Die Grundlehren der
mathematischen Wissenschaften in Einzeldarstellungen
Band 184

Murray Rosenblatt

Markov Processes. Structure and Asymptotic Behavior

Murray Rosenblatt

Markov Processes. Structure and Asymptotic Behavior

Springer-Verlag Berlin Heidelberg New York 1971

Prof. Murray Rosenblatt
University of California, San Diego, La Jolla, California

AMS Subject Classifications (1970)
Primary 60J05, 60F05
Secondary 60B99, 60G10, 60G50, 60J10, 60J15, 60J20, 60G25

ISBN 3-540-05480-4 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-05480-4 Springer-Verlag New York Heidelberg Berlin

 Library of Congress Catalog Card Number 70-161441. Printed in Germany. Monophoto typesetting and offset printing: Zechnersche Buchdruckerei, Speyer. Bookbinding: Konrad Triltsch, Graphischer Betrieb, Würzburg

Preface

This book is concerned with a set of related problems in probability theory that are considered in the context of Markov processes. Some of these are natural to consider, especially for Markov processes. Other problems have a broader range of validity but are convenient to pose for Markov processes. The book can be used as the basis for an interesting course on Markov processes or stationary processes. For the most part these questions are considered for discrete parameter processes, although they are also of obvious interest for continuous time parameter processes. This allows one to avoid the delicate measure theoretic questions that might arise in the continuous parameter case. There is an attempt to motivate the material in terms of applications. Many of the topics concern general questions of structure and representation of processes that have not previously been presented in book form. A set of notes comment on the many problems that are still left open and related material in the literature. It is also hoped that the book will be useful as a reference to the reader who would like an introduction to these topics as well as to the reader interested in extending and completing results of this type.

The first chapter deals with some basic properties of Markov processes as well as a number of illustrations. The limiting behavior of Markov chains with stationary transition mechanism is dealt with. There are remarks on independent random variables and the "theory of errors" as a motivation for classical limit theorems. The simplest continuous parameter processes, the Poisson and Wiener (or Brownian motion) processes are introduced. A generalization of the classical result of Polya on recurrence for random walks on commutative countable groups is given.

A number of models in statistical mechanics, "learning" theory in psychology, and in statistical economics are discussed in the second chapter. The object is to show how Markovian-like models arise in a variety of applications and how often questions concerning collapsing of the state space can arise. The concept of ergodicity is already very important in statistical mechanics. A desire to retain at least an approximate version of the Markov property as well as interest in a central limit theorem for dependent processes are evident in the heuristic discussion of the foundations of "non-equilibrium" statistical mechanics.

The third chapter considers a number of results on functions of Markov processes. Conditions under which a function of a Markov process is still Markovian and the relationship between the Chapman-Kolmogorov equation and the Markov property are examined. Functions of Markov processes are generally not Markovian. It is of interest to ask when a finite state process is a function of a finite state Markov chain. An interesting algebraic treatment of this problem is given.

The fourth chapter deals with ergodic problems and first discusses the restriction of a Markov process to a subset of the state space. The L^1 ergodic theorem due to Chacon and Ornstein is presented. The concepts of ergodicity and mixing are introduced and illustrated. Many of the results in "ergodic theory" assume the existence of an invariant measure. Conditions of a topological character on the transition operator that insure the existence of an invariant measure are introduced. Finally, results are obtained on the asymptotic behavior of unaveraged powers of the transition probability operator. Such results can be related to a prediction problem for Markov processes.

Chapter 5 is devoted to random walks or, more properly, convolution of regular measures on compact groups and semigroups. A limit theorem of P. Lévy dealing with the circle group is introduced to motivate the development that then follows. The uniform or Haar measure on a compact group is discussed as the limit law of a convolution sequence of measures. The corresponding type of limit law for convolution sequences of measures on a compact semigroup is an idempotent measure under convolution. The so-called "Rees-Suschkewitsch" theorem describing the structure of compact semigroups is developed in order to characterize the idempotent measures on a compact semigroup. The results are illustrated in the case of semigroups of $n \times n$ (n finite) transition probability matrices. These are perhaps the simplest types of limit theorems for products of independent, identically distributed operators which may not commute.

