ip_{ij}^{(n)}(h) \tag{6.14}$$

converge for a dense set of $t$, in the manner of Helly's selection principle. This is carried out by Jurkat [5] with a further refinement.

While this method has analytic power, it is unfortunately devoid of probabilistic meaning at the moment. We shall sketch two different approaches based on considerations of sample functions.

Since (6.9) is obtained by analyzing the first entrance into the final state $j$, it is natural to reflect upon the last exit from the initial state $i$. Let us define on $\Delta_i$:

$$\gamma_i(t, \omega) = \sup\{s : 0 \leqq s \leqq t, x(s, \omega) = i\}. \tag{6.15}$$

For each fixed $t$ this is a random variable but clearly it is not optional in any sensible way: to determine if $\gamma_i(t, \omega) \leqq s$ we must know $x(\cdot, \omega)$ up to the time $t$. On the other hand, its distribution function is easily written down, if $0 \leqq s < t$,

$$\Gamma_i(s, t) \overset{\text{def}}{=} \mathbf{P}_i\{\gamma_i(t, \omega) \leqq s\} = \sum_k p_{ik}(s)[1 - F_{ki}(t - s)]. \tag{6.16}$$

Furthermore, for every $j \neq i$ we have

$$\Gamma_{ij}(s, t) \overset{\text{def}}{=} \mathbf{P}_i\{\gamma_i(t, \omega) \leqq s; x(t, \omega) = j\} = \sum_k p_{ik}(s)\,{}_ip_{kj}(t - s) \tag{6.17}$$

so that, for $0 \leqq s < t$, we have

$$\Gamma_i(s, t) = \sum_{j \neq i} \Gamma_{ij}(s, t). \tag{6.18}$$

For $s = t$ the above equation becomes false. We have

$$p_{ij}(t) = \int_0^t \mathbf{P}_i\{x(t, \omega) = j | \gamma_i(t, \omega) = s\}\, d_s \Gamma_i(s, t). \tag{6.19}$$

Now the salient fact here is that the conditional probability in (6.19) turns out to be a function of $t - s$ only, while the distribution function $\Gamma_i(s, t)$ has a density function which is the product of a function of $t - s$ and one of $s$ only.

To demonstrate these facts by our first method, we decompose the sample functions $x(\cdot, \omega)$ with $x(0, \omega) = i$ and $x(t, \omega) = j$ into subsets according to the location of $\gamma_i(t, \omega)$. To be precise, for each $n$ let $\gamma_i^{(n)}(t, \omega)$ be the unique dyadic number $(\nu - 1)2^{-n}$ such that

$$x[(\nu - 1)2^{-n}, \omega] = i \quad \text{and} \quad x(u, \omega) \neq i, \qquad \nu 2^{-n} \leqq u \leqq t. \tag{6.20}$$

We have $\lim_{n\to\infty} \gamma_i^{(n)}(t, \omega) = \gamma_i(t, \omega)$ by separability, and consequently

$$\begin{aligned} p_{ij}(t) &= \lim_{n\to\infty} \sum_{\nu \leqq 2^n t} \mathbf{P}_i\{\gamma_i^{(n)}(t, \omega) = (\nu - 1)2^{-n}; x(t, \omega) = j\} \\ &= \lim_{n\to\infty} \sum_{\nu \leqq 2^n t} p_{ii}((\nu - 1)2^{-n}) \sum_k p_{ik}(2^{-n})\, {}_ip_{kj}(t - \nu 2^{-n}). \end{aligned} \tag{6.21}$$

The last written sum may be exhibited as

$$\int_0^t \phi_{ij}^{(n)}(t - s)\, d\, \pi_i^{(n)}(s) \tag{6.22}$$

where

$$\pi_i^{(n)}(s) = \sum_{\nu \leqq 2^n s} p_{ii}((\nu - 1)2^{-n})2^{-n}, \qquad \phi_{ij}^{(n)}(s) = 2^n \sum_k p_{ik}(2^{-n})\, {}_ip_{kj}(s). \tag{6.23}$$

Clearly,

$$\lim_{n\to\infty} \pi_i^{(n)}(t) = \int_0^t p_{ii}(s)\, ds. \tag{6.24}$$

Hence it remains to show that $\phi_{ij}^{(n)}(s)$ converges *uniformly* in every finite interval to $g_{ij}(s)$ in order to obtain in the limit the desired formula:

$$p_{ij}(t) = \int_0^t g_{ij}(t - s) p_{ii}(s)\, ds. \tag{6.25}$$

By the definition of $\phi_{ij}^{(n)}(s)$,

$$\sum_j g_{ij}(s)\, {}_ip_{jk}(t) = g_{ik}(s + t), \tag{6.26}$$

and so by lemma 1 all $q_{ij}$ are continuous in $\mathbf{T}$. The convergence of $\phi_{ij}^{(n)}$ follows from properties of taboo probability functions, only the uniformity causes some technical difficulty. This plan of attack has been carried out in detail in [1]. The purpose of the résumé above is to show the basic probabilistic idea underlying this method.

