BAYESIAN INFERENCE IN STATISTICAL ANALYSIS

George E.P. Box
University of Wisconsin

George C. Tiao
University of Chicago

Wiley Classics Library Edition Published 1992

A Wiley-Interscience Publication
JOHN WILEY AND SONS, INC.
New York / Chichester / Brisbane / Toronto / Singapore

This page intentionally left blank

BAYESIAN INFERENCE IN STATISTICAL ANALYSIS

This page intentionally left blank

BAYESIAN INFERENCE IN STATISTICAL ANALYSIS

George E.P. Box
University of Wisconsin

George C. Tiao
University of Chicago

Wiley Classics Library Edition Published 1992

A Wiley-Interscience Publication
JOHN WILEY AND SONS, INC.
New York / Chichester / Brisbane / Toronto / Singapore

Wiley Classics Library Edition Published 1992

Library of Congress Cataloging-in-Publication Data

Box, George E. P.
Bayesian inference in statistical analysis / George E.P. Box, George C. Tiao. — Wiley classics library ed.
p. cm.
Originally published: Reading, Mass. : Addison-Wesley Pub. Co., c1973.
"A Wiley-Interscience publication."
Includes bibliographical references and indexes.
ISBN 0-471-57428-7
1. Mathematical statistics. I. Tiao, George C., 1933- II. Title.
[QA276.B677 1992]
519.5'4—dc20 92-3745
CIP

10 9 8 7

To BARBARA, HELEN, and HARRY

This page intentionally left blank

PREFACE

The object of this book is to explore the use and relevance of Bayes' theorem to problems such as arise in scientific investigation in which inferences must be made concerning parameter values about which little is known *a priori*.

In Chapter 1 we discuss some important general aspects of the Bayesian approach, including: the role of Bayesian inference in scientific investigation, the choice of prior distributions (and, in particular, of noninformative prior distributions), the problem of nuisance parameters, and the role and relevance of sufficient statistics.

In Chapter 2, as a preliminary to what follows, a number of standard problems concerned with the comparison of location and scale parameters are discussed. Bayesian methods, for the most part well known, are derived there which closely parallel the inferential techniques of sampling theory associated with t-tests, F-tests, Bartlett's test, the analysis of variance, and with regression analysis. These techniques have long proved of value to the practicing statistician and it stands to the credit of sampling theory that it has produced them. It is also encouraging to know that parallel procedures may, with at least equal facility, be derived using Bayes' theorem. Now, practical employment of such techniques has uncovered further inferential problems, and attempts to solve these, using sampling theory, have had only partial success. One of the main objectives of this book, pursued from Chapter 3 onwards, is to study some of these problems from a Bayesian viewpoint. In this we have in mind that the value of Bayesian analysis may perhaps be judged by considering to what extent it supplies insight and sensible solutions for what are known to be awkward problems.

The following are examples of the further problems considered:

1. How can inferences be made in small samples about parameters for which no parsimonious set of sufficient statistics exists?
2. To what extent are inferences about means and variances sensitive to departures from assumptions such as error Normality, and how can such sensitivity be reduced?
3. How should inferences be made about variance components?
4. How and in what circumstances should mean squares be pooled in the analysis of variance?

5. How can information be pooled from several sources when its precision is not exactly known, but can be estimated, as, for example, in the "recovery of interblock information" in the analysis of incomplete block designs?
6. How should data be transformed to produce parsimonious parametrization of the model as well as to increase sensitivity of the analysis?

The main body of the text is an investigation of these and similar questions with appropriate analysis of the mathematical results illustrated with numerical examples. We believe that this (1) provides evidence of the value of the Bayesian approach, (2) offers useful methods for dealing with the important problems specifically considered and (3) equips the reader with techniques which he can apply in the solution of new problems.

There is a continuing commentary throughout concerning the relation of the Bayes results to corresponding sampling theory results. We make no apology for this arrangement. In any scientific discussion alternative views ought to be given proper consideration and appropriate comparisons made. Furthermore, many readers will already be familiar with sampling theory results and perhaps with the resulting problems which have motivated our study.

This book is principally a bringing together of research conducted over the years at Wisconsin and elsewhere in cooperation with other colleagues, in particular David Cox, Norman Draper, David Lund, Wai-Yuan Tan, and Arnold Zellner. A list of the consequent source references employed in each chapter is given at the end of this volume.

An elementary knowledge of probability theory and of standard sampling theory analysis is assumed, and from a mathematical viewpoint, a knowledge of calculus and of matrix algebra. The material forms the basis of a two-semester graduate course in Bayesian inference; we have successfully used earlier drafts for this purpose. Except for perhaps Chapters 8 and 9, much of the material can be taught in an advanced undergraduate course.

We are particularly indebted to Fred Mosteller and James Dickey, who patiently read our manuscript and made many valuable suggestions for its improvement, and to Mukhtar Ali, Irwin Guttman, Bob Miller, and Steve Stigler for helpful comments. We also wish to record our thanks to Biyi Afonja, Yu-Chi Chang, William Cleveland, Larry Haugh, Hiro Kanemasu, David Pack, and John MacGregor for help in checking the final manuscript, to Mary Esser for her patience and care in typing it, and to Greta Ljung and Johannes Ledolter for careful proofreading.

The work has involved a great deal of research which has been supported by the Air Force Office of Scientific Research under Grants AF-AFOSR-1158-66, AF-49(638) 1608 and AF-AFOSR 69-1803, the Office of Naval Research under Contract ONR-N-00014-67-A-0128-0017, the Army Office of Ordnance Research under Contract DA-ARO-D-31-124-G917, the National Science Foundation under Grant GS-2602, and the British Science Research Council.

The manuscript was begun while the authors were visitors at the Graduate School of Business, Harvard University, and we gratefully acknowledge support from the Ford Foundation while we were at that institution. We must also express our gratitude for the hospitality extended to us by the University of Essex in England when the book was nearing completion.

We are grateful to Professor E. S. Pearson and the Biometrika Trustees, to the editors of *Journal of the American Statistical Association* and *Journal of the Royal Statistical Society Series B*, and to our coauthors David Cox, Norman Draper, David Lund, Wai-Yuan Tan, and Arnold Zellner for permission to reprint condensed and adapted forms of various tables and figures from articles listed in the principal source references and general references. We are also grateful to Professor O. L. Davies and to G. Wilkinson of the Imperial Chemical Industries, Ltd., for permission to reproduce adapted forms of Tables 4.2 and 6.3 in *Statistical Methods in Research and Production*, 3rd edition revised, edited by O. L. Davies.

We acknowledge especial indebtedness for support throughout the whole project by the Wisconsin Alumni Research Foundation, and particularly for their making available through the University Research Committee the resources of the Wisconsin Computer Center.

Madison, Wisconsin G.E.P.B.
August 1972 G.C.T.

This page intentionally left blank

CONTENTS

CHAPTER 1

NATURE OF BAYESIAN INFERENCE

1.1 INTRODUCTION AND SUMMARY

Opinion as to the value of Bayes' theorem as a basis for statistical inference has swung between acceptance and rejection since its publication in 1763. During periods when it was thought that alternative arguments supplied a satisfactory foundation for statistical inference Bayesian results were viewed, sometimes condescendingly, as an interesting but mistaken attempt to solve an important problem. When subsequently it was found that initially unsuspected difficulties accompanied the alternatives, interest was rekindled. Bayes' mode of reasoning, finally buried on so many occasions, has recently risen again with astonishing vigor.

In addition to the present growing awareness of possible deficiencies in the alternatives, three further factors account for the revival. First, the work of a number of authors, notably Fisher, Jeffreys, Barnard, Ramsey, De Finetti, Savage, Lindley, Anscombe and Stein, has, although not always directed to that end, helped to clarify and overcome some of the philosophical and practical difficulties.

Second, while other inferential theories had yielded nice solutions in cases where rather special assumptions such as Normality and independence of errors could be made, in other cases, and particularly where no sufficient statistics existed, the solutions were often unsatisfactory and messy. Although it is true that these special assumptions covered a number of situations of scientific interest, it would be idle to pretend that the set of statistical problems whose solution has been or will be needed by the scientific investigator coincides with the set of problems thus amenable to convenient treatment. Data gathering is frequently expensive compared with data analysis. It is sensible then that hard-won data be inspected from many different viewpoints. In the selection of viewpoints, Bayesian methods allow greater emphasis to be given to scientific interest and less to mathematical convenience.

Third, the nice solutions based on the rather special assumptions have been popular for another reason—they were easy to compute. This consideration has much less force now that the desk calculator is no longer the most powerful instrument for executing statistical analysis. Suppose, using a desk calculator, it takes five hours to perform a data analysis appropriate to the assumption that errors are Normal and independent, then the five hundred hours it might take to explore less restrictive assumptions could be prohibitive. By contrast, the use of an electronic computer can so reduce the time base that, with general programs available, the wider analysis can be almost as immediate and economic as the more restricted one.

Scientific investigation uses statistical methods in an iteration in which controlled data gathering and data analysis alternate. Data analysis is a subiteration in which inference from a tentatively entertained model alternates with criticism of the conditional inference by inspection of residuals and other means. Statistical inference is thus only one of the responsibilities of the statistician. It is however an important one. Bayesian inference alone seems to offer the possibility of sufficient flexibility to allow reaction to scientific complexity free from impediment from purely technical limitation.

A prior distribution, which is supposed to represent what is known about unknown parameters before the data is available, plays an important role in Bayesian analysis. Such a distribution can be used to represent prior knowledge or relative ignorance. In problems of scientific inference we would usually, were it possible, like the data "to speak for themselves." Consequently, it is usually appropriate to conduct the analysis as if a state of relative ignorance existed *a priori*. In this book, therefore, extensive use is made of "noninformative" prior distributions and very little of informative priors. The aim is to obtain an inference which would be appropriate for an unprejudiced observer. The understandable uneasiness felt by some statisticians about the use of prior distributions is often associated with the fear that the prior may dominate and distort "what the data are trying to say." We hope to show by the examples in this book that, by careful choice of model structure and appropriate noninformative priors, Bayesian analysis can produce the reverse of what is feared. It can permit the data to comment on dubious aspects of a model in a manner not otherwise possible.

The usefulness of a theory is customarily assessed by tentatively adopting it, and then considering whether its consequences agree with common sense, and whether they provide insight where common sense fails. It was in this spirit that some years ago the authors with others began research in applications of the Bayesian theory of inference. A series of problems were selected in the solution of which difficulties or inconsistencies had been encountered with other approaches. Because Bayesian analysis of these problems has seemed consistently helpful and interesting, we believe it is now appropriate to bring this and other related work together, and to consider its wider aspects.

The objective of this book, therefore, is to explore Bayesian inference in statistical analysis. The book consists of ten chapters. Chapter 1 discusses the role of statistical inference in scientific investigation. In the light of that discussion the nature of Bayesian inference, including the choice of noninformative prior distributions, is considered. The chapter ends with an account of the role and relevance of sufficient statistics, and discusses the problem of nuisance parameters.

In Chapter 2 a number of standard Normal theory inference problems concerning location and scale parameters are considered. Bayes' solutions are given which closely parallel sampling theory techniques† associated with t-tests, F-tests, the analysis of variance and regression analysis. While these procedures have long proved valuable to practising statisticians, efforts to extend them in important directions using non-Bayesian theories have met serious difficulties. An advantage of the Bayes approach is that it can be used to explore the consequences of any type of probability model, without restriction to those having special mathematical forms. Thus, in Chapter 3 the problem of making inferences about location parameters is considered for a wider class of parent probability models of which the Normal distribution is a member. In this framework, we show how it is possible to assess to what extent inferences about location parameters are sensitive to departures from Normality. Further, it is shown how we can use the evidence from the data to make inferences about the form of the parent distributions of the observations. The analysis is extended in Chapter 4 to the problem of comparing variances.

Chapters 5 and 6 discuss various random effect and mixed models associated with hierarchical and cross classification designs. With sampling theory, one experiences a number of difficulties in estimating means and variance components in these models. Notably one encounters problems of negative variance estimates, of eliminating nuisance parameters, of constructing confidence intervals, and of pooling variance estimates. Analysis, from a Bayesian standpoint, is much more tractable, and in particular provides an interesting and sensible solution to the pooling dilemma.

Chapter 7 deals with two further important problems in the analysis of variance. The first concerns the estimation of means in the one-way classification. When it is sensible to regard such means as themselves a sample from a population, the appropriate Bayesian analysis shows that there are then *two* sources of information about the means and appropriately combines them. The chapter ends with a discussion of the recovery of interblock information in the balanced incomplete block design model. This is again a problem in which two sources of information need to be appropriately combined and for which the sampling theory solution is unsatisfactory.

In Chapters 8 and 9 a general treatment of linear and nonlinear Normal multivariate models is given. While Bayesian results associated with standard linear models are discussed, particular attention is given to the problem of estimating common location parameters from several equations. The latter problem is of considerable practical importance, but is difficult to tackle by sampling theory methods, and has not previously received much attention.

---

† We shall assume in this book that the reader has some familiarity with standard ideas of the sampling theory approach explained for example in Mood and Graybill (1963) and Hogg and Craig (1970).

Finally, in Chapter 10, we consider the important problem of data transformation from a Bayesian viewpoint. The problem is to select a transformation which, so far as possible, achieves Normality, homogeniety of variance, and simplicity of the expectation function in the transformed variate.

A bald statement of a mathematical expression, however correct, frequently fails to produce understanding. Many Bayesian results are of particular interest because they seem to provide a kind of higher intuition. Mathematical results which at first seemed puzzling have later been seen to provide a maturer kind of common sense. For this reason, throughout this book, individual mathematical formulae are carefully analyzed and illustrated with examples and diagrams. Also, appropriate approximations are developed when they provide deeper understanding of a situation, or where they simplify calculation. For the convenience of the reader a number of short summaries of formulas and calculations are given in appropriate places.

1.1.1 The Role of Statistical Methods in Scientific Investigation

Statistical methods are tools of scientific investigation. Scientific investigation is a controlled learning process in which various aspects of a problem are illuminated as the study proceeds. It can be thought of as a major iteration within which secondary iterations occur. The major iteration is that in which a tentative conjecture suggests an experiment, appropriate analysis of the data so generated leads to a modified conjecture, and this in turn leads to a new experiment, and so on. An idealization of this process is seen in Fig. 1.1.1, involving an alternation between *conjecture* and *experiment* carried out via experimental *design* and data *analysis*.† As indicated by the zig-zag line at the bottom of the figure, most investigations involve not one but a number of alternations of this kind.

An efficient investigation is one where convergence to the objective occurs as quickly and unambiguously as possible. A basic determinant of efficiency, which we must suppose is outside the control of the statistician, is the originality, imagination, and subject matter knowledge of the investigator. Apart from this vital determining factor, however, efficiency is decided by the appropriateness and force of the methods of design and analysis employed. In moving from conjecture to experimental data, (D), experiments must be designed which make best use of the experimenter's current state of knowledge and which best illuminate his conjecture. In moving from data to modified conjecture, (A), data must be analyzed so as to accurately present information in a manner which is readily understood by the experimenter.

† The words design and experiment are broadly interpreted here to refer to any data gathering process. In an economic study, a conjecture might lead the investigator to study the functional relationship between money supply and interest rate. The difficult decision as to what types of money supply and interest rate data to use, here constitutes the design. In social studies a particular sample survey might be the experiment.

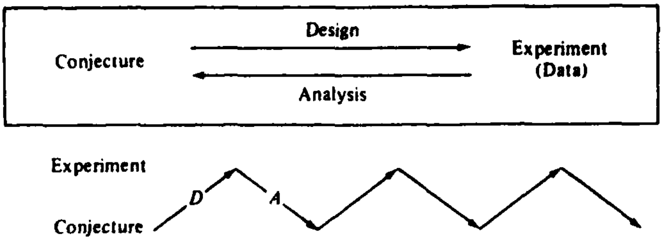

Fig. 1.1.1 Iterative process of scientific investigation (the alternation between conjecture and experiment).

A full treatise on the use of statistical methods in scientific investigation therefore would necessarily include consideration of statistical design as well as statistical analysis. The aims of this book are, however, much more limited. We shall not discuss experimental design, and will be concerned only with one aspect of statistical analysis, namely, *statistical inference*.

1.1.2 Statistical Inference as one Part of Statistical Analysis

For illustration, suppose we were studying the useful life of batteries produced by a particular machine. It might be appropriate to assume tentatively that the observed lives of batteries coming from the machine were distributed independently and Normally about some mean θ with variance σ^2. The probability distribution of a projected sample of n observations $\mathbf{y}' = (y_1, \ldots, y_n)$ would then be

$$p(\mathbf{y} \mid \theta, \sigma^2) \propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t - \theta)^2\right], \qquad -\infty < y_t < \infty. \quad (1.1.1)$$

Given the value of the parameters θ and σ^2, this expression permits the calculation of the probability density $p(\mathbf{y} \mid \theta, \sigma^2)$ associated with any *hypothetical* data set $\mathbf{y}$ *before* any data is taken. For statistical analysis this is, in most cases, the converse of what is needed. The analyst already has the data but he does not know θ and σ^2. He can, however, use $p(\mathbf{y} \mid \theta, \sigma^2)$ indirectly to make *inferences* about the values of θ and σ^2, given the n data values.

Two of the methods by which this may be attempted employ

a. Sampling Theory,

b. Bayes' Theorem.

We now give a brief description of each of these approaches using the Normal probability model (1.1.1) for illustration.

Sampling Theory Approach

In this approach inferences are made by directing attention to a reference set of hypothetical data vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_j, \ldots$ which could have been generated by the probability model $p(\mathbf{y} \mid \theta_0, \sigma_0^2)$ of (1.1.1), where θ_0 and σ_0^2 are the hypothetical true values of θ and σ^2. Estimators $\hat{\theta}(\mathbf{y})$ and $\hat{\sigma}^2(\mathbf{y})$, which are functions of the data vector $\mathbf{y}$, are selected. By imagining values $\hat{\theta}(\mathbf{y}_j)$ and $\hat{\sigma}^2(\mathbf{y}_j)$ to be calculated for each hypothetical data vector $\mathbf{y}_j$, reference sets are generated for $\hat{\theta}(\mathbf{y})$ and $\hat{\sigma}^2(\mathbf{y})$. Inferences are then made by comparing the values of $\hat{\theta}(\mathbf{y})$ and $\hat{\sigma}_2(\mathbf{y})$ actually observed with their "sampling distributions" generated by the reference sets.

The functions $\hat{\theta}(\mathbf{y})$ and $\hat{\sigma}^2(\mathbf{y})$ are usually chosen so that the sampling distributions of the extimators $\hat{\theta}(\mathbf{y}_j)$ and $\hat{\sigma}^2(\mathbf{y}_j)$ are, in some sense, concentrated as closely as possible about the true values θ_0 and σ_0. To provide some idea of how far away from the true values the calculated quantities $\hat{\theta}(\mathbf{y})$ and $\hat{\sigma}^2(\mathbf{y})$ might be, *confidence intervals* are calculated. For example, the $1-\alpha$ confidence interval for θ would be of the form

$$\bar{\theta}_1(\mathbf{y}) < \theta < \bar{\theta}_2(\mathbf{y}),$$

where $\bar{\theta}_1(\mathbf{y})$ and $\bar{\theta}_2(\mathbf{y})$ would be functions of $\mathbf{y}$, chosen so that in repeated sampling the computed confidence intervals included the value θ_0, a proportion $1-\alpha$ of the time.

Bayesian Approach

In a Bayesian approach, a different line is taken. As part of the model a *prior* distribution $p(\theta, \sigma^2)$ is introduced. This is supposed to express a state of knowledge or ignorance about θ and σ^2 before the data are obtained. Given the prior distribution, the probability model $p(\mathbf{y} \mid \theta, \sigma^2)$ and the data $\mathbf{y}$, it is now possible to calculate the probability distribution $p(\theta, \sigma^2 \mid \mathbf{y})$ of θ and σ^2, given the data $\mathbf{y}$. This is called the *posterior* distribution of θ and σ^2. From this distribution inferences about the parameters are made.

1.1.3 The Question of Adequacy of Assumptions

Consider the battery-life example and suppose $n = 20$ observations are available. Then, whichever method of inference is used, *conditional on the assumptions* we can summarize all the information in the 20 data values in terms of inferences about just *two* parameters, θ and σ^2.

The inferences are, in particular, conditional on the adequacy of the probability model in (1.1.1). It is not difficult, however, to imagine situations in which this model, and therefore the associated inferences, could be inadequate. It might happen, for example, that during the period of observation, a quality characteristic x of a chemical additive, used in making the batteries, could vary and could cause, via an approximate linear relationship, a corresponding change in the mean life time of the batteries. In this case, a more appropriate model might be

$$p(\mathbf{y} \mid \mathbf{x}, \sigma^2, \theta_1, \theta_2) \propto \sigma^{-20} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{20}(y_t - \theta_1 - \theta_2 x_t)^2\right],$$

$$-\infty < y_t < \infty. \qquad (1.1.2)$$

ABCDEFGH798765432

This page intentionally left blank