The sixth chapter deals with nonlinear one-sided representations of Markov processes in terms of independent random variables. Such a treatment is motivated in part by prediction problems. A brief discussion of a corresponding linear representation (the Wold representation) in the linear prediction problem is given. N. Wiener dealt with such nonlinear representations in his book *Nonlinear Methods in Random Theory* where he discussed coding and decoding. Rather complete results are obtained for finite state Markov chains. Partial results are obtained for real-valued Markov processes. A relation between such representations and the isomorphism problem for stationary processes is briefly indicated.

The last chapter is concerned with conditions for the validity of central limit theorems for Markov processes. Cogburn's condition of uniform ergodicity and its relation to infinitely divisible laws as limiting laws for partial sums of stationary Markov variables are given. Uniform (or strong) mixing is also introduced. Both conditions are examined in the case of Markov processes. Finally a central limit theorem is obtained using a condition similar to uniform mixing.

A series of notes that relate the material in the text to relevant results in the literature for Markov processes and more general processes are given after each chapter. Open questions that are of interest are also discussed. It is hoped that the text will be appropriate for an audience with a general mathematical outlook as well as for those with a probabilistic (or statistical) orientation. Readers with a good strong mathematical interest and background whose primary concern is in such areas as statistical physics, mathematical economics, or learning theory should find ideas and methods that are relevant. Several of the questions examined illustrate the interplay of concepts from probability theory with other areas of mathematics. The appendices give a discussion of those ideas from other mathematical areas that bear on the material developed in the text.

The chapters are divided into numbered sections. Sections and formulas cited contain just enough information to identify them. For example, the formulas numbered 2, 1.2, 3.1.2 and referred to in section 6.3 (third section of the sixth chapter) are formula 2 of the same section, formula 2 of section 1 of the same chapter, and formula 2 of section 3.1, respectively.

I am grateful to the John Simon Guggenheim Foundation for its support in the year 1965—66 and the Office of Naval Research for funding of much of the research on which this book is based. I am indebted to Peter Bickel, D. Brillinger, J. L. Doob, John Evans, D. Rosenblatt, T. C. Sun and K. Wickwire for their helpful comments and suggestions. Lastly, my thanks go to Olive Lee and Lillian Johnson for typing and correcting the manuscript, and to Diana Marcus for help in proofreading.

La Jolla, California 1971 Murray Rosenblatt

Table of Contents

Chapter I

Basic Notions and Illustrations

Chapter II

Remarks on Some Applications

Chapter III

Functions of Markov Processes

Chapter IV

Ergodic and Prediction Problems

Chapter V

Random Walks and Convolution on Groups and Semigroups

Chapter VI

Nonlinear Representations in Terms of Independent Random Variables

Chapter VII

Mixing and the Central Limit Theorem

Chapter I

Basic Notions and Illustrations

0. Summary

The probability space for a one sided Markov process with stationary transition mechanism is set up in the discrete and continuous time parameter case under appropriate conditions in section 1. The extension (if possible) to a two-sided process is discussed, as well as the Chapman-Kolmogorov equation for first order transition probabilities. A number of illustrative examples are taken up in the following sections. The asymptotic properties of transition probabilities for Markov chains (Markov processes with a countable state space) are considered in section 2. This motivates in part the later development of an ergodic theorem (in Chapter 4 section 2) for Markov processes with a general state space. The classical example of a sequence of independent random variables is taken up in section 3. There is a brief discussion of the theory of errors and then a derivation of the Poisson approximation to the Binomial distribution and the normal approximation to the distribution of a sum of independent random variables, both with error terms. The theorems on the Poisson and normal approximation are not only of independent interest but are also used later in Chapter 7 section 1 to obtain a remarkable result of Kolmogorov on the approximation of the distribution of a sum of independent and identically distributed random variables by an infinitely divisible distribution with error term. A brief discussion of the continuous parameter Poisson and Wiener (Brownian motion) processes is given in section 4. The classical result of Polya on recurrence of one and two dimensional and nonrecurrence of three dimensional random walks is given in section 5. A generalization (due to Dudley) for random walks on countable Abelian groups is then developed.