Our second method shows promise of general applicability, being inherent in the nature of the stochastic scheme of things. It is that of reversing the direction of time, or retracing the process. Formally let $U \in \mathbf{T}^0$ and define

$$z^U(t, \omega) = x(U - t, \omega), \qquad 0 \leqq t \leqq U. \tag{6.27}$$

The new process $\{z_t^U, 0 \leqq t \leqq U\}$ is Markovian with the state space $\mathbf{I}$, but has in general *nonstationary* transition probabilities. This is one difficulty to be faced in this approach, the other one being the dependence on $U$. But these difficulties may also give us new clues.

For the sake of simplicity let us suppose that $p_i = 1$. Then if $0 \leqq s \leqq t \leqq U$, we have

$$(6.28)\qquad p^U(s,t;j,i) \overset{\text{def}}{=} \mathbf{P}\{z^U(t,\omega) = i | z^U(s,\omega) = j\} = \frac{p_{ii}(U-t)}{p_{ij}(U-s)}\, p_{ij}(t-s).$$

The first entrance time distribution form $j$ to $i$, starting at time $s$, is also easily written down:

$$(6.29)\qquad F^U(s,t;j,i) \overset{\text{def}}{=} \mathbf{P}\{z^U(u,\omega) = i \quad \text{for some} \quad u \in [s,t] | z^U(s,\omega) = j\}$$
$$= 1 - \frac{1}{p_{ij}(U-s)} \sum_k p_{ik}(U-t)\,{}_ip_{kj}(t-s)$$
$$= 1 - \frac{1}{p_{ij}(U-s)} \Gamma_{ij}(U-t, U-s).$$

Now the reversed Markov chain (if the proper version is taken) also possesses a strong Markov property, a particular case of which is the first entrance formula generalizing (6.2),

$$(6.30)\qquad p^U(s,t;j,i) = \int_s^t p^U(u,t;i,i)\, d_u F^U(s,u;j,i).$$

For a proof of this see [2]. Substituting from (6.28) and (6.29) we obtain

$$(6.31)\qquad p_{ij}(t-s) = \int_t p_{ii}(t-u) \frac{d_u\, \Gamma_{ij}(U-u, U-s)}{p_{ii}(U-u)},$$

or

$$(6.32)\qquad p_{ij}(t) = \int_t^0 p_{ii}(t-u) \frac{d_u\, \Gamma_{ij}(U-s-u, U-s)}{p_{ii}(U-s-u)},$$

if $t \leqq U - s$. This being so it is reasonable to conjecture that the measures in $u$ generated by $\Gamma_{ij}(U-u, U)/p_{ii}(U-u)$ for different values of $U - u$ coincide, namely, there exists a nondecreasing function $G_{ij}$ on $\mathbf{T}$ such that

$$(6.33)\qquad \int_{u_2}^{u_1} \frac{d_u\, \Gamma_{ij}(U-u, U)}{p_{ii}(U-u)} = \int_{u_1}^{u_2} d\, G_{ij}(u)$$

for $0 \leqq u_1 \leqq u_2 \leqq U$. This is indeed true by a known, though formidable, theorem due to Titchmarsh [10], p. 328. (For a proof by real variable method, see Mikusinski [6], chapter 7.) We have by (6.17) and (6.6)

$$(6.34)\qquad \sum_j \Gamma_{ij}(U-u, U)\,{}_ip_{jk}(s) = \Gamma_{ik}(U-u, U+s).$$

