Subjective and Objective Bayesian Statistics

Principles, Models, and Applications

Second Edition

S. JAMES PRESS

with contributions by

SIDHARTHA CHIB
MERLISE CLYDE
GEORGE WOODWORTH
ALAN ZASLAVSKY

WILEY-INTERSCIENCE

A John Wiley & Sons, Inc., Publication
WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS


A complete list of the titles in this series appears at the end of this volume.
Subjective and Objective Bayesian Statistics

Principles, Models, and Applications

Second Edition

S. JAMES PRESS

with contributions by

SIDDHARTHA CHIB
MERLISE CLYDE
GEORGE WOODWORTH
ALAN ZASLAVSKY

A John Wiley & Sons, Inc., Publication
To my Family
G, D, S, and all the J’s

Reason, Observation, and Experience—The Holy Trinity of Science
—Robert G. Ingersoll (1833–1899)
This sketch of the person we believe to be Thomas Bayes was created by Rachel Tanur and is reproduced here by permission of her estate.
CONTENTS

Preface xxi
Preface to the First Edition xxv
A Bayesian Hall of Fame xxix

PART I. FOUNDATIONS AND PRINCIPLES 1

1. Background 3
   1.1 Rationale for Bayesian Inference and Preliminary Views of Bayes' Theorem, 3
   1.2 Example: Observing a Desired Experimental Effect, 8
   1.3 Thomas Bayes, 11
   1.4 Brief Descriptions of the Chapters, 13
      Summary, 15
      Exercises, 15
      Further Reading, 16

2. A Bayesian Perspective on Probability 17
   2.1 Introduction, 17
   2.2 Types of Probability, 18
      2.2.1 Axiom Systems, 18
      2.2.2 Frequency and Long-Run Probability, 19
      2.2.3 Logical Probability, 20
      2.2.4 Kolmogorov Axiom System of Frequency Probability, 20
Example 4.2 Quality Control in Manufacturing: Discrete Data and Discrete Parameter (Inference About a Proportion), 46

4.3.2 Bayes' Theorem for Discrete Data and Discrete Models, 48

4.4 Bayes' Theorem for Continuous Data and Discrete Parameter, 48

4.4.1 Interpretation of Bayes' Theorem for Continuous Data and Discrete Parameter, 48

Example 4.3 Inferring the Section of a Class from which a Student was Selected: Continuous Data and Discrete Parameter (Choosing from a Discrete Set of Models), 49

4.5 Bayes' Theorem for Discrete Data and Continuous Parameter, 50

Example 4.4 Quality Control in Manufacturing: Discrete Data and Continuous Parameter, 50

4.6 Bayes' Theorem for Continuous Data and Continuous Parameter, 53

Example 4.5 Normal Data: Unknown Mean, Known Variance, 54

Example 4.6 Normal Data: Unknown Mean, Unknown Variance, 58

Summary, 63

Exercises, 63

Further Reading, 66

Complement to Chapter 4: Heights of the Standard Normal Density, 66

5. Prior Distributions 73

5.1 Introduction, 70

5.2 Objective and Subjective Prior Distributions, 70

5.2.1 Objective Prior Distributions, 70

  Public Policy Priors, 71

  Principle of Insufficient Reason (Laplace), 71

5.2.2 Weighing the Use of Objective Prior Distributions, 72

  Advantages, 72

  Disadvantages, 73

5.2.3 Weighing the Use of Subjective Prior Distributions, 74

  Advantages, 74

  Example 5.1, 74

  Example 5.2, 74

  Disadvantages, 75

5.3 (Univariate) Prior Distributions for a Single Parameter, 75

5.3.1 Vague (Indifference, Default, Objective) Priors, 76

  Vague Prior Density for Parameter on \((-\infty, \infty)\), 78

  Vague Prior Density for Parameter on \((0, \infty)\), 78

5.3.2 Families of Subjective Prior Distributions, 79

  A. Natural Conjugate Families of Prior Distributions, 79

    Example 5.3 A Natural Conjugate Prior: Binomial Data, 80
B. Exponential Power Family (EPF) of Prior Distributions, 81
C. Mixture Prior Distribution Families, 82
   Example 5.4 (Binomial), 82
5.3.3 Data-Based Prior Distributions, 84
   A. Historical Priors, 84
   B. Sample Splitting Priors, 84
5.3.4 g-Prior Distributions, 85
5.3.5 Stable Estimation Prior Distributions, 85
5.3.6 Assessing Fractiles of Your Subjective Prior Probability Distribution, 86
   Assessment Steps, 86
5.4 Prior Distributions for Vector and Matrix Parameters, 86
5.4.1 Vague Prior Distributions for Parameters on (−∞, ∞), 86
5.4.2 Vague Prior Distributions for Parameters on (0, ∞), 87
5.4.3 Jeffreys' Invariant Prior Distribution: Objective Bayesian Inference in the Normal Distribution, 88
   Example 5.5 Univariate Normal Data (Both Parameters Unknown), 89
   A. Vague Prior Density, 89
   B. Jeffreys' Prior Density, 91
   Example 5.6 Multivariate Normal Data (Both Parameters Unknown), 92
5.4.4 Assessment of a Subjective Prior Distribution for a Group, 94
   Multivariate Subjective Assessment for a Group, 94
   Assessment Overview for a Group, 95
   Model for a Group, 95
   Multivariate Density Assessment for a Group, 95
   Normal Density Kernel, 96
   Summary of Group Assessment Approach, 97
   Empirical Application of Group Assessment: Probability of Nuclear War in the 1980s, 97
   Consistency of Response, 99
   Implications, 99
   Histogram, 102
   Smoothed Prior Density (Fitted), 102
   Qualitative Data Provided by Expert Panelists (Qualitative Controlled Feedback: Content Analysis, Ethnography), 103
   Psychological Factors Relating to Subjective Probability Assessments for Group Members (or Individuals), 105
   Biases, 106
   Conclusions Regarding Psychological Factors, 106
   Summary of Group Prior Distribution Assessment, 106
   Posterior Distribution for Probability of Nuclear War, 106
5.4.5 Assessing Hyperparameters of Multiparameter Subjective Prior Distributions, 107
PART II. NUMERICAL IMPLEMENTATION OF THE BAYESIAN PARADIGM

6. Markov Chain Monte Carlo Methods

Siddhartha Chib

6.1 Introduction, 119
6.2 Metropolis–Hastings (M–H) Algorithm, 121
   6.2.1 Example: Binary Response Data, 123
   Random Walk Proposal Density, 127
   Tailored Proposal Density, 128
6.3 Multiple-Block M–H Algorithm, 130
   6.3.1 Gibbs Sampling Algorithm, 132
6.4 Some Techniques Useful in MCMC Sampling, 135
   6.4.1 Data Augmentation, 136
   6.4.2 Method of Composition, 137
   6.4.3 Reduced Blocking, 138
   6.4.4 Rao–Blackwellization, 139
6.5 Examples, 140
   6.5.1 Binary Response Data (Continued), 140
   6.5.2 Hierarchical Model for Clustered Data, 142
6.6 Comparing Models Using MCMC Methods, 147

Summary, 148
Exercises, 149
Further Reading, 151

Complement A to Chapter 6: The WinBUGS Computer Program,
by George Woodworth, 153

Introduction, 154
The WinBUGS Programming Environment, 155
Specifying the Model, 155
Example 6.1 Inference on a Single Proportion, 155
Simple Convergence Diagnostics, 160
Example 6.2 Comparing Two Proportions, Difference, Relative Risk, Odds Ratio, 160
Advanced Tools: Loops, Matrices, Imbedded Documents, Folds, 163
Example 6.3 Multiple Logistic Regression, 164
Additional Resources, 168
Further Reading, 169
Complement B to Chapter 6: Bayesian Software, 169

7. Large Sample Posterior Distributions and Approximations 172

7.1 Introduction, 172
7.2 Large-Sample Posterior Distributions, 173
7.3 Approximate Evaluation of Bayesian Integrals, 176
  7.3.1 Lindley Approximation, 176
  7.3.2 Tierney–Kadane–Laplace Approximation, 179
  7.3.3 Naylor–Smith Approximation, 182
7.4 Importance Sampling, 184
  Summary, 185
  Exercises, 185
  Further Reading, 186

PART III. BAYESIAN STATISTICAL INFERENCE AND DECISION MAKING 189

8. Bayesian Estimation 191

8.1 Introduction, 191
8.2 Univariate (Point) Bayesian Estimation, 191
  8.2.1 Binomial Distribution, 192
    Vague Prior, 192
    Natural Conjugate Prior, 193
  8.2.2 Poisson Distribution, 193
    Vague Prior, 193
    Natural Conjugate Prior, 194
  8.2.3 Negative Binomial (Pascal) Distribution, 194
    Vague Prior, 195
    Natural Conjugate Prior, 195
  8.2.4 Univariate Normal Distribution (Unknown Mean but Known Variance), 195
    Vague (Flat) Prior, 196
    Normal Distribution Prior, 197
  8.2.5 Univariate Normal Distribution (Unknown Mean and Unknown Variance), 198
    Vague Prior Distribution, 199
    Natural Conjugate Prior Distribution, 201
8.3 Multivariate (Point) Bayesian Estimation, 203
8.3.1 Multinomial Distribution, 203
   Vague Prior, 204
   Natural Conjugate Prior, 204
8.3.2 Multivariate Normal Distribution with Unknown Mean
   Vector and Unknown Covariance Matrix, 205
   Vague Prior Distribution, 205
   Natural Conjugate Prior Distribution, 208

8.4 Interval Estimation, 208
8.4.1 Credibility Intervals, 208
8.4.2 Credibility Versus Confidence Intervals, 209
8.4.3 Highest Posterior Density Intervals and Regions, 210
   Formal Statement for HPD Intervals, 211

8.5 Empirical Bayes' Estimation, 212
8.6 Robustness in Bayesian Estimation, 214
   Summary, 215
   Exercises, 215
   Further Reading, 216

9. Bayesian Hypothesis Testing 217
9.1 Introduction, 217
9.2 A Brief History of Scientific Hypothesis Testing, 217
9.3 Problems with Frequentist Methods of Hypothesis Testing, 220
9.4 Lindley's Vague Prior Procedure for Bayesian Hypothesis Testing, 224
   9.4.1 The Lindley Paradox, 225
9.5 Jeffreys' Procedure for Bayesian Hypothesis Testing, 225
   9.5.1 Testing a Simple Null Hypothesis Against a Simple
      Alternative Hypothesis, 225
      Jeffreys' Hypothesis Testing Criterion, 226
      Bayes' Factors, 226
   9.5.2 Testing a Simple Null Hypothesis Against a Composite
      Alternative Hypothesis, 227
   9.5.3 Problems with Bayesian Hypothesis Testing with Vague
      Prior Information, 229
   Summary, 230
   Exercises, 231
   Further Reading, 231

10. Predictivism 233
10.1 Introduction, 233
10.2 Philosophy of Predictivism, 233
10.3 Predictive Distributions/Comparing Theories, 234
10.3.1 Predictive Distribution for a Discrete Random Variable, 235
   Discrete Data Example: Comparing Theories Using the Binomial Distribution, 235
10.3.2 Predictive Distribution for a Continuous Random Variable, 237
   Continuous Data Example: Exponential Data, 237
10.3.3 Assessing Hyperparameters from Predictive Distributions, 238
10.4 Exchangeability, 238
10.5 De Finetti's Theorem, 239
10.5.1 Summary, 239
10.5.2 Introduction and Review, 239
10.5.3 Formal Statement, 240
10.5.4 Density Form, 241
10.5.5 Finite Exchangeability and De Finetti's Theorem, 242
10.6 The De Finetti Transform, 242
   Example 10.1 Binomial Sampling Distribution with Uniform Prior, 242
   Example 10.2 Normal Distribution with Both Unknown Mean and Unknown Variance, 243
10.6.1 Maxent Distributions and Information, 244
   Shannon Information, 244
10.6.2 Characterizing \( h(x) \) as a Maximum Entropy Distribution, 247
   Arbitrary Priors, 251
10.6.3 Applying De Finetti Transforms, 252
10.6.4 Some Remaining Questions, 253
10.7 Predictive Distributions in Classification and Spatial and Temporal Analysis, 253
10.8 Bayesian Neural Nets, 254
   Summary, 257
   Exercises, 257
   Further Reading, 259

11. Bayesian Decision Making
11.1 Introduction, 264
11.1.1 Utility, 264
11.1.2 Concave Utility, 265
11.1.3 Jensen's Inequality, 266
11.1.4 Convex Utility, 266
11.1.5 Linear Utility, 266
11.1.6 Optimizing Decisions, 267
11.2 Loss Functions, 267
   11.2.1 Quadratic Loss Functions, 268
       Why Use Quadratic Loss?, 268
   11.2.2 Linear Loss Functions, 270
   11.2.3 Piecewise Linear Loss Functions, 270
   11.2.4 Zero/One Loss Functions, 272
   11.2.5 Linex (Asymmetric) Loss Functions, 274

11.3 Admissibility, 275
    Summary, 276
    Exercises, 277
    Further Reading, 279

PART IV. MODELS AND APPLICATIONS 281

12. Bayesian Inference in the General Linear Model 283
   12.1 Introduction, 283
   12.2 Simple Linear Regression, 283
       12.2.1 Model, 283
       12.2.2 Likelihood Function, 284
       12.2.3 Prior, 284
       12.2.4 Posterior Inferences About Slope Coefficients, 284
       12.2.5 Credibility Intervals, 285
       12.2.6 Example, 286
       12.2.7 Predictive Distribution, 287
       12.2.8 Posterior Inferences About the Standard Deviation, 288
   12.3 Multivariate Regression Model, 289
       12.3.1 The Wishart Distribution, 289
       12.3.2 Multivariate Vague Priors, 290
       12.3.3 Multivariate Regression, 290
       12.3.4 Likelihood Function, 291
       Orthogonality Property at Least-Squares Estimators, 291
       12.3.5 Vague Priors, 292
       12.3.6 Posterior Analysis for the Slope Coefficients, 292
       12.3.7 Posterior Inferences About the Covariance Matrix, 293
       12.3.8 Predictive Density, 293
   12.4 Multivariate Analysis of Variance Model, 294
       12.4.1 One-Way Layout, 294
       12.4.2 Reduction to Regression Format, 294
       12.4.3 Likelihood, 295
       12.4.4 Priors, 295
       12.4.5 Practical Implications of the Exchangeability Assumption in
           the MANOVA Problem, 296
       Other Implications, 296
12.4.6 Posterior, 297
  Joint Posterior, 297
  Conditional Posterior, 297
  Marginal Posterior, 298
12.4.7 Balanced Design, 298
  Case of $p = 1$, 299
  Interval Estimation, 299
12.4.8 Example: Test Scores, 299
  Model, 299
  Contrasts, 301
12.4.9 Posterior Distributions of Effects, 301
12.5 Bayesian Inference in the Multivariate Mixed Model, 302
  12.5.1 Introduction, 302
  12.5.2 Model, 303
  12.5.3 Prior Information, 305
    A. Nonexchangeable Case, 306
    B. Exchangeable Case, 306
  12.5.4 Posterior Distributions, 307
  12.5.5 Approximation to the Posterior Distribution of $B$, 309
  12.5.6 Posterior Means for $\Sigma, \Sigma_1, \ldots, \Sigma_c$, 311
  12.5.7 Numerical Example, 314
  Summary, 316
  Exercises, 316
  Further Reading, 318

13. Model Averaging

320

Merlise Clyde

13.1 Introduction, 320
13.2 Model Averaging and Subset Selection in Linear Regression, 321
13.3 Prior Distributions, 323
  13.3.1 Prior Distributions on Models, 323
  13.3.2 Prior Distributions for Model-Specific Parameters, 323
13.4 Posterior Distributions, 324
13.5 Choice of Hyperparameters, 325
13.6 Implementing BMA, 326
13.7 Examples, 326
  13.7.1 Pollution and Mortality, 326
  13.7.2 O-Ring Failures, 328
  Summary, 331
  Exercises, 332
  Further Reading, 334
14. Hierarchical Bayesian Modeling

Alan Zaslavsky

14.1 Introduction, 336
14.2 Fundamental Concepts and Nomenclature, 336
   14.2.1 Motivating Example, 336
   14.2.2 What Makes a Hierarchical Model?, 3337
      Multilevel Parameterization, 338
      Hierarchically Structured Data, 338
      Correspondence of Parameters to Population Structures, and
      Conditional Independence, 339
   14.2.3 Marginalization, Data Augmentation and Collapsing, 340
   14.2.4 Hierarchical Models, Exchangeability, and De Finetti's
      Theorem, 341
14.3 Applications and Examples, 341
   14.3.1 Generality of Hierarchical Models, 341
   14.3.2 Variance Component Models, 342
   14.3.3 Random Coefficient Models, Mixed Models, Longitudinal
      Data, 343
   14.3.4 Models with Normal Priors and Non-Normal
      Observations, 344
   14.3.5 Non-Normal Conjugate Models, 345
14.4 Inference in Hierarchical Models, 345
   14.4.1 Levels of Inference, 345
   14.4.2 Full Bayes' Inference, 346
   14.4.3 Priors for Hyperparameters of Hierarchical Models, 347
14.5 Relationship to Non-Bayesian Approaches, 348
   14.5.1 Maximum Likelihood Empirical Bayes and Related
      Approaches, 348
   14.5.2 Non-Bayesian Theoretical Approaches: Stein Estimation, Best
      Linear Unbiased Predictor, 349
   14.5.3 Contrast to Marginal Modeling Approaches with Clustered
      Data, 350
14.6 Computation for Hierarchical Models, 351
   14.6.1 Techniques Based on Conditional Distributions: Gibbs
      Samplers and Data Augmentation, 351
   14.6.2 Techniques Based on Marginal Likelihoods, 352
14.7 Software for Hierarchical Models, 352
   Summary, 353
   Exercises, 353
   Further Reading, 356
15. **Bayesian Factor Analysis** 359

15.1 Introduction, 359
15.2 Background, 359
15.3 Bayesian Factor Analysis Model for Fixed Number of Factors, 361
   15.3.1 Likelihood Function, 361
   15.3.2 Priors, 362
   15.3.3 Joint Posteriors, 363
   15.3.4 Marginal Posteriors, 363
   15.3.5 Estimation of Factor Scores, 364
   15.3.6 Historical Data Assessment of $F$, 364
   15.3.7 Vague Prior Estimator of $F$, 364
   15.3.8 Large Sample Estimation of $F$, 365
   15.3.9 Large Sample Estimation of $f_i$, 366
   15.3.10 Large Sample Estimation of the Elements of $f_i$, 366
   15.3.11 Estimation of the Factor Loadings Matrix, 367
   15.3.12 Estimation of the Disturbance Covariance Matrix, 365
   15.3.13 Example, 368
15.4 Choosing the Number of Factors, 372
   15.4.1 Introduction, 372
   15.4.2 Posterior Odds for the Number of Factors: General Development, 376
   15.4.3 Likelihood Function, 377
   15.4.4 Prior Densities, 378
   15.4.5 Posterior Probability for the Number of Factors, 379
   15.4.6 Numerical Illustrations and Hyperparameter Assessment, 380
      Data Generation, 380
      Results, 381
   15.4.7 Comparison of the Maximum Posterior Probability Criterion with AIC and BIC, 382
15.5 Additional Model Considerations, 382
   Summary, 384
   Exercises, 384
   Further Reading, 385
   Complement to Chapter 15: Proof of Theorem 15.1, 387

16. **Bayesian Inference in Classification and Discrimination** 391

16.1 Introduction, 391
16.2 Likelihood Function, 392
16.3 Prior Density, 393
16.4 Posterior Density, 393
16.5 Predictive Density, 393
16.6 Posterior Classification Probability, 395
16.7 Example: Two Populations, 396
16.8 Second Guessing Undecided Respondents: An Application, 397
   16.8.1 Problem, 397
   Solution, 397
16.8.2 Example, 399
16.9 Extensions of the Basic Classification Problem, 399
   16.9.1 Classification by Bayesian Clustering, 399
   16.9.2 Classification Using Bayesian Neural Networks and Tree-Based Methods, 400
   16.9.3 Contextual Bayesian Classification, 401
   16.9.4 Classification in Data Mining, 402
Summary, 402
Exercises, 403
Further Reading, 404

APPENDICES

Description of Appendices 407
Appendix 1. Bayes, Thomas, 409
   Hilary L. Seal
Appendix 2. Thomas Bayes. A Bibliographical Note, 415
   George A. Barnard
Appendix 3. Communication of Bayes’ Essay to the Philosophical Transactions of the Royal Society of London, 419
   Richard Price
Appendix 4. An Essay Towards Solving a Problem in the Doctrine of Chances, 423
   Reverend Thomas Bayes
Appendix 5. Applications of Bayesian Statistical Science, 449
Appendix 6. Selecting the Bayesian Hall of Fame, 456
Appendix 7. Solutions to Selected Exercises. 459

Bibliography 523
Subject Index 543
Author Index 553
This Page intentionally left blank
PREFACE

This second edition is intended to be an introduction to Bayesian statistics for students and research workers who have already been exposed to a good preliminary statistics and probability course, probably from a frequentist viewpoint, but who have had a minimal exposure to Bayesian theory and methods. We assume a mathematical level of sophistication that includes a good calculus course and some matrix algebra, but nothing beyond that. We also assume that our audience includes those who are interested in using Bayesian methods to model real problems, in areas that range across the disciplines.

This second edition is really a new book. It is not merely the first edition with a few changes inserted; it is a completely restructured book with major new chapters and material.

The first edition to this book was completed in 1988. Since then the field of Bayesian statistical science has grown so substantially that it has become necessary to rewrite the story in broader terms to account for the changes that have taken place, both in new methodologies that have been developed since that time, and in new techniques that have emerged for implementing the Bayesian paradigm. Moreover, as the fields of computer science, numerical analysis, artificial intelligence, pattern recognition, and machine learning have also made enormous advances in the intervening years, and because their interfaces with Bayesian statistics have steadily increased, it became important to expand our story to include, at least briefly, some of those important interface topics, such as data mining tree models and Bayesian neural networks. In addition, as the field of Bayesian statistics has expanded, the applications that have been made using the Bayesian approach to learning from experience and analysis of data now span most of the disciplines in the biological, physical, and social sciences. This second edition attempts to tell the broader story that has developed.

One direction of growth in Bayesian statistics that has occurred in recent years resulted from the contributions made by Geman and Geman (1984), Tanner and Wong (1987), and Gelfand and Smith (1990). These papers proposed a new method, now called Markov chain Monte Carlo (or just MCMC), for applying and imple-
menting Bayesian procedures numerically. The new method is computer intensive and involves sampling by computer (so-called Monte Carlo sampling) from the posterior distribution to obtain its properties. Usually, Bayesian modeling procedures result in ratios of multiple integrals to be evaluated numerically. Sometimes these multiple integrals are high dimensional. The results of such Bayesian analysis are wonderful theoretically because they arise from a logical, self-consistent, set of axioms for making judgments and decisions. In the past, however, to evaluate such ratios of high-dimensional multiple integrals numerically it was necessary to carry out tedious numerical computations that were difficult to implement for all but the very computer-knowledgeable researcher. With a computer environment steadily advancing from the early 1980s, and with the arrival of computer software to implement the MCMC methodology, Bayesian procedures could finally be implemented rapidly, and accurately, and without the researcher having to possess a sophisticated understanding of numerical methods.

In another important direction of growth of the field, Bayesian methodology has begun to recognize some of the implications of the important distinction between subjective and objective prior information. This distinction is both philosophical and mathematical. When information based upon underlying theory or historical data is available (subjective prior information), the Bayesian approach suggests that such information be incorporated into the prior distribution for use in Bayesian analysis. If families of prior distributions are used to capture the prior knowledge, such prior distributions will contain their own parameters (called hyperparameters) that will need to be assessed on the basis of the available information. For example, many surveys are carried out on the same topic year after year, so that results obtained in earlier years can be used as a best guess for what is likely to be obtained in a new survey in the current year. Such “best available” information can be incorporated into a prior distribution. Such prior distributions are always proper (integrate or sum to one), and so behave well mathematically. A Bayesian analysis using such a prior distribution is called subjective Bayesian analysis.

In some situations, however, it is difficult to specify appropriate subjective prior information. For example, at the present time, there is usually very little, if any, prior information about the function of particular sequences of nucleotide base pairs in the DNA structure of the human genome. In such situations it is desirable to have meaningful ways to begin the Bayesian learning updating process. A prior distribution adopted for such a situation is called objective, and an analysis based upon such an objective prior distribution is called an objective Bayesian analysis. Such analyses serve to provide benchmark statistical inferences based upon having inserted as little prior information as possible, prior to taking data. Objective prior distributions correspond to “knowing little” prior to taking data. When such prior distributions are continuous, it is usually the case that these (improper) prior distributions do not integrate to one (although acceptable posterior distributions that correspond to these improper prior distributions must integrate to one). Sometimes, in simple cases, posterior inferences based upon objective prior distributions will result in inferences that correspond to those arrived at by frequentist means. The field has begun to focus on the broader implications of the similarities and differences between subjective
and objective types of information. We treat this important topic in this edition, and recognize its importance in the title of the book. In many applications of interest, there is not enough information in a problem for classical inference to be carried out. So some researchers resort to subjective Bayesian inference out of necessity. The subjective Bayesian approach is adopted because it is the most promising way to introduce sufficient additional information into the problem so that a real solution can be found.

In earlier years, it was difficult to take into account uncertainty about which model to choose in a Bayesian analysis of data. Now we are learning how to incorporate such uncertainty into the analysis by using Bayesian model averaging. Moreover, we have been learning how to use Bayesian modeling in a hierarchical way to represent nested degrees of uncertainty about a problem. A whole new framework for exploratory factor analysis has been developed based upon the Bayesian paradigm. These topics are new and are discussed in this edition.

In this edition, for the first time, we will present an extensive listing, by field, of some of the broad-ranging applications that have been made of the Bayesian approach.

As Bayesian statistical science has developed and matured, its principal founders and contributors have become apparent. To record and honor them, in this edition we have included a Bayesian Hall of Fame, which we developed by means of a special opinion poll taken among senior Bayesian researchers. Following the table of contents is a collection of the portraits and brief biographies of these most important contributors to the development of the field, and there is an appendix devoted to an explanation of how the members of the Hall of Fame were selected.

The first edition of this book contained eight chapters and four appendices; this edition contains 16 chapters, generally quite different from those in the first edition, and seven appendices. The current coverage reflects not only the addition of new topics and the deletion of some old ones, but also the expansion of some previously covered topics into greater depth, and more domains. In addition, there are solutions to some of the exercises.

This second edition has been designed to be used in a year-long course in Bayesian statistics at the senior undergraduate or graduate level. If the academic year is divided into semesters, Chapters 1–8 can be covered in the first semester and Chapters 9–16 in the second semester. If the academic year is divided into quarters, Chapters 1–5 (Part I) can be covered in the fall quarter, Chapters 6–11 (Parts II and III) in the winter quarter, and Chapters 12–16 (Part IV) in the spring quarter.

Three of the sixteen chapters of this second edition have been written with the assistance of four people: Chapter 6 by Professor Siddhartha Chib of Washington University; Complement A to Chapter 6 by Professor George Woodworth of the University of Iowa; Chapter 13 by Professor Merlise Clyde of Duke University; and Chapter 14 by Professor Alan Zaslavsky of Harvard University. I am very grateful for their help. Much of Appendix 7 was written with the help of my former students, Dr. Thomas Ferryman, Dr. Mahmood Gharnsary, and Ms. Dawn Kummer. I am also grateful to Stephen Quigley of John Wiley and Sons, Inc., who encouraged me to prepare this second edition, and to Heather Haselkorn of Wiley, who helped and
prodded me until it was done. Dr. Judith Tanur helped me to improve the exposition and to minimize the errors in the manuscript. The remaining errors are totally my responsibility. I am grateful to Rachel Tanur for her sketch of Thomas Bayes at the beginning of the book. Her untimely death prevented her from her intention of also sketching the scientists who appear in the Bayesian Hall of Fame. Dr. Linda Penas solved some of our more complex LaTex editorial problems, while Ms. Peggy Franklin typed some of the chapters in LaTex with indefatigable patience and endurance.

S. JAMES PRESS

Oceanside, CA

September, 2002
This book is intended to be an introduction to Bayesian statistics for students and research workers who have already been exposed to a good preliminary statistics and probability course from a classical (frequentist) point of view but who have had minimal exposure to Bayesian theory and methods. We assume a mathematical level of sophistication that includes a good calculus course and some matrix algebra but nothing beyond that. We also assume that our audience includes those who are interested in using Bayesian methods to model real problems in the various scientific disciplines. Such people usually want to understand enough of the foundational principles so that they will (1) feel comfortable using the procedures, (2) have no compunction about recommending solutions based upon these procedures to decision makers, and (3) be intrigued enough to go to referenced sources to seek additional background and understanding. For this reason we have tried to maximize interpretation of theory and have minimized our dependence upon proof of theorems.

The book is organized in two parts of four chapters each; in addition, the back of the book contains appendixes, a bibliography, and separate author and subject indexes. The first part of the book is devoted to theory; the second part is devoted to models and applications. The appendixes provide some biographical material about Thomas Bayes, along with a reproduction of Bayes's original essay.

Chapter I shows that statistical inference and decision making from a Bayesian point of view is based upon a logical, self-consistent system of axioms; it also shows that violation of the guiding principles will lead to "incoherent" behavior, that is, behavior that would lead to economically unsound decisions in a risky situation.

Chapter II covers the basic principles of the subject. Bayes's theorem is presented for both discrete and absolutely continuous random variables.

We discuss Bayesian estimation, hypothesis testing, and decision theory. It is here that we introduce prior distributions, Bayes' factors, the important theorem of de Finetti, the likelihood principle, and predictive distributions.
Chapter III includes various methods for approximating the sometimes complicated posterior distributions that result from applications of the Bayesian paradigm. We present large-sample theory results as well as Laplacian types of approximations of integrals (representing posterior densities). We will show how importance sampling as well as simulation of distributions can be used for approximation of posterior densities when the dimensions are large. We will also provide a convenient up-to-date summary of the latest Bayesian computer software available for implementation.

Chapter IV shows how prior distributions can be assessed subjectively using a group of experts. The methodology is applied to the problem of using a group of experts on strategic policy to assess a multivariate prior distribution for the probability of nuclear war during the decade of the 1980s.

Chapter V is concerned with Bayesian inference in both the univariate and multivariate regression models. Here we use vague prior distributions, and we apply the notion of predictive distributions to predicting future observations in regression models.

Chapter VI continues discussion of the general linear model begun in Chapter V, only here we show how to carry out Bayesian analysis of variance and covariance in the multivariate case. We will invoke the de Finetti notion of exchangeability (of the population mean vector distributions).

Chapter VII is devoted to the theory and application of Bayesian classification and discrimination procedures. The methodology is illustrated by applying it to the sample survey problem of second guessing "undecided" respondents.

Chapter VIII presents a case study of how disputed authorship of some of the Federalist papers was resolved by means of a Bayesian analysis.

The book is easily adapted to a one- or two-quarter sequence or to a one-semester, senior level, or graduate course in Bayesian statistics. The first two chapters and the appendixes could easily fill the first quarter, with Chapters III–VIII devoted to the second quarter. In a one-quarter or one-semester course, certain sections or chapters would need to be deleted; which chapters or sections to delete would depend upon the interests of the students and teacher in terms of the balance desired between (1) theory and (2) models and applications.

The book represents an expansion of a series of lectures presented in South Australia in July 1984 at the University of Adelaide. These lectures were jointly sponsored by the Commonwealth Scientific and Industrial Research Organization (CSIRO), Division of Mathematics and Statistics and by the University of Adelaide's Departments of Economics and Statistics. I am grateful to Drs. Graham Constantine, William Davis, and Terry Speed, all of CSIRO, for their stimulating comments on the original lecture material, for their encouragement and support, and for planting the seeds from which this monograph grew. I am grateful to Dr. John Darroch, Dr. Alastair Fischer, Dr. Alan James, Dr. W. N. Venables, and to other participants of the lecture series for their stimulating questions that helped to put the book into perspective. Dr. John Pratt and Dr. S. L. Zabell helped to clarify the issues about de Finetti's theorem in Section 2.9.3, and Dr. S. K. Sinha suggested an example used in Section 2.7.1. Dr. Persi Diaconis and Dr. Richard Jeffrey presented stimulating discussions.
about randomness, exchangeability, and some of the foundational issues of the subject in a seminar at Stanford University during winter quarter of 1984–1985, a sabbatical year the author spent visiting Stanford University. I am deeply grateful to Drs. Harry Roberts and Arnold Zellner for exposing me to Bayesian ideas. Dr. Stephen Fienberg provided encouragement and advice regarding publishing the manuscript. I am also grateful to Dr. Stephen Fienberg, Dr. Ingram Olkin, and an anonymous publisher’s referee for many helpful suggestions for improving the presentation. I am very grateful for suggestions made by Dr. Judith Tanur who read the entire manuscript; to Dr. Ruben Klein who read Chapters I and II; and to Drs. Frederick Mosteller and David Wallace who read Chapter VIII. I also wish to thank graduate students, James Bentley, David Guy, William Kemple, Thomas Lucas, and Hamid Namini whose questions about the material during class prompted me to revise and clarify various issues. Mrs. Peggy Franklin is to be congratulated for her outstanding typing ability and for her forbearance in seeing me through the many iterations that the manuscript underwent. We think we have eliminated most, if not all, errors in the book, but readers could help the author by calling any additional ones they find to his attention.

S. James Press

Riverside, California
January, 1989
A BAYESIAN HALL OF FAME

Bayes, Thomas
1701–1761

DeFinetti, Bruno
1906–1985

DeGroot, Morris
1931–1989

Jeffreys, Harold
1891–1989

Lindley, Dennis V.
1923–

Savage, Leonard J.
1917–1971