1. Markov Processes and Transition Probability Functions

Markov processes are structurally the simplest models of dependent random behavior through time that have been dealt with. Our concern will be with Markov processes whose generating mechanism is stable through time. To avoid complications let us assume that observations of some phenomenon are made at times $n = 0, 1, 2, \ldots$. Let Ω (the state space) be a space of points x representing the possible observations at any given fixed time. The possible events for which a probability is well defined will be the elements of a Borel field $\mathscr{A}$ of subsets of Ω. The stable generating mechanism for a Markov process is given by its *transition probability function* $P(x, A)$ which is assumed to be $\mathscr{A}$-measurable as a function of x for each set (or event) A in $\mathscr{A}$ and a probability measure on the Borel field $\mathscr{A}$ for each x in Ω. Intuitively, $P(x, A)$ represents the probability that the observation at time $n+1$ of the Markov process will fall in the set A given that, at time n, observation x was made. With an initial probability measure μ (at time 0) on the Borel field $\mathscr{A}$, a discrete time parameter Markov process with initial distribution μ and stationary transition probability function $P(\cdot, \cdot)$ can be constructed as follows. For any finite collection of sets $A_0, A_1, \ldots, A_n \in \mathscr{A}$ let

$$P_\mu(A_0 \times A_1 \times \cdots \times A_n) = P_\mu(x_0 \in A_0, \ldots, x_n \in A_n) \tag{1}$$

$$= \int_{A_0} \mu(dx_0) \int_{A_1} P(x_0, dx_1) \cdots \int_{A_{n-1}} P(x_{n-2}, dx_{n-1}) P(x_{n-1}, A_n).$$

An extension theorem of C. Ionescu-Tulcea (see Appendix 3) can be used to extend this set function to a measure P_μ on the Borel field $\mathscr{A}_\infty$ generated by sets of the form $A_0 \times A_1 \times \cdots \times A_n$ on the space Ω_∞ of points $(x_0, x_1, x_2, \ldots) = \omega$.

This probability measure P_μ describes the relative likelihood of observing the different possible trajectories $\omega = (x_0, x_1, x_2, \ldots)$ of the random system being studied through time. The observation on the system at time n is given by the n^{th} coordinate function or random variable $X_n(\omega) = x_n$ and the random process is written $\{X_n\} = \{X_n(\omega); n = 0, 1, \ldots\}$. The theorem of Ionescu-Tulcea can be applied to define a one-sided Markov process $\{X_n(\omega); n = 0, 1, \ldots\}$ with transition probability $P(\cdot, \cdot)$ without any additional conditions in the discrete time parameter case we are now considering. The extension theorem of Kolmogorov (see Appendix 3) can also be used to define a Markov process with transition probability function $P(\cdot, \cdot)$ if additional conditions of a mixed topological and measure theoretic character are imposed on the transition probability function $P(\cdot, \cdot)$ and an initial or marginal probability measure. However, the Kolmogorov extension theorem is especially useful in constructing two-sided Markov processes $\{X_n(\omega); n = \ldots, -1, 0, 1, \ldots,\}$ or continuous time parameter Markov processes and whenever we apply the Kolmogorov theorem we shall implicitly assume that the conditions required for its application are satisfied. The extension theorem of Ionescu-

Tulcea is not useful in the construction of continuous time parameter Markov processes.

Higher step transition probability functions $P_n(\cdot,\cdot)$ can be generated from the one-step transition probability function $P(\cdot,\cdot)$ by the following recursive procedure:

$$P_{n+1}(x,A) = \int_{\Omega} P(x,dy)\, P_n(y,A), \quad (n=1,2,\ldots) \tag{2}$$

The relation

$$P_{n+m}(x,A) = \int_{\Omega} P_n(x,dy)\, P_m(y,A) \quad (n,m=1,2,\ldots,) \tag{3}$$

follows from (2) and is commonly called the Chapman-Kolmogorov equation.

