An Introduction to

FINITE MARKOV PROCESSES

S R Adke

S M Manjunath

A Halsted Press Book

An Introduction to FINITE MARKOV PROCESSES

S. R. ADKE
University of Poona, Pune
India

S. M. MANJUNATH
Bangalore University, Bangalore
India

JOHN WILEY & SONS
New York Chichester Brisbane Toronto Singapore

Published in the Western Hemisphere
by Halsted Press, A Division of
John Wiley & Sons, Inc., New York

Library of Congress Cataloging in Publication Data

Printed in India at Gajendra Printing Press, Delhi.

PREFACE

The discrete state-space, continuous parameter Markov process has been used extensively to construct stochastic models in a variety of disciplines like biology, chemistry, electrical engineering, medicine, physics, sociology etc. Most of the books dealing with discrete Markov processes concentrate either on the exposition of their mathematical properties e.g., Chung (1967), Cinlar (1975), Feller (1972), Iosifescu (1980), or provide their introductory properties e.g., Bhat (1972), Karlin and Taylor (1975), Medhi (1982), Parzen (1962). A scientist who employs a finite Markov process to construct a stochastic model has to know, not only the basic assumptions and their consequences, but also the statistical properties and procedures of statistical inference for such processes. This book is intended to meet this requirement.

We discuss the theoretical properties of a finite state-space, continuous parameter Markov process in the first three chapters. Thus chapter 1 provides the basic definitions and establishes the equivalence of various versions of the Markov property. The analytic properties of the transition probabilities, their evaluation and the properties of sample functions are discussed in chapter 2. A detailed discussion of the classification of states of a discrete Markov process is provided in chapter 3 which also discusses the asymptotic behaviour of the transition probabilities. The probability generating function, moment generating function and the first two moments of random variables like duration of stay in a particular state, transition counts for reducible and irreducible Markov processes are discussed in chapter 4. The last and the longest chapter, chapter 5, develops the asymptotic properties of maximum likelihood estimators as well as asymptotic theory of likelihood ratio tests of hypotheses. Finally we provide thirty problems which are constructed on the basis of research papers, some old and some very recent. These problems are based on concepts introduced in two or more chapters and we have therefore listed them at the end of the book.

It is expected that our reader is familiar with basic concepts of probability theory as discussed in Bhat (1981) or in chapters III, IV and VII of Loeve (1968). A good acquaintance with matrix algebra and the contents of chapter V and VI of Rao (1973) is desirable for understanding our chapter 5 on Statistical Inference. This book can be used for a one semester course for advanced level graduate/post graduate students.

Each chapter in the book has been divided into sections which are serially numbered. The definitions, equations, lemmas, theorems, corollaries and examples are all serially numbered in each section. A reference to equation (c) in a section is to the c-th equation of the same section. The equation (b.c) stands for the c-th equation in b-th section of the same chapter and the equation (a.b.c) refers to c-th equation of b-th section of the a-th chapter. The same scheme applies to definitions, lemmas etc.

The book was initiated and almost completed when one of the authors (SMM) was visiting the Department of Statistics, University of Poona as a Teacher Fellow from Bangalore University, Bangalore. We thank Dr A. V. Kharshikar and Dr M. S. Prasad for the discussions we had with them. We also thank Mrs. A. V. Sabane for her careful typing of the manuscript.

University of Poona,
Pune - 7, India
May, 1984.

S. R. ADKE
S. M. MANJUNATH

CONTENTS

CHAPTER 1

DISCRETE MARKOV PROCESSES : DEFINITIONS

1. INTRODUCTION

A large number of natural phenomena which evolve in time have indeterministic components. Such phenomena can be described in terms of an infinite collection of random variables (r.v.s) $X(t)$, $t \in T$, all defined on the same probability space $(\Omega, \mathbb{F}, P)$ and indexed by a parameter t taking values in an infinite index set T; i.e., in terms of the stochastic process $\{X(t), t \in T\}$.

