



---

# 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*

In recognition of the importance of preserving what has been written, it is a policy of John Wiley & Sons, Inc., to have books of enduring value published in the United States printed on acid-free paper, and we exert our best efforts to that end.

Copyright © 1973 by George E.P. Box.

Wiley Classics Library Edition Published 1992

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

***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

I. 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*  
*August 1972*

G.E.P.B.  
G.C.T.

This page intentionally left blank

# CONTENTS

<b>Chapter 1</b>	<b>Nature of Bayesian Inference</b>	
1.1	Introduction and summary . . . . .	1
1.1.1	The role of statistical methods in scientific investigation . . . . .	4
1.1.2	Statistical inference as one part of statistical analysis . . . . .	5
1.1.3	The question of adequacy of assumptions . . . . .	6
1.1.4	An iterative process of model building in statistical analysis . . . . .	7
1.1.5	The role of Bayesian analysis . . . . .	9
1.2	Nature of Bayesian inference . . . . .	10
1.2.1	Bayes' theorem . . . . .	10
1.2.2	Application of Bayes' theorem with probability interpreted as frequencies . . . . .	12
1.2.3	Application of Bayes' theorem with subjective probabilities . . . . .	14
1.2.4	Bayesian decision problems . . . . .	19
1.2.5	Application of Bayesian analysis to scientific inference . . . . .	20
1.3	Noninformative prior distributions . . . . .	25
1.3.1	The Normal mean $\theta$ ( $\sigma^2$ known) . . . . .	26
1.3.2	The Normal standard deviation $\sigma$ ( $\theta$ known) . . . . .	29
1.3.3	Exact data translated likelihoods and noninformative priors . . . . .	32
1.3.4	Approximate data translated likelihood . . . . .	34
1.3.5	Jeffreys' rule, information measure, and noninformative priors . . . . .	41
1.3.6	Noninformative priors for multiple parameters . . . . .	46
1.3.7	Noninformative prior distributions: A summary . . . . .	58
1.4	Sufficient statistics . . . . .	60
1.4.1	Relevance of sufficient statistics in Bayesian inference . . . . .	63
1.4.2	An example using the Cauchy distribution . . . . .	64
1.5	Constraints on parameters . . . . .	67
1.6	Nuisance parameters . . . . .	70
1.6.1	Application to robustness studies . . . . .	70
1.6.2	Caution in integrating out nuisance parameters . . . . .	71
1.7	Systems of inference . . . . .	72
1.7.1	Fiducial inference and likelihood inference . . . . .	73
	<b>Appendix A1.1</b> Combination of a Normal prior and a Normal likelihood . . . . .	74
<b>Chapter 2</b>	<b>Standard Normal Theory Inference Problems</b>	
2.1	Introduction . . . . .	76
2.1.1	The Normal distribution . . . . .	77
2.1.2	Common Normal-theory problems . . . . .	79
2.1.3	Distributional assumptions . . . . .	80