Under certain circumstances, the study of a Markov process with stationary transition mechanism can be extended backward in time. This can be done if there exists a sequence of probability measures μ_n, $n=0,\pm 1,\ldots$, on $\mathscr{A}$ such that $\mu_0=\mu$ and

$$\int \mu_n(dx)\, P(x,A) = \mu_{n+1}(A), A \in \mathscr{A}.$$

Equation (1) with $\mu=\mu_m$ is used to define the probability of sets $A_m \times \cdots \times A_n$, $-\infty < m < n < \infty$. The Kolmogorov extension theorem can then be employed to set up a measure on the space of points $(\ldots, x_{-1}, x_0, x_1, \ldots) = \omega$ describing the history of a system from the infinite past to the infinite future. A random process described by the probability measure is now written $\{X_n\} = \{X_n(\omega);\ n=0, \pm 1, \ldots\}$ with $X_n(\omega) = x_n$ as before. Whether the process is one-sided or two-sided, a *shift transformation* τ corresponding to a forward time shift can be introduced. In the one-sided case, $\omega = (x_0, x_1, \ldots)$ and $(\tau\omega)_n = x_{n+1}$. An inverse τ^{-1} is not always well-defined. In the two-sided case,

$$\omega = (\ldots, x_{-1}, x_0, x_1, \ldots)$$

with $(\tau\omega)_n = x_{n+1}$ and the inverse τ^{-1} is always defined. In both the one-sided and two-sided situations $\mathscr{A}_\infty$ and Ω_∞ will be used to denote the Borel field and space of infinite sequences, respectively. If the probability measure μ is *invariant with respect to* $P(\cdot,\cdot)$

$$\int \mu(dx)\, P(x,A) = \mu(A), \tag{4}$$

then the process can clearly be extended backward in time using the Kolmogorov extension theorem. Then for any event $C \in \mathscr{A}_\infty$,

$$P_\mu(\tau C) = P_\mu(C).$$

If a process is two-sided, $\tau^{-1} C$ is defined and

$$P_\mu(\tau^{-1} C) = P_\mu(C).$$

Such processes are called *stationary Markov processes* because their probability structure is invariant with respect to time translation.

The Borel field $\mathscr{A}_\infty$ was constructed so that it is exactly the Borel field generated by the random variables $\{X_n\}$. Let $\mathscr{A}_m^n$ be the Borel field generated by the random variables $X_k, m \leqslant k \leqslant n$. Notice that $\mathscr{A}_\infty = \mathscr{A}_0^\infty$ in the case of a one-sided process and $\mathscr{A}_\infty = \mathscr{A}_{-\infty}^\infty$ for a two-sided process. $\mathscr{A}_m^n$ carries the information given by the random variables $X_k, m \leqslant k \leqslant n$. Let $\mathscr{B}_m$ be the Borel field generated by $X_k, k \leqslant m$, and $\mathscr{F}_n$ the Borel field generated by $X_k, k \geqslant n$. $\mathscr{B}_m$ and $\mathscr{F}_n$ are the backward and forward Borel fields relative to times m and n, respectively. Sometimes we will write $\mathscr{A}_n$ in place of $\mathscr{A}_n^n$.

Suppose P is some probability measure on the Borel field $\mathscr{A}_\infty$ of points ω of Ω_∞. Given any event C of $\mathscr{A}_\infty$ and a sub-Borel field $\mathscr{B} \subset \mathscr{A}_\infty$ let $P(C|\mathscr{B})(\omega)$ denote the Radon-Nikodym derivative of the measure $P(C \cap B)$ with respect to $P(B)$, $B \in \mathscr{B}$, as measures on $\mathscr{B}$. The derivative is $\mathscr{B}$-measurable and is called the conditional probability of C given the Borel field $\mathscr{B}$. Since the first measure is absolutely continuous with respect to the second,

$$P(C \cap B) = \int_B P(C|\mathscr{B})(\omega) P(d\omega), \quad (B \in \mathscr{B}.)$$

If $\mathscr{B}$ is generated by a family of random variables, then intuitively the conditioning is with respect to the family of random variables. The Markov property can now be simply given for such a general process on the space of sequences Ω_∞. Let F be any event of $\mathscr{F}_n$, $n \geqslant m$. The process is *Markovian* if

$$P(F|\mathscr{B}_m)(\omega) = P(F|\mathscr{A}_m^m)(\omega)$$

for any such event. The conditional probability $P(F|\mathscr{B}_m)(\omega)$ is $\mathscr{A}_m^m$-measurable and so depends only on the past information given at the last time that a completely specified observation is made on the process. It can be shown that the Markov property also may be written

$$P(B|\mathscr{F}_n)(\omega) = P(B|\mathscr{A}_n^n)(\omega)$$

for any event $B \in \mathscr{B}_m$, $m \leqslant n$. The Markov property is independent of time direction. Under fairly broad conditions the probability of an event $A_0 \times A_1 \times \cdots \times A_n$ can be written in terms of an initial distribution and one-step transition probabilities

