

Statistical Inference for Stochastic Processes

B.L.S. PRAKASA RAO

Statistical Inference for Stochastic Processes

ISHWAR V. BASAWA

*Department of Mathematics
La Trobe University
Australia*

and

B.L.S. PRAKASA RAO

*Indian Statistical Institute
New Delhi Centre,
India*



ACADEMIC PRESS London New York Toronto Sydney San Francisco 1980

A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS INC. (LONDON) LTD.
24/28 Oval Road,
London NW1 7DX

United States Edition Published by
ACADEMIC PRESS INC.
111 Fifth Avenue
New York, New York 10003

Copyright © 1980 by
ACADEMIC PRESS INC. (LONDON) LTD.

All Rights Reserved

No part of this book may be reproduced in any form by photostat, microfilm, or any other means, without written permission from the publishers.

British Library Cataloguing in Publication Data

Basawa, I V

Statistical inference for stochastic processes.
(Probability and mathematical statistics).

1. Stochastic processes

I. Prakasa Rao, B. L. S. II. Series

519.2 QA274 79-50533

ISBN 0-12-080250-3

Printed in Great Britain by Whitstable Litho Ltd., Whitstable, Kent

Preface

The current decade has witnessed an exponential growth of literature on stochastic processes. The preoccupation with model-building has resulted in a large number of theoretical models which seem to lack immediate appeal from a statistical point of view. As a result there appears to be an ever-widening gap between theory and applications in the area of stochastic modelling. The aim of this monograph is to attempt to reduce this gap by directing the interest of future researchers to the inference aspects of stochastic processes. Our choice of topics and approaches used has been partly constrained by the current trends in the literature and partly by our own interest in the area. We have attempted to put on record the work depicting the current state of the art in the general area of inference for stochastic processes as it appears to us.

Many of the results presented here will be appearing in book form for the first time. This is a research monograph written for specialists in the common area of stochastic processes and theoretical statistics; hopefully, it will appeal to both of these broad groups. Advanced graduate and M.Sc. level students should find the book useful; a seminar-type course based on a selected list of topics from the monograph seems feasible for students working towards their Ph.D.s in the North American system.

The topics in the book have been divided into three parts. Part I discusses some standard (special) models in six chapters, in addition to the introductory Chapter 0. In this part, the main ideas and methods are emphasized, avoiding

technicalities as far as possible (within the constraint of maintaining a reasonable level of accuracy). In places, heuristic derivations and sketches of proofs rather than rigorous proofs are given in order not to interrupt the flow of the main arguments. The readers of Part I are expected to be familiar with *classical stochastic processes* at the level of Cox and Miller (1965) or Karlin and Taylor (1975) and with *Mathematical Statistics* at the level of Cramer (1946). Some familiarity with weak convergence (Billingsley, 1968), and martingale limit results (summarized in Appendix 1) will be an advantage.

Part II consists of three chapters on the theory of inference for general processes in discrete and continuous time including diffusion processes. Part III surveys in three chapters recent results on Bayesian, non-parametric and sequential procedures. The treatment of the subject in Parts II and III is more rigorous than that in Part I, and the theory is emphasized here rather than the methods. Familiarity with the *theory of stochastic processes* roughly at the level of Gikhman and Skorokhod (1969) and with *probability theory* at the level of Breiman (1968) is desirable for a smooth reading of most of the topics in Parts II and III.

All the chapters (except Chapters 0 and 1) contain, at the end, "Complements" which are intended to lead the interested reader to various special problems, extensions, etc. for which space could not be found in the text. Appropriate references to most of these excursions are given where solutions could be found. Bibliographical notes at the end of each chapter provide directly related references rather than attempting to trace the origin of individual results. Special topics and results relevant to some of the material in the text are briefly discussed in four appendices at the end of the book.

Any individual chapter can be read independently of the other chapters; the notation, and contents of each chapter are essentially self-contained. Some frequently used notation and abbreviations are listed on page xiii.

It is a pleasure to extend our thanks to the Departments of Statistics at La Trobe and Sheffield Universities, Department of Operations Research at Cornell University, and the Indian Statistical Institute, New Delhi Centre, for the support received during the various stages of the preparation of the manuscript. We would like to record our appreciation of Ms Kirsty Healey of the Academic Press London for her efficient work in getting the manuscript ready for the printers.

We are greatly indebted to all the authors whose work is discussed in this monograph and we extend our special thanks to those who kindly supplied us with the preprints and reprints of their work.

February 1979.