The classical theory of Statistics mainly deals with experiments which are repeatable under identical conditions. The outcomes of such experiments can be modelled in a fairly satisfactory manner by a sequence $\{X_n, n = 1, 2, \ldots\}$ of independent and identically distributed r.v.s. The possibility of dependence between successive r.v.s led Markov (1906) to introduce " chains " of random variables which are now well-known as Markov chains. The sequence of independent and identically distributed r.v.s and the Markov chain are stochastic processes with the set of non-negative integers as the index set. However, the following examples illustrate that it is necessary to consider stochastic processes indexed by a linearly ordered index set like the non-negative half of the real line.

Example 1 : Consider a telephone exchange with a finite number M of channels. Calls arrive at the exchange at random instants of time and a call is connected only if a free channel is available. The availability of a channel depends on the durations of conversations which are also of a random nature. Let $X(t)$ denote the number of busy channels at the instant or the epoch t, $0 \leq t < \infty$. The fluctuations in the random number $X(t)$ of busy channels as t progresses on $[0, \infty)$ are of interest to the design and maintenance engineers. One is obviously dealing with the stochastic process $\{X(t), 0 \leq t < \infty\}$ with $[0, \infty)$ as the index set and $\{0,1,2,..., M\}$ as the set of possible values of $X(t)$, $t \in [0, \infty)$. This example was first discussed by Kolmogorov (1931).

Example 2 : A patient suffering from a disease like cancer can be in a number of different states. The initial state E_1 is the state in which a person is identified as a cancer patient. Depending on the state of his health and the treatment received by him, the patient may move to state E_2 of recovery or to one of the terminal states E_3 or E_4 representing the death of the patient due to cancer or some other cause respectively. A patient may also oscillate between E_1 and E_2 before he is finally claimed by E_3 or E_4. The state of the patient can be described in terms of the random variable $X(t)$ which equals 1, 2, 3 or 4 according as the state of the patient is E_1, E_2, E_3 or E_4. We are thus dealing with the stochastic process $\{X(t), 0 \leq t < \infty\}$ with $X(t)$ taking values in the set $\{1, 2, 3, 4\}$. This example was discussed by Fix and Neyman (1951).

Example 3 : More recently, Wasserman (1980) has discussed the development of relationships between a group of say M persons. In his most general model, Wasserman defines $X_{ij}(t)$ to be one or zero according as a ' relationship ' exists or does not exist at epoch t between the i-th and j-th individuals of the group, $i \neq j$, $i, j = 1, ..., M$. By convention we take $X_{ii}(t) \equiv 0$, $t \in [0, \infty)$. The entire social network of relationships at any instant t of time can be represented by the M x M matrix $\underline{X}(t) = ((X_{ij}(t)))$ of binary elements. The total number of possible binary matrices representing the social network at any instant t is $2^{M(M-1)}$. Wasserman is thus dealing with a matrix-valued stochastic process $\{\underline{X}(t), 0 \leq t < \infty\}$ indexed by the continuous time parameter $t \in [0,\infty)$

As indicated earlier, a stochastic process $\{X(t), t \in T\}$ is an infinite collection of random variables defined on the same probability space $(\Omega, \mathbb{F}, P)$. We shall take the index set T to be either the set Z^+ of non-negative integers or the non-negative half $\mathbb{R}^+ = [0, \infty)$ of the real line $\mathbb{R}$. The parameter t is usually referred to as the time parameter and a point of the index set is called an epoch. The union of range-spaces of the random variables $X(t), t \in T$, is called the state-space S of the stochastic process. If $X(t) = j \in S$, then we say that the stochastic process is in state j at epoch t.

The state-spaces of the stochastic processes in the three examples given above are finite. A process with a finite state-space will be called a finite stochastic process. This book deals with finite stochastic processes which have the Markov property. Loosely speaking, a process has the Markov property if the knowledge of its state at an epoch t is sufficient to determine the probability distribution of $X(u), u > t$; any additional information about $X(s), s < t$, being irrelevant.

In section 2, we define and describe some properties of a Markov chain $\{X_n, n \in Z^+\}$ which are needed in the study of finite Markov processes. The Poisson process and a first definition of a Markov process are introduced in section 3. In section 4, we discuss the different definitions of a Markov process and establish their mutual equivalence and equivalence with the definition introduced in section 3. The strong Markov property is described in section 5. In the last section 6 of this chapter we discuss some properties of the finite dimensional distributions of a finite Markov process.

2. MARKOV CHAINS

Most of the classical theory of Statistics deals with a sequence

$\{X_n, n \in Z^+\}$ of independent and identically distributed (i.i.d.) r.v.s. Suppose the r.v.s X_n, $n \in Z^+$ are non-negative and integer valued. They are said to be identically distributed iff for every $n \in Z^+$,

$$\Pr[X_n = j] = p_j, \qquad j = 0, 1, 2, \ldots,$$

where $p_j \geq 0$, $j \in Z^+$ and $\sum_{j=0}^{\infty} p_j = 1$. They are independently distributed iff for every $n \geq 2$,

$$\Pr[X_1 = j_1, \ldots, X_n = j_n] = \prod_{r=1}^{n} \Pr[X_r = j_r]$$

for all $j_1, \ldots, j_n \in Z^+$.

A Markov chain constitutes the first weakening of the assumption of i.i.d. nature of the r.v.s. Suppose then that $\{X_n, n \in Z^+\}$ is a sequence of discrete r.v.s taking values in a finite or a countably infinite subset S of the real line $\mathbb{R}$. Such a sequence is said to constitute a Markov chain with state-space S iff for every $n \geq 1$, $j_0, j_1, \ldots, j_{n+1} \in S$,

$$\Pr[X_{n+1} = j_{n+1} \mid X_0 = j_0, \ldots, X_n = j_n]$$

$$= \Pr[X_{n+1} = j_{n+1} \mid X_n = j_n] \quad , \qquad (1)$$

whenever the conditioning event $[X_0 = j_0, \ldots, X_n = j_n]$ on the left has positive probability. Here we initiate the sequence at n=0 rather than the usual n = 1, because in the study of stochastic processes, it is customary to treat n as a time parameter and to regard X_0 as the r.v. representing state of the initial state of the process. It is for this reason that the distribution of X_0 is usually known as the initial distribution.

Suppose that the initial distribution of a Markov chain $\{X_n, n \in Z^+\}$ is specified by

$$Pr[X_o = j] = a_j, \; j \in S, \; a_j \geq 0, \quad \sum_{j \in S} a_j = 1 ,$$

and let

$$a(n, j, k) = Pr[X_{n+1} = k \mid X_n = j] , \; j, k \in S$$

where $a(n, j, k) \geq 0$ for each $n \in Z^+$ and

$$\sum_{k \in S} a(n, j, k) = 1 , \quad j \in S .$$

A knowledge of $\{a_j, j \in S\}$ and $\{a(n,j,k), n \in Z^+, j, k \in S\}$ enables us to specify the joint distribution of $X_o, X_1, \ldots, X_n$ for each $n \in Z^+$ since

$$\begin{aligned} Pr[X_o = j_o, \ldots, X_n = j_n] &= Pr[X_o = j_o] \prod_{r=0}^{n-1} Pr[X_{r+1} = j_{r+1} \mid X_o = j_o, \ldots, X_r = j_r] \\ &= a_{j_o} \prod_{r=0}^{n-1} a(r, j_r, j_{r+1}) \end{aligned} \tag{2}$$

by virtue of the defining equation (1) of the Markov chain.

Example 1 : Ehrenfest Model of Diffusion.