Hence for $0 \leqq u_2 \leqq u_1 \leqq U - s$,

$$(6.35)\qquad \int_{u_2}^{u_1} \sum_j \frac{d_u\, \Gamma_{ij}(U-u, U)}{p_{ii}(U-u)}\,{}_ip_{jk}(s) = \int_{u_2}^{u_1} \frac{d_u\, \Gamma_{ik}(U-u, U+s)}{p_{ii}(U-u)}$$
$$= \int_{s+u_2}^{s+u_1} \frac{d_u\, \Gamma_{ik}(U+s-u, U+s)}{p_{ii}(U+s-u)}$$
$$= \int_{s+u_2}^{s+u_1} \frac{d_u\, \Gamma_{ik}(U-u, U)}{p_{ii}(U-u)},$$

where the last equation follows from (6.33). This is equivalent to

$$\sum_j [G_{ij}(u_2) - G_{ij}(u_1)]_i p_{jk}(s) = G_{ik}(s + u_2) - G_{ik}(s + u_1). \tag{6.36}$$

Thus by lemma 2 (see the remark there) each $G_{ij}$ has a continuous derivative $g_{ij}$ in $\mathbf{T}$ satisfying (6.26). Substituting back into (6.32) we obtain (6.25).

Incidentally, we have shown that $\Gamma_{ij}(U - u, U)$ has a derivative with respect to $u$ in $[0, U]$ which is equal to $p_{ii}(U - u)g_{ij}(u)$, verifying the remark after (6.19). This can also be deduced from (6.25) and (6.26) since by (6.17) we have

$$\begin{aligned} \Gamma_{ij}(s, t) &= \sum_k \int_0^s p_{ii}(u) g_{ik}(s - u)_i p_{kj}(t - s)\, du \\ &= \int_0^s p_{ii}(u) g_{ij}(t - u)\, du. \end{aligned} \tag{6.37}$$

Summing (6.25) over all $j \neq i$, we see that

$$1 - p_{ii}(t) = \int_0^t g_i(t - s) p_{ii}(s)\, ds, \tag{6.38}$$

where $g_i = \sum_{j \neq i} g_{ij}$. This integral equation for $p_{ii}$ can be made as the starting point of another proof of Ornstein's theorem [8] on the continuous differentiability in $\mathbf{T}^0$ of all $p_{ij}$ of a transition matrix. Such a proof is given by Jurkat [5] without the use of (1.4). He has also indicated a proof which is based on (6.9) instead of (6.25). It can be shown moreover that the series in (5.14) and (5.19) converge absolutely in $\mathbf{T}^0$ [without the condition (1.4)] and so does that in (5.12) if (1.4) is assumed. However the following problem is open: if $i$ is instantaneous, is it true that

$$\lim_{t \downarrow 0} p'_{ii}(t) = -q_i = -\infty? \tag{6.39}$$

The answer is "yes" if $i$ is stable, as a consequence of (5.11) and the existence of $r_{ii}(0+) = 0$. This problem is particularly interesting since almost every sample function $x(\cdot, \omega)$ with $x(0, \omega) = i$ "oscillates tremendously" at $t = 0$, while it is not even known if $p_{ii}$ is monotone in a neighborhood of zero.

I take this opportunity to correct an oversight (p. 270 of [1], lines 4 to 5) brought to my attention by Reuter. For every $i$ and $j$, we have

$$\lim_{t \to \infty} p'_{ij}(t) = 0. \tag{6.40}$$

This follows by fixing a positive $t$ in equation (27) there, let $s \to \infty$ according to theorem II. 10.1, and use the inequality in (28) to justify uniform convergence with respect to $s$. The existence of the limit in (6.38) implies that it is equal to zero.

## 7. The minimal chain

Returning to section 5, we now wish to study what happens at the exact moment of exit from a stable state $i$. Noting that (4.10) remains in force at $t = 0$ but, instead of (5.10), we have by Fatou's lemma

$$\mathbf{P}\{y(0, \omega) \in \mathbf{I} | \Omega_\alpha\} = \sum_j r_{ij}(0) \leqq 1. \tag{7.1}$$

Since

$$r_{ij}(0) = \frac{(1 - \delta_{ij})q_{ij}}{q_i} \tag{7.2}$$

by (5.2) and (5.11), this amounts to the easy analytic result

$$\sum_{j \neq i} q_{ij} \leqq q_i. \tag{7.3}$$

If strict inequality holds above, then with a probability equal to $1 - \sum_j r_{ij}(0) > 0$ we have $y(0, \omega) = \infty$. We recall that $\infty$ is the "point at infinity" adjoined to compactify $\mathbf{I}$ to render the process separable. For a general optional $\alpha$ and the post-$\alpha$ process $\{y_t\}$, $y(0, \omega) = \infty$ if and only if $\lim_{t \downarrow \alpha(\omega)} x(t, \omega) = \infty$, on account of (1.8). On the set of $\omega$ for which this is true the process $\{y_t\}$ does not have an initial distribution (on $\mathbf{I}$), and is a Markov chain only in $\mathbf{T}^0$; see (2) of section 4. It is important to note that