Ishwar V. Basawa
B. L. S. Prakasa Rao

List of Notation

$L_n(\theta; x^n)$ or $L_{x^n}(\theta)$	likelihood function based on the sample $x^n = (x_1, x_2, \dots, x_n)$
$\mathcal{X}, \Omega, \mathcal{Y}$	sample space
Θ	parameter space
$\mathcal{F}, \mathcal{A}, \mathcal{B}$	σ -algebras
\mathbb{R}^p	p -dimensional Euclidean space
\mathbb{R}, \mathbb{R}^1	real line
E	expectation
var	variance
cov	covariance
\mathbf{a}	column vector
\mathbf{a}'	transpose of \mathbf{a}
$V_n = O_p(1)$	for every $\varepsilon > 0$, there exists $A_\varepsilon > 0$ and $N_\varepsilon > 0$ such that for $n > N_\varepsilon$, $P[V_n V_n' > A_\varepsilon] < \varepsilon$.
$V_n = o_p(1)$	V_n tends to zero in probability as $n \rightarrow \infty$
χ_A	indicator function of set A
\sum	convergence in law, distribution, weak convergence
$P(\cdot)$	probability of the event in parens
\xrightarrow{p}	convergence in probability

$\xrightarrow{\text{a.s.}}$	convergence almost surely
$p\text{-}\lim$	limit in probability
$N(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$N_p(\mu, \Sigma)$	p -dimensional multivariate normal distribution with mean vector μ and covariance matrix Σ
$f_n \cong g_n$	$\frac{f_n}{g_n} \rightarrow 1$ as $n \rightarrow \infty$
$\ \cdot\ $	ordinary Euclidean norm on \mathbb{R}^p
$\langle \mathbf{x}, \mathbf{y} \rangle$ or (\mathbf{x}, \mathbf{y})	inner product of \mathbf{x} and \mathbf{y} in \mathbb{R}^p
$\mu \ll \nu$	μ is absolutely continuous with respect to ν
$\mu \sim \nu$	μ and ν are mutually absolutely continuous
$\det(\mathbf{A})$	determinant of a square matrix \mathbf{A}
i.i.d.	independent and identically distributed
MLE*	maximum likelihood estimator
LEE*	likelihood equation estimator
AMLE	approximate maximum likelihood estimator
MMLE	modified maximum likelihood estimator
UMP	uniformly most powerful
UMPU	uniformly most powerful unbiased
DF	distribution free
S-E	spherically exchangeable
H-L	Hodges-Lehmann
UMPDF	uniformly most powerful distribution free
MPDF	most powerful distribution free
SPRT	sequential probability ratio test
GSPRT	generalized sequential probability ratio test
GSP	generalized stochastic process
LAN	local asymptotic normality
SII	stationary independent increment

(* In a majority of cases treated in the book an MLE and a consistent LEE are identical. For this reason the two terms will frequently be used synonymously in the text.)

Contents

<i>Preface</i>	xi
<i>List of Notation</i>	xiii
Chapter 0 Introductory Examples of Stochastic Models	
Example 1. A Random Walk Model for Neuron Firing	6
Example 2. Chain Binomial Models in Epidemiology	2
Example 3. A Population Growth Model	3
Example 4. A Spatial Model for Plant Ecology	4
Example 5. A Cluster Process for Population Settlements	4
Example 6. A Model in Population Genetics	5
Example 7. A Storage Model	5
Example 8. A Compound Poisson Model for Insurance Risk	6
Example 9. System Reliability Models	6
Example 10. A Model for Cell Kinetics	6
Example 11. Queueing Models for Telephone Calls	7
Example 12. Clustering Splitting Model for Animal Behaviour	8
Example 13. Prediction of Economic Time Series	8
Example 14. Signal Estimation	9
 Bibliographical Notes	 9

PART I SPECIAL MODELS**Chapter 1 Basic Principles and Methods of Statistical Inference**

1. Introduction	10
2. The Likelihood Function and Sufficient Statistics	10
3. Frequency Approach	12
4. The Bayesian Approach	13
5. Asymptotic Inference	14
6. Nonparametric Methods	18
7. Sequential Methods	19

Bibliographical Notes	19
-----------------------	----

Chapter 2 Branching Processes

1. Introduction	20
2. The Galton–Watson Process	20
3. The Markov Branching Process	29

Bibliographical Notes	32
Complements	33

Chapter 3 Simple Linear Models

1. Introduction	35
2. Prediction	36
3. Filtering Problem	38
4. Parameter Estimation	41
5. Further Topics	47

Bibliographical Notes	49
Complements	50

Chapter 4 Discrete Markov Chains

1. Introduction	52
2. Finite Markov Chains	52
3. A Macro Model (Finite State Space)	72
4. Grouped Markov Chains (Finite State Space)	74
5. Countable State Space	77

