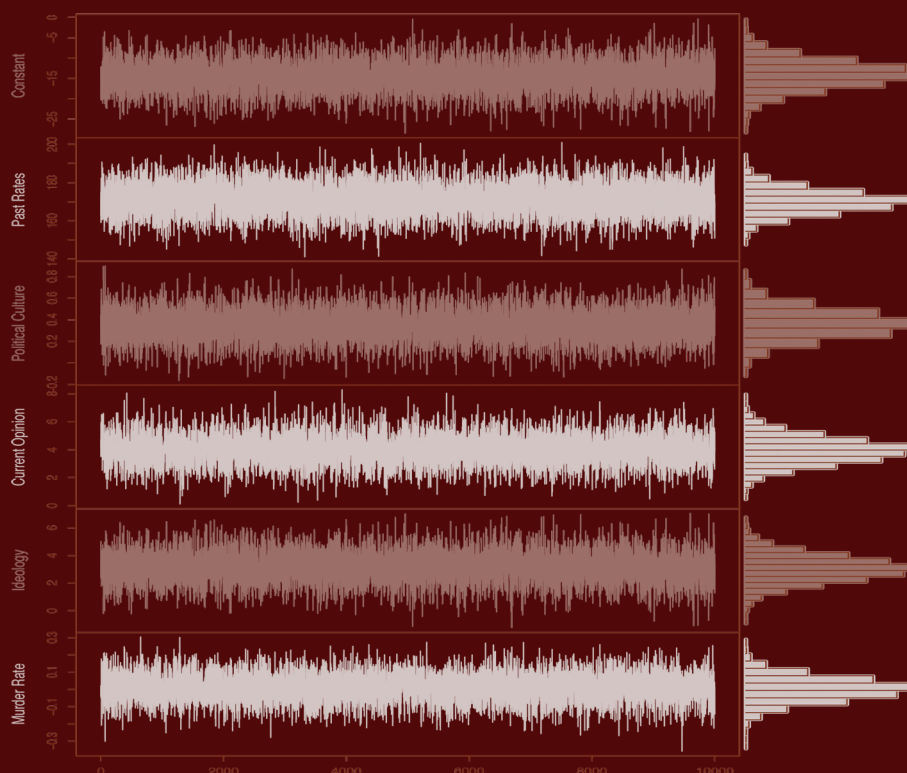


Chapman & Hall/CRC  
Statistics in the Social and Behavioral Sciences Series

# Bayesian Methods

A Social and Behavioral  
Sciences Approach

Third Edition



Jeff Gill



CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK



# Bayesian Methods

A Social and  
Behavioral Sciences  
Approach

Third Edition

Chapman & Hall/CRC  
**Statistics in the Social and Behavioral Sciences Series**

**Series Editors**

**Jeff Gill**

Washington University, USA

**Steven Heeringa**

University of Michigan, USA

**Wim van der Linden**

CTB/McGraw-Hill, USA

**J. Scott Long**

Indiana University, USA

**Tom Snijders**

Oxford University, UK  
University of Groningen, NL

**Aims and scope**

Large and complex datasets are becoming prevalent in the social and behavioral sciences and statistical methods are crucial for the analysis and interpretation of such data. This series aims to capture new developments in statistical methodology with particular relevance to applications in the social and behavioral sciences. It seeks to promote appropriate use of statistical, econometric and psychometric methods in these applied sciences by publishing a broad range of reference works, textbooks and handbooks.

The scope of the series is wide, including applications of statistical methodology in sociology, psychology, economics, education, marketing research, political science, criminology, public policy, demography, survey methodology and official statistics. The titles included in the series are designed to appeal to applied statisticians, as well as students, researchers and practitioners from the above disciplines. The inclusion of real examples and case studies is therefore essential.

## Published Titles

### **Analyzing Spatial Models of Choice and Judgment with R**

*David A. Armstrong II, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, and Howard Rosenthal*

### **Analysis of Multivariate Social Science Data, Second Edition**

*David J. Bartholomew, Fiona Steele, Irini Moustaki, and Jane I. Galbraith*

### **Latent Markov Models for Longitudinal Data**

*Francesco Bartolucci, Alessio Farcomeni, and Fulvia Pennoni*

### **Statistical Test Theory for the Behavioral Sciences**

*Dato N. M. de Gruijter and Leo J. Th. van der Kamp*

### **Multivariable Modeling and Multivariate Analysis for the Behavioral Sciences**

*Brian S. Everitt*

### **Multilevel Modeling Using R**

*W. Holmes Finch, Jocelyn E. Bolin, and Ken Kelley*

### **Bayesian Methods: A Social and Behavioral Sciences Approach, Third Edition**

*Jeff Gill*

### **Multiple Correspondence Analysis and Related Methods**

*Michael Greenacre and Jorg Blasius*

### **Applied Survey Data Analysis**

*Steven G. Heeringa, Brady T. West, and Patricia A. Berglund*

### **Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists**

*Herbert Hoijtink*

### **Generalized Structured Component Analysis: A Component-Based Approach to Structural Equation Modeling**

*Heungsun Hwang and Yoshio Takane*

### **Statistical Studies of Income, Poverty and Inequality in Europe: Computing and Graphics in R Using EU-SILC**

*Nicholas T. Longford*

### **Foundations of Factor Analysis, Second Edition**

*Stanley A. Mulaik*

### **Linear Causal Modeling with Structural Equations**

*Stanley A. Mulaik*

### **Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data**

*Robert M. O'Brien*

### **Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis**

*Leslie Rutkowski, Matthias von Davier, and David Rutkowski*

### **Generalized Linear Models for Categorical and Continuous Limited Dependent Variables**

*Michael Smithson and Edgar C. Merkle*

### **Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys**

*Guo-Liang Tian and Man-Lai Tang*

### **Computerized Multistage Testing: Theory and Applications**

*Duanli Yan, Alina A. von Davier, and Charles Lewis*



Chapman & Hall/CRC  
Statistics in the Social and Behavioral Sciences Series

# Bayesian Methods

A Social and  
Behavioral Sciences  
Approach

Third Edition

**Jeff Gill**

Washington University  
St. Louis, Missouri, USA



**CRC Press**

Taylor & Francis Group  
Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business  
A CHAPMAN & HALL BOOK

CRC Press  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC  
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works  
Version Date: 20141105

International Standard Book Number-13: 978-1-4398-6249-0 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

**Visit the Taylor & Francis Web site at**  
**<http://www.taylorandfrancis.com>**

**and the CRC Press Web site at**  
**<http://www.crcpress.com>**



*This book is again dedicated to Jack Gill, who is still there when I need him.*



---

# Contents

List of Figures	xix
List of Tables	xxi
Preface to the Third Edition	xxv
Preface to the Second Edition	xxix
Preface to the First Edition	xxxvii
<b>1 Background and Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 General Motivation and Justification . . . . .	4
1.3 Why Are We Uncertain about Uncertainty? . . . . .	7
1.3.1 Required Probability Principles . . . . .	8
1.4 Bayes' Law . . . . .	9
1.4.1 Bayes' Law for Multiple Events . . . . .	12
1.5 Conditional Inference with Bayes' Law . . . . .	15
1.5.1 Statistical Models with Bayes' Law . . . . .	17
1.6 Science and Inference . . . . .	20
1.6.1 The Scientific Process in Our Social Sciences . . . . .	20
1.6.2 Bayesian Statistics as a Scientific Approach to Social and Behavioral Data Analysis . . . . .	22
1.7 Introducing Markov Chain Monte Carlo Techniques . . . . .	24
1.7.1 Simple Gibbs Sampling . . . . .	25
1.7.2 Simple Metropolis Sampling . . . . .	27
1.8 Historical Comments . . . . .	30
1.9 Exercises . . . . .	33
<b>2 Specifying Bayesian Models</b>	<b>37</b>
2.1 Purpose . . . . .	37
2.2 Likelihood Theory and Estimation . . . . .	37
2.3 The Basic Bayesian Framework . . . . .	40
2.3.1 Developing the Bayesian Inference Engine . . . . .	40
2.3.2 Summarizing Posterior Distributions with Intervals . . . . .	42

2.3.2.1	Bayesian Credible Intervals . . . . .	43
2.3.2.2	Bayesian Highest Posterior Density Intervals . . . . .	46
2.3.3	Quantile Posterior Summaries . . . . .	48
2.3.4	Beta-Binomial Model . . . . .	49
2.4	Bayesian “Learning” . . . . .	53
2.5	Comments on Prior Distributions . . . . .	56
2.6	Bayesian versus Non-Bayesian Approaches . . . . .	57
2.7	Exercises . . . . .	61
2.8	Computational Addendum: R for Basic Analysis . . . . .	66
<b>3</b>	<b>The Normal and Student’s-<i>t</i> Models</b>	<b>69</b>
3.1	Why Be Normal? . . . . .	69
3.2	The Normal Model with Variance Known . . . . .	70
3.3	The Normal Model with Mean Known . . . . .	72
3.4	The Normal Model with Both Mean and Variance Unknown . . . . .	74
3.5	Multivariate Normal Model, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ Both Unknown . . . . .	76
3.6	Simulated Effects of Differing Priors . . . . .	80
3.7	Some Normal Comments . . . . .	82
3.8	The Student’s- <i>t</i> Model . . . . .	83
3.9	Normal Mixture Models . . . . .	87
3.10	Exercises . . . . .	89
3.11	Computational Addendum: Normal Examples . . . . .	93
3.11.1	Normal Example with Variance Known . . . . .	93
3.11.2	Bivariate Normal Simulation Example . . . . .	94
3.11.3	Multivariate Normal Example, Health Data . . . . .	95
<b>4</b>	<b>The Bayesian Prior</b>	<b>97</b>
4.1	A Prior Discussion of Priors . . . . .	97
4.2	A Plethora of Priors . . . . .	98
4.3	Conjugate Prior Forms . . . . .	100
4.3.1	Example: Conjugacy in Exponential Specifications . . . . .	100
4.3.2	The Exponential Family Form . . . . .	101
4.3.3	Limitations of Conjugacy . . . . .	104
4.4	Uninformative Prior Distributions . . . . .	104
4.4.1	Uniform Priors . . . . .	105
4.4.2	Jeffreys Prior . . . . .	107
4.4.2.1	Bernoulli Trials and Jeffreys Prior . . . . .	108
4.4.2.2	Other Forms of Jeffreys Priors . . . . .	109
4.4.2.3	Jeffreys Prior in the Multiparameter Case . . . . .	110
4.4.3	Reference Priors . . . . .	111
4.4.4	Improper Priors . . . . .	113

4.5	Informative Prior Distributions . . . . .	114
4.5.1	Power Priors . . . . .	115
4.5.2	Elicited Priors . . . . .	116
4.5.2.1	The Community of Elicited Priors . . . . .	118
4.5.2.2	Simple Elicitation Using Linear Regression . . . . .	119
4.5.2.3	Variance Components Elicitation . . . . .	121
4.5.2.4	Predictive Modal Elicitation . . . . .	123
4.5.2.5	Prior Elicitation for the Normal-Linear Model . . . . .	126
4.5.2.6	Elicitation Using a Beta Distribution . . . . .	127
4.5.2.7	Eliciting Some Final Thoughts on Elicited Priors . . . . .	128
4.6	Hybrid Prior Forms . . . . .	129
4.6.1	Spike and Slab Priors for Linear Models . . . . .	130
4.6.2	Maximum Entropy Priors . . . . .	131
4.6.3	Histogram Priors . . . . .	133
4.7	Nonparametric Priors . . . . .	133
4.8	Bayesian Shrinkage . . . . .	136
4.9	Exercises . . . . .	138
<b>5</b>	<b>The Bayesian Linear Model</b>	<b>145</b>
5.1	The Basic Regression Model . . . . .	145
5.1.1	Uninformative Priors for the Linear Model . . . . .	147
5.1.2	Conjugate Priors for the Linear Model . . . . .	151
5.1.3	Conjugate Caveats for the Cautious and Careful . . . . .	154
5.2	Posterior Predictive Distribution for the Data . . . . .	155
5.3	Linear Regression with Heteroscedasticity . . . . .	161
5.4	Exercises . . . . .	165
5.5	Computational Addendum . . . . .	169
5.5.1	Palm Beach County Normal Model . . . . .	169
5.5.2	Educational Outcomes Model . . . . .	171
5.5.3	Ancient China Conflict Model . . . . .	172
<b>6</b>	<b>Assessing Model Quality</b>	<b>175</b>
6.1	Motivation . . . . .	175
6.1.1	Posterior Data Replication . . . . .	177
6.1.2	Likelihood Function Robustness . . . . .	180
6.2	Basic Sensitivity Analysis . . . . .	181
6.2.1	Global Sensitivity Analysis . . . . .	181
6.2.1.1	Specific Cases of Global Prior Sensitivity Analysis . . . . .	182
6.2.1.2	Global Sensitivity in the Normal Model Case . . . . .	182
6.2.1.3	Example: Prior Sensitivity in the Analysis of the 2000 U.S. Election in Palm Beach County . . . . .	183

6.2.1.4	Problems with Global Sensitivity Analysis . . . . .	183
6.2.2	Local Sensitivity Analysis . . . . .	184
6.2.2.1	Normal-Normal Model . . . . .	185
6.2.2.2	Local Sensitivity Analysis Using Hyperparameter Changes . . . . .	186
6.2.3	Global and Local Sensitivity Analysis with Recidivism Data . . . . .	187
6.3	Robustness Evaluation . . . . .	189
6.3.1	Global Robustness . . . . .	190
6.3.2	Local Robustness . . . . .	192
6.3.2.1	Bayesian Linear Outlier Detection . . . . .	193
6.3.3	Bayesian Specification Robustness . . . . .	195
6.4	Comparing Data to the Posterior Predictive Distribution . . . . .	196
6.5	Simple Bayesian Model Averaging . . . . .	198
6.6	Concluding Comments on Model Quality . . . . .	200
6.7	Exercises . . . . .	202
<b>7</b>	<b>Bayesian Hypothesis Testing and the Bayes Factor</b>	<b>207</b>
7.1	Motivation . . . . .	207
7.2	Bayesian Inference and Hypothesis Testing . . . . .	209
7.2.1	Problems with Conventional Hypothesis Testing . . . . .	209
7.2.1.1	One-Sided Testing . . . . .	211
7.2.1.2	Two-Sided Testing . . . . .	214
7.2.2	Attempting a Bayesian Approximation to Frequentist Hypothesis Testing . . . . .	215
7.3	The Bayes Factor as Evidence . . . . .	216
7.3.1	Bayes Factors for a Mean . . . . .	218
7.3.2	Bayes Factors for Difference of Means Test . . . . .	219
7.3.3	Bayes Factor for the Linear Regression Model . . . . .	219
7.3.4	Bayes Factors and Improper Priors . . . . .	223
7.3.4.1	Local Bayes Factor . . . . .	224
7.3.4.2	Intrinsic Bayes Factor . . . . .	225
7.3.4.3	Partial Bayes Factor . . . . .	226
7.3.4.4	Fractional Bayes Factor . . . . .	226
7.3.4.5	Redux . . . . .	227
7.3.5	Two-Sided Hypothesis Tests and Bayes Factors . . . . .	228
7.3.6	Challenging Aspects of Bayes Factors . . . . .	229
7.4	The Bayesian Information Criterion (BIC) . . . . .	231
7.5	The Deviance Information Criterion (DIC) . . . . .	233
7.5.1	Some Qualifications . . . . .	236
7.6	Comparing Posterior Distributions with the Kullback-Leibler Distance . . . . .	237
7.7	Laplace Approximation of Bayesian Posterior Densities . . . . .	239
7.8	Exercises . . . . .	243

<b>8</b>	<b>Bayesian Decision Theory</b>	<b>247</b>
8.1	Introducing Decision Theory . . . . .	247
8.2	Basic Definitions . . . . .	249
8.2.1	Personal Preference . . . . .	250
8.2.2	Rules, Rules, Rules . . . . .	250
8.2.3	Lots of Loss . . . . .	251
8.2.4	Risky Business . . . . .	253
8.2.4.1	Notes on Bayes Rules . . . . .	256
8.2.5	Minimax Decision Rules . . . . .	258
8.3	Regression-Style Models with Decision Theory . . . . .	259
8.3.1	Prediction from the Linear Model . . . . .	261
8.4	James-Stein Estimation . . . . .	262
8.5	Empirical Bayes . . . . .	267
8.6	Exercises . . . . .	271
<b>9</b>	<b>Monte Carlo and Related Iterative Methods</b>	<b>275</b>
9.1	Background . . . . .	275
9.2	Basic Monte Carlo Integration . . . . .	277
9.3	Rejection Sampling . . . . .	280
9.3.1	Continuous Form with Bounded Support . . . . .	281
9.3.2	Continuous Form with Unbounded Support . . . . .	284
9.4	Classical Numerical Integration . . . . .	288
9.4.1	Newton-Cotes . . . . .	289
9.4.1.1	Riemann Integrals . . . . .	289
9.4.1.2	Trapezoid Rule . . . . .	289
9.4.1.3	Simpson's Rule . . . . .	290
9.5	Gaussian Quadrature . . . . .	292
9.5.1	Redux . . . . .	295
9.6	Importance Sampling and Sampling Importance Resampling . . . . .	296
9.6.1	Importance Sampling for Producing HPD Regions . . . . .	301
9.7	Mode Finding and the EM Algorithm . . . . .	302
9.7.1	Deriving the EM Algorithm . . . . .	304
9.7.2	Convergence of the EM Algorithm . . . . .	307
9.7.3	Extensions to the EM Algorithm . . . . .	313
9.7.4	Additional Comments on EM . . . . .	315
9.7.5	EM for Exponential Families . . . . .	316
9.8	Survey of Random Number Generation . . . . .	320
9.9	Concluding Remarks . . . . .	322
9.10	Exercises . . . . .	323
9.11	Computational Addendum: R Code for Importance Sampling . . . . .	328

<b>10 Basics of Markov Chain Monte Carlo</b>	<b>333</b>
10.1 Who Is Markov and What Is He Doing with Chains? . . . . .	333
10.1.1 What Is a Markov Chain? . . . . .	334
10.1.2 A Markov Chain Illustration . . . . .	335
10.1.3 The Chapman-Kolmogorov Equations . . . . .	338
10.1.4 Marginal Distributions . . . . .	339
10.2 General Properties of Markov Chains . . . . .	339
10.2.1 Homogeneity . . . . .	340
10.2.2 Irreducibility . . . . .	340
10.2.3 Recurrence . . . . .	340
10.2.4 Stationarity . . . . .	341
10.2.5 Ergodicity . . . . .	342
10.3 The Gibbs Sampler . . . . .	343
10.3.1 Description of the Algorithm . . . . .	343
10.3.2 Handling Missing Dichotomous Data with the Gibbs Sampler . . . .	345
10.3.3 Summary of Properties of the Gibbs Sampler . . . . .	353
10.4 The Metropolis-Hastings Algorithm . . . . .	353
10.4.1 Background . . . . .	354
10.4.2 Description of the Algorithm . . . . .	354
10.4.3 Metropolis-Hastings Properties . . . . .	356
10.4.4 Metropolis-Hastings Derivation . . . . .	356
10.4.5 The Transition Kernel . . . . .	358
10.4.6 Example: Estimating a Bivariate Normal Density . . . . .	359
10.5 The Hit-and-Run Algorithm . . . . .	360
10.6 The Data Augmentation Algorithm . . . . .	362
10.7 Historical Comments . . . . .	367
10.7.1 Full Circle? . . . . .	368
10.8 Exercises . . . . .	368
10.9 Computational Addendum: Simple R Graphing Routines for MCMC . . .	375
<b>11 Implementing Bayesian Models with Markov Chain Monte Carlo</b>	<b>377</b>
11.1 Introduction to Bayesian Software Solutions . . . . .	377
11.2 It's Only a Name: BUGS . . . . .	378
11.3 Model Specification with BUGS . . . . .	380
11.3.1 Model Specification . . . . .	383
11.3.2 Running the Model in WinBUGS . . . . .	385
11.3.3 Running the Model in JAGS . . . . .	388
11.4 Differences between WinBUGS and JAGS Code . . . . .	392
11.5 Technical Background about the Algorithm . . . . .	401
11.6 Epilogue . . . . .	408
11.7 Exercises . . . . .	408



<b>12 Bayesian Hierarchical Models</b>	<b>413</b>
12.1 Introduction to Multilevel Specifications . . . . .	413
12.2 Basic Multilevel Linear Models . . . . .	414
12.3 Comparing Variances . . . . .	416
12.4 Exchangeability . . . . .	420
12.5 Essential Structure of the Bayesian Hierarchical Model . . . . .	425
12.5.1 A Poisson-Gamma Hierarchical Specification . . . . .	427
12.6 The General Role of Priors and Hyperpriors . . . . .	434
12.7 Bayesian Multilevel Linear Regression Models . . . . .	436
12.7.1 The Bayesian Hierarchical Linear Model of Lindley and Smith . . .	436
12.8 Bayesian Multilevel Generalized Linear Regression Models . . . . .	441
12.9 Exercises . . . . .	446
12.10 Computational Addendum . . . . .	451
12.10.1 R Function for importing BUGS output . . . . .	451
<b>13 Some Markov Chain Monte Carlo Theory</b>	<b>453</b>
13.1 Motivation . . . . .	453
13.2 Measure and Probability Preliminaries . . . . .	453
13.3 Specific Markov Chain Properties . . . . .	455
13.3.1 $\psi$ -Irreducibility . . . . .	455
13.3.2 Closed and Absorbing Sets . . . . .	455
13.3.3 Homogeneity and Periodicity . . . . .	455
13.3.4 Null and Positive Recurrence . . . . .	456
13.3.5 Transience . . . . .	456
13.3.6 Markov Chain Stability . . . . .	457
13.3.7 Ergodicity . . . . .	458
13.4 Defining and Reaching Convergence . . . . .	458
13.5 Rates of Convergence . . . . .	460
13.6 Implementation Concerns . . . . .	464
13.6.1 Mixing . . . . .	466
13.6.2 Partial Convergence for Metropolis-Hastings . . . . .	467
13.6.3 Partial Convergence for the Gibbs Sampler . . . . .	469
13.7 Exercises . . . . .	471
<b>14 Utilitarian Markov Chain Monte Carlo</b>	<b>475</b>
14.1 Objectives . . . . .	475
14.2 Practical Considerations and Admonitions . . . . .	476
14.2.1 Starting Points . . . . .	476
14.2.2 Thinning the Chain . . . . .	477
14.2.3 The Burn-In Period . . . . .	478
14.3 Assessing Convergence of Markov Chains . . . . .	479