$${}_kP(X_{k+1} \in A|\mathscr{A}_k^k)(\omega)$$

just as in (1) except that the transition probability function may not be stationary. In (1) it is rather curious and perhaps unesthetic that even though the process has stationary transition function in the forward direction, with time reversed the process may not have a stationary transition mechanism. Occasionally we will have to deal with derived Markov processes not having a stationary transition mechanism. The probability structure of such a process on sets of the form $A_0 \times A_1 \times \cdots \times A_n$ is then given by a sequence ${}_kP(\cdot,\cdot)$, $k = 0, 1, \ldots$, of one-step transition probability functions describing the transition from state x at time k into set A at time $k+1$ as follows

$$P_\mu(A_0 \times A_1 \times \cdots \times A_n) = P_\mu(x_0 \in A_0, \ldots, x_n \in A_n) \tag{5}$$
$$= \int_{A_0} \mu(dx_0) \int_{A_1} {}_0P(x_0, dx_1) \cdots \int_{A_{n-1}} {}_{n-2}P(x_{n-2}, dx_{n-1})\, {}_{n-1}P(x_{n-1}, A_n).$$

P_μ is then extended to a probability measure on the Borel field $\mathscr{A}_\infty$ generated by sets of the form $A_0 \times A_1 \times \cdots \times A_n$ just as in the case of a stationary transition probability function.

A transition probability function $P(x, A)$ induces an operator T taking probability measures μ on $\mathscr{A}$ into probability measures μT.

$$v(A) = (\mu T)(A) = \int \mu(dx) P(x, A) \tag{6}$$

on $\mathscr{A}$. This is also true for finite measures μ, that is, measures such that $\mu(\Omega) < \infty$. The operation (6) need not be well-defined for σ-finite measures μ. The transition probability function also induces an operator taking bounded functions into bounded functions. We again use the letter T to denote this operator. T acts on the left as an operator on measures; on the right it is to be considered as an operator on $\mathscr{A}$ measurable functions

$$(Tf)(x) = \int P(x, dy) f(y). \tag{7}$$

The Hölder inequality implies that

$$|(Tf)(x)|^p \leqslant \int P(x, dy) |f(y)|^p = (T|f|^p)(x), \quad (\infty > p \geqslant 1.) \tag{8}$$

Given a σ-finite measure μ let $L^p(d\mu)$ denote the set of $\mathscr{A}$ measurable functions f whose p^{th} $(1 \leqslant p < \infty)$ absolute mean with respect to μ is finite

$$\int |f(x)|^p \mu(dx) < \infty.$$

Inequality (8) implies that if $f \in L^p(dv)$, $v = \mu T$, then $Tf \in L^p(d\mu)$ with

$$\|Tf\|_{\mu,p} \leqslant \|f\|_{v,p}.$$

Here $\|f\|_{\mu,p}$ is the norm given by

$$\|f\|_{\mu,p} = \{\int |f(x)|^p \mu(dx)\}^{1/p}.$$

Further, there is equality in (8) if and only if for almost every x (with respect to μ), f is constant almost everywhere with respect to $P(x, \cdot)$.

The construction of Markov processes with continuous time parameter proceeds in a similar manner. As before, the state space Ω consists of points x representing a possible observation at any given fixed time t.

Here t may be the set of all real numbers or the set of all nonnegative real numbers. For convenience, we shall consider the set of all nonnegative real numbers. The events (at a fixed time) for which probabilities are well defined will be elements of a Borel field $\mathscr{A}$ of Ω. The stable generating mechanism for the Markov process is now given by a transition probability function $P_t(x,A)$, $0 \leqslant t < \infty$, which is $\mathscr{A}$-measurable as a function of x for each set (or event) $A \in \mathscr{A}$ and a probability measure on $\mathscr{A}$ for each x. We now have to *assume* that $P_t(x,A)$ satisfies

$$\int P_t(x,dy) P_\tau(y,A) = P_{t+\tau}(x,A),$$

for $t,\tau \geqslant 0$. This is called the Chapman-Kolmogorov equation as it was in the discrete time parameter case. It is natural to assume that

$$P_0(x,A) = \delta(x,A) = \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ otherwise} \end{cases}$$