Bibliographical Notes	79
Complements	79

Chapter 5 Markov Chains in Continuous Time

1. Introduction	81
2. Finite Markov Chains	81
3. Queueing Models	83
4. Pure Birth Process	86
5. The Birth and Death Process	90

Bibliographical Notes	95
Complements	95

Chapter 6 Simple Point Processes

1. Introduction	98
2. Homogeneous Poisson Process	98
3. Non-homogeneous Poisson Process	102
4. Compound Poisson Process	105
5. Further Topics	109

Bibliographical Notes	117
Complements	117

PART II GENERAL THEORY

Chapter 7 Large Sample Theory for Discrete Parameter Stochastic Processes

1. Introduction	119
2. Estimation	119
3. Efficient Tests of Simple Hypotheses	137
4. Large Sample Tests	139
5. Optimal Asymptotic Tests of Composite Hypotheses	147
6. Further Topics	160

Bibliographical Notes	161
Complements	162

Chapter 8 Large Sample Theory for Continuous Parameter Stochastic Processes

1. Introduction	166
2. Observable Coordinates	166
3. The General Problem	168

4. Testing Hypotheses	170
5. Estimation	181
6. Estimating the Infinitesimal Generator for a Continuous Time Finite State Markov Process	190
7. Further Topics	198
Bibliographical Notes	198
Complements	199

Chapter 9 Diffusion Processes

1. Introduction	201
2. Diffusion Processes	203
3. Absolute Continuity of Measures for Diffusion Processes	205
4. Parameter Estimation in a Linear Stochastic Differential Equation	210
5. Asymptotic Likelihood Theory for Multidimensional Diffusion Processes	218
6. Hypotheses Testing for Parameters of Diffusion Processes	228
7. Sequential Estimation of the Parameters of a Diffusion Process	233
8. Sequential Test for Diffusion Processes	238
9. Bayes Estimation for Diffusion Processes	241
10. Further Topics	247
Bibliographical Notes	250
Complements	250

PART III FURTHER APPROACHES

Chapter 10 Bayesian Inference for Stochastic Processes

1. Introduction	255
2. Preliminaries	256
3. The Bernstein–Von Mises Theorem	257
4. Asymptotic Behaviour of Bayes Estimators	264
5. Bayesian Testing	279
6. Further Topics	284
7. Proof of Tightness of Processes	286
Bibliographical Notes	289
Complements	289

Chapter 11 Nonparametric Inference for Stochastic Processes

- | | |
|--|-----|
| 1. Introduction | 294 |
| 2. Nonparametric Estimation for Stochastic Processes | 294 |
| 3. Nonparametric Tests for Stochastic Processes | 324 |
| 4. Further Topics | 338 |

Bibliographical Notes	340
Complements	340

Chapter 12 Sequential Inference for Stochastic Processes

- | | |
|--|-----|
| 1. Introduction | 343 |
| 2. Sequential Estimation for Stochastic Processes | 344 |
| 3. Sequential Tests of Hypotheses for Stochastic Processes | 356 |
| 4. Sequential Tests for the Drift of Wiener Process | 373 |
| 5. Further Topics | 379 |

Bibliographical Notes	380
Complements	380

Appendix 1 Martingales

- | | |
|--|-----|
| 1. Martingales and Limit Theorems | 383 |
| 2. A Random Central Limit Theorem for Martingales | 388 |
| 3. Embedding Submartingales in Wiener Process with Drift | 391 |
| 4. Structure of Continuous Parameter Martingales | 393 |

Appendix 2 Stochastic Differential Equations

- | | |
|---|-----|
| 1. Stochastic Integrals | 396 |
| 2. Central Limit Theorem for Vector-valued Stochastic Integrals | 405 |
| 3. Stochastic Differential Equations | 408 |

Appendix 3 Proof of Sudakov's Lemma (Theorem 2.3) of Chapter 12 411**Appendix 4 Generalized Functions and Generalized Stochastic Processes** 413**References** 415**Index** 433

Chapter 0

Introductory Examples of Stochastic Models

Stochastic models are being increasingly used in scientific research in an ever-widening spectrum of disciplines. In this introductory chapter we describe some of their applications. The examples given below are by no means exhaustive nor are they intended to be a representative selection of possible fields of application or the processes to be discussed in this monograph. Their purpose is merely to highlight the diversity of applications.

EXAMPLE 1. A RANDOM WALK MODEL FOR NEURON FIRING

The neuron fires when the membrane potential reaches a critical threshold value, say C . Excitatory and inhibitory impulses are the inputs for the neuron; these impulses arrive according to a Poisson process. Each excitatory impulse increases, and each inhibitory impulse decreases, the membrane potential by a random quantity x with a common probability density $p(x)$. After each firing the membrane potential is reset at zero and the process is repeated. A typical realization of the model is given in Fig. 1.

Let $\{Y_1, Y_2, \dots\}$ denote the times at which the neuron fires. The process of

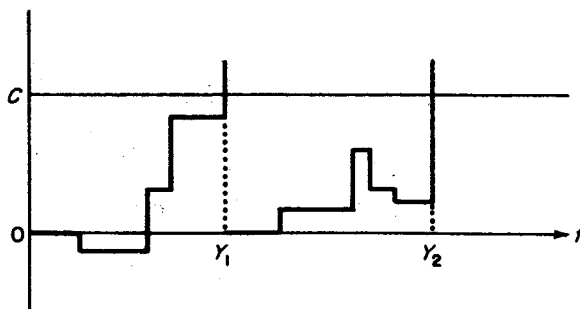


Fig. 1. Realization of a random walk model for neuron firing.

inter-spike intervals $\{Y_1, Y_2 - Y_1, Y_3 - Y_2, \dots\}$ is of interest to the neurologist. Various questions such as the moments, correlation structure, and joint distributions of these intervals are relevant. This model is an example of a *point process*. Statistical questions related to some simple point processes are discussed in Chapter 6.

EXAMPLE 2. CHAIN BINOMIAL MODELS IN EPIDEMIOLOGY

Suppose at time $t=0$ there are $S(0)$ susceptibles and $I(0)$ infectives. After a certain latent period of the infection, considered here as a unit of time, some of the susceptibles are infected; thus, at time $t=1$, the initial $S(0)$ susceptibles split into two groups: those who are infected, $I(1)$ in number say, and the remaining susceptibles say, $S(1)$ in number. The process continues until

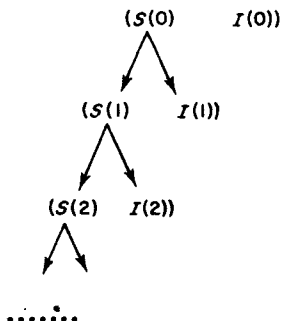


Fig. 2. Chain binomial model.

there are no more infectives in the population. The scheme is indicated in Fig. 2. Note that, in general, $S(t) = S(t+1) + I(t+1)$, ($t=0, 1, 2, \dots$). Assuming that the probability of a susceptible being infected is p we have, for the Greenwood Model,

$$P(S(t+1) = s(t+1) | S(t) = s(t)) = \binom{s(t)}{s(t) - s(t+1)} p^{s(t) - s(t+1)} (1-p)^{s(t+1)}.$$

The process $\{S(t), t=0, 1, 2, \dots\}$ is a Markov chain with the transition probabilities given above. Given $S(0)$, these transition probabilities specify the process $\{S(t)\}$ completely.

In the Greenwood Model, the number of infectives at time t is assumed not to influence the probability of infection during $(t, t+1)$. However, we can remove this restriction by viewing $\{(S(t), I(t))\}$ as a bivariate Markov chain with transition probabilities

$$\begin{aligned} P(I(t+1) = i(t+1), S(t+1) = s(t) - i(t+1) | I(t) = i(t), S(t) = s(t)) \\ = \binom{s(t)}{s(t+1)} (1-p)^{i(t)s(t+1)} \{1 - (1-p)^{i(t)}\}^{s(t) - s(t+1)} \end{aligned}$$

This now becomes the Reed-Frost Model. Markov chains provide a powerful tool in model building. The present example is a typical application of Markov chains in epidemiology. Problems of inference for discrete time Markov chains are discussed in Chapter 4.

EXAMPLE 3. A POPULATION GROWTH MODEL

Suppose an organism produces a random number, say Y , of offspring with $p_k = P(Y=k)$, ($k=0, 1, 2, \dots$), $\sum_0^\infty p_k = 1$. Each offspring in turn produces organisms independently according to the same distribution $\{p_k\}$. If $Z(t)$ denotes the population size at the t th generation, ($t=0, 1, 2, \dots$) it is seen that $\{Z(t)\}$ is a Markov chain with transition probabilities given by

$$P(Z(t) = j | Z(t-1) = i) = P(Y_1 + Y_2 + \dots + Y_i = j)$$

where Y_1, Y_2, \dots are independent identically distributed random variables with probability distribution $\{p_k\}$. The model just described is the so-called Galton-Watson branching process which is studied extensively in the literature. This model is an important example of a non-ergodic type process for which certain difficulties in extending the usual large-sample methods of inference arise. These are discussed in detail in Chapter 2.

EXAMPLE 4. A SPATIAL MODEL FOR PLANT ECOLOGY

Consider a lattice of sites (points) labelled by integer pairs (i, j) over some region. Let X_{ij} denote the random variable associated with the plant located at the site (i, j) . If we are interested in a study of the pattern of infection in an array of plants we may set $X_{ij} = 1$ if the plant located at (i, j) is infected, and zero otherwise. It may be reasonable to assume that the conditional distribution of X_{ij} given all other site values depends only on the four nearest neighbour sites, namely, on $X_{i-1,j}$, $X_{i+1,j}$, $X_{i,j-1}$, and $X_{i,j+1}$. This is a natural extension of the Markovian assumption to the spatial process $\{X_{ij}\}$. Spatial processes based on the "nearest-neighbour" formulation (as in this example) are of a recent origin. Work on statistical procedures relevant for such models is still at a rudimentary stage, at the time of writing. Spatial models will not be discussed in this monograph.

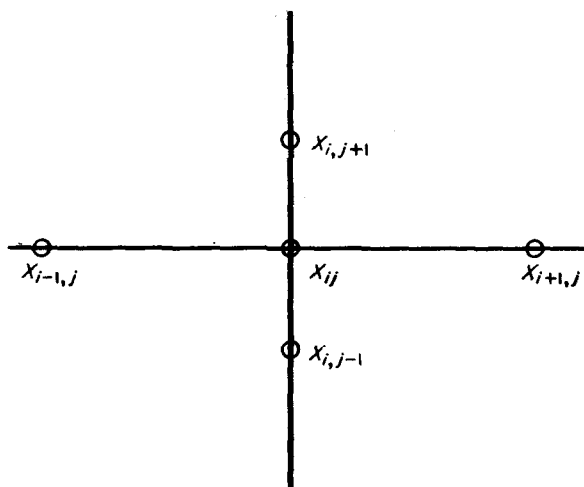


Fig. 3. Spatial ecological model.

EXAMPLE 5. A CLUSTER PROCESS FOR POPULATION SETTLEMENTS

An important area of research in geography is the study of spacings between settlements of human population and of sizes of settlements. It is frequently assumed that the settlements are located according to a spatial

Poisson process; thus one assumes that

$$P(\text{one settlement in an area } \delta A) = \lambda \delta A + o(\delta A)$$

$$P(\text{two or more settlements in } \delta A) = o(\delta A),$$

$$P(\text{no settlements in } \delta A) = 1 - \lambda \delta A + o(\delta A).$$

Then, it follows that the number of settlements in any given area A is a Poisson random variable with mean λA . Given the locations of the settlements the size of any settlement is assumed to follow a logarithmic series growth distribution. One then obtains a negative binomial distribution for the size of the population in a given area. This model is extensively used in geography, but will not be discussed any further in this monograph.

EXAMPLE 6. A MODEL IN POPULATION GENETICS

Consider a population of $2N$ genes each of which belongs to one of the two genotypes, say A and B . Let $X(t)$ denote the proportion of type A genes in the i th generation. Assuming that the total number of genes remains the same from generation to generation, and neglecting selection and mutation effects, the genes in the $(t+1)$ th generation may be assumed to be a random sample of size $2N$ of genes in the t th generation. The sequence of random variables $\{X(t)\}$, $t=1, 2, \dots$, forms a Markov chain. Conditionally on $X(t-1)=x$, $2NX(t)$ will be a binomial random variable with success probability x and index $2N$. By taking suitable limits one can approximate this Markov chain by a continuous time Markov process with a continuous state space $[0, 1]$. Such an approximation is an example of a diffusion process. Diffusion processes are discussed in detail in Chapter 9.

EXAMPLE 7. A STORAGE MODEL

Let $X(t)$ denote the annual random input during the years $(t, t+1)$, M the annual non-random release at the end of each year, and $Z(t)$ the content of the dam after the release. Then, it can be seen that

$$Z(t+1) = \min\{Z(t) + X(t), K\} - \min\{Z(t) + X(t), M\}$$

where K is the capacity of the dam, and $t=0, 1, 2, \dots$. If the inputs $\{X(t)\}$ are assumed to be independent random variables, the sequence $\{Z(t)\}$ will form a Markov chain. If, however, $\{X(t)\}$ is assumed to follow a Markov chain, $\{X(t), Z(t)\}$ can be viewed as a bivariate Markov chain. Statistical questions specifically related to storage models as such will not be covered in the book; however, discrete Markov chains will be studied in Chapter 4.