$$\mathbf{P}\{y(t, \omega) \in \mathbf{I} | \Omega_\alpha\} = 1, \qquad t \in \mathbf{T}^0, \tag{7.4}$$

is part of the assertion of the strong Markov property. The above conclusions may be stated as follows: at the first exit time $\alpha(\omega)$ from the stable state $i$, the probability of a *pseudojump to* $j$ ($\neq i$) is $r_{ij}(0) = q_{ij}/q_i$, and the probability of a *pseudojump to* $\infty$ is $1 - \sum_{j \neq i} (q_{ij}/q_i)$. We say "pseudojump" rather than "jump," since if $j$ is instantaneous the sample function does not have a jump in the usual sense but shows the following behavior,

$$\varliminf_{t \uparrow \alpha(\omega)} x(t, \omega) = j < \infty = \varlimsup_{t \downarrow \alpha(\omega)} x(t, \omega). \tag{7.5}$$

We have thus a complete analysis of the first discontinuity of a sample function which starts at a stable state. To continue this process, we shall assume that all states are stable and that equality holds in (7.1) or (7.3) so that a pseudojump to $j$ is a genuine jump and the possibility of a pseudojump to $\infty$ is excluded. Finally we suppose that there is no *absorbing state* to omit trivial modifications. These assumptions are summed up as follows:

$$0 < q_i = \sum_{j \neq i} q_{ij} < \infty, \qquad i \in \mathbf{I}. \tag{7.6}$$

The preceding analysis then implies, by an induction on the number of jumps, that there are infinitely many jumps of the sample function

$$\tau_1(\omega) < \cdots < \tau_n(\omega) < \cdots. \tag{7.7}$$

Let us put also $\tau_0(\omega) = 0$ and

$$\chi_n(\omega) = x(\tau_n(\omega), \omega) = \lim_{t \downarrow \tau_n(\omega)} x(t, \omega). \tag{7.8}$$

It is easy to verify that each $\tau_n$ is optional with $\mathbf{P}\{\Omega_{\tau_n}\} = 1$. (One may use in this connection theorem II. 15.1 of [1], but that is not necessary.) It follows from (1) of section 4 that

(7.9) $\mathbf{P}\{\chi_{n+1}(\omega) = i_{n+1}|\chi_\nu(\omega) = i_\nu, 0 \leqq \nu \leqq n\} = \mathbf{P}\{\chi_{n+1}(\omega) = i_{n+1}|\chi_n(\omega) = i_n\}$.

Applying the preceding analysis of the first discontinuity to the post-$\tau_n$ process, we see that the right member of (7.9) is equal to $r_{i_n i_{n+1}}(0)$. Hence $\{\chi_n, n \in N\}$ is a discrete parameter homogeneous Markov chain with the one-step transition matrix $(r_{ij}(0))$. Furthermore it follows from (5.3) applied to the post-$\tau_n$ process that

$$\mathbf{P}\{\tau_{n+1}(\omega) - \tau_n(\omega) \leqq t|\chi_\nu(\omega), 0 \leqq \nu \leqq n-1;\, \chi_n(\omega) = i\} = e_{q_i}(t). \tag{7.10}$$

Now let

$$\tau_\infty(\omega) = \lim_{n\to\infty} \tau_n(\omega). \tag{7.11}$$

Then it is clear from the definition that for almost all $\omega$,

$$\tau_\infty(\omega) = \sup \{t : x(\cdot, \omega) \text{ has only jumps in } (0, t)\}. \tag{7.12}$$

We call $\tau_\infty$ the *first infinity* of the M.C. Since $\{\omega : \tau_n(\omega) < t\} \in \mathfrak{F}_t$ by the definition of optionality, we have $\{\omega : \tau_\infty(\omega) < t\} \in \mathfrak{F}_t$ by (7.11). Hence $\tau_\infty$ is optional. Let

$$L_i(t) = \mathbf{P}_i\{\tau_\infty(\omega) \leqq t\}. \tag{7.13}$$

Let $\Theta(t_1, t_2)$ denote the set $\{\omega : x(\cdot, \omega)$ has only jumps in $(t_1, t_2)\}$. For any $\Lambda \in \mathfrak{F}_t$ we have, using the optionality of $\tau_\infty$,

$$\begin{aligned}\mathbf{P}_i\{\tau_\infty(\omega) \geqq t + t'|\Lambda;\, \tau_\infty(\omega) \geqq t;\, x(t, \omega) = j\} \\ = \mathbf{P}_i\{\Theta(t, t+t')|x(t, \omega) = j\} \\ = \mathbf{P}_j\{\Theta(0, t')\} = \mathbf{P}_j\{\tau_\infty(\omega) \geqq t'\}.\end{aligned} \tag{7.14}$$

Consider a new process $\{\bar{x}_t\}$, $t \in \mathbf{T}^0$ or $\mathbf{T}$ as in $\{x_t\}$, defined as

$$\bar{x}(t, \omega) = \begin{cases} x(t, \omega) & \text{if } t < \tau_\infty(\omega), \\ \infty & \text{if } t \geqq \tau_\infty(\omega). \end{cases} \tag{7.15}$$

We have then, if $i_\nu \in \mathbf{I}$, $1 \leqq \nu \leqq n+1$,

$$\begin{aligned}&\mathbf{P}\{\bar{x}(t_{n+1}, \omega) = i_{n+1}|\bar{x}(t_\nu, \omega) = i_\nu,\, 1 \leqq \nu \leqq n\} \\ &= \mathbf{P}\{x(t_{n+1}, \omega) = i_{n+1};\, \tau_\infty(\omega) > t_{n+1}|x(t_\nu, \omega) = i_\nu,\, 1 \leqq \nu \leqq n;\, \tau_\infty(\omega) > t_n\} \\ &= \mathbf{P}\{x(t_{n+1}, \omega) = i_{n+1};\, \Theta(t_n, t_{n+1})|x(t_n, \omega) = i_n\} \\ &= \mathbf{P}_{i_n}\{x(t_{n+1} - t_n, \omega) = i_{n+1};\, \Theta(0, t_{n+1} - t_n)\}.\end{aligned} \tag{7.16}$$

If we put

$$\bar{p}_{ij}(t) = \mathbf{P}_i\{x(t, \omega) = j;\, \Theta(0, t)\}, \tag{7.17}$$

$$\bar{p}_{i\infty}(t) = 1 - \sum_j \bar{p}_{ij}(t) = L_i(t), \qquad \bar{p}_{\infty,j}(t) = \delta_{\infty j}, \tag{7.18}$$

then the last probability in (7.16) is $\bar{p}_{i_n i_{n+1}}(t_{n+1} - t_n)$ and the calculation shows that $\{\bar{x}_t\}$ is a M.C. Its state space is $\bar{\mathbf{I}}$ and its transition matrix is $(\bar{p}_{ij})$ with $i$ and $j$ in $\bar{\mathbf{I}}$, provided that $\mathbf{P}\{\tau_\infty(\omega) = \infty\} < 1$, or equivalently that at least one $L_i$ is not identically zero. Otherwise $\{\bar{x}_t\}$ coincides with $\{x_t\}$.

The process $\{\bar{x}_t\}$ will be called the *minimal chain* associated with the given $\{x_t\}$. Our discussion in this section amounts to a probabilistic construction of the matrix $(\bar{p}_{ij})$, called the *minimal solution* corresponding to $Q = (q_{ij})$. We omit further properties of this matrix which will not be explicitly used below. But we note the following equation which follows from our analysis of the first discontinuity,

$$L_i(t) = \int_0^t e^{-q_i s} \sum_{j \neq i} q_{ij} L_j(t - s)\, ds. \tag{7.19}$$

Differentiating, we have

$$l_i(t) \stackrel{\text{def}}{=} L_i'(t) = \sum_j q_{ij} L_j(t). \tag{7.20}$$

Thus $L_i$ has a continuous derivative. Introducing the Laplace transform

$$\hat{l}_i(\lambda) = \int_0^\infty e^{-\lambda t} l_i(t)\, dt \tag{7.21}$$

and writing $\hat{l}(\lambda)$ for the column vector $\{\hat{l}_i(\lambda)\}$, we may put (7.20) in the form

$$(\lambda I - Q)\hat{l}(\lambda) = 0. \tag{7.22}$$

## 8. Beyond the first infinity

We continue to assume (7.6). The first infinity $\tau_\infty$ clearly depends on the initial distribution of $\{x_t, t \in \mathbf{T}\}$. Let $\tau_\infty^i$ be the restriction of $\tau_\infty$ on the set $\Delta_i$. We rewrite (5.6) as

$$\begin{aligned} p_{jk}(t) &= \mathbf{P}_j\{\tau_\infty^j(\omega) > t;\, x(t, \omega) = k\} \\ &\qquad + \int_0^t \mathbf{P}_j\{x(t, \omega) = k | \tau_\infty^j(\omega) = s\}\, d\mathbf{P}_j\, \{\tau_\infty^j(\omega) \leqq s\} \\ &= \bar{p}_{jk}(t) + \int_0 \xi_{jk}(s, t) l_j(s)\, ds. \end{aligned} \tag{8.1}$$