and this will be taken for granted. Let μ be an initial probability measure (at time 0) on $\mathscr{A}$. A Markov process with continuous time parameter $\{X(t);\ 0 \leqslant t < \infty\}$ with initial distribution μ and stationary transition probability function $P_t(x,A)$ is constructed very much as it was in the discrete parameter case. Given any finite collection of events

$$A_0, A_1, \ldots, A_n \in \mathscr{A} \text{ and } t_i > 0$$

let

$$P_\mu(x(0) \in A_0, x(t_1) \in A_1, \ldots, x(t_1 + \cdots + t_n) \in A_n)$$
$$= \int_{A_0} \mu(dx_0) \int_{A_1} P_{t_1}(x_0, dx_1) \cdots \int_{A_{n-1}} P_{t_{n-1}}(x_{n-2}, dx_{n-1}) P_{t_n}(x_{n-1}, A_n).$$

The Kolmogorov extension theorem can be used under appropriate conditions on μ and $P_t(\cdot,\cdot)$ (see Appendix 3) to extend this set function to a measure P_μ on the Borel field generated by sets of the form

$$\{x(0) \in A_0, x(t_1) \in A_1, \ldots, x(t_1 + \cdots + t_n) \in A_n\},\ t_1, t_2, \ldots, t_n > 0$$

on the space of points $(x(t); 0 \leqslant t < \infty) = \omega$. Notice that the points of this space are functions, so that a measure is being constructed on a function space. Each point $\omega = (x(t); 0 \leqslant t < \infty)$ represents a possible trajectory (as a function of time) of the system whose probability structure is described by the Markov process. If there is a family of probability measures μ_t on $\mathscr{A}$, $-\infty < t < \infty$, such that

$$\mu_t(A) = \int \mu_\tau(dx)\ P_{t-\tau}(x,A)$$

for all pairs t, τ with $-\infty < \tau < t < \infty$, then one can construct a Markov process whose instantaneous distribution at time t is given by μ_t. Simply set

$$P(x(t_0) \in A_0, \ldots, x(t_n) \in A_n)$$
$$= \int_{A_0} \mu_{t_0}(dx_0) \int_{A_1} P_{t_1 - t_0}(x_0, dx_1) \cdots \int_{A_{n-1}} P_{t_{n-1} - t_{n-2}}(x_{n-2}, dx_{n-1})\ P_{t_n - t_{n-1}}(x_{n-1}, A_n)$$

for any finite collection of sets $A_0, \ldots, A_n \in \mathscr{A}$ and real numbers $t_0 < t_1 < \cdots < t_n$. Extend this set function to a measure P on the Borel field generated by the sets of the form $\{x(t_0) \in A_0, \ldots, x(t_n) \in A_n\}$, $t_0 < t_1 < \cdots < t_n$, on the space of points $(x(t); -\infty < t < \infty) = \omega$. We shall write the process as $\{X(t)\} = \{X(t)(\omega); -\infty < t < \infty\}$ with $X(t)(\omega) = x(t)$.

2. Markov Chains

The simplest interesting case of a Markov process is the Markov chain, a Markov process with a countable number of states that we shall label by the integers 1, 2, 3, ... for convenience. Let the one-step transition probability from state j to state k be $p_{j,k} = p_{j,k}^{(1)} \geqslant 0$, $\sum_k p_{j,k} = 1$. The $(n+1)$-step transition probability from state j to state k is given recursively by

$$p_{j,k}^{(n+1)} = \sum_l p_{j,l} p_{l,k}^{(n)}, \quad (n = 1, 2, \ldots) \tag{1}$$

The transition function is sometimes conveniently represented by the matrix $P^{(n)} = (p_{i,j}^{(n)}; i, j = 1, 2, \ldots) = P^n$, $n = 1, 2, \ldots$, and the Chapman-Kolmogorov equation is then simply given by

