Parametric Statistical Inference

Basic Theory and Modern Approaches

by

SHELEMYAHU ZACKS

Parametric Statistical Inference

Basic Theory and Modern Approaches

by

SHELEMYAHU ZACKS

State University of New York at Binghamton

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon of Canada, Suite 104, 150 Consumers Road, Willowdale, Ontario M2J 1P9, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., P.O. Box 544, Potts Point, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
FEDERAL REPUBLIC OF GERMANY	Pergamon Press GmbH, 6242 Kronberg-Taunus, Hammerweg 6, Federal Republic of Germany

First edition 1981

British Library Cataloguing in Publication Data

Zacks, Shelemyahu
Parametric statistical inference.
- (International series in nonlinear mathematics; vol.4). - (Pergamon international library).
1. Mathematical statistics
2. Probabilities
I. Title II. Series
519.5'4 QA276 80-41715

ISBN 0-08-026468-9 (Hardcover)

In order to make this volume available as economically and as rapidly as possible the author's typescript has been reproduced in its original form. This method has its typographical limitations but it is hoped that they in no way distract the reader.

Printed and bound in Great Britain by
William Clowes (Beccles) Limited, Beccles and London

PREFACE

The present textbook is aimed at the population of senior undergraduate and graduate students in statistics, or mathematical sciences, who have previously taken an introductory course in probability and statistics. The objective of the present book is to present the basic theory and some of the recent developments in a framework of a two-semester intermediate level course on the theory of statistics. The mathematical prerequisites for this course are only advanced calculus and matrix algebra.

It has been a common experience that students taking an intermediate theory course often lack the proficiency required in the theory of statistical distributions, despite the fact that they have had a previous course in probability and statistics. For this reason the beginning of the book (Chapter 2) is devoted to the development of the required tools in the area of statistical theory. This chapter is followed by a discussion in Chapter 3 of the notions of statistical information, sufficient statistics, completeness, exponential families and the important information functions: the Fisher and the Kullback-Leibler. Chapter 4 is presenting the classical theory of testing hypotheses. Chapter 5 is devoted to estimation theory. A whole spectrum of estimation approaches is developed, from unbiased estimation to robust estimation, including jackknifing, maximum-likelihood, equivariant, moment-equations and pre-test estimators. In Chapter 6 the concept of relative efficiency of estimators in small and large samples is discussed. Chapter 7 is concerned with confidence and tolerance intervals. Finally, in Chapter 8, the theory of statistical inference is represented from the point of view of decision theory and the Bayesian approach. Many examples are completely developed in the text and numerous problems are given at the end of each chapter. The text was tried in various graduate courses at Case Western University, Tel Aviv University, The Technion and at Virginia Polytechnic Institute and State University. The experience at all of these institutions has been very encouraging, despite the fact that the students were heterogeneous in their background. The students who master the material in the present book are well prepared for further advanced studies and for research in applied statistics. The material in the present text is as important for the curriculum in applied statistics as it is for that in theoretical statistics.

The author wishes to acknowledge the help provided by many of his colleagues and graduate students. In particular, Professor Robert Berk provided excellent comments on an early draft, when we both spent a turmoiled year at Tel Aviv University. In addition, Professors Dan Anbar, Chris Tsokos, and Micha Yadin read the text carefully and provided valuable comments. The author wishes to express his sincere thanks to Tel Aviv University, Case Western Reserve University, and Virginia Polytechnic Institute and State University for providing the conditions and assistance during the years in which the book was written. Special acknowledgment is given to the Office of Naval Research for continuing research contracts, resulting in many papers, from which examples were drawn. Finally, the author is grateful to Miss Judy Galliher who typed the final version of the text with devotion and a lot of intelligence.