In general $\xi_{jk}(s, t)$ is not a function of $t - s$ only, in other words [see (4) of section 4] the two fields $\mathfrak{F}_{\tau_\infty}$ and $\mathfrak{F}'_{\tau_\infty}$ are not necessarily independent. (The statement to the contrary effect on p. 235 of [1] is erroneous.) Now for an ordinary state $i \in \mathbf{I}$ we have as an easy generalization of (6.2),

$$p_{jk}(t) = {}_ip_{jk}(t) + \int_0^t p_{ik}(t - s)\, dF_{ji}(s). \tag{8.2}$$

If we replace the $i$ above by $\infty$ and revert to our previous notation this would become, by analogy,

$$p_{jk}(t) = \bar{p}_{jk}(t) + \int_0^t \xi_k(t - s)\, dL_j(s). \tag{8.3}$$

Thus $\xi_{jk}(s, t)$ should not only be a function of $t - s$ only but also be independent of $j$. The second assertion would mean the extension of the Markov property to where $x(\tau_\infty(\omega), \omega) = \infty$, which is not asserted by the strong Markov property. The failure of (8.3) in general shows that the so-called *fictitious state* $\infty$ cannot be treated like a single ordinary state, and calls for a recompactification of **I**. To illustrate the idea and to speak only heuristically, if only a finite number $m$ of adjoined (fictitious) states $\infty^{(\nu)}$, $1 \leqq \nu \leqq m$, are needed, the situation should be as follows. To each $\infty^{(\nu)}$ corresponds an *atomic almost closed set* (see section I. 17 of [1]) $A^{(\nu)}$ of the *jump chain* $\{\chi_n, n \in \mathbf{N}\}$ in section 7, such that $x(\tau_\infty(\omega), \omega) = \infty^{(\nu)}$ iff $\chi_n(\omega) \in A^{(\nu)}$ for all sufficiently large $n$. Let the restriction of $\tau_\infty$ on the set $x(\tau_\infty(\omega), \omega) = \infty^{(\nu)}$ be $\tau_\infty^{(\nu)}$, and let the corresponding post-$\tau_\infty$ process be $\{y_t^{(\nu)}, t \in \mathbf{T}^0\}$. We put

$$L_j^{(\nu)}(t) = \mathbf{P}_j\{\tau_\infty^{(\nu)}(\omega) \leqq t\}, \tag{8.4}$$

$$\xi_k^{(\nu)}(t) = \mathbf{P}\{y^{(\nu)}(t, \omega) = k | \Omega_{\tau_{\infty^{(\nu)}}}\}. \tag{8.5}$$

Then we should have

$$p_{jk}(t) = \bar{p}_{jk}(t) + \sum_{\nu=1}^{m} \int_0^t \xi_k^{(\nu)}(t - s)\, dL_j^{(\nu)}(s) \tag{8.6}$$

as an improvement on (8.1). Note that each $L_j^{(\nu)}$ satisfies the same equation (7.20) as $L_j$ and $\sum_{\nu=1}^{m} L_j^{(\nu)} = L_j$.

In some sense the heuristic equation (8.6) must be contained in results proved by Feller [3] by function-analytic methods. But the precise identification of the probabilistic quantities is not clear to us and in any case no probabilistic proof seems known.

If there is only one bounded nonnegative solution $\hat{l}(\lambda)$ of (7.22), apart from a scalar factor (function of $\lambda$), then $m = 1$ in (8.6) and the resulting equation (8.3) can be easily proved (see Reuter [9]). It follows from (4) of section 4 that in this case $\mathfrak{F}_{\tau_\infty}$ and $\mathfrak{F}'_{\tau_\infty}$ are independent. By (2) of section 4, we have

$$\xi_k(s + t) = \sum_j \xi_j(s) p_{jk}(t). \tag{8.7}$$

Hence every $\xi_j$ is continuous in **T** by lemma 1. Substituting from (8.3), we have

$$\xi_k(s + t) = \sum_j \xi_j(s)\bar{p}_{jk}(t) + \int_0^t \xi_k(t - u)\, d_u\Big[\sum_j \xi_j(s) L_j(u)\Big]. \tag{8.8}$$