Suppose two urns U_1 and U_2 contain a total of M balls. At each trial, a ball is selected at random out of the M balls, independently of the results of the earlier trials and of the number of balls in U_1 and U_2 at the specific trial. The selected ball is transferred from the urn it is in, to the other urn. Let X_n denote the number of balls in urn U_1 at the end of the n-th trial. Observe that

$$X_{n+1} = X_n + \Delta_{n+1} ,$$

where $\Delta_{n+1} = \pm 1$ according as the selected ball is from U_2 or U_1. It is easy to check that the state-space of the sequence $\{X_n, n \in Z^+\}$ is $S = \{0, 1, \ldots, M\}$ and that

$$\Pr[X_{n+1} = j_{n+1} \mid X_o = j_o, \ldots, X_n = j_n]$$

$$= \begin{cases} 1 - j_n/M , & \text{if } j_{n+1} = j_n+1,\ j_n = 0,1,\ldots, M-1, \\ j_n/M , & \text{if } j_{n+1} = j_n-1,\ j_n = 1,2,\ldots, M, \\ 0 , & \text{otherwise,} \end{cases}$$

$$= \Pr[X_{n+1} = j_{n+1} \mid X_n = j_n] .$$

Thus $\{X_n, n \in Z^+\}$ is a Markov chain with state space $S = \{0, 1, \ldots, M\}$.

In this example observe that for fixed $j, k \in S$,

$$\Pr[X_{n+1} = k \mid X_n = j] = \begin{cases} 1-j/M, & k = j+1,\ j = 0,1,\ldots, M-1, \\ j/M , & k = j-1,\ j = 1,2,\ldots, M, \\ 0 , & \text{otherwise} ; \end{cases}$$

for all $n \geq 0$. In other words, the conditional probability of the event $[X_{n+1} = k]$, given the event $[X_n = j]$, does not depend on $n \in Z^+$. This observation leads to the following

Definition 1 : A Markov chain $\{X_n, n \in Z^+\}$ is said to have stationary transition probabilities or to be time homogeneous if the conditional probability $\Pr[X_{n+1} = k \mid X_n = j]$ does not depend on $n \in Z^+$ for all $j, k \in S$.

Hereinafter, unless otherwise specified, a Markov chain will be assumed to have stationary transition probabilities. Let

$$p_{jk} = \Pr[X_{n+1} = k \mid X_n = j], \quad j, \quad k \in S ,$$

which is the probability of a transition from state j to state k in one step and is therefore called a one-step transition probability. These probabilities obviously satisfy the following conditions :

$$p_{jk} \geq 0, \ j, \ k \in S , \quad \sum_{k \in S} p_{jk} = 1, \quad j \in S .$$

If the state-space S is finite, we may take $S = \{1, 2, \ldots, M\}$ without loss of generality and arrange p_{jk} in the form of a square matrix $P = ((p_{jk}))$ of order M, whose (j, k) - element in the j-th row and k-th column is p_{jk} . The matrix P of one-step transition probabilities is thus a stochastic matrix in the sense of the following

Definition 2 : A square matrix A of order M is a stochastic matrix if all its elements a_{jk} are non-negative and for each

$$j = 1, \ldots, M, \quad \sum_{k=1}^{M} a_{jk} = 1.$$

We denote the n-step transition probability

$\Pr[X_{m+n} = k \mid X_m = j]$, $j, k \in S$, $n \in Z^+$, by $p_{jk}^{(n)}$, with the obvious convention that

$$p_{jk}^{(o)} = \delta_{jk} = \begin{cases} 1, & \text{if } j = k , \\ 0, & \text{otherwise} , \end{cases}$$

and $p_{jk}^{(1)} = p_{jk}$. Observe that

$$p_{jk}^{(n+1)} = Pr[X_{m+n+1} = k \mid X_m = j]$$

$$= \sum_{r \in S} Pr[X_{m+n} = r \mid X_m = j] . Pr[X_{m+n+1} = k \mid X_{m+n} = r, X_m = j] \tag{3}$$

$$= \sum_{r \in S} p_{jr}^{(n)} \; p_{rk} , \tag{4}$$

where (3) is a consequence of the theorem of total probabilities and (4) is a consequence of the Markovian nature of the sequence $\{X_n, n \in Z^+\}$. The equation (4) provides us with a recursive way of calculating the n-step transition probabilities.

One can use the above argument also to establish that

$$p_{jk}^{(m+n)} = \sum_{r \in S} p_{jr}^{(m)} \; p_{rk}^{(n)} = \sum_{r \in S} p_{jr}^{(n)} \; p_{rk}^{(m)} \tag{5}$$

for all $m, n \in Z^+$, $j, k \in S$. These equations play an important role in the study of Markov chains and are called the Chapman - Kolmogorov equations.

If $P^{(n)} = ((p_{jk}^{(n)}))$ denotes the matrix of n-step transition probabilities, then the Chapman-Kolmogorov equations (5) become

$$P^{(n+m)} = P^{(n)} P^{(m)} = P^{(m)} P^{(n)} \tag{6}$$

in matrix notation. It is obvious that $P^{(n)}$ is also a stochastic matrix and that in fact it is the n-th power P^n of the matrix P of one-step transition probabilities. In this connection one may observe that by virtue of equation (2) , the initial distribution and the one-step transition probabilities of a Markov chain determine the joint distribution of $X_o, \ldots, X_n$ for all $n \in Z^+$ and hence as a consequence, the joint distribution of any finite subset of the sequence $\{X_n, n \in Z^+\}$.

An important aspect of the study of a Markov chain is the classification of its states. We proceed to describe the classification of the states and refer to Chung (1967) for details.

A state j <u>leads</u> to a state k, $j \to k$, if there exists an $n \in Z^+$ such that $p_{jk}^{(n)} > 0$. Thus, by definition, every state leads to itself as $p_{jj}^{(o)} = 1$. Two states j and k <u>communicate</u> if they lead to each other. A proper subset C of the state-space is a <u>closed</u> set if $j \in C$, $k \notin C$, implies that $p_{jk}^{(n)} = 0$ for all $n \in Z^+$. The state space is closed by definition. A closed set C is <u>minimal closed</u> if no proper subset of C is closed. A Markov chain is <u>irreducible</u> iff its state-space is minimal closed. If for a state j, $p_{jj} = 1$, then j is said to be an <u>absorbing state.</u> It is easy to verify that for an absorbing state j, $p_{jk}^{(n)} \equiv 0$ for all $k \neq j$ and all $n \in Z^+$.

A state j is an <u>essential</u> state if it communicates with every state it leads to; i.e. if $p_{jk}^{(n)} > 0$ implies the existence of an $m \in Z^+$ such that $p_{kj}^{(m)} > 0$. A state which is not essential is an inessential state.

Example 2 : Suppose $\{X_n, n \in Z^+\}$ is a Markov chain with one-step transition probability matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/3 & 2/3 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} .$$

It is easy to check that in this chain, state 4 leads to states 1, 2 and 3 but that state 1 does not lead to state 4. Hence state 4 is inessential. States 2 and 3 communicate with each other and are essential states forming a minimal closed set. The state 1 is an absorbing state.

Observe that in Example 1, if we take M = 1, we get the trivial Markov chain on $\{0, 1\}$ with $p_{01} = p_{10} = 1$, so that

$$p_{00}^{(2n)} = 1 , \quad p_{00}^{(2n+1)} = 0 , \quad n \in Z^+ .$$

Thus the greatest common divisor (g.c.d.) of the set

$\{n \mid p_{00}^{(n)} > 0 , n \geq 1\}$ is 2 and therefore state zero may be said to be periodic with period 2. More generally, a state j is periodic with period d, if the g.c.d. of the set $\{n \mid p_{jj}^{(n)} > 0 , n \geq 1\}$ is $d > 1$. A state j for which d = 1 is called an aperiodic state.

Further classification of states of a Markov chain is based on the probability of its return to its original state. Define

$$f_{jk}^{(1)} = p_{jk}$$