$$P^{(n+m)} = P^{n+m} = P^{(n)} P^{(m)} = P^n P^m . \tag{2}$$

The chain is said to be *irreducible* if every state can be reached from any other state with positive probability in a finite number of steps, that is, given any pair of states j and k there is an integer $n = n(j, k) > 0$ such that $p_{j,k}^{(n)} > 0$. A state j is *recurrent* if $\sum_{n=1}^{\infty} p_{j,j}^{(n)} = \infty$ and *transient* if $\sum_{n=1}^{\infty} p_{j,j}^{(n)} < \infty$. The state j is recurrent (transient) if the mean number of returns to the state j from the present to the infinite future is infinite (finite). The state j is *periodic with period s* if s is the greatest common divisor of the integers n for which $p_{j,j}^{(n)} > 0$. In the case of an irreducible chain, if one state is recurrent (periodic with period ρ) then all the states are recurrent (periodic with period ρ). Assume that j is recurrent with k any other state. Since the chain is irreducible there are integers r, s such that $p_{j,k}^{(r)}, p_{k,j}^{(s)} > 0$. Therefore

$$\sum_{n=1}^{\infty} p_{k,k}^{(n)} \geqslant \sum_m p_{k,j}^{(s)} p_{j,j}^{(m)} p_{j,k}^{(r)} = \infty$$

and k is recurrent. It immediately follows that if one state in an irreducible chain is transient, all the states are transient. Assume that state j is periodic with period ρ and that k is any other state. Again let r and s be integers such that $p_{j,k}^{(r)}, p_{k,j}^{(s)} > 0$. Since $r+s$ is divisible by ρ and

$$\begin{aligned} p_{j,j}^{(r+s+m)} &\geqslant p_{j,k}^{(r)} p_{k,k}^{(m)} p_{k,j}^{(s)}, \\ p_{k,k}^{(r+s+m)} &\leqslant p_{k,j}^{(s)} p_{j,j}^{(m)} p_{j,k}^{(r)}, \end{aligned} \tag{3}$$

it follows that k is periodic with period ρ.

Let $f_{j,k}^{(n)}=P\{X_m(\omega)\neq k, 0<m<n, X_n(\omega)=k\,|\,X_0(\omega)=j\}$ be the conditional probability of going from j to k for the first time in precisely n steps. The conditional probabilities $f_{j,k}^{(n)}$ and $p_{j,k}^{(n)}$ satisfy the equation

$$p_{j,k}^{(n)}=f_{j,k}^{(n)}+\sum_{m=1}^{n-1} f_{j,k}^{(m)}\,p_{k,k}^{(n-m)}. \tag{4}$$

Introducing the generating functions

$$F_{j,k}(s)=\sum_{n=1}^{\infty} f_{j,k}^{(n)} s^n,$$

$$G_{j,k}(s)=\delta_{j,k}+\sum_{n=1}^{\infty} p_{j,k}^{(n)} s^n$$

we see that

$$G_{j,j}(s)=(1-F_{j,j}(s))^{-1},\ G_{i,j}(s)=F_{i,j}(s)\,G_{j,j}(s),\ i\neq j. \tag{5}$$

The state j is transient if and only if $G_{jj}(1)<\infty$ or equivalently $F_{jj}(1)<1$. Let

$$\mu_j=\sum_{n=1}^{\infty} n f_{j,j}^{(n)}=F'_{j,j}(1), \tag{6}$$

$$\lim_{s\to 1-}(1-s)\,G_{j,j}(s)=\frac{1}{\mu_j}=u_j.$$

Now

$$F_{k,k}(1)\leqslant F_{k,j}(1)\,F_{j,k}(1)+(1-F_{k,j}(1)).$$

If the chain is irreducible and recurrent $F_{k,k}(1)=1$, $F_{k,j}(1)>0$. Then $F_{k,j}(1)\leqslant F_{k,j}(1)\,F_{j,k}(1)$ which implies that $F_{j,k}(1)=1$. If j is a recurrent state, the number μ_j is called the *mean recurrence time* for the state j. A recurrent state j is called *positive recurrent* if $\mu_j<\infty$ and *null recurrent* if $\mu_j=\infty$. Equation (6) and inequalities (3) imply that *an irreducible chain with one state positive (null) recurrent has all its states positive (null) recurrent.*

If a stationary probability distribution $\{v_j\}$ exists then

$$v_j=\sum_i v_i p_{ij}$$

which implies that

$$s v_j=(1-s)\,G_{jj}(s)\sum_i v_i F_{ij}(s)$$

by (5). Letting $s\to 1-$ we have

$$v_j=u_j\sum_i v_i F_{ij}(1)$$