In analogy with (5.23), we put

$$S_\infty(\omega) = \{t : \overline{\lim_{s \to t}}\, x(s, \omega) = \infty\}. \tag{8.9}$$

We remark in this connection that $x(t, \omega)$ need not be $\infty$ even if $\lim_{s \uparrow t} x(s, \omega) = \infty$ or $\lim_{s \downarrow t} x(s, \omega) = \infty$ by (1.8); hence the obvious extension of (5.23) for $i = \infty$ is not adequate. Next, in analogy with (6.15) to (6.17) but for the post-$\tau_\infty$ process $\{y_t, t \in \mathbf{T}^0\}$, we put for $0 \leqq s \leqq t$:

$$\delta_\infty(t, \omega) = \sup\{s : 0 \leqq s \leqq t, y(s, \omega) \in S_\infty(\omega)\}, \tag{8.10}$$

$$\begin{aligned} \nabla_\infty(s, t) = \mathbf{P}\{\delta_\infty(t, \omega) \leqq s\} &= \sum_j \xi_j(s)[1 - L_j(t - s)] \\ &= 1 - \sum_j \xi_j(s) L_j(t - s), \end{aligned} \tag{8.11}$$

$$(8.12)\qquad \nabla_{\infty k}(s, t) = \mathbf{P}\{\delta_\infty(t, \omega) \leqq s; y(t, \omega) = k\} = \sum_j \xi_j(s)\bar{p}_{jk}(t).$$

The quantities in (8.11) and (8.12) occur on the right side of (8.8). Letting $s \downarrow 0$, it is easy to see that the limits below exist

$$(8.13)\qquad M(t) \overset{\text{def}}{=} \lim_{s \downarrow 0} \nabla_\infty(s, s + t) = \mathbf{P}\{\delta_\infty(t, \omega) = 0\},$$

$$(8.14)\qquad \eta_k(t) \overset{\text{def}}{=} \lim_{s \downarrow 0} \nabla_{\infty k}(s, s + t) = \mathbf{P}\{\delta_\infty(t, \omega) = 0; y(t, \omega) = k\}.$$

Thus $M(t)$ is the probability that the sample function $y(\cdot, \omega)$ has only jumps in $(0, t)$, while $\eta_k(t)$ is the probability that this is so and also $y(t, \omega) = k$. Letting $s \downarrow 0$ in (8.8), we obtain

$$(8.15)\qquad \xi_k(t) = \eta_k(t) + \int_0^t \xi_k(t - u)\, dM(u).$$

This is an integral equation of the renewal type for $\xi$ in terms of $\eta$. By definition we have

$$(8.16)\qquad \eta_k(s + t) = \sum_j \eta_j(s)\bar{p}_{jk}(t).$$

It follows by lemma 1 that $\eta_j(0)$ exist. Let

$$(8.17)\qquad \zeta_k(t) = \eta_k(t) - \sum_j \eta_j(0)\bar{p}_{jk}(t).$$

Then $\{\zeta_j\}$ satisfies the same equations (8.16) as $\eta_j$, and $\zeta_j(0) = 0$. Multiplying these equations by the "monotonicity factor" exp $(q_k t)$, see section 5, and then differentiating as in lemma 2, we obtain

$$(8.18)\qquad \zeta_k'(t) = \sum_j \zeta_j(s)\bar{p}_{jk}'(t - s)$$

for each $t$ and almost all $s \leqq t$. Using the second system of differential equations for $(\bar{p}_{jk})$ (see section II. 17 of [1]), we conclude that

$$(8.19)\qquad \zeta_k'(t) = \sum_j \zeta_j(t)q_{jk}.$$

Passing to Laplace transforms, the last equation may be written as [compare (7.22)]

$$(8.20)\qquad \hat{\zeta}(\lambda)(\lambda I - Q) = 0.$$

The above results check with those of Reuter [9] obtained by function-analytic methods. Unfortunately it does not represent the most general case treated by Reuter, because (8.15) reduces to a trivial identity when $M = \epsilon$, or equivalently when all $\eta_k(t) \equiv 0$. However, the following positive result may be stated.

*Unless the first infinity* $\tau_\infty(\omega)$ *is a limit point (from the right) of* $S_\infty(\omega)$ *with probability one, the equation* (8.15) *holds nonvacuously where* $\eta$ *and* $M$ *are defined by* (8.13) *and* (8.14), *and* (8.16) *holds.*

Suppose $M(0+) - M(0-) = \beta$ where $0 \leqq \beta < 1$ and let $\tilde{M} = M - \beta$.