14.3.1	Autocorrelation . . . . .	485
14.3.2	Graphical Diagnostics . . . . .	487
14.3.3	Standard Empirical Diagnostics . . . . .	493
14.3.3.1	The Geweke Time-Series Diagnostic . . . . .	494
14.3.3.2	Gelman and Rubin's Multiple Sequence Diagnostic . . . . .	496
14.3.3.3	The Heidelberger and Welch Diagnostic . . . . .	499
14.3.3.4	The Raftery and Lewis Integrated Diagnostic . . . . .	503
14.3.4	Summary of Diagnostic Similarities and Differences . . . . .	505
14.3.5	Other Empirical Diagnostics . . . . .	507
14.3.6	Why Not to Worry Too Much about Stationarity . . . . .	509
14.4	Mixing and Acceleration . . . . .	510
14.4.1	Reparameterization . . . . .	510
14.4.2	Grouping and Collapsing the Gibbs Sampler . . . . .	512
14.4.3	Adding Auxiliary Variables . . . . .	513
14.4.4	The Slice Sampler . . . . .	513
14.5	Chib's Method for Calculating the Marginal Likelihood Integral . . . . .	515
14.6	Rao-Blackwellizing for Improved Variance Estimation . . . . .	517
14.7	Exercises . . . . .	520
14.8	Computational Addendum: Code for Chapter Examples . . . . .	523
14.8.1	R Code for the Death Penalty Support Model . . . . .	523
14.8.2	JAGS Code for the Military Personnel Model . . . . .	524
<b>15</b>	<b>Markov Chain Monte Carlo Extensions</b>	<b>525</b>
15.1	Simulated Annealing . . . . .	525
15.1.1	General Points on Simulated Annealing . . . . .	529
15.1.2	Metropolis-Coupling . . . . .	530
15.1.3	Simulated Tempering . . . . .	531
15.1.4	Tempored Transitions . . . . .	532
15.1.5	Comparison of Algorithms . . . . .	533
15.1.6	Dynamic Tempered Transitions . . . . .	535
15.2	Reversible Jump Algorithms . . . . .	536
15.3	Perfect Sampling . . . . .	538
15.4	Hamiltonian Monte Carlo . . . . .	542
15.5	Exercises . . . . .	547
<b>Appendix A</b>	<b>Generalized Linear Model Review</b>	<b>553</b>
A.1	Terms . . . . .	553
A.1.1	The Linear Regression Model . . . . .	555
A.2	The Generalized Linear Model . . . . .	557
A.2.1	Defining the Link Function . . . . .	558
A.2.2	Deviance Residuals . . . . .	560

A.3	Numerical Maximum Likelihood . . . . .	564
A.3.1	Newton-Raphson and Root Finding . . . . .	564
A.3.1.1	Newton-Raphson for Statistical Problems . . . . .	566
A.3.1.2	Weighted Least Squares . . . . .	566
A.3.1.3	Iterative Weighted Least Squares . . . . .	567
A.4	Quasi-Likelihood . . . . .	568
A.5	Exercises . . . . .	572
<b>Appendix B</b>	<b>Common Probability Distributions</b>	<b>579</b>
<b>References</b>		<b>583</b>



---

## List of Figures

1.1	GIBBS SAMPLING, MARGINAL EXPONENTIALS . . . . .	27
1.2	METROPOLIS SAMPLING, BIVARIATE EXPONENTIAL . . . . .	31
2.1	POSTERIOR $\propto$ PRIOR $\times$ LIKELIHOOD . . . . .	41
2.2	CI FOR STATE DURATION TIME TO ADAPTATION, 1998-2005 . . . . .	44
2.3	BIMODAL DISTRIBUTION HIGHEST POSTERIOR DENSITY INTERVAL . . . . .	46
2.4	HPD REGION, STATE DURATION TIME TO ADAPTATION, 1998-2005 . . . . .	48
2.5	PRIOR AND POSTERIOR DISTRIBUTIONS IN THE BETA-BINOMIAL MODEL . . . . .	52
3.1	A MENAGERIE OF INVERSE GAMMA FORMS . . . . .	73
3.2	POSTERIOR DISTRIBUTION VERSUS LIKELIHOOD FUNCTION . . . . .	80
4.1	PRIOR FROM TRANSFORMATION . . . . .	106
4.2	BINOMIAL: JEFFREYS PRIOR . . . . .	109
4.3	FINDING THE BETA PARAMETERS . . . . .	125
4.4	$t$ RATIOS . . . . .	126
4.5	SPIKE AND SLAB PRIOR . . . . .	130
5.1	MCMC RUNNING MEANS, CHINESE CONFLICTS . . . . .	164
6.1	OUTCOME COMPARISON: OBSERVED VERSUS SIMULATED . . . . .	180
6.2	MARGINAL POSTERiors, PALM BEACH COUNTY MODEL . . . . .	184
6.3	EXPONENTIAL MODEL FOR $\theta$ WITH GAMMA PRIOR . . . . .	186
6.4	EXPONENTIAL MODEL SENSITIVITY . . . . .	187
6.5	COMPARISON OF POSTERiors, RECIDIVISM MODEL . . . . .	190
6.6	POSTERIOR PREDICTIVE DISTRIBUTION DRAWS, AFRICA GROWTH . . . . .	199
7.1	ONE-SIDED TESTING FOR THE FRENCH STRIKES DATA . . . . .	214
8.1	EMPIRICAL BAYES ESTIMATES VERSUS DATA VALUES . . . . .	270
9.1	REJECTION SAMPLING FOR BOUNDED FORMS . . . . .	283
9.2	REJECTION SAMPLING FOR UNBOUNDED FORMS . . . . .	285
9.3	REJECTION SAMPLING, OPTIMAL COVERAGE . . . . .	286
9.4	FOLDED NORMAL REJECTION SAMPLING EXAMPLE . . . . .	288

9.5	ILLUSTRATION OF NUMERICAL INTEGRATION . . . . .	292
9.6	CONTOURPLOT MATRIX, RURAL POVERTY IN TEXAS . . . . .	300
9.7	EM DEMONSTRATION: U.S. ENERGY CONSUMPTION . . . . .	313
10.1	GIBBS SAMPLING DEMONSTRATION . . . . .	346
10.2	GIBBS SAMPLING PATH FOR TERRORISM EVENTS . . . . .	351
10.3	AFGHAN FATALITIES POSTERIOR HISTOGRAM, HPD REGION . . . . .	352
10.4	METROPOLIS-HASTINGS, BIVARIATE NORMAL SIMULATION . . . . .	360
10.5	HIT-AND-RUN ALGORITHM DEMONSTRATION . . . . .	362
11.1	OECD EMPLOYMENT AND PRODUCTIVITY DATA . . . . .	382
11.2	ADAPTIVE REJECTION SAMPLING STEP . . . . .	403
11.3	ADAPTIVE REJECTION SAMPLING SQUEEZING . . . . .	404
12.1	POISSON-GAMMA HIERARCHICAL MODEL . . . . .	428
12.2	TRACEPLOT OF OUTPUT FOR GAMMA . . . . .	430
12.3	SMOOTHED POSTERIORES FOR $\alpha$ AND $\beta$ . . . . .	433
12.4	LAMBDA POSTERIOR SUMMARY, ITALY MODEL . . . . .	434
12.5	RUNNING MEANS, ITALY MODEL . . . . .	435
12.6	CORRELATION DISTRIBUTIONS FROM $W(\mathbf{\Sigma} \alpha, \beta)$ . . . . .	439
14.1	TRACEPLOT AND HISTOGRAM GRAPHICS, TOBIT MODEL . . . . .	489
14.2	TRACEPLOT AND HISTOGRAM GRAPHICS, MILITARY PERSONNEL MODEL . . . . .	490
14.3	TRACEPLOTS OF THE SAME CHAIN WITH DIFFERENT SPANS . . . . .	491
14.4	RUNNING MEAN DIAGNOSTIC, TOBIT MODEL . . . . .	492
14.5	RUNNING MEAN DIAGNOSTIC, MILITARY PERSONNEL MODEL . . . . .	493
14.6	TRACEPLOT AND HISTOGRAM, LAST 2,000 ITERATIONS . . . . .	502
14.7	A COMPARISON OF GIBBS SAMPLING FOR NORMALS . . . . .	511
14.8	SLICE SAMPLER OUTPUT . . . . .	515
15.1	SIMULATED ANNEALING APPLIED TO THE WITCH'S HAT DISTRIBUTION . . . . .	531
15.2	A HIGHLY MULTIMODAL BOUNDED SURFACE . . . . .	534
15.3	A COMPARISON OF TRAVELS . . . . .	535
15.4	SAMPLES FROM HAMILTONIAN MONTE CARLO . . . . .	547

---

## List of Tables

2.1	STATE DURATION TIME TO ADAPTATION, 1998-2005 . . . . .	45
2.2	HPD REGIONS: PREDICTED 1982 VOTE PERCENTAGES FOR FIVE MAJOR PARTIES . . . . .	60
3.1	POSTERIOR SUMMARIES, NORMAL MODELS . . . . .	79
3.2	A BIVARIATE NORMAL, 1000 SIMULATIONS . . . . .	81
3.3	ANOTHER BIVARIATE NORMAL, 1000 SIMULATIONS . . . . .	82
3.4	POSTERIOR SUMMARY: NATIONAL IQ SCORES . . . . .	86
4.1	SOME EXPONENTIAL FAMILY FORMS AND THEIR CONJUGATE PRIORS . .	103
4.2	SOME JEFFREYS PRIORS FOR COMMON FORMS . . . . .	110
5.1	PBC DATA VARIABLE DEFINITIONS . . . . .	149
5.2	POSTERIOR: LINEAR MODEL, UNINFORMATIVE PRIORS . . . . .	150
5.3	LINEAR REGRESSION MODEL, PRIOR COMPARISON . . . . .	153
5.4	POSTERIOR: LINEAR MODEL, CONJUGATE PRIORS . . . . .	155
5.5	POSTERIOR: MEIER REPLICATION MODEL . . . . .	159
5.6	POSTERIOR: INTERACTION MODEL . . . . .	160
5.7	HETEROSCEDASTIC MODEL, ANCIENT CHINESE CONFLICTS . . . . .	165
6.1	MODEL SUMMARY, ABORTION ATTITUDES IN BRITAIN . . . . .	179
6.2	RECIDIVISM BY CRIME, OKLAHOMA, 1/1/85 TO 6/30/99 . . . . .	188
6.3	$\lambda$ POSTERIOR SUMMARY, RECIDIVISM MODELS . . . . .	189
6.4	OKLAHOMA RECIDIVISM 95% HPD REGIONS, 1/1/85 TO 6/30/99 . . .	192
6.5	SUB-SAHARAN AFRICA AVERAGE ECONOMIC GROWTH RATE, 1990-1994	198
7.1	FRENCH COAL STRIKES, BY YEAR . . . . .	213
7.2	1964 ELECTORAL DATA . . . . .	221
7.3	1964 ELECTION SURVEY, GROUPED DATA . . . . .	222
7.4	AVAILABLE MODEL SPECIFICATIONS, THAI MIGRATION . . . . .	235
7.5	MODEL COMPARISON, THAI MIGRATION . . . . .	236
7.6	MARTIKAINEN <i>et al.</i> (2005) MORTALITY PARAMETERS . . . . .	238
7.7	MARTIKAINEN <i>et al.</i> (2005) MORTALITY DIFFERENCES . . . . .	239
8.1	PREDICTION RISK: EDUCATION OUTCOMES MODEL . . . . .	262

8.2	MLE AND JAMES-STEIN ESTIMATES OF SOCIAL SPENDING . . . . .	266
9.1	1990-1993 W.EUROPE ETHNIC/MINORITY POPULATIONS . . . . .	280
9.2	MONTE CARLO QUANTILES FOR PARETO PARAMETER $\alpha$ . . . . .	281
9.3	LOGIT REGRESSION MODEL: POVERTY IN TEXAS . . . . .	299
9.4	EM OUTPUT: U.S. ENERGY CONSUMPTION . . . . .	311
9.5	MODEL RESULTS, MORE DEVELOPED AREAS . . . . .	320
10.1	COUNT OF TERRORISM INCIDENTS IN BRITAIN, 1970-2004 BY ROWS . .	347
10.2	GIBBS SAMPLER DRAWS FOR $k$ PARAMETER . . . . .	349
10.3	TERRORISM IN GREAT BRITAIN MODEL, POSTERIOR SUMMARY . . . . .	350
10.4	NATO FATALITIES IN AFGHANISTAN, 10/01 TO 1/07 . . . . .	350
10.5	AFGHAN FATALITIES POSTERIOR . . . . .	352
11.1	OECD PROTECTION VERSUS PRODUCTIVITY . . . . .	381
11.2	OECD MODEL RESULTS . . . . .	389
11.3	LOBBYING INFLUENCE RESULTS . . . . .	399
11.4	PRESIDENTIAL ELECTION MODEL RESULTS . . . . .	407
12.1	DEPRESSION TREATMENT BY CLINIC, CONTROL GROUP . . . . .	424
12.2	ITALIAN MARRIAGE RATES PER 1,000, 1936-1951 . . . . .	430
12.3	POSTERIOR SUMMARY STATISTICS, ITALY MODEL . . . . .	432
12.4	POSTERIOR SUMMARY STATISTICS, ECONOMIC INDICATORS MODEL . . .	441
12.5	POSTERIOR SUMMARY STATISTICS, DRUG USE BY ARRESTEES MODEL .	444
12.6	CONTRACEPTION DATA, DEVELOPING COUNTRIES (BY SIZE) . . . . .	446
12.7	RACIAL COMPOSITION OF U.S. FEDERAL AGENCIES (1998) . . . . .	450
14.1	POSTERIOR, TOBIT MODEL . . . . .	482
14.2	PROPORTIONAL CHANGES IN EASTERN EUROPEAN MILITARIES, 1948- 1983 . . . . .	483
14.3	POSTERIOR, MILITARY PERSONNEL MODEL . . . . .	485
14.4	CORRELATIONS AND AUTOCORRELATIONS, TOBIT MODEL . . . . .	487
14.5	CORRELATIONS AND AUTOCORRELATIONS, MILITARY PERSONNEL MODEL	488
14.6	GEWEKE DIAGNOSTIC, TOBIT MODEL . . . . .	495
14.7	GEWEKE DIAGNOSTIC, MILITARY PERSONNEL MODEL . . . . .	495
14.8	GELMAN AND RUBIN DIAGNOSTIC, TOBIT MODEL . . . . .	498
14.9	GELMAN AND RUBIN DIAGNOSTIC, MILITARY PERSONNEL MODEL . . . .	499
14.10	DIAGNOSTIC SCORECARD, TOBIT MODEL . . . . .	506
14.11	DIAGNOSTIC SCORECARD, MILITARY PERSONNEL MODEL . . . . .	507
A.1	NATURAL LINK FUNCTIONS FOR COMMON SPECIFICATIONS . . . . .	559
A.2	SUMMED DEVIANCES FOR COMMON SPECIFICATIONS . . . . .	561
A.3	POISSON MODEL OF MILITARY COUPS . . . . .	563



A.4	NORMALIZING CONSTANTS AND VARIANCE FUNCTIONS . . . . .	570
-----	--	-----



---

# *Preface to the Third Edition*

---

## **General Comments**

Welcome to the third edition of BMSBSA. When the first edition appeared in 2002 Bayesian methods were still considered a bit exotic in the social sciences. Many distrusted the use of prior distributions and some mysterious simulation process that involved non-iid sampling and uncertainty about convergence. The world is completely different now, and Bayesian modeling has become standard and MCMC is well-understood and trusted. Of course it helps that Moore's Law (doubling of computing power every two years, presumably until we reach the 7 nanometer threshold) continues without pause making our computers notably faster and allowing longer sampling procedures and parallel process without an agonizingly long wait. In this context the third edition spends less time justifying procedures and more time providing implementation details on these procedures. This is also an opportunity to expand the set of examples.

---

## **Changes from the Second Edition**

As expected there are a number of additions in the new edition. First, and most laboriously, the number of exercises has been doubled such that there are now twenty in each chapter. The former exercises are now the odd-numbered exercises with the answer key being fully publicly distributed (not just to instructors). The new exercises emphasize recent developments in Bayesian inference and Bayesian computing as a way to include more modern material. All of the chapters have been refreshed, although some more than others. The basic material has not changed in over a century, so there is no need to dramatically alter basic material. Conversely, Bayesian stochastic simulation (MCMC) has undergone dramatic developments in the last decade, including having become routine in applied settings. New MCMC material includes Hamiltonian Monte Carlo and expanded model implementation. Second, there are two new chapters. A chapter on Bayesian decision theory is long overdue, and this is now Chapter 8. It includes discussion of both Bayesian and frequentist decision theory since this is where the two paradigms are most intertwined. Included topics are: loss

functions, risk, decision rules, and regression-model applications. The chapter finishes with two important topics not featured in previous editions: James-Stein estimation and empirical Bayes. While empirical Bayes was briefly covered in the second edition, its connection with James-Stein estimation was ignored. This section now covers the important topics in this area and provides Bayesian context. Also new is a chapter on practical implementation of MCMC methods (Chapter 11). This covers mechanical issues with the BUGS language, including calling the software from R. The goal is to provide a detailed introduction to the essential software for running Bayesian hierarchical regression models. Relatedly the chapter on hierarchical models is greatly expanded. This is an area of great applied interest right now and provides a strong motivation for the Bayesian paradigm. Finally, on a more practical side, there is a wealth of new examples and applications. These are chosen from a variety of social science disciplines and are intended to illustrate the key principles of the relevant chapter. In addition, the BaM package in R that accompanies this manuscript has been greatly expanded with new datasets and new code. This includes new procedures for calling BUGS packages from R.

---

## Course Plans

The recommended course plans remain essentially the same as outlined in the preface to the second edition. The one critical difference is adding Chapter 11 (Implementing Bayesian Models with Markov Chain Monte Carlo) to a basic course or comprehensive course. The longer length of the text means that not all chapters are practical in a one-semester course. For a standard introductory Bayesian social science graduate course, the most succinct set of chapters are:

- ▷ Chapter 1: Background and Introduction
- ▷ Chapter 2: Specifying Bayesian Models
- ▷ Chapter 3: The Normal and Students'- $t$  Models
- ▷ Chapter 4: The Bayesian Prior
- ▷ Chapter 5: The Bayesian Linear Model
- ▷ Chapter 10: Basics of Markov Chain Monte Carlo
- ▷ Chapter 11: Implementing Bayesian Models with Markov Chain Monte Carlo
- ▷ Chapter 12: Bayesian Hierarchical Models
- ▷ Chapter 14: Utilitarian Markov Chain Monte Carlo.

This assumes some knowledge of basic Monte Carlo methods of the students. Chapter 11 and Chapter 14 could also be assigned reading rather than part of lectures since the focus is on very practical concerns.

---

## Support

As done in the last two editions of this text, there is a dedicated website provided to support readers: <http://stats.wustl.edu/BMSBSA3>. This site has software, errata, comments, and the answer key for odd-numbered exercises. All of the code is also provided in the associated R package, **BaM**, which has been substantially updated to include new code and data. Where possible **BUGS** code is included in this package. Note that in many cases the code relies on multiple R packages, as well as stand-alone software such as **JAGS** and **WinBUGS**, so changes over time may introduce incompatibilities that need to be worked out. In many cases this amounts to downloading the most recent version of some software. Relevant updates will be posted at the dedicated webpage when they come to my attention.

---

## Acknowledgments

Since this edition has been seven years in the making, there are many people to thank. First, I appreciate the support from the Department of Political Science, the Division of Biostatistics, and the Department of Surgery (Public Health Sciences) at Washington University for being supportive home environments. This includes support from the Transdisciplinary Research on Energetics and Cancer (TREC) grant and Graham Colditz and Sarah Gehlert. I also thank the Summer School in Social Science Data Analysis at The University of Essex leadership, Thomas Plümper and Vera Troeger, since some of the datasets and software were developed in the process of teaching there. In particular, new ideas for the chapter on hierarchical models were created in the welcoming community that is Essex. The manuscript was finished while on sabbatical Spring 2014 at the University of Minnesota Division of Biostatistics (with additional financial support from the Departments of Political Science, Psychology, Sociology, and Statistics). I thank Brad Carlin, Sudipto Banerjee, John Freeman, and Niels Waller for making this happen and providing a pleasant (but really cold!) environment to wrap up this project. Finally, I acknowledge the support of NSF grants: DMS-0631632 and SES-0631588, which are my last in cooperation with George Casella. Numerous people have sent useful comments, notably Gary McDonald, Kentaro Fukumoto, Ben Begozzi, Patrick Brandt, Yuta Kamahara, Bruce Desmarais, Skyler Cranmer, Ryan Bakker, and Gary in particular was relentless in his comments (in a good way!). In the Spring of 2011 twelve graduate students at Washington University participated in a weekly reading group around the manuscript giving useful and insightful comments on methods, computing, and even the writing. I thank Peter Casey, Adriana Crespo-Tenorio, Constanza Figueroa Schibber, Morgan Hazelton, Rachael Hinkley, Chia-yi Lee, Michael Nelson, Christopher Pope, Elizabeth Rose, and Alicia Uribe. All of these stu-

dents have finished their studies and gone on to successful careers. This is the first edition of this book after George Casella left us. He made a big impact on the first two issues as my neighbor, close friend, colleague, and coauthor. As Christian Robert states, George simply made everyone's work better for those fortunate to be around him. This work is certainly better because of George's influence.

