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

WILEY-INTERSCIENCE
A John Wiley & Sons, Inc., Publication
# CONTENTS

<table>
<thead>
<tr>
<th>Preface</th>
<th>xxvi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface to the First Edition</td>
<td>xxv</td>
</tr>
<tr>
<td>A Bayesian Hall of Fame</td>
<td>xxix</td>
</tr>
</tbody>
</table>

## PART I. FOUNDATIONS AND PRINCIPLES  

1. **Background**  
   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**  
   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
2.2.5 Savage System of Axioms of Subjective Probability, 21
2.2.6 Rényi Axiom System of Probability, 22
2.3 Coherence, 24
  2.3.1 Example of Incoherence, 24
2.4 Operationalizing Subjective Probability Beliefs, 25
  2.4.1 Example of Subjective Probability Definition and Operationalization, 26
2.5 Calibration of Probability Assessors, 26
2.6 Comparing Probability Definitions, 27
  Summary, 28
Complement to Chapter 2: The Axiomatic Foundation of Decision making of L. J. Savage, 29
Utility Functions, 30
Exercises, 30
Further Reading, 31

3. The Likelihood Function 34
  3.1 Introduction, 34
  3.2 Likelihood Function, 34
  3.3 Likelihood Principle, 35
  3.4 Likelihood Principle and Conditioning, 36
  3.5 Likelihood and Bayesian Inference, 37
  3.6 Development of the Likelihood Function Using Histograms and Other Graphical Methods, 38
  Summary, 39
  Exercises, 39
  Further Reading, 40

4. Bayes' Theorem 41
  4.1 Introduction, 41
  4.2 General Form of Bayes' Theorem for Events, 41
    4.2.1 Bayes' Theorem for Complementary Events, 42
    4.2.2 Prior Probabilities, 42
    4.2.3 Posterior Probabilities, 42
    4.2.4 Odds Ratios, 42
      Example 4.1 Bayes' Theorem for Events: DNA Fingerprinting, 43
  4.3 Bayes' Theorem for Discrete Data and Discrete Parameter, 45
    4.3.1 Interpretation of Bayes' Theorem for Discrete Data and Discrete Parameter, 45
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 (−∞, ∞), 78

Vague Prior Density for Parameter on (0, ∞), 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
CONTENTS

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 $(-\infty, \infty)$, 86
  5.4.2 Vague Prior Distributions for Parameters on $(0, \infty)$, 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. Jeffrey' 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
Maximum Entropy (Maxent) Prior Distributions (Minimum Information Priors), 108

5.5 Data-Mining Priors, 108
5.6 Wrong Priors, 110
Summary, 110
Exercises, 111
Further Reading, 113

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
CONTENTS

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  

*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

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 $f_j$, 366
  15.3.10 Large Sample Estimation of the Elements of $f_j$, 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

16.1 Introduction, 391
16.2 Likelihood Function, 392
16.3 Prior Density, 393
16.4 Posterior Density, 393
16.5 Predictive Density, 393
CONTENTS

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
  16.8.2 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