Solving (8.15) for $\xi_k$, we have

$$(1-\beta)\,\xi_k(t) = \int_0^t \eta_k(t-u)\,dN(u), \tag{8.21}$$

where

$$N = \sum_{n=0}^{\infty} \tilde{M}^{n*} \tag{8.22}$$

in a notation similar to (5.16). If $\beta = 0$ then $N$ is the distribution of the time between two successive points of $S_\infty(\omega)$, necessarily isolated. If $\sum_k \eta_k(0) = 1$, then this must be the case and we have the so-called "instant return from infinity" of Doob (theorem II. 19.4 of [1]). In this case we have $\zeta_k(t) = 0$ for all $k$. The other extreme is where $\zeta_k(t) \equiv \eta_k(t)$ for all $k$ and only a gradual descent from infinity is possible.

The random variable $\delta_\infty(t, \cdot)$ is the last exit time from $\infty$ in $(0, t)$ for the post-$\tau_\infty$ process. If we consider a similar random variable $\gamma_\infty(t, \cdot)$ obtained by replacing $y$ with $x$ in (8.10), we are led naturally to the consideration of the following quantity, for $0 \leqq s \leqq t$,

$$\Phi_{ik}(s, t-s) \stackrel{\text{def}}{=} \mathbf{P}_i\{\gamma_\infty(\omega) \leqq s;\, x(t,\omega) = k\} = \sum_k p_{ij}(s)\bar{p}_{jk}(t-s) \tag{8.23}$$

and dually

$$\Psi_{ik}(s, t-s) \stackrel{\text{def}}{=} \mathbf{P}_i\{\tau_\infty(\omega) \geqq s;\, x(t,\omega) = k\} = \sum_j \bar{p}_{ij}(s)p_{jk}(t-s). \tag{8.24}$$

Clearly for each $t$, $\Phi_{ik}(s, t)$ is nondecreasing and $\Psi_{ik}(s, t)$ is nonincreasing in $s$. We have

$$\Phi_{ik}(0, t) = \Psi_{ik}(t, 0) = \bar{p}_{ik}(t) \tag{8.25}$$

and

$$\Phi_{ik}(t, 0) = \Psi_{ik}(0, t) = p_{ik}(t). \tag{8.26}$$

This remains true if $\bar{p}_{ij}$ in (8.23) and (8.24) is replaced by $\tilde{p}_{ij}$ such that $(\tilde{p}_{ij})$ is a substochastic transition matrix and

$$\tilde{p}_{ij}(t) \leqq p_{ij}(t) \tag{8.27}$$

for all $i$ and $j$ in I. However, there are analytical difficulties if we try to differentiate $\Phi_{ik}(s, t)$ or $\Psi_{ik}(s, t)$ with respect to $s$. Neveu [7] overcomes these difficulties by going to Laplace transforms and we refer to his paper for further results.

## REFERENCES

[1] K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, Berlin, Göttingen, Heidelberg, Springer, 1960.

[2] ———, "On last exit times," *Illinois J. Math.*, Vol. 4 (1960), pp. 629–639.

[3] W. Feller, "On boundaries and lateral conditions for the Kolmogoroff differential equations," *Ann. of Math.*, Vol. 65 (1957), pp. 527–570.

[4] E. HILLE and R. S. PHILLIPS, *Functional Analysis and Semi-groups*, American Mathematical Society Colloquium Publications, Vol. 31, Providence, American Mathematical Society, 1957.

[5] W. B. JURKAT, "On the analytic structure of semi-groups of positive matrices," *Math. Z.*, Vol. 73 (1960), pp. 346–365.

[6] J. MIKUSINSKI, *Operational Calculus*, New York, Pergamon Press, 1959.

[7] J. NEVEU, "Lattice methods and submarkovian processes," *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1961, Vol. 2, pp. 347–391.

[8] D. ORNSTEIN, "The differentiability of transition functions," *Bull. Amer. Math. Soc.*, Vol. 66 (1960), pp. 36–39.

[9] G. E. H. REUTER, "Denumerable Markov processes II," *J. London Math. Soc.*, Vol. 34 (1959), pp. 81–91.

[10] E. C. TITCHMARSH, *Introduction to the Theory of Fourier Integrals*, Oxford, Clarendon Press, 1937.