---

# *Preface to the Second Edition*

---

## Starters

Wow, over five years have elapsed since the first edition appeared. Bayesian methods in the social sciences have grown and changed dramatically. This is a positive and encouraging development. When I was writing the first version I would often get questions from social science colleagues about why I would write on a seemingly obscure branch of statistics. This is clearly no longer an issue and Bayesian approaches appear to have a prominent role in social science methodology. I hope that the first edition contributed to this development.

Bayesian methods continue to become more important and central to statistical analysis, broadly speaking. Seemingly, no issue of the *Journal of the American Statistical Association* arrives without at least one Bayesian application or theoretical development. While this upward trend started in the 1990s after we discovered Markov chain Monte Carlo hiding in statistical physics, the trend accelerates in the 21st century. A nice foretelling is found in the 1999 *Science* article by David Malakoff, “Bayesian Boom,” complete with anecdotes about popular uses in biology and computing as well as quotes from John Geweke. Back in 1995, the Bayesian luminary Bruno de Finetti predicted that by the year 2020 we would see a paradigm shift to Bayesian thinking (quoted in Smith [1995]). I believe we are fully on track to meet this schedule.

Bayesian computing is broader and more varied than it was at the writing of the first edition. In addition to BUGS and WinBUGS, we now routinely use MCMCpack, JAGS, openbugs, bayesm, and even the new SAS MCMC procedure. The diagnostic routines in R, BOA, and CODA continue to be useful and are more stable than they were. Of the course the lingua franca of R is critical, and many researchers use C or C++ for efficiency. Issues of statistical computing remain an important component of the book. It is also necessary to download and use the R packages CODA and BOA for MCMC diagnostics.

Bayesian approaches are also increasingly popular in related fields not directly addressed in this text. There is now an interesting literature in archaeology that is enjoyable to read (Reese 1994, Freeman 1976, Laxton *et al.* 1994), and the best starting point is the seminal paper by Litton and Buck (1995) that sets the agenda for Bayesian archaeometrics. Researchers in this area have also become frustrated with the pathetic state of the null hypothesis significance test in the social and behavioral sciences (Cowgill 1977). One area where

Bayesian modeling is particularly useful in archaeological forensics, where researchers make adult-age estimates of early humans (Lucy *et al.* 1996, Aykroyd *et al.* 1999).

---

## Changes from the First Edition

A reader of the first edition will notice many changes in this revision. Hopefully these constitute improvements (they certainly constituted a lot of work). First, the coverage of Markov chain Monte Carlo is greatly expanded. The reason for this is obvious, but bears mentioning. Modern applied Bayesian work is integrally tied to stochastic simulation and there are now several high-quality software alternatives for implementation. Unfortunately these solutions can be complex and the theoretical issues are often demanding. Coupling this with easy-to-use software, such as **WinBUGS** and **MCMCpack**, means that there are users who are unaware of the dangers inherent in MCMC work. I get a fair number of journal and book press manuscripts to review supporting this point. There is now a dedicated chapter on MCMC theory covering issues like ergodicity, convergence, and mixing. The last chapter is an extension of sections from the first edition that now covers in greater detail tools like: simulated annealing (including its many variants), reversible jump MCMC, and coupling from the past. Markov chain Monte Carlo research is an incredibly dynamic and fast growing literature and the need to get some of these ideas before a social science audience was strong. The reader will also note a substantial increase on MCMC examples and practical guidance. The objective is to provide detailed advice on day-to-day issues of implementation. Markov chain Monte Carlo is now discussed in detail in the first chapter, giving it the prominent position that it deserves. It is my belief that Gibbs sampling is as fundamental to estimation as maximum likelihood, but we (collectively) just do not realize it yet. Recall that there was about 40 years between Fisher's important papers and the publication of Birnbaum's Likelihood Principle. This second edition now provides a separate chapter on Bayesian linear models. Regression remains the favorite tool of quantitative social scientists, and it makes sense to focus on the associated Bayesian issues in a full chapter. Most of the questions I get by email and at conferences are about priors, reflecting sensitivity about how priors may affect final inferences. Hence, the chapter on forms of prior distributions is longer and more detailed. I have found that some forms are particularly well-suited to the type of work that social and behavioral researchers do. One of the strengths of Bayesian methods is the ease with which hierarchical models can be specified to recognize different levels and sources in the data. So there is now an expanded chapter on this topic alone, and while Chapter 12 focuses exclusively on hierarchical model specifications, these models appear throughout the text reflecting their importance in Bayesian statistics.

Additional topics have crept into this edition, and these are covered at varied levels from a basic introduction to detailed discussions. Some of these topics are older and well-



known, such as Bayesian time-series, empirical Bayes, Bayesian decision theory, additional prior specifications, model checking with posterior data prediction, the deviance information criterion (DIC), methods for computing highest posterior density (HPD) intervals, convergence theory, metropolis-coupling, tempering, reversible jump MCMC, perfect sampling, software packages related to BUGS, and additional models based on normal and Student's-t assumptions.

Some new features are more structural. There is now a dedicated R package to accompany this book, **BaM** (for “**B**ayesian **M**ethods”). This package includes data and code for the examples as well as a set of functions for practical purposes like calculated HPD intervals. These materials and more associated with the book are available at the dedicated Washington University website: <http://stats.wustl.edu/BMSBSA>. The second edition includes three appendices covering basic maximum likelihood theory, distributions, and BUGS software. These were moved to separate sections to make referencing easier and to preserve the flow of theoretical discussions. References are now contained in a single bibliography at the end for similar reasons. Some changes are more subtle. I've changed all instances of “noninformative” to “uninformative” since the first term does not really describe prior distributions. Markov chain Monte Carlo techniques are infused throughout, befitting their central role in Bayesian work. Experience has been that social science graduate students remain fairly tepid about empirical examples that focus on rats, lizards, beetles, and nuclear pumps. Furthermore, as of this writing there is no other comprehensive Bayesian text in the social sciences, outside of economics (except the out-of-print text by Phillips [1973]).

---

## Road Map

To begin, the prerequisites remain the same. Readers will need to have a basic working knowledge of linear algebra and calculus to follow many of the sections. My math text, *Essential Mathematics for Political and Social Research* (2006), provides an overview of such material. Chapter 1 gives a brief review of the probability basics required here, but it is certainly helpful to have studied this material before. Finally, one cannot understand Bayesian modeling without knowledge of maximum likelihood theory. I recognize graduate programs differ in their emphasis on this core material, so Appendix A covers these essential ideas.

The second edition is constructed in a somewhat different fashion than the first. The most obvious difference is that the chapter on generalized linear models has been recast as an appendix, as mentioned. Now the introductory material flows directly into the construction of basic Bayesian statistical models and the procession of core ideas is not interrupted by a non-Bayesian discussion of standard models. Nonetheless, this material is important to have close at hand and hopefully the appendix approach is convenient. Another notable

change is the “promotion” of linear models to their own chapter. This material is important enough to stand on its own despite the overlap with Bayesian normal and Student’s-t models. Other organization changes are found in the computational section where considerable extra material has been added, both in terms of theory and practice. Markov chain Monte Carlo set the Bayesians free, and remains an extremely active research field. Keeping up with this literature is a time-consuming, but enjoyable, avocation.

There are a number of ways that a graduate course could be structured around this text. For a basic-level introductory course that emphasizes theoretical ideas, the first seven chapters provide a detailed overview without considering many computational challenges. Some of the latter chapters are directed squarely at sophisticated social scientists who have not yet explored some of the subtle theory of Markov chains. Among the possible structures, consider the following curricula.

### **Basic Introductory Course**

- ▷ Chapter 1: **Background and Introduction**
- ▷ Chapter 2: **Specifying Bayesian Models**
- ▷ Chapter 3: **The Normal and Student’s-t Models**
- ▷ Chapter 5: **The Bayesian Linear Model**
- ▷ Chapter 12: **Bayesian Hierarchical Models**

### **Thorough Course without an Emphasis on Computing**

- ▷ Chapter 1: **Background and Introduction**
- ▷ Chapter 2: **Specifying Bayesian Models**
- ▷ Chapter 3: **The Normal and Student’s-t Models**
- ▷ Chapter 5: **The Bayesian Linear Model**
- ▷ Chapter 4: **The Bayesian Prior**
- ▷ Chapter 6: **Assessing Model Quality**
- ▷ Chapter 7: **Bayesian Hypothesis Testing and the Bayes Factor**
- ▷ Chapter 12: **Bayesian Hierarchical Models**

### **A Component of a Statistical Computing Course**

- ▷ Chapter 2: **Specifying Bayesian Models**
- ▷ Chapter 9: **Monte Carlo and Related Iterative Methods**
- ▷ Chapter 10: **Basics of Markov Chain Monte Carlo**
- ▷ Chapter 13: **Some Markov Chain Monte Carlo Theory**

- ▷ Chapter 14: **Utilitarian Markov Chain Monte Carlo**
- ▷ Chapter 15: **Markov Chain Monte Carlo Extensions**

### A Component of an Estimation Course

- ▷ A: **Generalized Linear Model Review**
- ▷ Chapter 1: **Background and Introduction**
- ▷ Chapter 2: **Specifying Bayesian Models**
- ▷ Chapter 5: **The Bayesian Linear Model**
- ▷ Chapter 7: **Bayesian Hypothesis Testing and the Bayes Factor**

Of course I am eager to learn about how instructors use these chapters independent of any advice here.

---

## Acknowledgments

So many people have commented on this edition, the previous edition, related papers, associated conference presentations, and classes taught from the book that I am unlikely to remember them all. Apologies to anyone left out from this list. This edition received three detailed, formal reviews from Patrick Brandt, Andrew Gelman, and Andrew Martin. Their comments were invaluable and dramatically improved the quality of this work.

A substantial part of the writing of this book was done during the 2006-2007 academic year while I was Visiting Professor at Harvard University's Department of Government and Fellow at the Institute for Quantitative Social Science. I thank Gary King for inviting me into that dynamic and rewarding intellectual environment. The Fall semester of that year I taught a graduate seminar titled *Bayesian Hierarchical Modeling*, which enabled me to produce and distribute chapter material on a weekly basis. Participants in the seminar provided excellent critiques of the principles and exposition. These students and guests included: Justin Grimmer, Jonathan Harris, Jennifer Katkin, Elena Llaudet, Clayton Nall, Emre Ozaltin, Lindsay Page, Omar Wasow, Lefteris Anastasopoulos, Sheldon Bond, Janet Lewis, Serban Tanasa, and Lynda Zhang. The Teaching Fellows for this seminar were Skyler Cranmer and Andrew Thomas, who were instrumental in improving various technical sections. I also thank Jens Hainmueller, Dominik Hangartner, and Holger Lutz Kern for productive discussions of statistical computing during the Spring semester of that year.

Since 2000 I have taught a course based on this book at the Inter-University Consortium for Political and Social Research (ICPSR) Summer Program at the University of Michigan. Many of the highly motivated students in this program had constructive comments on the

material. I also benefited immensely from interactions with colleagues and administrators in the program, including Bob Andersen, David Armstrong, Ken Bollen, Dieter Burrell, John Fox, Charles Franklin, Hank Heitowit, Bill Jacoby, Jim Johnson, Dean Lacy, Scott Long, Jim Lynch, Tim McDaniel, Sandy Schneider, Bob Stine, and Lee Walker. The three teaching assistants over this period were incredibly helpful in developing homework and computer lab assignments: Ryan Bakker, Leslie Johns, and Yu-Sung Su. Overall, ICPSR is unmatched as a high-intensity teaching and learning Summer experience.

I have tried to record those making comments on the first edition or the manuscript version of the second edition. In addition to those already mentioned, useful critiques came from: Attic Access, Larry Bartels, Neal Beck, Jack Buckley, Sid Chib, Boyd Collier, Skyler J. Cranmer, Chris Dawes, Daniel J. Denis, Persi Diaconis, Alexis Dinno, Hanni Doss, George Duncan, James Fowler, Justin Gross, Josue Guzman, Michael Herrmann, Jim Hobert, Kosuke Imai, Brad Jones, Lucas Leemann, Jane Li, Rod Little, John Londregan, Enrico Luparini, Jonathan Nagler, Shunsuke Narita, Keith T. Poole, Kevin Quinn, Rafal Raciborski, Michael Smithson, John Sprague, Rick Waterman, Bruce Western, and Chris Zorn. I also want to give a special thank you to my friend and coauthor George Casella who has provided irreplaceable brotherly guidance over the last eight years.

The research on Dirichlet process priors (Chapter 15) was supported by National Science Foundation grants DMS-0631632 and SES-0631588. My work on elicited priors (Chapter 4) was helped by research with Lee Walker that appeared in the *Journal of Politics*. The discussion of dynamic tempered transitions (Chapter 15) draws from an article written with George Casella that appeared in *Political Analysis*. Comments from the editors, Bill Jacoby and Bob Erikson, made these works better and more focused. The education policy example (Chapter 5) using a Bayesian linear model is drawn from an article I wrote with my former graduate student Kevin Wagner. The discussion of convergence in Chapter 13 benefited from my recent article on this topic in *Political Analysis*. Finally, while there is no direct overlap, my understanding of detailed statistical computing principles benefited from the book project with Micah Altman and Michael McDonald.

---

## References

- Aykroyd, Robert G., Lucy, David, Pollard, Mark A., and Roberts, Charlotte A. (1999). Nasty, Brutish, But Not Necessarily Short: A Reconsideration of the Statistical Methods Used to Calculate Age At Death From Adult Human Skeletal and Dental Age Indicators. *American Antiquity* **64**, 55-70.
- Cowgill, G. L. (1977). The Trouble With Significance Tests and What We Can Do About It. *Philosophical Transactions of the Royal Society* **327**, 331-338.
- Freeman, P. R. (1976). A Bayesian Approach to the Megalithic Yard. *Journal of the Royal Statistical Association, Series A* **139**, 279-295.

- Gill, Jeff. (2006). *Essential Mathematics for Political and Social Research*. Cambridge, England: Cambridge University Press.
- Laxton, R. R., Cavanaugh, W. G., Litton, C. D., Buck, C. E., and Blair, R. (1994). The Bayesian Approach to Archaeological Data Analysis: An Application of Change-Point Analysis to Prehistoric Domes. *Archeologia e Calcolatori* **5**, 53-68.
- Litton, C. D. and Buck, C. E. (1995). The Bayesian Approach to the Interpretation of Archaeological Data. *Archaeometry* **37**, 1-24.
- Lucy, D., Aykroyd, R. G., Pollard, A. M., and Solheim, T. (1996). A Bayesian Approach to Adult Human Age Estimation from Dental Observations by Johanson's Age Changes. *Journal of Forensic Sciences* **41**, 189-194.
- Malakoff, David. (1999). Bayes Offers a 'New' Way to Make Sense of Numbers. *Science* **286**, 1460-1464.
- Phillips, L. D. (1973). *Bayesian Statistics for Social Scientists*. Thomas Nelson and Sons, London.
- Reese, R. (1994). Are Bayesian Statistics Useful to Archaeological Reasoning? *Antiquity* **68**, 848-850.
- Smith, A. F. M. (1995). A Conversation with Dennis Lindley. *Statistical Science* **10**, 305-319.



---

# *Preface to the First Edition*

---

## Contextual Comments

This book is intended to fill a void. There is a reasonably wide gap between the background of the median empirically trained social or behavioral scientist and the full weight of Bayesian statistical inference. This is unfortunate because, as we will see in the forthcoming chapters, there is much about the Bayesian paradigm that suits the type of data and data analysis performed in the social and behavioral sciences. Consequently, the goal herein is to bridge this gap by connecting standard maximum likelihood inference to Bayesian methods by emphasizing linkages between the standard or classical approaches and full probability modeling via Bayesian methods.

This is far from being an exclusively theoretical book. I strongly agree that “theoretical satisfaction and practical implementation are the twin ideals of coherent statistics” (Lindley 1980), and substantial attention is paid to the mechanics of putting the ideas into practice. Hopefully the extensive attention to calculation and computation basics will enable the interested readers to immediately try these procedures on their own data. Coverage of various numerical techniques from detailed posterior calculations to computational-numerical integration is extensive because these are often the topics that separate theory and realistic practice.

The treatment of theoretical topics in this work is best described as “gentle but rigorous”: more mathematical derivation details than related books, but with more explanation as well. This is not an attempt to create some sort of “Bayes-Lite” or “Bayes for Dummies” (to paraphrase the popular self-help works). Instead, the objective is to provide a Bayesian methods book tailored to the interests of the social and behavioral sciences. It therefore features data that these scholars care about, focuses more on the tools that they are likely to require, and speaks in a language that is more compatible with typical prerequisites in associated departments.

There is also a substantial effort to put the development of Bayesian methods in a historical context. To many, the principles of Bayesian inference appear to be something that “came out of left field,” and it is important to show that not only are the fundamentals of Bayesian statistics older than the current dominant paradigms, but that their history and development are actually closely intertwined.

## Outline of the Book

This book is laid out as follows. Chapter 1 gives a high-level, brief introduction to the basic philosophy of Bayesian inference. I provide some motivations to justify the time and effort required to learn a new way of thinking about statistics through Bayesian inference. Chapter 2 (*now Appendix A*) provides the necessary technical background for going on: basic likelihood theory, the generalized linear model, and numerical estimation algorithms.

Chapter 1 describes the core idea behind Bayesian thinking: updating prior knowledge with new data to give the *posterior distribution*. Examples are used to illustrate this process and some historical notes are included. The normal model and its relatives are no less important in Bayesian statistics than in non-Bayesian statistics, and Chapter 3 outlines the key basic normal models along with extensions.

Specifying prior distributions is a key component of the Bayesian inference process and Chapter 4 goes through the typology of priors. The Bayesian paradigm has a cleaner and more introspective approach to assessing the quality of fit and robustness of researcher-specified models, and Chapter 6 outlines procedures so that one can test the performance of various models. Chapter 7 is a bit more formal about this process; it outlines a number of ways to explicitly test models against each other and to make decisions about unknown parameters.

The most modern component of this book begins with Chapter 9, which is an introduction to Monte Carlo and related methods. These topics include the many varieties of numerical integration and importance sampling, and culminating with the EM algorithm. While none of these tools are exclusively Bayesian in nature, Bayesians generally make more use of them than others. Chapter 10 formally introduces Markov chain Monte Carlo (MCMC). These are the tools that revolutionized Bayesian statistics and led to the current renaissance. This chapter includes both theoretical background on Markov chains as well as practical algorithmic details. Chapter 12 discusses hierarchical models that give the Bayesian researcher great flexibility in specifying models through a general framework. These models often lead to the requirement of MCMC techniques and the examples in this chapter are illustrated with practical computing advice. Finally, Chapter 11 (*in the first edition*) discusses necessary details about the mechanics of running and testing MCMC inferences.

The structure of each chapter is reasonably uniform. The basic ideas are enumerated early in the chapter and several of the chapters include an advanced topics section to further explore ideas that are unlikely to be of interest to every reader or to the first-time reader. All chapters have exercises that are intended to give practice developing the central ideas of each topic, including computer-based assignments.

There are unfortunately several topics that I have not had the space to cover here. Foremost is Bayesian decision theory. Many social and behavioral scientists do not operate



in a data-analytic environment where the explicit cost of making a wrong decision can be quantified and incorporated into the model. This may be changing and there are a number of areas that are currently oriented toward identifying loss and risk, such as applied public policy. In the meantime, readers who are focused accordingly are directed to the books by Berger (1985), Winkler (1972), Robert (2001), and the foundational work of Wald (1950). The second major topic that is mentioned only in passing is the growing area of empirical Bayes. The best introduction is the previously noted text of Carlin and Louis (2001). See also the extensive empirical Bayes reference list in Section 12.4. I would very much have liked to cover the early, but exciting developments in perfect sampling (coupling from the past). See the original work by Propp and Wilson (1996).

Bayesian game theory is an important topic that has been omitted. Some of the better known citations are Raiffa (1982), Blackwell and Girshick (1954), Savage (1954), and Bayarri and DeGroot (1991). The Bayesian analysis of survival data as a distinct subspecialty is somewhat understudied. The recent book by Ibrahim, Chen, and Sinha (2001) goes a long way toward changing that. Chapter 10 provides the essentials for understanding Markov chains in general. The study of Markov chains extends well beyond basic MCMC and the mathematical references that I often find myself reaching for are Meyn and Tweedie (1993), Norris (1997), and Nummelin (1984). The Bayesian hierarchical models covered in Chapter 12 naturally and easily extend into meta-analysis, a subject well-covered in the social sciences by Cooper and Hedges (1994), Hunter and Schmidt (1990), and Lipsey and Wilson (2001).

---

## Background and Prerequisites

This is not a book for a first-semester course in social and behavioral statistics. Instead, it is intended to extend the training of graduate students and researchers who have already experienced a one-year (roughly) sequence in social statistics. Therefore good prerequisites include intermediate-level, regression-oriented texts such as Fox (1997), Gujarati (1995), Hanushek and Jackson (1977), Harrell (2001), Neter *et al.* (1996), and Montgomery *et al.* (2001). Essentially it is assumed that the reader is familiar with the basics of the linear model, simple inference, multivariate specifications, and some nonlinear specifications.

A rudimentary understanding of matrix algebra is required, but this does not need to go beyond the level of Chapter 1 in Greene (2000), or any basic undergraduate text. The essential manipulations that we will use are matrix multiplication, inversion, transposition, and segmentation. The calculus operations done here are more conceptual than mechanical; that is, it is more important to understand the meaning of differentiation and integration operations rather than to be an expert on the technicalities. A knowledge at the level of Kleppner and Ramsey's (1985) self-teaching primer is sufficient to follow the calculations.

The core philosophical approach taken with regard to model specification comes from the generalized linear model construct of Nelder and Wedderburn (1972), elaborated in McCullagh and Nelder (1989). This is an integrated theoretical framework that unifies disparate model specifications by re-expressing models based on making the appropriate choice of model configuration based on the structure of the outcome variable and the nature of the dispersion. This fundamental way of thinking is independent of whether the model is Bayesian (see Dey, Ghosh, and Mallick 2000) or classical (see Fahrmeir and Tutz 2001).

---

## Software

The concepts and procedures in this book would be of little practical value without a means of directly applying them. Consequently, there is an emphasis here on demonstrating ideas with statistical software. *All* code in R and BUGS and *all* data are posted at the dedicated webpage:

<http://web.clas.ufl.edu/~jgill/BMSBSA>.

A great deal of the material in this book focuses on developing examples using the R and BUGS statistical packages. Not only are these extremely high-quality analytical tools, they are also widely distributed free of charge.

It is hard to overstate the value of the R statistical environment. R is the Open Source implementation of the S statistical language (from AT&T-Bell Labs), which has become the *de facto* standard computing language in academic statistics because of its power, flexibility, and sense of community. R was initially written by Robert Gentleman and Ross Ihak at the University of Auckland, but is now supported by a growing group of dedicated scholars. An important aspect of R is the user community itself, and the user-written packages have been shown to be an effective way for scholars to share and improve new methods.

The homesite for R (see the details in Chapter 2, *now Appendix A*), contains documentation on installation and learning the language. In addition, because R is “non-unlike” S, any published book on S-Plus will be useful. The standard text for statistical modeling in S is the work of Venables and Ripley (1999). The forthcoming book by Fox (2002) is a particularly helpful and well-written introduction to doing applied work in S. In addition, an increasing number of applied methodology books that feature the S language have appeared, and I try to keep up with these on a webpage:

<http://web.clas.ufl.edu/~jgill/s-language.help.html>.

Any applied Bayesian today that wants to feel good about the state of the world with regard to software need only look at Press’ 1980 summary of available Bayesian analysis programs. This is a disparate, even tortured, list of mainframe-based programs that generally only implement one or two procedures each and require such pleasantries as “Raw data on

Paper tape.” In contrast, the BUGS package makes Bayesian analysis using MCMC pleasant and engaging by taking the odious mechanical aspects away from the user, allowing one to focus on the more interesting aspects of model specification and testing. This unbelievable gift to the Bayesian statistical community was developed at the MRC Biostatistics Unit in Cambridge:

<http://www.mrc-bsu.cam.ac.uk/bugs/>.

---

## Acknowledgments

I am indebted to many people for criticisms, suggestions, formal reviews, and influential conversations. These include Alan Agresti, Micah Altman, Attic Access, Sammy Barkin, Neal Beck, Jim Booth, Brad Carlin, George Casella, Lauren Cowles, John Fox, Wayne Francis, Charles Franklin, Malay Ghosh, Hank Heitowit, Jim Hobert, Bill Jacoby, Renee Johnson, Gary King, Andrew Martin, Michael Martinez, Mike McDonald, Ken Meier, Elias Moreno, Brad Palmquist, Kevin Quinn, Christian Robert, Stephen Sen, Jason Wittenberg, Sam Wu, Chris Zorn, and anonymous reviewers. I am especially grateful to George Casella for his continued advice and wisdom; I’ve learned as much sitting around George’s kitchen table or trying to keep up with him running on the beach as I have in a number of past seminars. Andrew Martin also stands out for having written detailed and indispensable reviews (formally and informally) of various drafts and for passing along insightful critiques from his students.

A small part of the material in Chapter 7 was presented as a conference paper at the Fifth International Conference on Social Science Methodology in Cologne, Germany, October 2000. Also, some pieces of Chapter 10 were taken from another conference paper given at the Midwestern Political Science Association Annual Meeting, Chicago, April 2001.

This book is a whole lot better due to input from students at the University of Florida and the University of Michigan Summer Program, including: Javier Arpricio-Castello, Jorge Aragon, Jessica Archer, Sam Austin, Ryan Bakker, David Conklin, Jason Gainous, Dukhong Kim, Eduardo Leoni, Carmela Lutmar, Abdel-hameed Hamdy Nawar, Per Simonnson, Jay Tate, Tim McGraw, Nathaniel Seavy, Lee Walker, and Natasha Zharinova. For research assistance, I would also like to thank Ryan Bakker, Kelly Billingsley, Simon Robertshaw, and Nick Theobald.

I would like to also thank my editor, Bob Stern, at Chapman & Hall for making this process much more pleasant than it would have been without his continual help. This volume was appreciatively produced and delivered “camera-ready” with L<sup>A</sup>T<sub>E</sub>X using the  $\mathcal{A}\mathcal{M}\mathcal{S}$  packages, pstricks, and other cool typesetting macros and tools from the T<sub>E</sub>X world. I cannot imagine the bleakness of a life restricted to the low-technology world of word processors.

## References

- Bayarri, M. J. and DeGroot, M. H. (1991). What Bayesians Expect of Each Other. *Journal of the American Statistical Association* **86**, 924-932.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second Edition. New York: Springer-Verlag.
- Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*. New York: Wiley.
- Carlin, B. P. and Louis, T. A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*. Second Edition. New York: Chapman & Hall.
- Cooper, H. and Hedges, L. (eds.) (1994). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.
- Dey, D. K., Ghosh, S. K., and Mallick, B. K. (2000). *Generalized Linear Models: A Bayesian Perspective*. New York: Marcel Dekker.
- Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Second Edition. New York: Springer.
- Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.
- Fox, J. (2002). *An R and S-Plus Companion to Applied Regression*. Thousand Oaks, CA: Sage.
- Greene, W. (2000). *Econometric Analysis*. Fourth Edition. Upper Saddle River, NJ: Prentice Hall.
- Gujarati, D. N. (1995). *Basic Econometrics*. New York: McGraw-Hill.
- Hanushek, E. A. and Jackson, J. E. (1977). *Statistical Methods for Social Scientists*. San Diego: Academic Press.
- Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag.
- Hunter, J. E. and Schmidt, F. L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Thousand Oaks, CA: Sage.
- Ibrahim, J. G., Chen, M-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer-Verlag.
- Kleppner, D. and Ramsey, N. (1985). *Quick Calculus: A Self-Teaching Guide*. New York: Wiley Self Teaching Guides.
- Lindley, D. V. (1980). Jeffreys's Contribution to Modern Statistical Thought. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*. Arnold Zellner (ed.). Amsterdam: North Holland.
- Lipsey, M. W. and Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Second Edition. New York: Chapman & Hall.
- Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. New York: Springer-Verlag.
- Montgomery, D. C. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to Linear Regression Analysis*. Third Edition. New York: John Wiley & Sons.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). "Generalized Linear Models." *Journal of the Royal Statistical Society, Series A* **135**, 370-85.
- Neter, J., Kutner, M. H., Nachtsheim, C., and Wasserman, W. (1996). *Applied Linear Regression Models*. Chicago: Irwin.
- Norris, J. R. (1997). *Markov Chains*. Cambridge: Cambridge University Press.

- Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge: Cambridge University Press.
- Press, S. J. (1980). Bayesian Computer Programs. In *Studies in Bayesian Econometrics and Statistics*, S. E. Fienberg and A. Zellner, eds., Amsterdam: North Holland, pp. 429-442..
- Propp, J. G. and Wilson, D. B. (1996). Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics. *Random Structures and Algorithms* **9**, 223-252.
- Raiffa, H. (1982). *The Art and Science of Negotiation*. Cambridge: Cambridge University Press.
- Robert, C. P. (2001). *The Bayesian Choice: A Decision Theoretic Motivation*. Second Edition. New York: Springer-Verlag.
- Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.
- Venables, W. N. and Ripley, B. D. (1999). *Modern Applied Statistics with S-Plus*, Third Edition. New York: Springer-Verlag.
- Wald, A. (1950). *Statistical Decision Functions*. New York: Wiley.
- Winkler, R. L. (1972). *Introduction to Bayesian Inference and Decision*. New York: Holt, Rinehart, and Winston.



# Chapter 1

---

## *Background and Introduction*

---

### 1.1 Introduction

Vitriolic arguments about the merits of Bayesian versus classical approaches seem to have faded into a quaint past of which current researchers in the social sciences are, for the most part, blissfully unaware. In fact, it almost seems odd well into the 21st century that deep philosophical conflicts dominated the last century on this issue. What happened? Bayesian methods always had a natural underlying advantage because all unknown quantities are treated probabilistically, and this is the way that statisticians and applied statisticians really prefer to think. However, without the computational mechanisms that entered into the field we were stuck with models that couldn't be estimated, prior distributions (distributions that describe what we know before the data analysis) that incorporated uncomfortable assumptions, and an adherence to some bankrupt testing notions. Not surprisingly, what changed all this was a dramatic increase in computational power and major advances in the algorithms used on these machines. We now live in a world where there are very few model limitations, other than perhaps our imaginations. We therefore live in world now where researchers are for the most part comfortable specifying Bayesian and classical models as it suits their purposes.

It is no secret that Bayesian methods require a knowledge of classical methods as well as some additional material. Most of this additional material is either applied calculus or statistical computing. That is where this book comes in. The material here is intended to provide an introduction to Bayesian methods all the way from basic concepts through advanced computational material. Some readers will therefore be primarily interested in different sections. Also it means that this book is not strictly a textbook, a polemic, nor a research monograph. It is intended to be all three.

Bayesian applications in medicine, the natural sciences, engineering, and the social sciences have been increasing at a dramatic rate since the middle of the early 1990s. Interestingly, the Mars Rovers are programmed to think Bayesianly while they traverse that planet. Currently seismologists perform Bayesian updates of aftershocks based on the main-shock and previous patterns of aftershocks in the region. Bayesian networks are built in computational biology, and the forefront of quantitative research in genomics is now firmly Bayesian.

So why has there been a noticeable increase in interest in Bayesian statistics? There are actually several visible reasons. *First*, and perhaps most critically, society has radically increased its demand for statistical analysis of all kinds. A combined increase in clinical trials, statistical genetics, survey research, general political studies, economic analysis, government policy work, Internet data distribution, and marketing research have led to golden times for applied statisticians. *Second*, many introspective scholars who seriously evaluate available paradigms find that alternatives to Bayesian approaches are fraught with logical inconsistencies and shortcomings. *Third*, until recent breakthroughs in statistical computing, it was easy to specify realistic Bayesian statistical models that could not provide analytically tractable summary results.

There is therefore ample motivation to understand the basics of Bayesian statistical methodology, and not just because it is increasingly important in mainstream analysis of data. The Bayesian paradigm rests on a superior set of underlying assumptions and includes procedures that allow researchers to include reliable information in addition to the sample, to talk about findings in intuitive ways, and to set up future research in a coherent manner. At the core of the data-analytic enterprise, these are key criteria to producing useful statistical results.

Statistical analysis is the process of “data reduction” with the goal of separating out underlying systematic effects from the noise inherent in all sets of observations. Obviously there is a lot more to it than that, but the essence of what we do is using models to distill findings out of murky data. There are actually three general steps in this process: collection, analysis, and assessment. For most people, data collection is not difficult in that we live in an age where data are omnipresent. More commonly, researchers possess an abundance of data and seek meaningful patterns lurking among the various dead-ends and distractions. Armed with a substantive theory, many are asking: what should I do now? Furthermore, these same people are often frustrated when receiving multiple, possibly conflicting, answers to that question.

Suppose that there exists a model-building and data analysis process with the following desirable characteristics:

- ▷ overt and clear model assumptions,
- ▷ a principled way to make probability statements about the real quantities of theoretical interest,
- ▷ an ability to update these statements (i.e., learn) as new information is received,
- ▷ systematic incorporation of previous knowledge on the subject,
- ▷ missing information handled seamlessly as part of the estimation process,
- ▷ recognition that population quantities can be changing over time rather than forever fixed,
- ▷ the means to model all data types including hierarchical forms,
- ▷ straightforward assessment of both model quality and sensitivity to assumptions.



As the title of this book suggests, the argument presented here is that the practice of Bayesian statistics possesses all of these qualities. Press (1989) adds the following practical advantages to this list:

- ▷ it often results in shorter confidence/credible intervals,
- ▷ it often gives smaller model variance,
- ▷ predictions are usually better,
- ▷ “proper” prior distributions (Chapter 4) give models with good frequentist properties,
- ▷ reasonably “objective” assumptions are available,
- ▷ hypotheses can be tested without pre-determination of testing quality measures.

This text will argue much beyond these points and assert that the type of data social and behavioral scientists routinely encounter makes the Bayesian approach ideal in ways that traditional statistical analysis cannot match. These natural advantages include avoiding the assumption of infinite amounts of forthcoming data, recognition that fixed-point assumptions about human behavior are dubious, and a direct way to include existing expertise in a scientific field.

What reasons are there for *not* worrying about Bayesian approaches and sticking with the, perhaps more comfortable, traditional mindset? There are several reasons why a reader may not want to worry about the principles in this text for use in their research, including:

- ▷ their population parameters of interest are truly fixed and unchanging under all realistic circumstances,
- ▷ they do *not* have any prior information to add to the model specification process,
- ▷ it is necessary for them to provide statistical results as if data were from a *controlled experiment*,
- ▷ they care more about “significance” than effect size,
- ▷ computers are slow or relatively unavailable for them,
- ▷ they prefer very automated, “cookbook” type procedures.

So why do so-called classical approaches dominate Bayesian usage in the social and behavioral sciences? There are several reasons for this phenomenon. *First*, key figures in the development of modern statistics had strong prejudices against aspects of Bayesian inference for narrow and subjective reasons. *Second*, the cost of admission is higher in the form of additional mathematical formalism. *Third*, until recently realistic model specifications sometimes led to unobtainable Bayesian solutions. *Finally*, there has been a lack of methodological introspection in a number of disciplines. The primary mission of this text is to make the second and third reasons less of a barrier through accessible explanation, detailed examples, and specific guidance on calculation and computing.

It is important to understand that the Bayesian way does not mean throwing away one’s comfortable tools, and it is not itself just another tool. Instead it is a way of *organizing* one’s

toolbox and is also a way of doing statistical work that has sharply different philosophical underpinnings. So adopting Bayesian methods means keeping the usual set of methods, such as linear regression, ANOVA, generalized linear models, tabular analysis, and so on. In fact, many researchers applying statistics in the social sciences are not actually frequentists since they cannot assume an *infinite* stream of iid (independent and identically distributed) data coming from a controlled experimental setup. Instead, most of these analysts can be described as “likelihoodists,” since they obtain one sample of observational data that is contextual and will not be repeated, then perform standard likelihood-based (Fisherian) inference to get coefficient estimates.

Aside from underlying philosophical differences, many readers will be comforted in finding that Bayesian and non-Bayesian analyses often agree. There are two important instances where this is *always* true. *First*, when the included prior information is very uninformative (there are several ways of providing this), summary statements from Bayesian inference will match likelihood point estimates. Therefore a great many researchers are Bayesians who do not know it yet. *Second*, when the data size is very large, the form of the prior information used does not matter and there is agreement again. Other circumstances also exist in which Bayesian and non-Bayesian statistical inferences lead to the same results, but these are less general than the two mentioned. In addition to these two important observations, all hierarchical models are overtly Bayesian since they define distributional assumptions at levels. These are popular models due to their flexibility with regard to the prevalence of different levels of observed aggregation in the same dataset. We will investigate Bayesian hierarchical models in Chapter 12.

We will now proceed to a detailed justification for the use of modern Bayesian methods.

---

## 1.2 General Motivation and Justification

With Bayesian analysis, assertions about unknown model parameters are not expressed in the conventional way as single point estimates along with associated reliability assessed through the standard null hypothesis significance test. Instead the emphasis is on making probabilistic statements using distributions. Since Bayesians make no fundamental distinction between unobserved data and unknown parameters, the world is divided into: immediately available quantities, and those that need to be described probabilistically. Before observing some data, these descriptions are called *prior distributions*, and after observing the data these descriptions are called *posterior distributions*. The quality of the modeling process is the degree to which a posterior distribution is more informed than a prior distribution for some unknown quantity of interest. Common descriptions of posterior distributions include standard quantile levels, the probability of occupying some affected region of the

sample space, the predictive quantities from the posterior, and Bayesian forms of confidence intervals called credible intervals.

It is important to note here that the pseudo-frequentist null hypothesis significance test (NHST) is not just sub-optimal, it is *wrong*. This is the dominant hypothesis testing paradigm as practiced in the social sciences. Serious problems include: a logical inconsistency coming from probabilistic modus tollens, confusion over the order of the conditional probability, chasing significance but ignoring effect size, adherence to completely arbitrary significance thresholds, and confusion about the probability of rejection. There is a general consensus amongst those that have paid attention to this issue that the social sciences have been seriously harmed by the NHST since it has led to fixations with counting stars on tables rather than looking for effect sizes and general statistical reliability. See the recent discussions in Gill (1999) and Denis (2005) in particular for detailed descriptions of these problems and how they damage statistical inferences in the social sciences. Serious criticism of the NHST began shortly after its creation in the early 20th century by textbook writers who blended Fisherian likelihoodism with Neyman and Pearson frequentism in an effort to offend neither warring and evangelical camp. An early critic of this unholy union was Rozeboom (1960) who noticed its “strangle-hold” on social science inference. In 1962 Arthur Melton wrote a parting editorial in the *Journal of Experimental Psychology* revealing that he had held authors to a 0.01 p-value standard: “In editing the *Journal* there has been a strong reluctance to accept and publish results related to the principal concern of the research when those results were significant at the .05 level, whether by one- or two-tailed test!” This had the effect of accelerating the criticism and led to many analytical and soul-searching articles discussing the negative consequences of this procedure in the social sciences, including: Barnett (1973), Berger (2003), Berger, Boukai, and Wang (1997), Berger and Sellke (1987), Berkhardt and Schoenfeld (2003), Bernardo (1984), Brandstätter (1999), Carver (1978, 1993), Cohen (1962, 1977, 1988, 1992, 1994), Dar, Serlin and Omer (1994), Falk and Greenbaum (1995), Gigerenzer (1987, 1998a, 1998b, 2004), Gigerenzer and Murray (1987), Gliner, Leech and Morgan (2002), Greenwald (1975), Greenwald, *et al.* (1996), Goodman (1993, 1999), Haller and Krauss (2002), Howson and Urbach (1993), Hunter (1997), Hunter and Schmidt (1990), Kirk (1996), Krueger (2001), Lindsay (1995), Loftus (1991, 1993), Macdonald (1997), McCloskey and Ziliak (1996), Meehl (1978, 1990, 1997), Moran and Soloman (2004), Morrison and Henkel (1969, 1970), Nickerson (2000), Oakes (1986), Pollard (1993), Pollard and Richardson (1987), Robinson and Levin (1997), Rosnow and Rosenthal (1989), Schmidt (1996), Schmidt and Hunter (1977), Schervish (1996), Sedlmeier and Gigerenzer (1989), Thompson (1996, 1997, 2002a, 2002b, 2004), Wilkinson (1977), Ziliak and McCloskey (2007). And this is only a small sample of the vast literature describing the NHST as bankrupt. Conveniently some of the more influential articles listed above are reprinted in Harlow *et al.* (1997). We will return to this point in Chapter 7 (starting on page 209) in the discussion of Bayesian hypothesis testing and model comparison.

This focus on distributional inference leads to two key assumptions for Bayesian work.

*First*, a specific parametric form is assumed to describe the distribution of the data given parameter values. Practically, this is used to construct a *likelihood function* ( $A$ ) to incorporate the contribution of the full sample of data. Note that this is an inherently parametric setup, and although nonparametric Bayesian modeling is a large and growing field, it exists beyond the scope of the basic setup. *Second*, since unknown parameters are treated as having distributional qualities rather than being fixed, an assumed prior distribution on the parameters of interest unconditional on the data is given. This reflects either uncertainty about a truly fixed parameter or recognition that the quantity of interest actually behaves in some stochastic fashion.

With those assumptions in hand, the essentials of Bayesian thinking can be stated in three general steps:

1. Specify a probability model that includes some prior knowledge about the parameters for unknown parameter values.
2. Update knowledge about the unknown parameters by conditioning this probability model on observed data.
3. Evaluate the fit of the model to the data and the sensitivity of the conclusions to the assumptions.

Notice that this process does not include an unrealistic and artificial step of making a contrived decision based on some arbitrary quality threshold. The value of a given Bayesian model is instead found in the description of the distribution of some parameter of interest in probabilistic terms. Also, there is nothing about the process contained in the three steps above that cannot be repeated as new data are observed. It is often convenient to use the conventional significance thresholds that come from Fisher, but Bayesians typically do not ascribe any major importance to being barely on one side or the other. That is, Bayesian inference often prescribes something like a 0.05 threshold, but it is rare to see work where a 0.06 finding is not taken seriously as a likely effect.

Another core principle of the Bayesian paradigm is the idea that the data are fixed once observed. Typically (but not always) these data values are assumed to be *exchangeable*; the model results are not changed by reordering the data values. This property is more general than, and implied by, the standard assumption that the data are *independent and identically distributed* (iid): independent draws from the same distribution, and also implies a common mean and variance for the data values (Leonard and Hsu 1999, p.1). Exchangeability allows us to say that the data generation process is conditional on the unknown model parameters in the same way for every data value (de Finetti 1974, Draper *et al.* 1993, Lindley and Novick 1981). Essentially this is a less restrictive version of the standard iid assumption. Details about the exchangeability assumption are given in Chapter 12. We now turn to a discussion of probability basics as a precursor to Bayesian mechanics.

### 1.3 Why Are We Uncertain about Uncertainty?

The fundamental principles of probability are well known, but worth repeating here. Actually, it is relatively easy to *intuitively* define the properties of a probability function: (1) its range is bounded by zero and one for all the values in the applicable domain, (2) it sums or integrates to one over this domain, and (3) the sum or integral of the functional outcome (probability) of disjoint events is equal to the functional outcome of the union of these events. These are the Kolmogorov axioms (1933), and are given in greater detail in Gill (2006), Chapter 7. The real problem lies in describing the actual meaning of probability statements. This difficulty is, in fact, at the heart of traditional disagreements between Bayesians and non-Bayesians.

The frequentist statistical interpretation of probability is that it is a limiting relative frequency: the long-run behavior of a nondeterministic outcome or just an observed proportion in a population. This idea can be traced back to Laplace (1814), who defined probability as the number of successful events out of trials observed. Thus if we could simply repeat the experiment or observe the phenomenon enough times, it would become apparent what the future probability of reoccurrence will be. This is an enormously useful way to think about probability but the drawback is that frequently it is not possible to obtain a large number of outcomes from exactly the same event-generating system (Kendall 1949, Plackett 1966).

A competing view of probability is called “subjective” and is often associated with the phrase “degree of belief.” Early proponents included Keynes (1921) and Jeffreys (1961), who observed that two people could look at the same situation and assign different probability statements about future occurrences. This perspective is that probability is *personally* defined by the conditions under which a person would make a bet or assume a risk in pursuit of some reward. Subjective probability is closely linked with the idea of decision-making as a field of study (see, for instance, Bernardo and Smith [1994, Chapter 2]) and the principle of selecting choices that maximize personal utility (Berger 1985).

These two characterizations are necessarily simplifications of the perspectives and de Finetti (1974, 1975) provides a much deeper and more detailed categorization, which we will return to in Chapter 12. To de Finetti, the ultimate arbiter of subjective probability assignment is the conditions under which individuals will wager their own money. In other words, a person will not violate a personal probability assessment if it has financial consequences. Good (1950) makes this idea more axiomatic by observing that people have personal probability assessments about many things around them rather than just one, and in order for these disparate comparative statements to form a *body of beliefs* they need to be free of contradictions. For example, if a person thinks that *A* is more likely to occur than *B*, and *B* is more likely to occur than *C*, then this person cannot coherently believe that *C* is more likely than *A* (transitivity). Furthermore, Good adds the explicitly Bayesian idea that people are constantly updating these personal probabilities as new information is observed,

although there is evidence that people have subadditive notions of probability when making calculations (the probability of some event plus the probability of its complement do not add to certainty).

The position underlying nearly all Bayesian work is the subjective probability characterization, although there have been many attempts to “objectify” Bayesian analysis (see Chapter 4). Prior information is formalized in the Bayesian framework and this prior information can be subjective in the sense that the researcher’s experience, intuition, and theoretical ideas are included. It is also common to base the prior information on previous studies, experiments, or just personal observations and this process is necessarily subject to a limited (although possibly large) number of observations rather than the infinite number assumed under the frequentist view. We will return to the theme of subjectivity contained in prior information in Chapter 4 and elsewhere, but the principal point is that *all* statistical models include subjective decisions, and therefore we should *ceteris paribus* prefer one that is the most explicit about assumptions. This is exactly the sense that the Bayesian prior provides readers with a specific, formalized statement of currently assumed knowledge in probabilistic terms.

### 1.3.1 Required Probability Principles

There are some simple but important principles and notational conventions that must be understood before proceeding. We will not worry too much about measure theory until Chapter 13, and the concerned reader is directed to the first chapter of any mathematical statistics text or the standard reference works of Billingsley (1995), Chung (1974), and Feller (1990, Volumes 1 and 2). Abstract events are indicated by capital Latin letters:  $A$ ,  $B$ ,  $C$ , etc. A probability function corresponding to some event  $A$  is always indicated by  $p(A)$ . The complement of the event  $A$  is denoted  $A^c$ , and it is a consequence of Kolmogorov’s axioms listed above that  $p(A^c) = 1 - p(A)$ . The union of two events is indicated by  $A \cup B$  and the intersection by  $A \cap B$ . For any two events:  $p(A \cup B) = p(A) + p(B) - p(A \cap B)$ . If two events are independent, then  $p(A \cap B) = p(A)p(B)$ , but not necessarily the converse (the product relationship does not imply independence).

Central to Bayesian thinking is the idea of conditionality. If an event  $B$  is material to another event  $A$  in the sense that the occurrence or non-occurrence of  $B$  affects the probability of  $A$  occurring, then we say that  $A$  is conditional on  $B$ . It is a basic tenet of Bayesian statistics that we update our probabilities as new relevant information is observed. This is done with the definition of conditional probability given by:  $p(A|B) = p(A \cap B)/p(B)$ , which is read as “the probability of  $A$  given  $B$  is equal to the probability of  $A$  and  $B$  divided by the probability of  $B$ .”

In general the quantities of interest here are random variables rather than the simple discrete events above. A random variable  $X$  is defined as a measurable function from a probability space to a state space. This can be defined very technically (Shao 2005, p.7), but for our purposes it is enough to understand that the random variable connects

possible occurrences of some data value with a probability structure that reflects the relative “frequency” of these occurrences. The function is thus defined over a specific state space of all possible realizations of the random variable, called support. Random variables can be discrete or continuous. For background details see Casella and Berger (2002), Shao (2005), or the essay by Doob (1996). The expression  $p(X \cap Y)$  is usually denoted as  $p(X, Y)$ , and is referred to as the *joint distribution* of random variables  $X$  and  $Y$ . Marginal distributions are then simply  $p(X)$  and  $p(Y)$ . Restating the principle above in this context, for two independent random variables the joint distribution is just the product of the marginals,  $p(X, Y) = p(X)p(Y)$ . Typically we will need to integrate expressions like  $p(X, Y)$  to get marginal distributions of interest. Sometimes this is done analytically, but more commonly we will rely on computational techniques.

We will make extensive use of expected value calculations here. Recall that if a random variable  $X$  is distributed  $p(X)$ , the expected value of some function of the random variable,  $h(X)$ , is

$$E[h(X)] = \begin{cases} \sum_{i=1}^k h(x_i)p(x_i) & k\text{-category discrete case} \\ \int_{\mathcal{X}} h(X)p(X)dx & \text{continuous case.} \end{cases} \quad (1.1)$$

Commonly  $h(X) = X$ , and we are simply concerned with the expectation of the random variable itself. In the discrete case this is a very intuitive idea as the expected value can be thought of as a probability-weighted average over possible events. For the continuous case there are generally limits on the integral that are dictated by the support of the random variable, and sometimes these are just given by  $[-\infty, \infty]$  with the idea that the PDF (probability density function) indicates zero and non-zero regions of density. Also  $p(X)$  is typically a conditional statement:  $p(X|\theta)$ . For the  $k \times 1$  vector  $\mathbf{X}$  of discrete random variables, the expected value is:  $E[\mathbf{X}] = \sum \mathbf{X}p(\mathbf{X})$ . With the expected value of a function of the continuous random vector, it is common to use the *Riemann-Stieltjes integral* form (found in any basic calculus text):  $E[f(\mathbf{X})] = \int_{\mathcal{X}} f(\mathbf{X})dF(\mathbf{X})$ , where  $F(\mathbf{X})$  denotes the joint distribution of the random variable vector  $\mathbf{X}$ . The principles now let us look at Bayes’ Law in detail.

---

## 1.4 Bayes’ Law

The Bayesian statistical approach is based on updating information using what is called Bayes’ Law (and synonymously Bayes’ Theorem) from his famous 1763 essay. The Reverend Thomas Bayes was an amateur mathematician whose major contribution (the others remain rather obscure and do not address the same topic) was an essay found and published two years after his death by his friend Richard Price. The enduring association of an important branch of statistics with his name actually is somewhat of an exaggeration of

the generalizeability of this work (Stigler 1982). Bayes was the first to explicitly develop this famous law, but it was Laplace (1774, 1781) who (apparently independently) provided a more detailed analysis that is perhaps more relevant to the practice of Bayesian statistics today. See Stigler (1986) for an interesting historical discussion and Sheynin (1977) for a detailed technical analysis. Like Bayes, Laplace assumed a uniform distribution for the unknown parameter, but he worried much less than Bayes about the consequences of this assumption. Uniform prior distributions are simply “flat” distributions that assign equal probability for every possible outcome.

Suppose there are two events of interest  $A$  and  $B$ , which are not independent. We know from basic axioms of probability that the conditional probability of  $A$  given that  $B$  has occurred is given by:

$$p(A|B) = \frac{p(A, B)}{p(B)}, \quad (1.2)$$

where  $p(A|B)$  is read as “the probability of  $A$  given that  $B$  has occurred,  $p(A, B)$  is the “the probability that both  $A$  and  $B$  occur” (i.e., the joint probability) and  $p(B)$  is just the unconditional probability that  $B$  occurs. Expression (1.2) gives the probability of  $A$  after some event  $B$  occurs. If  $A$  and  $B$  are independent here then  $p(A, B) = p(A)p(B)$  and (1.2) becomes uninteresting.

We can also define a different conditional probability in which  $A$  occurs first:

$$p(B|A) = \frac{p(B, A)}{p(A)}. \quad (1.3)$$

Since the probability that  $A$  and  $B$  occur is the same as the probability that  $B$  and  $A$  occur ( $p(A, B) = p(B, A)$ ), then we can rearrange (1.2) and (1.3) together in the following way:

$$p(A, B) = p(A|B)p(B)$$

$$p(B, A) = p(B|A)p(A)$$

$$p(A|B)p(B) = p(B|A)p(A)$$

$$p(A|B) = \frac{p(A)}{p(B)}p(B|A). \quad (1.4)$$

The last line is the famous Bayes’ Law. This is really a device for “inverting” conditional probabilities. Notice that we could just as easily produce  $p(B|A)$  in the last line above by moving the unconditional probabilities to the left-hand side in the last equality.

We can also use Bayes’ Law with the use of odds, which is a common way to talk about uncertainty related to probability. The odds of an event is the ratio of the probability of an event happening to the probability of the event not happening. So for the event  $A$ , the odds of this event is simply:

$$Odds = \frac{p(A)}{1 - p(A)} = \frac{p(A)}{p(\neg A)}, \quad (1.5)$$



which is this ratio expressed in two different ways. Note the use of “ $\neg A$ ” for “not  $A$ ,” which is better notation when the complement of  $A$  isn’t specifically defined and we care only that event  $A$  did not happen. Since this statement is not conditional on any other quantity, we can call it a “prior odds.” If we make it conditional on  $B$ , then it is called a “posterior odds,” which is produced by multiplying the prior odds by the reverse conditional with regard to  $B$ :

$$\frac{p(A|B)}{p(\neg A|B)} = \frac{p(A)}{p(\neg A)} \frac{p(B|A)}{p(B|\neg A)}. \quad (1.6)$$

The last ratio,  $p(B|A)/p(B|\neg A)$  is the “likelihood ratio” for  $B$  under the two conditions for  $A$ . This is actually the ratio of two expressions of Bayes’ Law in the sense of (1.4), which we can see with the introduction of the ratio  $p(B)/p(B)$ :

$$\frac{p(A|B)}{p(\neg A|B)} = \frac{p(A)/p(B)}{p(\neg A)/p(B)} \frac{p(B|A)}{p(B|\neg A)}. \quad (1.7)$$

This ratio turns out to be very useful in Bayesian model consideration since it implies a test between the two states of nature,  $A$  and  $\neg A$ , given the observation of some pertinent information  $B$ .

■ **Example 1.1: Testing with Bayes’ Law.** How is this useful? As an example, hypothetically assume that 2% of the population of the United States are members of some extremist Militia group ( $p(M) = 0.02$ ), a fact that some members might attempt to hide and therefore not readily admit to an interviewer. A survey is 95% accurate on positive Classification,  $p(C|M) = 0.95$ , (“sensitivity”) and the unconditional probability of classification (i.e., regardless of actual militia status) is given by  $p(C) = 0.05$ . To illustrate how  $p(C)$  is really the normalizing constant obtained by accumulating over all possible events, we will stipulate the additional knowledge that the survey is 97% accurate on negative classification,  $p(C^c|M^c) = 0.97$  (“specificity”). The unconditional probability of classifying a respondent as a militia member results from accumulation of the probability across the sample space of survey events using the Total Probability Law:  $p(C) = p(C \cap M) + p(C \cap M^c) = p(C|M)p(M) + [1 - p(C^c|M^c)]p(M^c) = (0.95)(0.02) + (0.03)(0.98) \cong 0.05$ .

Using Bayes’ Law, we can now derive the probability that someone positively classified by the survey as being a militia member really *is* a militia member:

$$p(M|C) = \frac{p(M)}{p(C)} p(C|M) = \frac{0.02}{0.05} (0.95) = 0.38. \quad (1.8)$$

The startling result is that although the probability of correctly classifying an individual as a militia member given they really are a militia member is 0.95, the probability that an individual really is a militia member given that they are positively classified is only 0.38.

The highlighted difference here between the order of conditional probability is often substantively important in a policy or business context. Consider the problem of

designing a home pregnancy test. Given that there exists a fundamental business trade-off between the reliability of the test and the cost to consumers, no commercially viable product will have perfect or near-perfect test results. In designing the chemistry and packaging of the test, designers will necessarily have to compromise between the probability of **P**regnancy given positive **T**est results,  $p(\mathbf{PR}|\mathbf{T})$ , and the probability of positive test results given pregnancy,  $p(\mathbf{T}|\mathbf{PR})$ . Which one is more important? Clearly, it is better to maximize  $p(\mathbf{T}|\mathbf{PR})$  at the expense of  $p(\mathbf{PR}|\mathbf{T})$ , as long as the reduction in the latter is reasonable: it is preferable to give a higher number of false positives, sending women to consult their physician to take a more sensitive test, than to fail to notify many pregnant women. This reduces the possibility that a woman who does not realize that she is pregnant might continue unhealthy practices such as smoking, drinking, or maintaining a poor diet. Similarly, from the perspective of general public health, it is better to have preliminary tests for deadly contagious diseases designed to be similarly conservative with respect to false positives.

### 1.4.1 Bayes' Law for Multiple Events

It would be extremely limiting if Bayes' Law only applied to two alternative events. Fortunately the extension to multiple events is quite easy. Suppose we observe some data **D** and are interested in the relative probabilities of three events *A*, *B*, and *C* conditional on these data. These might be rival hypotheses about some social phenomenon for which the data are possibly revealing. Thinking just about event *A*, although any of the three could be selected, we know from Bayes' Law that:

$$p(A|\mathbf{D}) = \frac{p(\mathbf{D}|A)p(A)}{p(\mathbf{D})}. \quad (1.9)$$

We also know from the Total Probability Law and the definition of conditional probability that:

$$\begin{aligned} p(\mathbf{D}) &= p(A \cap \mathbf{D}) + p(B \cap \mathbf{D}) + p(C \cap \mathbf{D}) \\ &= p(\mathbf{D}|A)p(A) + p(\mathbf{D}|B)p(B) + p(\mathbf{D}|C)p(C). \end{aligned} \quad (1.10)$$

This means that if we substitute the last line into the expression for Bayes' Law, we get:

$$p(A|\mathbf{D}) = \frac{p(\mathbf{D}|A)p(A)}{p(\mathbf{D}|A)p(A) + p(\mathbf{D}|B)p(B) + p(\mathbf{D}|C)p(C)}, \quad (1.11)$$

which demonstrates that the conditional distribution for any of the rival hypotheses can be produced as long as there exist unconditional distributions for the three rival hypotheses,  $p(A)$ ,  $p(B)$ , and  $p(C)$ , and three statements about the probability of the data given these three hypotheses,  $p(\mathbf{D}|A)$ ,  $p(\mathbf{D}|B)$ ,  $p(\mathbf{D}|C)$ . The first three probability statements are called prior distributions because they are unconditional from the data and therefore presumably determined before observing the data. The second three probability statements are merely

PDF (probability density function) or PMF (probability mass function) statements in the conventional sense. All this means that a posterior distribution,  $p(A|\mathbf{D})$ , can be determined through Bayes' Law to look at the weight of evidence for any one of several rival hypotheses or claims.

There is a more efficient method for making statements like (1.11) when the number of outcomes increases. Rather than label the three hypotheses as we have done above, let us instead use  $\theta$  as an unknown parameter whereby different regions of its support define alternative hypotheses. So statements may take the form of "Hypothesis A:  $\theta < 0$ ," or any other desired statement. To keep track of the extra outcome, denote the three hypotheses as  $\theta_i$ ,  $i = 1, 2, 3$ . Now (1.11) is given more generally for  $i = 1, 2, 3$  as:

$$p(\theta_i|\mathbf{D}) = \frac{p(\mathbf{D}|\theta_i)p(\theta_i)}{\sum_{j=1}^3 p(\mathbf{D}|\theta_j)p(\theta_j)} \quad (1.12)$$

for the posterior distribution of  $\theta_i$ . This is much more useful and much more in line with standard Bayesian models in the social and behavioral sciences because it allows us to compactly state Bayes' Law for any number of discrete outcomes/hypotheses, say  $k$  for instance:

$$p(\theta_i|\mathbf{D}) = \frac{p(\theta_i)p(\mathbf{D}|\theta_i)}{\sum_{j=1}^k p(\theta_j)p(\mathbf{D}|\theta_j)}. \quad (1.13)$$

Consider also that the denominator of this expression averages over the  $\theta$  variables and therefore just produces the *marginal distribution of the sample data*, which we could overtly label as  $p(\mathbf{D})$ . Doing this provides a form that very clearly looks like the most basic form of Bayes' Law:  $p(\theta_i|\mathbf{D}) = p(\theta_i)p(\mathbf{D}|\theta_i)/p(\mathbf{D})$ . We can contrast this with the standard likelihood approach in the social sciences (King 1989, p.22), which overtly ignores information available through a prior and has no use for the denominator above:  $L(\hat{\theta}|y) \propto p(y|\hat{\theta})$ , in King's notation using proportionality since the objective is simply to find the mode and curvature around this mode, thus making constants unimportant. Furthermore, in the continuous case, where the support of  $\theta$  is over some portion of the real line, and possibly all of it, the summation in (1.13) is replaced with an integral. The continuous case is covered in the next chapter.

■ **Example 1.2: Monty Hall.** The well-known Monty Hall problem (Selvin 1975) can be analyzed using Bayes' Law. Suppose that you are on the classic game show *Let's Make a Deal* with its personable host Monty Hall, and you are to choose one of three doors, A, B, or C. Behind two of the doors are goats and behind the third door is a new car, and each door is equally likely to award the car. Thus, the probabilities of selecting the car for each door at the beginning of the game are simply:

$$p(A) = \frac{1}{3}, \quad p(B) = \frac{1}{3}, \quad p(C) = \frac{1}{3}.$$

After you have picked a door, say A, before showing you what is behind that door Monty opens another door, say B, revealing a goat. At this point, Monty gives you

the opportunity to switch doors from A to C if you want to. What should you do? The psychology of this approach is to suggest the idea to contestants that they must have picked the correct door and Monty is now trying to induce a change. A naïve interpretation is that you should be indifferent to switching due to a perceived probability of 0.5 of getting the car with either door since there are two doors left. To see that this is false, recall that Monty is not a benign player in this game. He is deliberately trying to deny you the car. Therefore consider his probability of opening door B. Once you have picked door A, success is clearly conditional on what door of the three possibilities actually provides the car since Monty has this knowledge and the contestant does not. After the first door selection, we can define the three conditional probabilities as follows:

---

The probability that Monty opens door B,  
given the car is behind A:  $p(B_{\text{Monty}}|A) = \frac{1}{2}$

The probability that Monty opens door B,  
given the car is behind B:  $p(B_{\text{Monty}}|B) = 0$

The probability that Monty opens door B,  
given the car is behind C:  $p(B_{\text{Monty}}|C) = 1.$

---

Using the definition of conditional probability, we can derive the following three joint probabilities:

$$p(B_{\text{Monty}}, A) = p(B_{\text{Monty}}|A)p(A) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

$$p(B_{\text{Monty}}, B) = p(B_{\text{Monty}}|B)p(B) = 0 \times \frac{1}{3} = 0$$

$$p(B_{\text{Monty}}, C) = p(B_{\text{Monty}}|C)p(C) = 1 \times \frac{1}{3} = \frac{1}{3}.$$

Because there are only three possible events that cover the complete sample space, and these events are non-overlapping (mutually exclusive), they form a partition of the sample space. Therefore the sum of these three events is the unconditional probability of Monty opening door B, which we obtain with the Total Probability Law:

$$\begin{aligned} p(B_{\text{Monty}}) &= p(B_{\text{Monty}}, A) + p(B_{\text{Monty}}, B) + p(B_{\text{Monty}}, C) \\ &= \frac{1}{6} + 0 + \frac{1}{3} = \frac{1}{2}. \end{aligned}$$

Now we can apply Bayes' Law to obtain the two probabilities of interest:

$$p(A|B_{\text{Monty}}) = \frac{p(A)}{p(B_{\text{Monty}})}p(B_{\text{Monty}}|A) = \frac{\frac{1}{3}}{\frac{1}{2}} \times \frac{1}{2} = \frac{1}{3}$$

$$p(C|B_{\text{Monty}}) = \frac{p(C)}{p(B_{\text{Monty}})}p(B_{\text{Monty}}|C) = \frac{\frac{1}{3}}{\frac{1}{2}} \times 1 = \frac{2}{3}.$$

Therefore you are twice as likely to win the car if you switch to door C! This example demonstrates that Bayes' Law is a fundamental component of probability calculations, and the principle will be shown to be the basis for an inferential system of statistical analysis. For a nice generalization to  $N$  doors, see McDonald (1999).

---

## 1.5 Conditional Inference with Bayes' Law

To make the discussion more concrete and pertinent, consider a simple problem in sociology and crime studies. One quantity of interest to policy-makers is the recidivism rate of prisoners released after serving their sentence. The quantity of interest is the probability of committing an additional crime and returning to prison. Notice that this is a very elusive phenomenon. Not only are there regional, demographic, and individualistic differences, but the aggregate probability is also constantly in flux, given entries and exits from the population as well as exogenous factors (such as the changing condition of the economy). Typically, we would observe a change in law or policy at the state or federal level, and calculate a point estimate from observed recidivism that follows.

Perhaps we should not assume that there is some fixed value of the recidivism probability,  $\mathcal{A}$ , and that it should be estimated with a single point, say  $\bar{\mathcal{A}}$ . Instead, consider this unknown quantity in probabilistic terms as the random variable  $A$ , which means conceptualizing a distribution for the probability of recidivism. Looking at data from previous periods, we might have some reasonable guess about the distribution of this probability parameter,  $p(A)$ , which is of course the prior distribution since it is not conditional on the information at hand,  $B$ .

In all parametric statistical inference, a model is proposed and tested in which an event has some probability of occurring given a specific value of the parameter. This is the case for both Bayesian and traditional approaches, and is just a recognition that the researcher must specify a data generation model. Let us call this quantity  $p(B|A)$ , indicating that for posited values of recidivism, we would expect to see a particular pattern of events. For instance, if recidivism suddenly became much higher in a particular state, then there might be pressure on the legislature to toughen sentencing and parole laws. This is a probability model and we do not need to have a specific value of  $A$  to specify a parametric form (i.e., PMF or PDF). Of course what we are really interested in is  $p(A|B)$ , the (posterior) distribution of  $A$  after having observed an event, which we obtain using Bayes' Law:  $p(A|B) = \frac{p(A)}{p(B)}p(B|A)$ . From a public policy perspective, this is equivalent to asking how do recidivism rates change for given statutes.

We are still missing one component of the right-hand-side of Bayes' Law here, the *unconditional* probability of generating the legal or policy event,  $p(B)$ . This is interpretable as the denominator of (1.13), but to a Bayesian this is an unimportant *probability* statement

since  $B$  has already been observed and therefore has probability one of occurring. Recall that for Bayesians, observed quantities are fixed and unobserved quantities are assigned probability statements. So there is no point in treating  $B$  probabilistically if the actual facts are sitting on our desk right now. This does not mean that everything is known about all possible events, missing events, or events occurring in the future. It just means that everything is known about *this* event. So the only purpose for  $p(B)$  in this context is to make sure that  $p(A|B)$  sums or integrates to one.

This last discussion suggests simply treating  $p(B)$  as a normalizing constant since it does not change the *relative* probabilities for  $A$ . Maybe this is a big conceptual leap, but if we could recover unconditional  $p(B)$  later, it is convenient to just use it then to make the conditional statement,  $p(A|B)$ , a properly scaled probability statement. So if  $p(A|B)$  summed or integrated to five instead of one, we would simply divide everywhere by five and lose nothing but the agony of carrying  $p(B)$  through the calculations. If we temporarily ignore  $p(B)$ , then:

$$p(A|B) \propto p(A)p(B|A), \quad (1.14)$$

where “ $\propto$ ” means “proportional to” (i.e., the *relative* probabilities are preserved). So the final estimated probability of recidivism (in our example problem) given some observed behavior, is proportional to prior notions about the distribution of the probability times the parametric model assumed to be generating the new observed event. The conditional probability of interest on the left-hand side of (1.14) is a balance between things we have already seen or believe about recidivism,  $p(A)$ , and the contribution from the new observation,  $p(B|A)$ . It is important to remember that there are occasions where the data are more influential than the prior and vice-versa. This is comforting since if the data are poor in size or information we want to rely more on prior knowledge, prior research, researcher or practitioner information and so on. Conversely, if the data are plentiful and highly informed, then we should not care much about the form of the prior information. Remarkably, the Bayesian updating process in (1.14) has this trade-off automatically built-in to the process.

As described, this is an ideal paradigm for inference in the social and behavioral sciences, since it is consentaneously desirable to build models that test theories with newly observed events or data, but also based on previous research and knowledge. We never start a data analysis project with absolutely no *a priori* notions whatsoever about the state of nature (or at least we should not!). This story actually gets better. As the number of events increases,  $p(B|A)$  becomes progressively more influential in determining  $p(A|B)$ . That is, the greater the number of our new observations, the less important are our previous convictions:  $p(A)$ . Also, if either of the two distributions,  $p(A)$  and  $p(B|A)$ , are widely dispersed relative to the other, then this distribution will have less of an impact on the final probability statement. We will see this principle detailed-out in Chapter 2. The natural weighting of these two distributions suitably reflects relative levels of uncertainty in the two quantities.

### 1.5.1 Statistical Models with Bayes' Law

The statistical role of the quantities in (1.14) has not yet been identified since we have been talking abstractly about “events” rather than conventional data. The goal of inference is to make claims about unknown quantities using data currently in hand. Suppose that we designate a generic Greek character to denote an unobserved parameter that is the objective of our analysis. As is typical in these endeavors, we will use  $\theta$  for this purpose. What we usually have available to us is generically (and perhaps a little vaguely) labeled  $\mathbf{D}$  for data. Therefore, the objective is to obtain a probabilistic statement about  $\theta$  given  $\mathbf{D}$ :  $p(\theta|\mathbf{D})$ .

Inferences in this book, and in the majority of Bayesian and non-Bayesian statistics, are made by first specifying a parametric model for the data generating process. This defines what the data should be expected to look like given a specific probabilistic function conditional on unknown variable values. These are the common probability density functions (continuous data) and probability mass functions (discrete data) that we already know, such as normal, binomial, chi-square, etc., denoted by  $p(\mathbf{D}|\theta)$ .

Now we can relate these two conditional probabilities using (1.14):

$$\pi(\theta|\mathbf{D}) \propto p(\theta)p(\mathbf{D}|\theta), \quad (1.15)$$

where  $p(\theta)$  is a formalized statement of the prior knowledge about  $\theta$  before observing the data. If we know little, then this prior distribution should be a vague probabilistic statement and if we know a lot then this should be a very narrow and specific claim. The right-hand side of (1.15) implies that the *post*-data inference for  $\theta$  is a compromise between prior information and the information provided by the new data, and the left-hand side of (1.15) is the posterior distribution of  $\theta$  since it provides the updated distribution for  $\theta$  after conditioning on the data.

Bayesians describe  $\pi(\theta|\mathbf{D})$  to readers via distributional summaries such as means, modes, quantiles, probabilities over regions, traditional-level probability intervals, and graphical displays. Once the posterior distribution has been calculated via (1.15), everything about it is known and it is entirely up to the researcher to highlight features of interest. Often it is convenient to report the posterior mean and variance in papers and reports since this is what non-Bayesians do by default. We can calculate the posterior mean using an expected value calculation, confining ourselves here to the continuous case:

$$E[\theta|\mathbf{D}] = \int_{-\infty}^{\infty} \theta \pi(\theta|\mathbf{D}) d\theta \quad (1.16)$$

and the posterior variance via a similar process:

$$\begin{aligned}
\text{Var}[\theta|\mathbf{D}] &= E[(\theta - E[\theta|\mathbf{D}])^2|\mathbf{D}] \\
&= \int_{-\infty}^{\infty} (\theta - E[\theta|\mathbf{D}])^2 \pi(\theta|\mathbf{D}) d\theta \\
&= \int_{-\infty}^{\infty} (\theta^2 - 2\theta E[\theta|\mathbf{D}] + E[\theta|\mathbf{D}]^2) \pi(\theta|\mathbf{D}) d\theta \\
&= \int_{-\infty}^{\infty} \theta^2 \pi(\theta|\mathbf{D}) d\theta - 2E[\theta|\mathbf{D}] \int_{-\infty}^{\infty} \theta \pi(\theta|\mathbf{D}) d\theta + \left( \int_{-\infty}^{\infty} \theta \pi(\theta|\mathbf{D}) d\theta \right)^2 \\
&= E[\theta^2|\mathbf{D}] - E[\theta|\mathbf{D}]^2
\end{aligned} \tag{1.17}$$

given some minor regularity conditions about switching the order of integration (see Casella and Berger 2002, Chapter 1). An obvious summary of this posterior would then be the vector  $(E[\theta|\mathbf{D}], \sqrt{\text{Var}[\theta|\mathbf{D}]})$ , although practicing Bayesians tend to prefer reporting more information.

Researchers sometimes summarize the Bayesian posterior distribution in a deliberately traditional, non-Bayesian way in an effort to communicate with some readers. The posterior mode corresponds to the maximum likelihood point estimate and is calculated by:

$$M(\theta) = \underset{\theta}{\text{argmax}} \pi(\theta|\mathbf{D}), \tag{1.18}$$

where  $\text{argmax}$  function specifies the value of  $\theta$  that maximizes  $\pi(\theta|\mathbf{D})$ . Note that the denominator of Bayes' Law is unnecessary here since the function has the same mode with or without including it. The accompanying measure of curvature (e.g., Fisher Information, defined in Appendix A) can be calculated with standard analytical tools or more conveniently from MCMC output with methods introduced in Chapter 9. The posterior median is a slightly less popular choice for a Bayesian point estimate, even though its calculation from MCMC output is trivial from just sorting empirical draws and determining the mid-point.

■ **Example 1.3: Summarizing a Posterior Distribution from Exponential Data.**

Suppose we had generic data,  $\mathbf{D}$ , distributed  $p(\mathbf{D}|\theta) = \theta e^{-\theta\mathbf{D}}$ , which can be either a single scalar or a vector for our purposes. Thus  $\mathbf{D}$  is exponentially distributed with the support  $[0:\infty)$ ; see Appendix A for details on this probability density function. We also need to specify a prior distribution for  $\theta$ :  $p(\theta) = 1$ , where  $\theta \in [0:\infty)$ . Obviously this prior distribution does not constitute a “proper” distribution in the Kolmogorov sense since it does not integrate to one (infinity, in fact). We should not let this bother us since this effect is canceled out due to its presence in both the numerator and denominator of Bayes' Law (a principle revisited in Chapter 4 in greater detail).



This type of prior is often used to represent high levels of prior uncertainty, although it is not completely *uninformative*. Using (1.15) now, we get:

$$\pi(\theta|\mathbf{D}) \propto p(\theta)p(\mathbf{D}|\theta) = (1)\theta e^{-\theta\mathbf{D}} = \theta e^{-\theta\mathbf{D}}. \quad (1.19)$$

This posterior distribution has mean:

$$E[\theta|\mathbf{D}] = \int_0^{\infty} (\theta) (\theta e^{-\theta\mathbf{D}}) d\theta = \frac{2}{\mathbf{D}^3}, \quad (1.20)$$

which is found easily using two iterations of integration-by-parts. Also, the expectation of  $\theta^2|\mathbf{D}$  is:

$$E[\theta^2|\mathbf{D}] = \int_0^{\infty} (\theta^2) (\theta e^{-\theta\mathbf{D}}) d\theta = \frac{6}{\mathbf{D}^4}, \quad (1.21)$$

which is found using three iterations of integration-by-parts now. So the posterior variance is:

$$\text{Var}[\theta|\mathbf{D}] = E[\theta^2|\mathbf{D}] - E[\theta|\mathbf{D}]^2 = 6\mathbf{D}^{-4} - 4\mathbf{D}^{-6}. \quad (1.22)$$

The notation would be slightly different if  $\mathbf{D}$  were a vector.

Using these quantities we can perform an intuitive Bayesian hypothesis test, such as asking what is the posterior probability that  $\theta$  is positive  $p(\theta|\mathbf{D}) > 0$ . In the context of a regression coefficient, this would be the probability that increases in the corresponding  $X$  explanatory variable have a positive affect on the  $Y$  outcome variable. Testing will be discussed in detail in Chapter 7. We can also use these derived quantities to create a Bayesian version of a confidence interval, the credible interval for some chosen  $\alpha$  level:

$$\left[ E[\theta|\mathbf{D}] - \sqrt{\text{Var}[\theta|\mathbf{D}]} f_{\alpha/2}, E[\theta|\mathbf{D}] + \sqrt{\text{Var}[\theta|\mathbf{D}]} f_{1-\alpha/2} \right], \quad (1.23)$$

where  $f_{\alpha/2}$  and  $f_{1-\alpha/2}$  are lower and upper tail values for some assumed or empirically observed distribution for  $\theta$  (Chapter 2).

The purpose of this brief discussion is to highlight the fact that conditional probability underlies the ability to update previous knowledge about the distribution of some unknown quantity. This is precisely in line with the iterative scientific method, which postulates theory improvement through repeated specification and testing with data. The Bayesian approach combines a formal structure of rules with the mathematical convenience of probability theory to develop a process that “learns” from the data. The result is a powerful and elegant tool for scientific progress in many disciplines.

## 1.6 Science and Inference

This is a book about the scientific process of discovery in the social and behavioral sciences. Data analysis is best practiced as a theory-driven exploration of collected observations with the goal of uncovering important and unknown effects. This is true regardless of academic discipline. Yet some fields of study are considered more rigorously analytical in this pursuit than others.

The process described herein is that of *inference*: making probabilistic assertions about unknown quantities. It is important to remember that “in the case of uncertain inference, however, the very uncertainty of uncertain predictions renders question of their proof or disproof almost meaningless” (Wilkinson 1977). Thus, confusion sometimes arises in the interpretation of the inferential process as a scientific, investigative endeavor.

### 1.6.1 The Scientific Process in Our Social Sciences

Are the social and behavioral sciences truly “scientific”? This is a question asked about fields such as sociology, political science, economics, anthropology, and others. It is not a question about whether serious, rigorous, and important work has been done in these endeavors; it is a question about the research process and whether it conforms to the empirico-deductive model that is historically associated with the natural sciences. From a simplistic view, this is an issue of the conformance of research in the social and behavioral sciences to the so-called scientific method. Briefly summarized, the scientific method is characterized by the following steps:

- ▷ Observe or consider some phenomenon.
- ▷ Develop a theory about the cause(s) of this phenomenon and articulate it in a specific hypothesis.
- ▷ Test this hypothesis by developing a model to fit experimentally generated or collected observational data.
- ▷ Assess the quality of the fit to the model and modify the theory if necessary, repeating the process.

This is sometimes phrased in terms of “prediction” instead of theory development, but we will use the more general term. If the scientific method as a process were the defining criterion for determining what is scientific and what is not, then it would be easy to classify a large proportion of the research activities in the social and behavioral sciences as scientific. However useful this typology is in teaching children about empirical investigation, it is a poor standard for judging academic work.

Many authors have posited more serviceable definitions. Braithwaite (1953, p.1) notes:

The function of a science, in this sense of the word, is to establish general laws covering the behavior of the empirical events or objects with which the science in question is concerned, and thereby to enable us to connect together our knowledge of the separately known events, and to make reliable predictions of events as yet unknown.

The core of this description is the centrality of empirical observation and subsequent accumulation of knowledge. Actually, “science” is the Latin word for knowledge. Legendary psychologist B. F. Skinner (1953, p.11) once observed that “science is unique in showing a cumulative process.” It is clear from the volume and preservation of published research that social and behavioral scientists *are* actively engaged in empirical research and knowledge accumulation (although the quality and permanence of this foundational knowledge might be judged to differ widely by field). So what is it about these academic pursuits that makes them only suspiciously scientific to some? The three defining characteristics about the *process* of scientific investigation are empiricism, objectivity, and control (Singleton and Straight 2004). This is where there is lingering and sometimes legitimate criticism of the social and behavioral sciences as being “unscientific.”

The social and behavioral sciences are partially empirical (data-oriented) and partially normative (value-oriented), the latter because societies develop norms about human behavior, and these norms permeate academic thought prior to the research process. For instance, researchers investigating the onset and development of AIDS initially missed the effects of interrelated social factors such as changes in behavioral risk factors, personal denial, and reluctance to seek early medical care on the progress of the disease as a sociological phenomenon (Kaplan *et al.* 1987). This is partially because academic investigators as well as health professionals made normative assumptions about individual responses to sociological effects. Specifically, researchers investigating human behavior, whether political, economic, sociological, psychological, or otherwise, cannot completely divorce their prior attitudes about some phenomenon of interest the way a physicist or chemist can approach the study of the properties of thorium: atomic number 90, atomic symbol Th, atomic weight 232.0381, electron configuration  $[Rn]7s^26d^2$ . This criticism is distinct from the question of objectivity; it is a statement that students of human behavior are themselves human.

We are also to some extent driven by the quality and applicability of our tools. Many fields have radically progressed after the introduction of new analytical devices. Therefore, some researchers may have a temporary advantage over others, and may be able to answer more complex questions: “It comes as no particular surprise to discover that a scientist formulates problems in a way which requires for their solution just those techniques in which he himself is especially skilled” (Kaplan 1964). The objective of this book is to “level the pitch” by making an especially useful tool more accessible to those who have thus far been accordingly disadvantaged.

## 1.6.2 Bayesian Statistics as a Scientific Approach to Social and Behavioral Data Analysis

The standard frequentist interpretation of probability and inference assumes an infinite series of trials, replications, or experiments using the same research design. The “objectivist” paradigm is typically explained and justified through examples like multiple tosses of a coin, repeated measurements of some physical quantity, or samples from some ongoing process like a factory output. This perspective, which comes directly from Neyman and Pearson (1928a, 1928b, 1933a, 1933b, 1936a, 1936b), and was formalized by Von Mises (1957) among others, is combined with an added Fisherian fixation with p-values in typical inference in the social and behavioral sciences (Gill 1999). Efron (1986), perhaps overly kindly, calls this a “rather uneasy alliance.”

Very few, if any, social scientists would be willing to seriously argue that human behavior fits this objectivist long-run probability model. Ideas like “personal utility,” “legislative ideal points,” “cultural influence,” “mental states,” “personality types,” and “principal-agent goal discrepancy” do not exist as parametrically uniform phenomena in some physically tangible manner. In direct contrast, the Bayesian or “subjective” conceptualization of probability is the degree of belief that the individual researcher is willing to personally assign and defend. This is the idea that an individual *personally* assigns a probability measure to some event as an expression of uncertainty about some event that may only be relevant to one observational situation or experiment.

The central idea behind subjective probability is the assignment of a prior probability based on what information one currently possesses and under what circumstances one would be willing to place an even wager. Naturally, this probability is updated as new events occur, therefore incorporating serial events in a systematic manner. The core disagreement between the frequentist notion of objective probability and the Bayesian idea of subjective probability is that frequentists see probability measure as a property of the outside world and Bayesians view probability as a personal internalization of observed uncertainty. The key defense of the latter view is the inarguable point that all statistical models are subjective: decisions about variable specifications, significance thresholds, functional forms, and error distributions are completely nonobjective.<sup>1</sup> In fact, there are instances when Bayesian subjectivism is more “objective” than frequentist objectivism with regard to the impact of irrelevant information and arbitrary decision rules (e.g., Edwards, Lindman, and Savage 1963, p.239).

---

<sup>1</sup>As a brief example, consider common discussions of reported analyses in social science journals and books that talk about reported model parameters being “of the wrong sign.” What does this statement mean? The author is asserting that the statistical model has produced a regression coefficient that is positive when it was *a priori* expected to be negative or vice versa. What is this statement in effect? It is a prior statement about knowledge that existed before the model was constructed. Obviously this is a form of the Bayesian prior without being specifically articulated as such.

Given the existence of subjectivity in all scientific data analysis endeavors,<sup>2</sup> one should prefer the inferential paradigm that gives the most *overt* presentation of model assumptions. This is clearly the Bayesian subjective approach since both prior information and posterior uncertainty are given with specific, clearly stated model assumptions. Conversely, frequentist models are rarely presented with caveats such as “Caution: the scientific conclusions presented here depend on repeated trials that were never performed,” or “Warning: prior assumptions made in this model are not discussed or clarified.” If there is a single fundamental scientific tenet that underlies the practice and reporting of empirical evidence, it is the idea that all important model characteristics should be provided to the reader. It is clear then which of the two approaches is more “scientific” by this criterion. While this discussion specifically contrasts Bayesian and frequentist approaches, likelihood inference is equally subjective in every way, and as already explained, ignores available information.

These ideas of what sort of inferences social scientists make are certainly not new or novel. There is a rich literature to support the notion that the Bayesian approach is more in conformance with widely accepted scientific norms and practices. Poirer (1988, p.130) stridently makes this point in the case of prior specifications:

I believe that subjective prior beliefs should play a *formal* role so that it is easier to investigate their impact on the results of the analysis. Bayesians must live with such honesty whereas those who introduce such beliefs informally need not.

The core of this argument is the idea that if the prior contains information that pertains to the estimation problem, then we are foolish to ignore it simply because it does not neatly fit into some familiar statistical process. For instance, Theil and Goldberger (1961) suggested “mixed” estimation some time ago, which is a way to incorporate prior knowledge about coefficients in a standard linear regression model by mixing earlier estimates into the estimation process and under very general assumptions is found to be simultaneously best linear unbiased with respect to both sample and prior information (see also Theil [1963]). This notion of combining information from multiple sources is not particularly controversial among statisticians, as observed by Samaniego and Reneau (1994, p.957):

If a prior distribution contains “useful” information about an unknown parameter, then the Bayes estimator with respect to that prior will outperform the best frequentist rule. Otherwise, it will not.

A more fundamental advantage to Bayesian statistics is that both prior and posterior parameter estimates are assumed to have a distribution and therefore give a more realistic picture of uncertainty that is also more useful in applied work:

With conventional statistics, the only uncertainty admitted to the analysis is

---

<sup>2</sup>See Press and Tanur (2001) for a fascinating account of the role of researcher-introduced subjectivity in a number of specific famous scientific breakthroughs, including discoveries by Galileo, Newton, Darwin, Freud, and Einstein.

sampling uncertainty. The Bayesian approach offers guidance for dealing with the myriad sources of uncertainty faced by applied researchers in real analyses.

Western (1999, p.20). Lindley (1986, p.7) expresses a more biting statement of preference:

Every statistician would be a Bayesian if he took the trouble to read the literature thoroughly and was honest enough to admit he might have been wrong.

This book rests on the perspective, sampled above, that the Bayesian approach is not only useful for social and behavioral scientists, but it also provides a more compatible methodology for analyzing data in the manner and form in which it arrives in these disciplines. As we describe in subsequent chapters, Bayesian statistics establishes a rigorous analytical platform with clear assumptions, straightforward interpretations, and sophisticated extensions. For more extended discussions of the advantages of Bayesian analysis over alternatives, see Berger (1986b), Dawid (1982), Efron (1986), Good (1976), Jaynes (1976), and Zellner (1985). We now look at how the Bayesian paradigm emerged over the last 250 years.

## 1.7 Introducing Markov Chain Monte Carlo Techniques

In this section we briefly discuss Bayesian computation and give a preview of later chapters. The core message is that these algorithms are relatively simple to understand in the abstract.

Markov chain Monte Carlo (MCMC) set the Bayesians free. Prior to 1990, it was relatively easy to specify an interesting and realistic model with actual data whereby standard results were unobtainable. Specifically, faced with a high dimension posterior resulting from a regression-style model, it was often very difficult or even impossible to perform multiple integration across the parameter space to produce a regression table of marginal summaries. The purpose of MCMC techniques is to replace this difficult analytical integration process with iterative work by the computer. When calculations similar to (1.16) are multidimensional, there is a need to summarize each marginal distribution to provide useful results to readers in a table or other format for journal submission. The basic principle behind MCMC techniques is that if an iterative chain of computer-generated values can be set up carefully enough, and run long enough, then *empirical* estimates of integral quantities of interest can be obtained from summarizing the observed output. If each visited multidimensional location is recorded as a row vector in a matrix, then the marginalization for some parameter of interest is obtained simply by summarizing the individual dimension down the corresponding column. So we replace an analytical problem with a sampling problem, where the sampling process has the computer perform the difficult and repetitive processes. This is an enormously important idea to Bayesians and to others since it frees researchers

from having to make artificial simplifications to their model specifications just to obtain describable results.

These Markov chains are successive quantities that depend probabilistically only on the value of their immediate predecessor: the *Markovian property*. In general, it is possible to set up a chain to estimate multidimensional probability structures (i.e., desired probability distributions), by starting a Markov chain in the appropriate sample space and letting it run until it settles into the target distribution. Then when it runs for some time confined to this particular distribution, we can collect summary statistics such as means, variances, and quantiles from the simulated values. This idea has revolutionized Bayesian statistics by allowing the empirical estimation of probability distributions that could not be analytically calculated.

### 1.7.1 Simple Gibbs Sampling

As a means of continuing the discussion about conditional probability and covering some basic principles of the R language, this section introduces an important, and frequently used Markov chain Monte Carlo tool, the Gibbs sampler. The idea behind a Gibbs sampler is to get a marginal distribution for each variable by iteratively conditioning on interim values of the others in a continuing cycle until samples from this process empirically approximate the desired marginal distribution. Standard regression tables that appear in journals are simply marginal descriptions. There will be much more on this topic in Chapter 10 and elsewhere, but here we will implement a simple but instructive example.

As outlined by Example 2 in Casella and George (1992), suppose that we have two conditional distributions, where they are conditional on each other such that the parameter of one is the variable of interest in the other:

$$f(x|y) \propto y \exp[-yx], \quad f(y|x) \propto x \exp[-xy], \quad 0 < x, y < B < \infty. \quad (1.24)$$

These conditional distributions are both exponential probability density functions (see Appendix B for details). The upper bound,  $B$ , is important since without it there is no finite joint density and the Gibbs sampler will not work. It is possible, but not particularly pleasant, to perform the correct integration steps to obtain the desired marginal distributions:  $f(x)$  and  $f(y)$ . Instead we will let the Gibbs sampler do the work computationally rather than us do it analytically.

The Gibbs sampler is defined by first identifying conditional distributions for each parameter in the model. These are conditional in the sense that they have dependencies on other parameters, and of course the data, which emerge from the model specification. The “transition kernel” for the Markov chain is created by iteratively cycling through these distributions, drawing values that are conditional on the latest draws of the dependencies. It is proven that this allows us to run a Markov chain that eventually settles into the desired limiting distribution that characterizes the marginals. In other language, it is an iterative process that cycles through conditional distributions until it reaches a stable status whereby

future samples characterize the desired distributions. The important theorem here assures us that when we reach this stable distribution, the autocorrelated sequence of values can be treated as an iid sample from the marginal distributions of interest. The amazing part is that this is accomplished simply by ignoring the time index, i.e., putting the values in a “bag” and just “shaking it up” to lose track of the order of occurrence. Gibbs sampling is actually even more general than this. Chib (1995) showed how Gibbs sampling can be used to compute the marginal distribution of the sample data, i.e., the denominator of (1.13), by using the individual parameter draws. This quantity is especially useful in Bayesian hypothesis testing and model comparison, as we shall see in Chapter 7. The second half of this text applies this tool and similar methods of estimation.

For two parameters,  $x$  and  $y$ , this process involves a starting point,  $[x_0, y_0]$ , and the cycles are defined by drawing random values from the conditionals according to:

$$\begin{array}{ll} x_1 \sim f(x|y_0), & y_1 \sim f(y|x_1) \\ x_2 \sim f(x|y_1), & y_2 \sim f(y|x_2) \\ x_3 \sim f(x|y_2), & y_3 \sim f(y|x_3) \\ : & : \\ : & : \\ x_m \sim f(x|y_{m-1}), & y_m \sim f(y|x_m). \end{array}$$

If we are successful, then after some reasonable period the values  $x_j, y_j$  are safely assumed to be empirical samples from the correct marginal distribution. There are many theoretical and practical concerns that we are ignoring here, and the immediate objective here is to give a rough overview.

The following steps indicate how the Gibbs sampler is set up and run:

- ▷ Set the initial values:  $B = 10$ , and  $m = 50,000$ .  $B$  is the parameter that ensures that the joint distribution is finite, and  $m$  is the desired number of generated values for  $x$  and  $y$ .
- ▷ Create  $x$  and  $y$  vectors of length  $m$  where the first value of each is a starting point uniformly distributed over the support of  $x$  and  $y$ , and all other vector values are filled in with unacceptable entries greater than  $B$ .
- ▷ Run the chain for  $m = 50,000 - 1$  iterations beginning at the starting points. At each iteration, fill-in and save only sampled exponential values that are less than  $B$ , repeating this sampling procedure until an acceptable value is drawn to replace the unacceptable  $B + 1$  in that position.
- ▷ Throw away some early part of the chain where it has not yet converged.
- ▷ Describe the marginal distributions of  $x$  and  $y$  with the remaining empirical values.

This leads to the following R code, which can be retyped verbatim, obtained from the book’s webpage, or the book’s R package **BaM** to replicate this example:

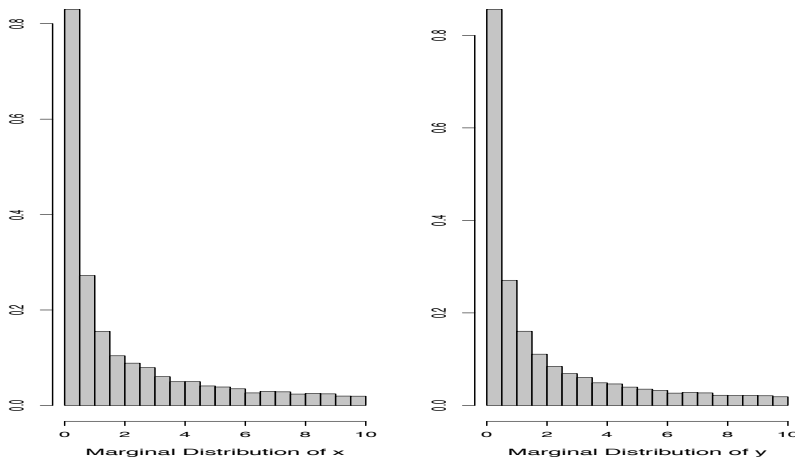


```

B <- 10; m <- 50000
gibbs.expo <- function(B,m) {
  x <- c(runif(1,0,B),rep((B+1),length=(m-1)))
  y <- c(runif(1,0,B),rep((B+1),length=(m-1)))
  for (i in 2:m) {
    while(x[i] > B) x[i] <- rexp(1,y[i-1])
    while(y[i] > B) y[i] <- rexp(1,x[i])
  }
  return(cbind(x,y))
}

gibbs.expo(B=5, m=500)

```



**FIGURE 1.1:** GIBBS SAMPLING, MARGINAL EXPONENTIALS

These samples are summarized by histograms of the empirical results for  $x$  and  $y$  in Figure 1.1, where  $m = 50,000$  samples are drawn and the first 40,000 are discarded (these are called “burn-in” values). It is clear from the figure that the marginal distributions are exponentially distributed. We can recover parameters by using the empirical draws to calculate sample statistics. This part of the MCMC process is actually quite trivial once we are convinced that there has been convergence of the Markov chain. In later chapters we will see this process in a more realistic, and therefore detailed, setting. This example is intended to give an indication of activities to come and to reinforce the linkage between Bayesian inference and modern statistical computing.

### 1.7.2 Simple Metropolis Sampling

Another Markov chain Monte Carlo tool with wide use is the Metropolis algorithm from statistical physics (Metropolis *et al.* 1953). The Metropolis algorithm is more flexible than

Gibbs sampling because it works with the joint distribution rather than a full listing of conditional distributions for the parameters in the model. As a result, many variations have been developed to satisfy particular sampling challenges posed by complicated models and ill-behaved target functions. Later chapters will cover these extensions in detail.

The essential idea behind the Metropolis algorithm is that, while we cannot easily generate values from the joint (posterior) distribution of interest, we can often find a “similar” distribution that is easy to sample from. Obviously we need to make sure that this alternative distribution is defined over the same support as the target distribution and that it does not radically favor areas of low density of this target. Once a candidate point in multivariate space has been produced by this candidate-generating distribution we will accept or reject it based upon characteristics of the target distribution. The algorithm is characterized by the following steps.

1. The candidate-generating distribution proposes that we move to some other point by drawing a point from *its* generating mechanism.

2. If this point produces a step on the target distribution that is of *higher* density, then we will always go there.

3. If this point produces a step on the target distribution that is of *lower* density, then we will go there probabilistically proportional to how much lower the step is in density.

Thus it is easy to see that the Markov chain “wanders around” the target density describing it as it goes and favoring higher density regions. The nice part is that the Markov chain will also explore other lower density regions as well, but with lower probability as we would want. Analogously, consider locking a house cat in large room with features that are attractive to cats (the high density regions of the posterior), and features that are unattractive to cats (the low density regions of the posterior). Anyone who has spent time with house cats can see at least some Markovian feature to their nature, as well as an innate curiosity. As our feline Markov chain wanders the room in a memory-less state, we record the coordinates of their travel. Over time we will find that the cat spends more time in the attractive areas, but still occasionally investigates the unattractive areas. If this attractiveness is proportional to density we want to describe, then the cat eventually produces description of the posterior distribution.

We can more precisely describe this algorithm. Suppose we have a two-dimensional target distribution,  $p(x, y)$ , which can be a posterior distribution from a Bayesian model, or any other form that is hard to marginalize, i.e., produce individual distributions  $p(x)$  and  $p(y)$ . A single Metropolis step is produced by:

1. Sample  $(x', y')$  from the candidate-generating distribution,  $q(x', y')$ .

2. Sample a value  $u$  from  $u[0 : 1]$ .

3. If

$$a(x', y'|x, y) = \frac{p(x', y')}{p(x, y)} > u$$

then accept  $(x', y')$  as the new destination.

4. Otherwise keep  $(x, y)$  as the new destination.

The result is a chain of values,  $[(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots]$ . Note that unlike the Gibbs sampler, the Metropolis algorithm does not necessarily have to move to a new position at each iteration, and the decision to stay put is considered a Markovian step to the current position (time is consumed by the step).

There are a few technical details that we will worry about in much more detail later beginning in Chapter 10. Often the candidate-generating distribution produces values conditional on the current position,  $q(x', y'|x, y)$ , but this is not strictly necessary. The basic version described here requires that the candidate-generating distribution be symmetrical in its arguments,  $q(x', y'|x, y) = q(x, y|x', y')$ . Also, the choice of candidate-generating distribution can be complicated by the need to match irregularities in the target distribution. Finally, it is important in real applications to run the Markov chain for some initial period to let it settle into the distribution of interest before recording values.

Consider a problem similar to that above, but where we have a joint distribution for the parameters and not the desired marginals (or conditionals as used in the Gibbs sampler). The bivariate exponential distribution for  $x, y \in [0:\infty]$  is given by the function:

$$p(x, y) = \exp[-(\lambda_1 + \lambda)x - (\lambda_2 + \lambda)y - \lambda \max(x, y)], \quad (1.25)$$

with non-negative parameters  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda$ . This model is common in reliability analysis (Marshall and Olkin 1967), where the interpretation is that the first two parameters are the event intensities for systems 1 and 2, and the non-subscripted parameter is the shared intensity between systems. In this literature events are usually machine failures, but for our purposes they can be death, graduation, cabinet dissolutions, divorce, cessation of war, and so on. In this example we have the parameters:

$$\lambda_1 = 0.5, \quad \lambda_2 = 0.5, \quad \lambda = 0.01, \quad B = \max(x) = \max(y) = 8,$$

which produces the bivariate distribution shown in the first panel of Figure 1.2. The maximum in the function makes it a little harder to analytically calculate marginal distributions with integration, so we might want to apply MCMC to save trouble. This is exactly analogous to the process where complicated Bayesian model specifications sometimes make it difficult to describe marginal posteriors for parameters of interest.

To implement the Metropolis algorithm we need a candidate-generating distribution from which to draw potential destinations for the Markov chain. Typically researchers look for some convenient distribution from the commonly used form since software such as R makes drawing values trivial. Here we will exploit the stipulated bounds on the problem and note that the bivariate exponential is enclosed in a big box with length and width equal to  $B = 8$  and maximum height equal to one from the form of (1.25). The process is further covered in Chapter 9, but note here that it is easy to draw points inside this box from scaled uniforms. Nicely, we do not have to rescale the distribution of  $q(x', y')$  because the values are drawn from this distribution but inserted into  $p()$ . It is important to note, without getting too far ahead of ourselves, that a better fitting candidate-generating distribution could be found and that drawing from uniform boxes is not particularly efficient.

To begin we define our function in R according to:

```
biv.exp <- function(x,y,L1,L2,L)
  exp( -(L1+L)*x - (L2+L)*y -L*max(x,y) )
```

So it will return density values for given  $(x, y)$  pairs and specific parameters. The candidate-generating function is:

```
cand.gen <- function(max.x,max.y)
  c(runif(1,0,max.x),runif(1,0,max.y))
```

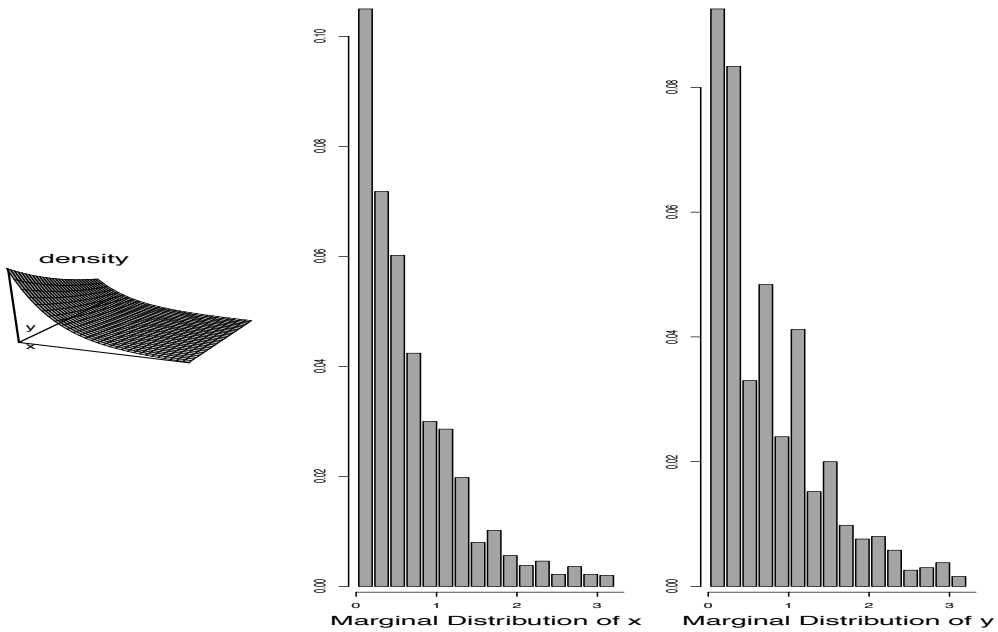
where we could have stipulated the  $B$  value but left the function slightly more general. Markov chains require starting positions and we arbitrarily select  $(x = 0.5, y = 0.5)$  here. The algorithm is now given to be the following R code, which (again) can be retyped verbatim to replicate the example:

```
m <-5000; x<-0.5; y<-0.5; L1<-0.5; L2<-0.5; L<-0.01; B<-8
for (i in 1:m) {
  cand.val <- cand.gen(B,B)
  a <- biv.exp(cand.val[1],cand.val[2],L1,L2,L)
    / biv.exp(x[i],y[i],L1,L2,L)
  if (a > runif(1)) {
    x <- c(x,cand.val[1])
    y <- c(y,cand.val[2])
  }
  else {
    x <- c(x,x[i])
    y <- c(y,y[i])
  }
}
```

The resulting values are shown by the histograms in the second and third panels of Figure 1.2, where the algorithm has been run for  $m = 5,000$  iterations but the first 3,000 are discarded. We could also simply summarize the resulting marginals for  $x$  and  $y$  empirically with means, quantiles, or other simple statistics. The Metropolis algorithm shown here will be expanded and generalized in Chapter 10 by loosening restrictions on the candidate-generating distribution and allowing for hybrid processes that accommodate difficult features in the target distribution. The two MCMC algorithms described here form the basis for all practical work needed to estimate complex Bayesian models in the social sciences.

## 1.8 Historical Comments

Statistics is a relatively new field of scientific endeavor. In fact, for much of its history it was subsumed to various natural sciences as a combination of foster-child and household



**FIGURE 1.2:** METROPOLIS SAMPLING, BIVARIATE EXPONENTIAL

maid: unwanted by its natural parents (mathematics and philosophy), yet necessary to clean things up. Beginning with the work of Laplace (1774, 1781, 1811), Gauss (1809, 1823, 1855), Legendre (1805), and de Morgan (1837, 1838, 1847), statistics began to emerge as a discipline worthy of study on its own merits. The first renaissance occurred around the turn of the last century due to the monumental efforts of Galton (1869, 1875, 1886, 1892), Fisher (1922, 1925a, 1925b, 1934), Neyman and (Egon) Pearson (1928a, 1928b, 1933a, 1933b, 1936a, 1936b), Gossett (as Student, 1908a, 1908b), Edgeworth (1892a, 1892b, 1893a, 1893b), (Karl) Pearson (1892, 1900, 1907, 1920), and Venn (1866). Left out of the twin intellectual developments of frequentist inference from Neyman and Pearson and likelihood inference from Fisher (see Chapter 2, Section 2.6 for details), was the Bayesian paradigm. Sir Thomas Bayes’ famous (and only) essay was published in 1763, two years after his death (Bayes chose to perish before publishing), suggesting to some that he was ambivalent about the approach of applying a uniform prior to a binomial probability parameter. This ingenious work unintendedly precipitated a philosophy about how researcher-specified theories are fit to empirical observations. Interestingly, it was not until the early 1950s that Bayesian statistics became a self-aware branch (Fienberg 2006).

Fisher in particular was hostile to the Bayesian approach and was often highly critical, though not always with substantiated claims: Bayesianism “which like an impenetrable jungle arrests progress towards precision of statistical concepts” (1922, p.311). Fisher also worked to discredit Bayesianism and inverse probability (Bayesianism with an assumed uniform prior) by pressuring peers and even misquoting other scholars (Zabell 1989). Yet Fisher (1935) develops *fiducial inference*, which is an attempt to apply inverse probability

without uniform priors, but this approach fails; Efron (1998, p.105) calls this “Fisher’s biggest blunder.” In fact, Lindley (1958) later proved that fiducial inference is consistent *only* when it is made equivalent to Bayesian inference with a uniform prior. The Neyman-Pearson paradigm was equally unkind to the development of Bayesian statistics, albeit on a less vindictive level. If one is willing to subscribe to the idea of an infinite sequence of samples, then the Bayesian prior is unimportant since the data will overwhelm this prior. Although there are scenarios where this is a very reasonable supposition, generally these are far more difficult to come by in the social and behavioral sciences.

Although Bayesianism had suffered “a nearly lethal blow” from Fisher and Neyman by the 1930s (Zabell 1989), it was far from dead. Scholars such as Jeffreys (1961), Good (1950), Savage (1954, 1962), de Finetti (1972, 1974, 1975), and Lindley (1961, 1965) re-activated interest in Bayesian methods in the middle of the last century in response to observed deficiencies in classical techniques. Lindley and Novick (1978, 1981) published important applied work in education psychology that carefully studied exchangeability and utility from a Bayesian perspective, and Novick *et al.* (1976) developed an early Bayesian software program for estimating simple models: CADA, Computer Assisted Data Analysis. Unfortunately many of the specifications developed by these modern Bayesians, while superior in theoretical foundation, led to mathematical forms that were intractable.<sup>3</sup> Fortunately, this problem has been largely resolved in recent years by a revolution in statistical computing techniques, and this has led to a second renaissance for the Bayesian paradigm (Berger 2001).

Markov chain Monte Carlo (MCMC) techniques solve a lingering problem in Bayesian analysis, and thus earn a special place in this work. Often Bayesian model specifications considered either interesting or realistic produced inference problems that were analytically intractable because they led to high-dimension integral calculations that were impossible to solve analytically. Previous numerical techniques for performing these integrations were often difficult and highly specialized tasks (e.g., Shaw 1988, Stewart and Davis 1986, van Dijk and Kloek 1982, Tierney and Kadane 1986). Beginning with the foundational work of Metropolis *et al.* (1953), Hastings (1970), Peskun (1973), Geman and Geman (1984), and the critical synthesizing essay of Gelfand and Smith (1990), there is now a voluminous literature on Markov chain Monte Carlo. In fact, modern Bayesian statistical practice is intimately and intrinsically tied to stochastic simulation techniques and as a result, these tools are an integral part of this book. We introduce these tools in this chapter in Section 1.7 and in much greater detail in Chapter 10.

Currently the most popular method for generating samples from posterior distributions using Markov chains is the **WinBUGS** program and its Unix-based precursor **BUGS** and the more recent functional equivalent **JAGS**. The name **BUGS** is a pseudo-acronym for *Bayesian inference Using Gibbs Sampling*, referring to the most frequently used method for producing Markov chains. In what constitutes a notable and noble contribution to the Bayesian

---

<sup>3</sup>This led one observer (Evans 1994) to compare Bayesians to “unmarried marriage guidance counselors.”

statistical world, the Community Statistical Research Project at the MRC Biostatistics Unit and the Imperial College School of Medicine at St. Mary's, London provide this high-quality software to users free of charge, and it can be downloaded from their web page: <http://www.mrc-bsu.cam.ac.uk/software/bugs/>. These authors have even made available extensive documentation at the same site by Spiegelhalter *et al.* (1996a, 1996b, 2000, 2012). Alternative ways to use WinBUGS with R as the interface are: **BRugs**, **rbugs**, and **R2WinBUGS**. There are also facilities for calling WinBUGS from **SAS**, **stata**, and **excel**. The **JAGS** program (Just Another Gibbs Sampler) is an engine for the BUGS language that has nearly the same structure as WinBUGS, with only a few syntactical differences. Authored by Martyn Plummer, it is extremely well-developed software that runs on non-windows platforms and is command-line driven rather than point-and-click. It can be downloaded at <http://www-ice.iarc.fr/~martyn/software/jags/>. There are also facilities for calling JAGS from R: **R2jags**, **Rjags**, and **runjags**. Most of the BUGS code in this text are run with JAGS from the command window. Other high-quality R packages using or providing MCMC computing include: **BMS**, **dclone**, **eco**, **glmdm**, **HI**, **lmm**, **MasterBayes**, **mcmc**, **MCMCglmm**, **MCMCpack**, **MNP**, **pscl**, **spBayes**, **tgp**, and **zic**. Of these **MCMCpack** is the most general, whereas most of the others are MCMC implementations to solve a specific problem. Given the rapid pace of R package development, this list is growing rapidly.

---

## 1.9 Exercises

- 1.1 Restate the three general steps of Bayesian inference from page 6 in your own words.
- 1.2 Given  $k$  possible disjoint (non-overlapping) events labeled:  $E_1, \dots, E_k$  where  $k$  could even be infinity, denote  $p(E_i)$  as the mapping from events  $E_i$  to  $[0:1]$  space. Write the Kolmogorov axioms of probability in technical detail.
- 1.3 Rewrite Bayes' Law when the two events are independent. How do you interpret this?
- 1.4 Equation (1.11) on page 12 showed that  $p(A|\mathbf{D}) = p(\mathbf{D}|A)p(A)/(p(\mathbf{D}|A)p(A) + p(\mathbf{D}|B)p(B) + p(\mathbf{D}|C)p(C))$ . Rewrite this expression for  $p(A|\mathbf{D})$  when there are arbitrary  $k \in \mathcal{I}^+$  events including  $A$ .
- 1.5 Suppose  $f(\theta|\mathbf{X})$  is the posterior distribution of  $\theta$  given the data  $\mathbf{X}$ . Describe the shape of this distribution when the mode,  $\underset{\theta}{\operatorname{argmax}} f(\theta|\mathbf{X})$ , is equal to the mean,  $\int_{\theta} \theta f(\theta|\mathbf{X}) d\theta$ .
- 1.6 The Rényi countable additivity axiom is defined by: (1) for any events  $E_1$  and  $E_2$ ,  $p(E_1|E_2) \geq 0$  (and reversed),  $p(E_i|E_i) = 1$ , (2) for disjoint sets  $E_1, \dots$  and

- another arbitrary event  $D$ ,  $p(\cup_{i=1}^{\infty} E_i | D) = \sum_{i=1}^{\infty} p(E_i | D)$ , and (3) for every subset of events,  $E_i, E_j, E_k$ , with  $E_j \subseteq E_k$  and  $p(E_j | E_k) > 0$ , we get  $p(E_i | E_j) = p(E_i, E_j | E_k) / p(E_j | E_k)$ . Show that the Kolmogorov axioms are a special case.
- 1.7 Using R run the Gibbs sampling function given on page 26. What effect do you see in varying the  $B$  parameter? What is the effect of producing 200 sampled values instead of 50,000?
  - 1.8 Some authors have objected to the uniform prior,  $p(\theta) = 1, \theta \in [0:1]$  to describe unknown probabilities in a binomial model and suggested instead the Haldane prior:  $p(\theta) \propto [\theta^{-1}(1 - \theta)]^{-1}$  (Haldane [1938], Novick and Hall [1965], Villegas [1977], ). Plot this prior and the uniform prior over  $[0:1]$  in the same graph.
  - 1.9 Rerun the Metropolis algorithm on page 30 in R but replacing the uniform generation of candidate values in `cand.gen` with a normal truncated to fit in the appropriate range. What differences do you observe?
  - 1.10 The Gibbs sampler described in Section 1.7.1 from Casella and George (1992) was originally done as follows: (1) set initial values for  $B = 5$ ,  $k = 15$ ,  $m = 5,000$ , and the set of accepted values  $(x, y)$  as an empty object, (2) run  $m$  chains of length  $k + 1$  where the first value is the uniformly distributed starting point  $[0 : B]$  and the rest are sampled conditional exponential values that are less than  $B$ , (3) save only the last value from the  $x$  and  $y$  series,  $x_{16}$  and  $y_{16}$  to the stored Markov chain until 5,000 of each are obtained. Implement this alternative algorithm in R and compare it to the output shown in Figure 1.1 on page 27.
  - 1.11 If  $p(\mathbf{D} | \theta) = 0.5$ , and  $p(\mathbf{D}) = 1$ , calculate the value of  $p(\theta | \mathbf{D})$  for priors  $p(\theta)$ ,  $[0.001, 0.01, 0.1, 0.9]$ .
  - 1.12 Buck, Cavanaugh, and Litton (1996) demonstrate the use of Bayesian statistics for radiocarbon dating of Early Bronze Age archaeological samples (seeds and bones) from St. Veit-Klinglberg, Austria. These ten age data points are produced by the Oxford accelerator dating facilities:

Context #	$\mu_i$	$\sigma_i$
758	3275	75
814	3270	80
1235	3400	75
493	3190	75
925	3420	65
923	3435	60
1168	3160	70
358	3340	80
813	3270	75
1210	3200	70



Given the model  $X_i|\mu_i, \sigma_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ , (1) calculate the probability that sample 358 originates between 3300 to 3400 years ago, (2) generate 10,000 samples from the distribution for sample 493 and sample 923 and plot a histogram of these in the same figure (side-by-side), (3) give the proportion of values that overlap, and (4) how do you interpret this overlap probabilistically with regard to the age of the samples?

- 1.13 Sometimes Bayesian results are given as *posterior odds ratios*, which for two possible alternative hypotheses is expressed as:

$$\text{odds}(\theta_1, \theta_2) = \frac{p(\theta_1|\mathbf{D})}{p(\theta_2|\mathbf{D})}.$$

If the prior probabilities for  $\theta_1$  and  $\theta_2$  are identical, how can this be re-expressed using Bayes' Law?

- 1.14 Using the posterior distribution in (1.19) on page 19, produce the posterior mean for  $\theta$  in (1.20) and the posterior variance for  $\theta$  in (1.21).
- 1.15 Suppose we had data,  $\mathbf{D}$ , distributed  $p(\mathbf{D}|\theta) = \theta e^{-\theta\mathbf{D}}$  as in Section 1.5.1 starting on page 18, but now  $p(\theta) = 1/\theta$ , for  $\theta \in (0:\infty)$ . Calculate the posterior mean.
- 1.16 Modify the Gibbs sampler in Section 1.7.1 starting on page 25 to sample from two mutually conditional gamma distributions instead of exponential distributions. The exponential distribution is a simplified form of the rate parameter gamma distribution where the first (shape) parameter is 1 (Appendix B, page 580). Set the two relevant shape parameters to values of your choosing  $\alpha > 1$ . Produce a graphs of the marginal draws.
- 1.17 Since the posterior distribution is a compromise between prior information and the information provided by the new data, then it is interesting to compare relative strengths. Perform an experiment where you flip a coin 10 times, recording the data as zeros and ones. Produce the posterior expected value (mean) for two priors on  $p$  (the probability of a heads): a uniform distribution between zero and one, and a beta distribution (Appendix B) with parameters  $[10, 1]$ . Which prior appears to influence the posterior mean more than the other?
- 1.18 Fisher (1930) defined fiducial inference for a parameter  $\theta$  with maximum likelihood estimate  $\hat{\theta}$  by making the CDF (cumulative distribution function) of  $\hat{\theta}|\theta$  uniform over  $[0:1]$ . Taking a derivative of this CDF with respect to  $\hat{\theta}$  then gives the associated PDF:

$$f(\hat{\theta}|\theta) = \left| \frac{d}{d\hat{\theta}} F(\hat{\theta}|\theta) \right|,$$

and taking a derivative of the CDF with respect to  $\theta$  produced what Fisher called the fiducial distribution of the parameter  $\theta$  given the statistic  $\hat{\theta}$ . Fisher (1956) later claimed that this approach “uses the observations only to change the logical status

of the parameter from one in which nothing is known of it, and no probability statement about it can be made, to the status of a random variable having a well defined distribution.” Show that this cannot be true and that this is not Bayesian inference.

- 1.19 The *Stopping Rule Principle* states that if a sequence of experiments (trials) is governed by a stopping rule,  $\eta$ , that dictates cessation, then inference about the unknown parameters of interest must depend on  $\eta$  only through the collected sample. Obvious stopping rules include setting the number of trials in advance (i.e., reaching  $n$  is the stopping rule), and stopping once a certain number of successes are achieved. Consider the following experiment and stopping rule. Standard normal distributed data are collected until the absolute value of the mean of the data exceeds  $1.96/\sqrt{n}$ . Explain why this fails as a non-Bayesian stopping rule for testing the hypothesis that  $\mu = 0$  (the underlying population is zero), but is perfectly acceptable for Bayesian inference.

- 1.20 The PDF of a  $C$ -finite mixture of normal distributions is given by:

$$f(y_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{c=1}^C \omega_c \mathcal{N}(y_i|\mu_c, \sigma_c^2),$$

where  $\omega_c$  is the weight placed on distribution  $c$  with  $\sum \omega_c = 1$ , and  $\mathcal{N}()$  denotes a normal distribution. For a mixture with means  $\boldsymbol{\mu} = [-3, 1, 4]$ , variances  $\boldsymbol{\Sigma} = [0.3, 1, 0.2]$ , and weights  $\Omega = [0.2, 0.5, 0.3]$  develop a simple Metropolis (symmetric) algorithm to get draws from this mixture using a Student's- $t$  candidate distribution where you determine a good choice for the degrees of freedom parameter,  $\nu$ . Describe the mixture distribution empirically with summary statistics and a graph from producing 10,000 draws and disposing of the first 5,000.

# References

## Preface to the Second Edition

Aykroyd, Robert G., Lucy, David, Pollard, Mark A., and Roberts, Charlotte A. (1999). Nasty, Brutish, But Not Necessarily Short: A Reconsideration of the Statistical Methods Used to Calculate Age At Death From Adult Human Skeletal and Dental Age Indicators. *American Antiquity* 64, 55-70.

Cowgill, G. L. (1977). The Trouble With Significance Tests and What We Can Do About It. *Philosophical Transactions of the Royal Society* 327, 331-338.

Freeman, P. R. (1976). A Bayesian Approach to the Megalithic Yard. *Journal of the Royal Statistical Association, Series A* 139, 279-295.

Gill, Jeff. (2006). *Essential Mathematics for Political and Social Research*. Cambridge, England: Cambridge University Press.

Laxton, R. R., Cavanaugh, W. G., Litton, C. D., Buck, C. E., and Blair, R. (1994). The Bayesian Approach to Archaeological Data Analysis: An Application of Change-Point Analysis to Prehistoric Domes. *Archeologia e Calcolatori* 5, 53-68.

Litton, C. D. and Buck, C. E. (1995). The Bayesian Approach to the Interpretation of Archaeological Data. *Archaeometry* 37, 1-24.

Lucy, D., Aykroyd, R. G., Pollard, A. M., and Solheim, T. (1996). A Bayesian Approach to Adult Human Age Estimation from Dental Observations by Johanson's Age Changes. *Journal of Forensic Sciences* 41, 189-194.

Malakoff, David. (1999). Bayes Offers a 'New' Way to Make Sense of Numbers. *Science* 286, 1460-1464.

Phillips, L. D. (1973). *Bayesian Statistics for Social Scientists*. Thomas Nelson and Sons, London.

Reese, R. (1994). Are Bayesian Statistics Useful to Archaeological Reasoning? *Antiquity* 68, 848-850.

Smith, A. F. M. (1995). A Conversation with Dennis Lindley. *Statistical Science* 10, 305-319.

## Preface to the First Edition

Bayarri, M. J. and DeGroot, M. H. (1991). What Bayesians Expect of Each Other. *Journal of the American Statistical Association* 86, 924-932.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second Edition. New York: Springer-Verlag.

Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*. New York: Wiley.

Carlin, B. P. and Louis, T. A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*. Second Edition. New York: Chapman & Hall.

Cooper, H. and Hedges, L. (eds.) (1994). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.

Dey, D. K., Ghosh, S. K., and Mallick, B. K. (2000). *Generalized Linear Models: A Bayesian Perspective*. New York: Marcel Dekker.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Second Edition. New York: Springer.

Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.

Fox, J. (2002). *An R and S-Plus Companion to Applied Regression*. Thousand Oaks, CA: Sage.

Greene, W. (2000). *Econometric Analysis*. Fourth Edition. Upper Saddle River, NJ: Prentice Hall.

Gujarati, D. N. (1995). *Basic Econometrics*. New York: McGraw-Hill.

Hanushek, E. A. and Jackson, J. E. (1977). *Statistical Methods for Social Scientists*. San Diego: Academic Press.

Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag.

Hunter, J. E. and Schmidt, F. L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Thousand Oaks, CA: Sage.

Ibrahim, J. G., Chen, M-H., and Sinha, D. (2001). Bayesian Survival Analysis. New York: SpringerVerlag.

Kleppner, D. and Ramsey, N. (1985). Quick Calculus: A Self-Teaching Guide. New York: Wiley Self Teaching Guides.

Lindley, D. V. (1980). Jeffreys's Contribution to Modern Statistical Thought. In Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys. Arnold Zellner (ed.). Amsterdam: North Holland.

Lipsey, M. W. and Wilson, D. B. (2001). Practical Meta-Analysis. Thousand Oaks, CA: Sage.

McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models. Second Edition. New York: Chapman & Hall.

Meyn, S. P. and Tweedie, R. L. (1993). Markov Chains and Stochastic Stability. New York: SpringerVerlag.

Montgomery, D. C. C., Peck, E. A., and Vining, G. G. (2001). Introduction to Linear Regression Analysis. Third Edition. New York: John Wiley & Sons.

Nelder, J. A. and Wedderburn, R. W. M. (1972). "Generalized Linear Models." Journal of the Royal Statistical Society, Series A 135, 370-85.

Neter, J., Kutner, M. H., Nachtsheim, C., and Wasserman, W. (1996). Applied Linear Regression Models. Chicago: Irwin.

Norris, J. R. (1997). Markov Chains. Cambridge: Cambridge University Press.

Nummelin, E. (1984). General Irreducible Markov Chains and Non-negative Operators. Cambridge: Cambridge University Press.

Press, S. J. (1980). Bayesian Computer Programs. In Studies in Bayesian Econometrics and Statistics, S. E. Fienberg and A. Zellner, eds., Amsterdam: North Holland, pp. 429-442..

Propp, J. G. and Wilson, D. B. (1996). Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics. Random Structures and Algorithms 9, 223-252.

Raiffa, H. (1982). The Art and Science of Negotiation. Cambridge: Cambridge University Press.

Robert, C. P. (2001). The Bayesian Choice: A Decision Theoretic Motivation. Second Edition. New York: Springer-Verlag.

Savage, L. J. (1954). The Foundations of Statistics. New York: Wiley.

Venables, W. N. and Ripley, B. D. (1999). Modern Applied Statistics with S-Plus, Third Edition. New York: Springer-Verlag.

Wald, A. (1950). Statistical Decision Functions. New York: Wiley.

Winkler, R. L. (1972). Introduction to Bayesian Inference and Decision. New York: Holt, Rinehart, and Winston.

---

# References

- Aarts, E. H. L. and Kors, T. J. (1989). *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. New York: John Wiley & Sons.
- Abraham, C. and Cadre, B. (2004). Asymptotic Global Robustness in Bayesian Decision Theory. *The Annals of Statistics* **32**, 1341-1366.
- Abramowitz, M. and Stegun, I. A. (eds.). (1977). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Mineola, NY: Dover Publications.
- Acton, F. S. (1996). *Real Computing Made Real: Preventing Errors in Scientific and Engineering Calculations*. Princeton: Princeton University Press.
- Adams, J. L. (1991). A Computer Experiment to Evaluate Regression Strategies. *Proceedings of the Computational Statistics Section of the American Statistical Association* **14**, 55-62.
- Adnan, V. E. and Raftery, A. E. (1986). Bayes Factors for Non-Homogeneous Poisson Processes with Vague Prior Information. *Journal of the Royal Statistical Society, Series B* **48**, 322-329.
- Agresti, A. (2002). *Categorical Data Analysis*. Second Edition. New York: John Wiley & Sons.
- Ahmed, S. E. and Reid, N. (2001). *Empirical Bayes and Likelihood Inference*. New York: Springer-Verlag.
- Ahrens, J. H. and Dieter, U. (1972). Computer Methods for Sampling from the Exponential and Normal Distributions. *Communications of the ACM* **15**, 873-882.
- Ahrens, J. H. and Dieter, U. (1973). Extensions of Forsythe's Method for Random Sampling from the Normal Distribution. *Mathematics for Computation* **27**, 927-937.
- Ahrens, J. H. and Dieter, U. (1974). Computer Methods for Sampling from Gamma, Beta, Poisson, and Binomial Distributions. *Computing* **12**, 223-246.
- Ahrens, J. H. and Dieter, U. (1980). Sampling from Binomial and Poisson Distributions: A Method with Bounded Computation Times. *Computing* **25**, 193-208.
- Ahrens, J. H. and Dieter, U. (1982). Generation of Poisson Deviates from Modified Normal Distributions. *ACM Transactions on Mathematical Software* **8**, 163-179.
- Aitkin, M. (1991). Posterior Bayes Factor. With Discussion. *Journal of the Royal Statistical Society, Series B* **53**, 111-142.
- Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *Proceedings of the 2nd International Symposium on Information Theory*, N. Petrov and F. Caski (eds.). Budapest: Akadémiai Kiadó, pp.176-723.
- Akaike, H. (1974). A New Look at Statistical Model Identification. *IEEE Transactions Automatic Control* **AU-19**, 716-722.
- Akaike, H. (1976). Canonical Correlation Analysis of Time Series and the Use of an Information Criterion. In *System Identification: Advances and Case Studies*, R. K. Mehra and D. G. Lainiotis (eds.). San Diego: Academic Press, pp.52-107.
- Akaike, H. (1980). The Interpretation of Improper Prior Distributions as Limits of Data-Dependent Proper Prior Distributions. *Journal of the Royal Statistical Society, Series B* **42**, 46-52.
- Albert, J. H. (1988). Computational Methods Using a Bayesian Hierarchical Generalized Linear Model. *Journal of the American Statistical Association* **83**, 1037-1044.

- Albert, J. H. (1990). *Bayesian Computation with R (Use R!)*. New York: Springer-Verlag.
- Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* **88**, 669-679.
- Albert, J. H. and Chib, S. (1995). Bayesian Residual Analysis for Binary Response Regression Models. *Biometrika* **82**, 747-759.
- Albert, J. H. and Pepple, P. A. (1989). A Bayesian Approach to Some Overdispersion Models. *Canadian Journal of Statistics* **17**, 333-344.
- Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology* **13**, 61-98.
- Altman, M., Gill, J. and McDonald, M. P. (2003). *Numerical Issues in Statistical Computing for the Social Scientist*. New York: John Wiley & Sons.
- Altman, M. and McDonald, M. P. (2001). Choosing Reliable Statistical Software. *PS: Political Science and Politics* **34**, 681-687.
- Altman, M. and McDonald, M. P. (2003). Replication with Attention to Numerical Accuracy. *Political Analysis* **11**, 302-307.
- Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
- Amit, Y. (1991). On Rates of Convergence of Stochastic Relaxation for Gaussian and non-Gaussian Distributions. *Journal of Multivariate Analysis* **38**, 82-99.
- Amit, Y. (1996). Convergence Properties of the Gibbs Sampler for Perturbations of Gaussians. *Annals of Statistics* **24**, 122-140.
- Amit, Y. and Grenander, U. (1991). Comparing sweep strategies for stochastic relaxation. *Journal of Multivariate Analysis* **37**, 197-222.
- Anderson, E. B. (1970). Sufficiency and Exponential Families for Discrete Sample Spaces. *Journal of the American Statistical Association* **65**, 1248-1255.
- Anderson, J. E. and Louis, T. A. (1996). Generating Pseudo-Random Variables From Mixture Models by Exemplary Sampling. *Journal of Statistical Computation and Simulation* **54**, 45-53.
- Andrews, D. F. (1974). A Robust Method for Multiple Linear Regression. *Technometrics* **16**, 523-531.
- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location*. Princeton: Princeton University Press.
- Andrews, D. F. and Mallows, C. L. (1974). Scale Mixtures of Normality. *Journal of the Royal Statistical Society, Series B* **36**, 99-102.
- Andrieu, C., