Lecture Notes in Statistics

Vol. 1: R. A. Fisher: An Appreciation. Edited by S. E. Fienberg and D. V. Hinkley. xi, 208 pages, 1980.

Vol. 2: Mathematical Statistics and Probability Theory. Proceedings 1978. Edited by W. Klonecki, A. Kozek, and J. Rosiński. xxiv, 373 pages, 1980.

Vol. 3: B. D. Spencer, Benefit-Cost Analysis of Data Used to Allocate Funds. viii, 296 pages, 1980.

Vol. 4: E. A. van Doorn, Stochastic Monotonicity and Queueing Applications of Birth-Death Processes. vi, 118 pages, 1981.

Vol. 5: T. Rolski, Stationary Random Processes Associated with Point Processes. vi, 139 pages, 1981.

Vol. 6: S. S. Gupta and D.-Y. Huang, Multiple Statistical Decision Theory: Recent Developments. viii, 104 pages, 1981.

Vol. 7: M. Akahira and K. Takeuchi, Asymptotic Efficiency of Statistical Estimators. viii, 242 pages, 1981.

Vol. 8: The First Pannonian Symposium on Mathematical Statistics. Edited by P. Révész, L. Schmetterer, and V. M. Zolotarev. vi, 308 pages, 1981.

Vol. 9: B. Jørgensen, Statistical Properties of the Generalized Inverse Gaussian Distribution. vi, 188 pages, 1981.

Vol. 10: A. A. McIntosh, Fitting Linear Models: An Application of Conjugate Gradient Algorithms. vi, 200 pages, 1982.

Vol. 11: D. F. Nicholls and B. G. Quinn, Random Coefficient Autoregressive Models: An Introduction. v, 154 pages, 1982.

Vol. 12: M. Jacobsen, Statistical Analysis of Counting Processes. vii, 226 pages, 1982.

Vol. 13: J. Pfanzagl (with the assistance of W. Wefelmeyer), Contributions to a General Asymptotic Statistical Theory. vii, 315 pages, 1982.

Vol. 14: GLIM 82: Proceedings of the International Conference on Generalised Linear Models. Edited by R. Gilchrist. v, 188 pages, 1982.

Vol. 15: K. R. W. Brewer and M. Hanif, Sampling with Unequal Probabilities. vii, 164 pages, 1983.

Vol. 16: Specifying Statistical Models: From Parametric to Non-Parametric, Using Bayesian or Non-Bayesian Approaches. Edited by J. P. Florens, M. Mouchart, J. P. Raoult, L. Simar, and A. F. M. Smith. xi, 204 pages, 1983.

Vol. 17: I. V. Basawa and D. J. Scott, Asymptotic Optimal Inference for Non-Ergodic Models. ix, 170 pages, 1983.

Lecture Notes in Statistics

Edited by D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, and K. Krickeberg

17

Ishwar V. Basawa
David John Scott

Asymptotic Optimal Inference for Non-ergodic Models

Springer-Verlag
New York Heidelberg Berlin

Ishwar V. Basawa
David John Scott
Department of Mathematical Statistics
La Trobe University
Bundoora, Australia 3083

AMS Subject Classification: 62G99

Library of Congress Cataloging in Publication Data
Basawa, Ishwar V.
Asymptotic optimal inference for non-ergodic models.
(Lecture notes in statistics; v. 17)
Bibliography: p.
Includes index.
1. Asymptotic efficiencies (Statistics) 2. Ergodic theory. 3. Estimation theory. I. Scott, David J. II. Title. III. Series: Lecture notes in statistics (Springer-Verlag);
QA276.8.B39 1983 519.5′44 82-19448

9 8 7 6 5 4 3 2 1

ISBN-13: 978-0-387-90810-6 e-ISBN-13: 978-1-4612-5505-5
DOI: 10.1007/978-1-4612-5505-5

PREFACE

This monograph contains a comprehensive account of the recent work of the authors and other workers on large sample optimal inference for non-ergodic models. The non-ergodic family of models can be viewed as an extension of the usual Fisher-Rao model for asymptotics, referred to here as an ergodic family. The main feature of a non-ergodic model is that the sample Fisher information, appropriately normed, converges to a non-degenerate random variable rather than to a constant. Mixture experiments, growth models such as birth processes, branching processes, etc., and non-stationary diffusion processes are typical examples of non-ergodic models for which the usual asymptotics and the efficiency criteria of the Fisher-Rao-Wald type are not directly applicable. The new model necessitates a thorough review of both technical and qualitative aspects of the asymptotic theory. The general model studied includes both ergodic and non-ergodic families even though we emphasise applications of the latter type.

The plan to write the monograph originally evolved through a series of lectures given by the first author in a graduate seminar course at Cornell University during the fall of 1978, and by the second author at the University of Munich during the fall of 1979. Further work during 1979-1981 on the topic has resolved many of the outstanding conceptual and technical difficulties encountered previously. While there are still some gaps remaining, it appears that the mainstream development in the area has now taken a more definite shape.

The introductory Chapter 0 gives an over-view of the main results in an informal setting. Chapters 1 and 2 then develop the new model and the related estimation theory more formally. Proofs of most of the results are included although we have tried to avoid heavy technicalities and over-generalisation. Chapters 3 and 4 are concerned with asymptotic tests for non-ergodic models and specific applications to mixture experiments. A more discursive treatment is given in these chapters although proofs of the main results are included. Chapter 5 is a brief introduction to non-local results. Finally, the appendices contain some relevant information on uniform convergence, and contiguity.

All references to the main sources are relegated to a final section in each chapter, except in Chapter 0 which contains a descriptive and partly historical account of the results.

A systematic treatment of Bayes, non-parametric and sequential methods for the non-ergodic family would be of interest. These topics are open for future research at the time of writing.

The book *Statistical Inference for Stochastic Processes* by Basawa and Prakasa Rao, Academic Press (1980), provides a background and collateral reference material for some of the topics treated. It is not however a superfluous prerequisite for reading this monograph.

There are a number of people we would like to thank for their assistance: Dr. P. Jeganathan kindly sent us a copy of his Ph.D. thesis and pre-prints of his work; Dr. A. R. Swenson also kindly provided us with a copy of his thesis and a pre-print of a paper; Dr. T. J. Sweeting helped us in correspondence concerning the subject, sent us a pre-print of a paper and suggested changes to the original draft; Dr. P. D. Feigin helped us greatly by carefully reading the original draft and providing numerous corrections and suggestions. We would like to thank Judy Stewart for her excellent typing of both drafts of the work and Irene Hudson and Richard Huggins for proof-reading the original typescript

NOTATION

The following notation and abbreviations are used throughout.

§, §§	section, sections
R^k	k-dimensional Euclidean space
$\mathcal{B}^k$	the k-dimensional Borel sets
μ_k	k-dimensional Lebesgue measure

These definitions apply for $k = \infty$ also. When $k = 1$ it is omitted.

B^c	the complement of the set B
$\overline{B}$	the closure of the set B
∂B	the boundary of the set B ; $B = \overline{B} \cap \overline{B^c}$
$\chi(A)$	the indicator function of the set A ; $\chi(A)(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$
$\mathcal{A} \times \mathcal{B}$	for σ-fields $\mathcal{A}$ and $\mathcal{B}$, the product σ-field
$\Vert a_{ij} \Vert_{i,j=1}^{k,\ell}$	the $(k \times \ell)$ matrix with (i,j)-entry a_{ij}
$\text{diag}\Vert a_i \Vert_{i=1}^{k}$	the $(k \times k)$ diagonal matrix with ith diagonal entry a_i
$\text{diag}\{a_1, a_2, \ldots, a_k\}$	same as $\text{diag}\Vert a_i \Vert_{i=1}^{k}$
$\text{tr}(A)$	the trace of the matrix A
x^T	the transpose of the vector or matrix x
$\lvert A \rvert$	for a matrix A , the matrix norm; $\lvert A \rvert = \{\text{tr}(A^T A)\}^{\frac{1}{2}}$
$A_n \to 0$	for a sequence of matrices $\{A_n\}$ convergence in the metric given by the matrix norm
$\lvert a \rvert$	for a vector a , the Euclidean norm

$a_n \to 0$	for a sequence of vectors $\{a_n\}$, convergence in the Euclidean metric
p.d.	positive definite
$\mathcal{L}(X\|P)$ or $\mathcal{L}(X)$	the distribution or law of the random variable or vector X when the probability measure is P; when P is omitted simply the law or distribution of X
$F_X(\cdot)$	the distribution function of the random variable or vector X when the argument is an (extended) real number or vector; the associated measure when the argument is a Borel set
	is distributed as
i.i.d.	independent and identically distributed
$N(\underset{\sim}{\mu},\Sigma)$	the multivariate normal distribution with mean vector μ and variance matrix Σ
$N(\mu,\sigma^2)$	the normal distribution with mean μ and variance σ^2.
E_θ	expectation when the underlying probability measure is P_θ
var_θ	variance when the underlying probability measure is P_θ
$P * Q$	the convolution of the two measures P and Q
$\\|P-Q\\|$	the L_1-norm of $P - Q$ for the two probability measures P and Q on $(\Omega,\mathcal{A})$ defined by $\\|P-Q\\| = 2\sup\{\|P(A)-Q(A)\|;\ A\in\mathcal{A}\}$

$\frac{dP}{dQ}$	the Radon-Nikodym derivative of P with respect to Q
a.c.	absolutely continuous
$\Rightarrow$	convergence in distribution or law or weak convergence
$\xrightarrow{p}$	convergence in probability
$\xrightarrow{L_2}$	convergence in mean square
$\xrightarrow{a.s.}$	almost sure convergence

The subscript c on a convergence symbol (e.g. $\Rightarrow_c$) indicates continuous convergence. The subscript u (e.g. $\Rightarrow_u$) denotes uniform convergence on all compact subsets. For a discussion of these concepts see Appendix (A.1).

$o_p(1)$	a term which converges to zero in probability
a.s.	almost surely
MLE	maximum likelihood estimator
LR	likelihood ratio
LAN	locally asymptotically normal
LAMN	locally asymptotically mixed normal
ULAMN	uniform locally asymptotically mixed normal
u.m.p. (u.)	uniformly most powerful (unbiased).

CONTENTS

Chapter 0. An Over-view

Chapter 1. A General Model and Its Local Approximation

Chapter 2. Efficiency of Estimation

Chapter 3. Optimal Asymptotic Tests

Chapter 4. Mixture Experiments and Conditional Inference