2.2	Inferences concerning a single mean from observations assuming common known variance . . . . .	82
2.2.1	An example . . . . .	82
2.2.2	Bayesian intervals . . . . .	84
2.2.3	Parallel results from sampling theory . . . . .	85
2.3	Inferences concerning the spread of a Normal distribution from observations having common known mean . . . . .	86
2.3.1	The inverted $\chi^2$ , inverted $\chi$ , and the log $\chi$ distributions . . . . .	87
2.3.2	Inferences about the spread of a Normal distribution . . . . .	89
2.3.3	An example . . . . .	90
2.3.4	Relationship to sampling theory results . . . . .	92
2.4	Inferences when both mean and standard deviation are unknown . . . . .	92
2.4.1	An example . . . . .	94
2.4.2	Component distributions of $p(\theta, \sigma   y)$ . . . . .	95
2.4.3	Posterior intervals for $\theta$ . . . . .	97
2.4.4	Geometric interpretation of the derivation of $p(\theta   y)$ . . . . .	98
2.4.5	Informative prior distribution of $\sigma$ . . . . .	99
2.4.6	Effect of changing the metric of $\sigma$ for locally uniform prior . . . . .	101
2.4.7	Elimination of the nuisance parameter $\sigma$ in Bayesian and sampling theories . . . . .	102
2.5	Inferences concerning the difference between two means . . . . .	103
2.5.1	Distribution of $\theta_2 - \theta_1$ when $\sigma_1^2 = \sigma_2^2$ . . . . .	103
2.5.2	Distribution of $\theta_2 - \theta_1$ when $\sigma_1^2$ and $\sigma_2^2$ are not assumed equal . . . . .	104
2.5.3	Approximations to the Behrens-Fisher distribution . . . . .	107
2.5.4	An example . . . . .	107
2.6	Inferences concerning a variance ratio . . . . .	109
2.6.1	H.P.D. intervals . . . . .	110
2.6.2	An example . . . . .	111
2.7	Analysis of the linear model . . . . .	113
2.7.1	Variance $\sigma^2$ assumed known . . . . .	115
2.7.2	Variance $\sigma^2$ unknown . . . . .	116
2.7.3	An example . . . . .	118
2.8	A general discussion of highest posterior density regions . . . . .	122
2.8.1	Some properties of the H.P.D. region . . . . .	123
2.8.2	Graphical representation . . . . .	124
2.8.3	Is $\theta_0$ inside or outside a given H.P.D. region? . . . . .	125
2.9	H.P.D. Regions for the linear model: a Bayesian justification of analysis of variance . . . . .	125
2.9.1	The weighing example . . . . .	126
2.10	Comparison of parameters . . . . .	127
2.11	Comparison of the means of $k$ Normal populations . . . . .	127
2.11.1	Choice of linear contrasts . . . . .	128
2.11.2	Choice of linear contrasts to compare location parameters . . . . .	131
2.12	Comparison of the spread of $k$ distributions . . . . .	132
2.12.1	Comparison of the spread of $k$ Normal populations . . . . .	132
2.12.2	Asymptotic distribution of $M$ . . . . .	134
2.12.3	Bayesian parallel to Bartlett's test . . . . .	136
2.12.4	An example . . . . .	136

2.13	Summarized calculations of various posterior distributions . . . . .	138
	<b>Appendix A2.1</b> Some useful integrals . . . . .	144
	<b>Appendix A2.2</b> Stirling's series . . . . .	146

**Chapter 3 Bayesian Assessment of Assumptions**

**1. Effect of Non-Normality on Inferences about a Population Mean with Generalizations**

3.1	Introduction . . . . .	149
	3.1.1 Measures of distributional shape, describing certain types of non-Normality . . . . .	149
	3.1.2 Situations where Normality would not be expected . . . . .	151
3.2	Criterion robustness and inference robustness illustrated using Darwin's data . . . . .	152
	3.2.1 A wider choice of the parent distribution . . . . .	156
	3.2.2 Derivation of the posterior distribution of $\theta$ for a specific symmetric parent . . . . .	160
	3.2.3 Properties of the posterior distribution of $\theta$ for a fixed $\beta$ . . . . .	162
	3.2.4 Posterior distribution of $\theta$ and $\beta$ when $\beta$ is regarded as a random variable . . . . .	164
	3.2.5 Marginal distribution of $\beta$ . . . . .	166
	3.2.6 Marginal distribution of $\theta$ . . . . .	167
	3.2.7 Information concerning the nature of the parent distribution coming from the sample . . . . .	169
	3.2.8 Relationship to the general Bayesian framework for robustness studies . . . . .	169
3.3	Approximations to the posterior distribution $p(\theta   \beta, \mathbf{y})$ . . . . .	170
	3.3.1 Motivation for the approximation . . . . .	170
	3.3.2 Quadratic approximation to $M(\theta)$ . . . . .	172
	3.3.3 Approximation of $p(\theta   \mathbf{y})$ . . . . .	174
3.4	Generalization to the linear model . . . . .	176
	3.4.1 An illustrative example . . . . .	178
3.5	Further extension to nonlinear models . . . . .	186
	3.5.1 An illustrative example . . . . .	187
3.6	Summary and discussion . . . . .	193
3.7	A summary of formulas for posterior distributions . . . . .	196
	<b>Appendix A3.1</b> Some properties of the posterior distribution $p(\theta   \beta, \mathbf{y})$ . . . . .	199
	<b>Appendix A3.2</b> A property of locally uniform distributions . . . . .	201

**Chapter 4 Bayesian Assessment of Assumptions**

**2. Comparison of Variances**

4.1	Introduction . . . . .	203
4.2	Comparison of two variances . . . . .	203
	4.2.1 Posterior distribution of $V = \sigma_1^2/\sigma_2^2$ for fixed values of $(\theta_1, \theta_2, \beta)$ . . . . .	204
	4.2.2 Relationship between the posterior distribution $p(V   \beta, \boldsymbol{\theta}, \mathbf{y})$ and sampling theory procedures . . . . .	206
	4.2.3 Inference robustness on Bayesian theory and sampling theory . . . . .	208

4.2.4	Posterior distribution of $\sigma_2^2/\sigma_1^2$ when $\beta$ is regarded as a random variable	209
4.2.5	Inferences about $V$ with $\beta$ eliminated	210
4.2.6	Posterior distribution of $V$ when $\theta_1$ and $\theta_2$ are regarded as random variables	211
4.2.7	Computational procedures for the posterior distribution $p(V   \beta, y)$	213
4.2.8	Posterior distribution of $V$ for fixed $\beta$ with $\theta_1$ and $\theta_2$ eliminated	215
4.2.9	Marginal distribution of $\beta$	217
4.2.10	Marginal distribution of $V$	218
4.3	Comparison of the variances of $k$ distributions	219
4.3.1	Comparison of $k$ variances for fixed values of $(\theta, \beta)$	219
4.3.2	Posterior distribution of $\beta$ and $\phi$	224
4.3.3	The situation when $\theta_1, \dots, \theta_k$ are not known	228
4.4	Inference robustness and criterion robustness	232
4.4.1	The analyst example	233
4.4.2	Derivation of the criteria	235
4.5	A summary of formulae for various prior and posterior distributions	236
Appendix A4.1	Limiting distributions for the variance ratio $V$ when $\beta$ approaches $-1$	241
<b>Chapter 5 Random Effect Models</b>		
5.1	Introduction	244
5.1.1	The analysis of variance table	245
5.1.2	Two examples	246
5.1.3	Difficulties in the sampling theory approach	248
5.2	Bayesian analysis of hierarchical classifications with two variance components	249
5.2.1	The likelihood function	250
5.2.2	Prior and posterior distribution of $(\theta, \sigma_1^2, \sigma_2^2)$	251
5.2.3	Posterior distribution of $\sigma_2^2/\sigma_1^2$ and its relationship to sampling theory results	253
5.2.4	Joint posterior distribution of $\sigma_1^2$ and $\sigma_2^2$	255
5.2.5	Distribution of $\sigma_1^2$	258
5.2.6	A scaled $\chi^{-2}$ approximation to the distribution of $\sigma_1^2$	261
5.2.7	A Bayesian solution to the pooling dilemma	264
5.2.8	Posterior distribution of $\sigma_2^2$	266
5.2.9	Computation of $p(\sigma_2^2   y)$	268
5.2.10	A summary of approximations to the posterior distributions of $(\sigma_1^2, \sigma_2^2)$	273
5.2.11	Derivation of the posterior distribution of variance components using constraints directly	273
5.2.12	Posterior distribution of $\sigma_1^2$ and a scaled $\chi^{-2}$ approximation	274
5.2.13	Use of features of posterior distribution to supply sampling theory estimates	276
5.3	A three-component hierarchical design model	276
5.3.1	The likelihood function	277
5.3.2	Joint posterior distribution of $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	278
5.3.3	Posterior distribution of the ratio $\sigma_2^2/\sigma_1^2$	280

5.3.4 Relative contribution of variance components . . . . . 283

5.3.5 Posterior distribution of  $\sigma_1^2$  . . . . . 286

5.3.6 Posterior distribution of  $\sigma_2^2$  . . . . . 290

5.3.7 Posterior distribution of  $\sigma_3^2$  . . . . . 290

5.3.8 A summary of the approximations to the posterior distributions of  $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$  . . . . . 291

5.4 Generalization to  $q$ -component hierarchical design model . . . . . 292

**Appendix A5.1** The posterior density of  $\sigma_1^2$  for the two component model expressed as a  $\chi^{-2}$  series . . . . . 294

**Appendix A5.2** Some useful integral identities . . . . . 295

**Appendix A5.3** Some properties of the posterior distribution of  $\sigma_2^2$  for the two-component model . . . . . 296

**Appendix A5.4** An asymptotic expansion of the distribution of  $\sigma_2^2$  for the two-component model . . . . . 298

**Appendix A5.5** A criticism of the prior distribution of  $(\sigma_1^2, \sigma_2^2)$  . . . . . 303

**Appendix A5.6** "Bayes" estimators . . . . . 304

**Appendix A5.7** Mean squared error of the posterior means of  $\sigma^{2p}$  . . . . . 315

**Chapter 6 Analysis of Cross Classification Designs**

6.1 Introduction . . . . . 317

    6.1.1 A car-driver experiment: Three different classifications . . . . . 317

    6.1.2 Two-way random effect model . . . . . 320

    6.1.3 Two-way mixed model . . . . . 323

    6.1.4 Special cases of mixed models . . . . . 325

    6.1.5 Two-way fixed effect model . . . . . 328

6.2 Cross classification random effect model . . . . . 329

    6.2.1 The likelihood function . . . . . 329

    6.2.2 Posterior distribution of  $(\sigma_\tau^2, \sigma_\rho^2, \sigma_\tau^2, \sigma_\rho^2)$  . . . . . 332

    6.2.3 Distribution of  $(\sigma_\tau^2, \sigma_\rho^2, \sigma_\tau^2)$  . . . . . 333

    6.2.4 Distribution of  $(\sigma_\tau^2, \sigma_\rho^2)$  . . . . . 334

    6.2.5 An illustrative example . . . . . 335

    6.2.6 A simple approximation to the distribution of  $(\sigma_\tau^2, \sigma_\rho^2)$  . . . . . 339

6.3 The additive mixed model . . . . . 341

    6.3.1 The likelihood function . . . . . 341

    6.3.2 Posterior distributions of  $(\theta, \phi, \sigma_\tau^2, \sigma_{\tau\epsilon}^2)$  ignoring the constraint  $C$  . . . . . 344

    6.3.3 Posterior distribution of  $\sigma_\tau^2$  and  $\sigma_\tau^2$  . . . . . 345

    6.3.4 Inferences about fixed effects for the mixed model . . . . . 346

    6.3.5 Comparison of two means . . . . . 346

    6.3.6 Approximations to the distribution  $p(\tau | y)$  . . . . . 351

    6.3.7 Relationship to some sampling theory results and the problem of pooling . . . . . 353

    6.3.8 Comparison of  $I$  means . . . . . 357

    6.3.9 Approximating the posterior distribution of  $V = m(\phi)/m_\epsilon$  . . . . . 360

    6.3.10 Summarized calculations . . . . . 361

6.4 The interaction model . . . . . 362

    6.4.1 The likelihood function . . . . . 362

    6.4.2 Posterior distribution of  $(\theta, \phi, \sigma_\tau^2, \sigma_{\tau\epsilon}^2, \sigma_{\tau\epsilon}^2)$  . . . . . 364

6.4.3	Comparison of $t$ means . . . . .	365
6.4.4	Approximating the distribution $p(\phi   y)$ . . . . .	365
6.4.5	An example . . . . .	367
<b>Chapter 7</b>	<b>Inference about Means with Information from more than One Source: One-Way Classification and Block Designs</b>	
7.1	Introduction . . . . .	369
7.2	Inferences about means for the one-way random effect model . . . . .	370
7.2.1	Various models used in the analysis of one-way classification . . . . .	370
7.2.2	Use of the random effect model in the study of means . . . . .	371
7.2.3	Posterior distribution of $\theta$ via intermediate distributions . . . . .	372
7.2.4	Random effect and fixed effect models for inferences about means . . . . .	377
7.2.5	Random effect prior versus fixed effect prior . . . . .	379
7.2.6	Effect of different prior assumptions on the posterior distribution of $\theta$ . . . . .	383
7.2.7	Relationship to sampling theory results . . . . .	388
7.2.8	Conclusions . . . . .	390
7.3	Inferences about means for models with two random components . . . . .	391
7.3.1	A general linear model with two random components . . . . .	393
7.3.2	The likelihood function . . . . .	393
7.3.3	Sampling theory estimation problems associated with the model . . . . .	396
7.3.4	Prior and posterior distributions of $(\sigma_e^2, \sigma_{be}^2, \theta)$ . . . . .	396
7.4	Analysis of balanced incomplete block designs . . . . .	397
7.4.1	Properties of the BIBD model . . . . .	397
7.4.2	A comparison of the one-way classification, RCBD and BIBD models . . . . .	400
7.4.3	Posterior distribution of $\phi = (\phi_1, \dots, \phi_{t-1})'$ for a BIBD . . . . .	403
7.4.4	An illustrative example . . . . .	405
7.4.5	Posterior distribution of $\sigma_{be}^2/\sigma_e^2$ . . . . .	407
7.4.6	Further properties of, and approximations to, $p(\phi   y)$ . . . . .	409
7.4.7	Summarized calculations for approximating the posterior distributions of $\phi$ . . . . .	415
7.4.8	Recovery of interblock information in sampling theory . . . . .	417
	<b>Appendix A7.1</b> Some useful results in combining quadratic forms. . . . .	418
<b>Chapter 8</b>	<b>Some Aspects of Multivariate Analysis</b>	
8.1	Introduction . . . . .	421
8.1.1	A general univariate model . . . . .	421
8.2	A general multivariate Normal model . . . . .	423
8.2.1	The likelihood function . . . . .	424
8.2.2	Prior distribution of $(\theta, \Sigma)$ . . . . .	425
8.2.3	Posterior distribution of $(\theta, \Sigma)$ . . . . .	427
8.2.4	The Wishart distribution . . . . .	427
8.2.5	Posterior distribution of $\theta$ . . . . .	428
8.2.6	Estimation of common parameters in a nonlinear multivariate model . . . . .	428
8.3	Linear multivariate models . . . . .	435
8.3.1	The use of linear theory approximations when the expectation is nonlinear in the parameters . . . . .	436
8.3.2	Special cases of the general linear multivariate model . . . . .	437

8.4	Inferences about $\theta$ for the case of a common derivative matrix $X$ . . . . .	438
8.4.1	Distribution of $\theta$ . . . . .	439
8.4.2	Posterior distribution of the means from a $m$ -dimensional Normal distribution . . . . .	440
8.4.3	Some properties of the posterior matrix-variate $t$ distribution of $\theta$ . . . . .	441
8.4.4	H.P.D regions of $\theta$ . . . . .	448
8.4.5	Distribution of $U(\theta)$ . . . . .	448
8.4.6	An approximation to the distribution of $U$ for general $m$ . . . . .	451
8.4.7	Inferences about a general parameter point of a block submatrix of $\theta$ . . . . .	452
8.4.8	An illustrative example . . . . .	453
8.5	Some aspects of the distribution of $\Sigma$ for the case of a common derivative matrix $X$ . . . . .	459
8.5.1	Joint distribution of $(\theta, \Sigma)$ . . . . .	460
8.5.2	Some properties of the distribution of $\Sigma$ . . . . .	461
8.5.3	An example . . . . .	464
8.5.4	Distribution of the correlation coefficient $\rho_{12}$ . . . . .	465
8.6	A summary of formulae and calculations for making inferences about $(\theta, \Sigma)$ . . . . .	470
Appendix A8.1	The Jacobians of some matrix transformations . . . . .	473
Appendix A8.2	The determinant of the information matrix of $\Sigma^{-1}$ . . . . .	475
Appendix A8.3	The normalizing constant of the $t_{mk}[\theta, (X'X)^{-1}, A, v]$ distribution . . . . .	475
Appendix A8.4	The Kronecker product of two matrices . . . . .	477

**Chapter 9 Estimation of Common Regression Coefficients**

9.1	Introduction: practical importance of combining information from different sources . . . . .	478
9.2	The case of $m$ -independent responses . . . . .	480
9.2.1	The weighted mean problem . . . . .	481
9.2.2	Some properties of $p(\theta   y)$ . . . . .	483
9.2.3	Compatibility of the means . . . . .	485
9.2.4	The fiducial distributions of $\theta$ derived by Yates and by Fisher . . . . .	487
9.3	Some properties of the distribution of the common parameters for two independent responses . . . . .	489
9.3.1	An alternative form for $p(\theta   y)$ . . . . .	489
9.3.2	Further simplifications for the distribution of $\theta$ . . . . .	491
9.3.3	Approximations to the distribution of $\theta$ . . . . .	494
9.3.4	An econometric example . . . . .	495
9.4	Inferences about common parameters for the general linear model with a common derivative matrix . . . . .	500
9.5	General linear model with common parameters . . . . .	501
9.5.1	The case $m = 2$ . . . . .	502
9.5.2	Approximations to the distribution of $\theta$ for $m = 2$ . . . . .	505
9.5.3	An example . . . . .	507

9.6	A summary of various posterior distributions for common parameters	511
<b>Appendix A9.1</b>	<b>Asymptotic expansion of the posterior distribution of <math>\theta</math> for two independent responses</b>	<b>515</b>
<b>Appendix A9.2</b>	<b>Mixed cumulants of the two quadratic forms <math>Q_1</math> and <math>Q_2</math></b>	<b>520</b>
<b>Chapter 10</b>	<b>Transformation of Data</b>	
10.1	Introduction	522
10.1.1	A factorial experiment on grinding	523
10.2	Analysis of the biological and the textile data	525
10.2.1	Some biological data	525
10.2.2	The textile data	527
10.3	Estimation of the transformation	530
10.3.1	Prior and posterior distributions of $\lambda$	531
10.3.2	The simple power transformation	533
10.3.3	The biological example	534
10.3.4	The textile example	536
10.3.5	Two parameter transformation	538
10.4	Analysis of the effects $\theta$ after transformation	539
10.5	Further analysis of the biological data	542
10.5.1	Successive constraints	545
10.5.2	Summary of analysis	548
10.6	Further analysis of the textile data	548
10.7	A summary of formulae for various prior and posterior distributions	550
	<b>Tables</b>	<b>554</b>
	<b>References</b>	<b>571</b>
	<b>Author Index</b>	<b>583</b>
	<b>Subject Index</b>	<b>585</b>

## CHAPTER I

# 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, homogeneity 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.

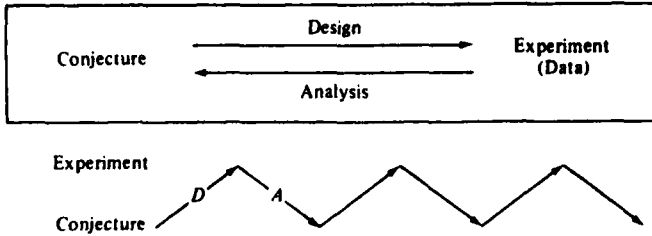


Fig. 1.1.1 Iterative process of scientific investigation (the alternation between conjecture and experiment).

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  $\theta$  with variance  $\sigma^2$ . The probability distribution of a projected sample of  $n$  observations  $y' = (y_1, \dots, y_n)$  would then be

$$p(y | \theta, \sigma^2) \propto \sigma^{-n} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2 \right], \quad -\infty < y_i < \infty. \quad (1.1.1)$$

Given the value of the parameters  $\theta$  and  $\sigma^2$ , this expression permits the calculation of the probability density  $p(y | \theta, \sigma^2)$  associated with any *hypothetical* data set  $y$  *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  $\theta$  and  $\sigma^2$ . He can, however, use  $p(y | \theta, \sigma^2)$  indirectly to make *inferences* about the values of  $\theta$  and  $\sigma^2$ , 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  $y_1, y_2, \dots, y_j, \dots$  which could have been generated by the probability model  $p(y | \theta_0, \sigma_0^2)$  of (1.1.1), where  $\theta_0$  and  $\sigma_0^2$  are the

hypothetical true values of  $\theta$  and  $\sigma^2$ . Estimators  $\hat{\theta}(y)$  and  $\hat{\sigma}^2(y)$ , which are functions of the data vector  $y$ , are selected. By imagining values  $\hat{\theta}(y_j)$  and  $\hat{\sigma}^2(y_j)$  to be calculated for each hypothetical data vector  $y_j$ , reference sets are generated for  $\hat{\theta}(y)$  and  $\hat{\sigma}^2(y)$ . Inferences are then made by comparing the values of  $\hat{\theta}(y)$  and  $\hat{\sigma}^2(y)$  actually observed with their "sampling distributions" generated by the reference sets.

The functions  $\hat{\theta}(y)$  and  $\hat{\sigma}^2(y)$  are usually chosen so that the sampling distributions of the estimators  $\hat{\theta}(y_j)$  and  $\hat{\sigma}^2(y_j)$  are, in some sense, concentrated as closely as possible about the true values  $\theta_0$  and  $\sigma_0$ . To provide some idea of how far away from the true values the calculated quantities  $\hat{\theta}(y)$  and  $\hat{\sigma}^2(y)$  might be, *confidence intervals* are calculated. For example, the  $1 - \alpha$  confidence interval for  $\theta$  would be of the form

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

where  $\bar{\theta}_1(y)$  and  $\bar{\theta}_2(y)$  would be functions of  $y$ , chosen so that in repeated sampling the computed confidence intervals included the value  $\theta_0$ , 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  $\theta$  and  $\sigma^2$  before the data are obtained. Given the prior distribution, the probability model  $p(y | \theta, \sigma^2)$  and the data  $y$ , it is now possible to calculate the probability distribution  $p(\theta, \sigma^2 | y)$  of  $\theta$  and  $\sigma^2$ , given the data  $y$ . This is called the *posterior* distribution of  $\theta$  and  $\sigma^2$ . 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,  $\theta$  and  $\sigma^2$ .

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

$$p(y | x, \sigma^2, \theta_1, \theta_2) \propto \sigma^{-20} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^{20} (y_i - \theta_1 - \theta_2 x_i)^2 \right],$$

$$-\infty < y_i < \infty. \quad (1.1.2)$$

- Covariance matrix (Normal), inference, 460, 461  
 latent roots and vectors, 459  
 Cross classification design, 319
- Decision theory, 19  
 Derivative matrix, 113  
 Design matrix, 421  
 Digamma function, 316  
 Double exponential distribution, 157  
 Double  $F$  distribution, 155
- Estimators  
 Bayes, 304  
 mean-squared error, 307  
 James–Stein problem, 388  
 variance components, 311, 313  
 Expectation function, 422  
 multivariate linear, 435  
 multivariate non-linear, 424  
 Exponential power distribution, 157  
 non-Normality (kurtosis) measure, 157  
 upper per cent points, 158
- $F$  distribution, 110  
 Fiducial inference, 73  
 Fixed effect models  
 cross classification, 318, 328  
 one-way classification, 370
- Gamma function, generalized, 427
- Hierarchical design, 244  
 H.P.D. (highest posterior density) region  
 general, 122  
 interval, 85  
 interval, standardized, 89
- Indicator variable, 113  
 Information matrix  
 definition, 53  
 Normal covariance matrix, 475  
 variance components, 252  
 Information measure, single parameter, 42  
 Integral identities  
 Dirichlet, 145  
 gamma, 144  
 Normal, multivariate, 145  
 $r$ , matrix-variate, 475  
 multivariate, 145  
 variance component distributions, 295
- Jacobians, matrix transformations, 473  
 Jeffreys' rule  
 single parameter, 42  
 multiparameter, 54
- Kronecker product, 477  
 Kurtosis, measure, 150  
 Kurtosis parameter (exponential power distribution), inference  
 prior distribution, 167  
 1-sample, 166, 167, 183, 187  
 $k$ -sample, unequal variances, 209, 217, 225, 231
- Likelihood, data translated, 32  
 approximate, 36, 41  
 multiparameter, 48  
 Likelihood function  
 definition, 10  
 role in Bayes' theorem, 10  
 standardized likelihood, 11  
 Likelihood inference, 73  
 Likelihood ratio criterion, comparison of  
 $k$  non-Normal variances, 221
- Linear model  
 non-Normal, 176  
 Normal, 46, 113  
 multivariate, common derivative matrix, 438  
 multivariate, general, 435  
 two random component, 393  
 Location parameter, definition, 18  
 Location-scale family distributions, 43  
 parameter inference, 43, 44, 58  
 Loss function, 308, 309, 313
- Matrix of independent variables, 113  
 Mean, definition, 76  
 Median, definition, 76  
 Mixed models  
 general, 324  
 additive, 325, 341  
 interaction, 326, 362  
 Monotone likelihood ratio, 351  
 Multinomial distribution, 55  
 parameter inference, 55
- Nonlinear model  
 non-Normal, 186  
 Normal, 422  
 linear approximation, 436  
 multivariate, 423

- Non-Normal means (exponential power parents), inference
  - 1-sample, 160, 162, 167, 171
  - $k$ -samples, unequal variances, 212, 229
- Normal distribution, 80
  - matrix-variate, 447
  - multivariate, 80, 115, 116
- Normal means, inference
  - fixed effect model ( $k$ -populations), 128, 377
  - multivariate model, 440
  - random effect model, 372, 376
  - random vs. fixed effect prior in high dimension, 379, 383
  - single mean, 28, 51, 82, 93, 95, 97
    - informative prior, 99
    - Normal prior, 18, 74
- Normal standard deviation, inference, 31, 87, 89, 93, 96
  - informative prior, 99
  - metric for locally uniform prior, 101
- Normal variance, inference (*see* Normal standard deviation)
- Nuisance parameter, 70
  - applied to robustness studies, 70, 169
- Observation, aberrant, 522
- Pascal distribution, 45
  - parameter inference, 45
- "Paired- $t$ " problem, 348
- Poisson distribution, 39
  - parameter inference, 40
- Pooling variance estimates, 249, 246, 289, 353
- Prior distribution
  - definition, 10
  - improper, 21
  - locally uniform, 23
  - noninformative, definition, 32
    - dependence on model, 44
    - location and scale parameters, 56
    - multiparameter, general, 54
    - summary of examples, 59
  - reference, 23
- Posterior distribution
  - definition, 10
  - parameters under constraints, 67
- Probability, subjective, 14
- Quadratic forms
  - combination of two, 418
  - mixed cumulants, 520
- Random effect models
  - hierarchical design, 244, 249, 277, 293
  - cross classification design, 318, 322, 329
  - one-way classification, 371
- Randomized (complete) block designs (*see* mixed model additive), 326
- Regression coefficients, inference
  - linear model, non-Normal, 177, 178
  - linear model, Normal, 48, 51, 115, 117, 125, 437
    - joint vs. marginal, 122
    - multivariate, common derivative matrix, 440, 441, 448, 452
    - two random components, 396
  - nonlinear model, non-Normal, 187
  - nonlinear model, Normal
    - multivariate, general, 427, 428
    - independent responses, 428
    - transformation of data, 531, 539
- Robustness
  - criterion, 152
  - inference, 153
    - variance ratio, 208
  - inference vs. criterion
    - mean, 153
    - variances, 232
- Sampling theory inference, 5, 72
- Scale parameter, definition, 49
- Scientific investigation, iterative process, 4
- Skewness, measure, 150
- Spread parameter, definition, 77
- Stable estimation, 22
- Statistical analysis, in scientific investigation, 5
- Statistical inference, general, 5
- Stirling's series, 147
- Student's  $t$  distribution (*see*  $t$  distribution)
- Sufficient statistics
  - additive mixed model, 341
  - definition, 61
  - exponential power distribution, variance, 206
  - hierarchical random effect models, 250, 278

- Normal distribution, 60
- Normal linear model, 115
- relevance in Bayesian inference, 63
- two-way random effect model, 331
- uniform (rectangular) distribution, 155
- $t$  distribution, 97
  - matric-variate, 441
  - multivariate, 117
  - product of multivariate, 481, 489
    - asymptotic expansion, 515
  - product of 2 univariate, 481
- $t^2$  criterion, non-Normal parent, 154
- Transformation of data
  - general model, 530
  - inference, 531, 539
  - logarithmic (textile data), 527, 536, 548
  - reciprocal (biological data), 525, 534, 542
  - square root (grinding experiment), 523
- $U$  distribution, 449
  - approximations, 451
- Uniformly most power similar test
  - two non-Normal variances, 207
- Variance components, definition, 245
- Variance components, inference
  - additive mixed model, 346
  - balanced incomplete blocks model, 407
  - interaction model, 364
  - $q$ -component hierarchical model, 293
  - 2-component hierarchical model
    - asymptotic expansion, 271, 298
    - individual components, 255, 258, 266, 272, 294, 296
    - noninformative prior, 251, 303
    - ratio, 253
    - sampling theory difficulties, 248
  - 3-component hierarchical model
    - individual components, 279, 286, 290
    - ratio, 280
    - relative contribution, 283
  - scaled  $\chi^{-2}$  approximation, 261, 274, 287, 291, 335, 353, 366
  - two-way random effect model, 332
- Weighted mean problem, 481
- Wishart distribution, 427
  - inverted, 460

This page intentionally left blank