Blacksburg, Virginia
July, 1980

Shelemyahu Zacks

CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 1
General Review

1.1 INTRODUCTION

The theory of statistical inference is developing very fast, with thousands of research papers and dozens of monographs and textbooks published every year. It is impossible therefore to give in one short volume a complete or even partial account of all the important results. The objective of the present volume is to discuss some fundamental results in certain subfields of statistical inference, in a manner which does not require much previous preparation. The emphasis in the present monograph is on modern trends and developments. Certain subject matters could be only mentioned or briefly discussed, mainly because they require advanced treatment. In order to make the development accessible to a wide population of readers, not necessarily mathematicians, we have adopted a level of exposition which does not require much more than advanced calculus and some linear algebra. The theoretical discussion is accompanied with many examples, which illustrate the theory in special cases of interest. On the whole there are over one hundred such examples with additional short illustrations and many problems at the end of each chapter. Hundreds of references are cited and listed at the end. In order to avoid the prerequisite of proficiency in the theory of statistical distributions and to establish some required tools, the second chapter is devoted to a brief discussion of statistical distributions and their properties. Chapter 3 is devoted to the problem of sufficient statistics and the information in samples. Chapter 4 provides some basic results from the theory of testing statistical hypotheses. In Chapter 5 we develop the classical theory of estimation. The efficiency of estimators and some large sample properties are discussed in Chapter 6. Confidence intervals can be considered to be the bridge between testing hypotheses and point estimation. We study topics from this area in Chapter 7. The development of the statistical theory of inference throughout Chapters 3-7 is from the classical point of view of the frequentist approach. Chapter 8 is devoted to the other point of view, namely the decision theoretical and Bayesian approach. In this chapter we again treat testing and estimation problems in the framework of statistical decision theory. Some results from the theory of sequential analysis are given in Chapter 4, 7 and 8. The treatment of this subject is, however, short and elementary since the subject of sequential analysis is strongly linked with the theory of optimal stopping. A good treatment of this theory requires advanced techniques. A short discussion of robust estimation procedures and adaptive techniques is provided in Chapter 5. There is no discussion, however, in the present book of design of experiments, multivariate analysis, time-series analysis, ranking and classification, stochastic approximation, statistical control theory, cluster analysis, pattern recognition and non-parametric techniques. All these are important subject areas of statistical research which are covered by specific monographs. The topics covered in the present volume are in a sense basic and common to many fields of statistical inference and thus provide a good preparation for further study. In the following sections we review the topics of the book chapter by chapter. This review is done in general terms, trying to avoid technicalities and to provide a general picture of the content of the book. To supplement the general discussion with more technical substance we provide at the beginning of each chapter a general introductory section, in which the topics of the chapter are introduced and discussed.

1.2 STATISTICAL MODELS, DISTRIBUTION FUNCTIONS AND THE ESSENCE OF STATISTICAL INFERENCE

Generally stating, statistical methods are needed when we study stochastic or random phenomena which are not completely predictable. More specifically, in almost all empirical studies there are uncontrollable elements of variability. For example, if we observe the average hourly wind velocity at a given meteorological station the variable under consideration is a stochastic, or random variable. We cannot generally predict with certainty what value this variable will assume in the next hour of observation. A complete characterization of the random variable being observed is given only if we can specify exactly the <u>distribution function</u> of this variable. The distribution function determines the probability that the observed random variable will assume values in specified intervals. Such a distribution function may depend on one or many parameters, which are related to the physical, biological or other type of system which governs the observed phenomenon. Such a mathematical description of an empirical phenomenon is always an ideal representation of the real world. We therefore call it a statistical or probabilistic <u>model</u>.

Statistical models depend generally on sets of assumptions. In the decision whether to apply a certain model to explain empirical observations we have first of all to question the relevance of the assumptions. If we do believe that the assumptions are reasonable we have to check whether the adopted model fits the observed data well. In Chapter 2 we study various types of distribution functions, set the assumptions needed for the applicability of these types as models and mention possible applications. The chapter is self-contained and can be studied without any prerequisite in probability or statistics. In addition to the basic properties of univariate and multivariate distributions the chapter contains material on conditional distributions, mixtures of distributions, the algebra of covariances, probability integral transforms, order statistics and limit theorems. The types of distributions considered are the Binomial, Hypergeometric, Poisson, Negative-Binomial, Rectangular, Beta, Gamma, Weibull, Extreme-Value, Normal, Multinomial, Multinormal, the order statistics, t-, F- and the distribution of the sample correlation. All these classes of distribution functions depend on one or several parameters. The parameters may vary over a specified range, called the <u>parameter space</u>. The collection of all the distribution functions of a given type, when the parameters vary over their possible range is called a <u>family</u>. Thus, we speak about the family of Binomial distributions, etc. Generally, in the statistical analysis of observed data, even if we apply a model specifying a certain family of distributions the actual values of the characterizing parameters are unknown. Assuming that the observations follow a distribution function belonging to the family specified by the model, the problems of statistical inference are those of deciding whether the actual parameters belong to a specified set in the parameter space or to estimate their actual values. These are problems of <u>parametric inference</u> known as testing hypotheses or estimation problems, respectively. One can consider also a different formulation of statistical inference problems, known as <u>statistical decision</u> problems. These are problems of choosing actions from specified sets of actions so that certain expected loss (or payoff) functions will be minimized (or maximized). The loss or payoff depends not only on the chosen action but also on the actual (true) distribution function. If this distribution is unknown observations are taken to supplement information on this distribution. The optimal choice of an action depends then on the observed values. The problem of statistical decision theory is to determine a proper function of the observations, called a <u>decision function</u>, according to which an optimal action can be taken.

In certain problems of statistical inference it is unwarranted to consider a model specifying a certain family of distributions of a known type or form. In other words, we may not have sufficient prior knowledge or evidence to adopt one family of distributions and not another. In such cases we may resort to more general methods of inference known as distribution-free methods. Distribution-free methods, generally require larger number of observations to attain certain precision than parametric methods. On the other hand, the distribution-free methods are independent of assumptions about the particular functional type of the actual distribution function. For this reason the distribution-free methods are often considerably more robust and sturdy than the parametric methods. If the parametric model is wrong the consequences may lead sometimes to substantial loss or error. Most of the discussion in the present monograph relates to parametric inference. A few short sections are devoted to distribution-free estimation. There are many papers in the literature concerning robust but parametric methods of estimating location parameters of symmetric distributions. For a survey of the important results on this topic see Hampel (1973) and the monograph of Andrews, Bickel, Hampel, Huber, Rogers and Tukey (1972).

Another class of general models is the class of non-parametric models. Many of non-parametric methods are based on the ranks of the observed variables in the sample. We refer the reader to the books of Hollander and Wolfe (1973), Lehmann (1975) or Gibbons (1976) for a study of these methods and for references to the rich literature.

1.3 THE INFORMATION IN SAMPLES AND SUFFICIENT STATISTICS

In the classical approach the information in the sample is given only by the sample values of the observable random variables $X_1, X_2, \ldots, X_n$. The Bayesian approach on the other hand allows also the addition of information extraneous to the sample (see Section 1.8). Let $\underset{\sim}{X}$ be the observable vector of the random variables to the sampled. Suppose that we consider a parametric model specified by a family $\mathcal{F}$ of distribution functions $F(x, \theta)$, where the parameter(s) θ belong(s) to a parameter space Θ. In relation to the family $\mathcal{F}$ a non-negative weight function, called the likelihood function, is defined over Θ. This is a function $L(\theta; \underset{\sim}{X})$ of θ and of $\underset{\sim}{X}$ which assigns for each value of θ a value proportional to the probability density function of $\underset{\sim}{X}$ under θ. Two likelihood functions $L_1(\theta; \underset{\sim}{X})$ and $L_2(\theta; \underset{\sim}{X})$ are called equivalent if $L_1(\theta; \underset{\sim}{X}) = C\, L_2(\theta; \underset{\sim}{X})$ for all $\theta \in \Theta$. C is some positive constant that may depend on $\underset{\sim}{X}$ but not on θ. We denote this equivalence by $\sim$. If two different samples yield vectors $\underset{\sim}{X}^{(1)}$ and $\underset{\sim}{X}^{(2)}$ such that $L(\theta; \underset{\sim}{X}^{(1)}) \sim L(\theta; \underset{\sim}{X}^{(2)})$ for all θ in Θ, then the two samples are said to have the same amount of information on the model (Basu, 1975). The question is whether we can reduce the data to some function or functions of the vector $\underset{\sim}{X}$ without losing information. Functions of the observed random vector $\underset{\sim}{X}$, which do not depend on unknown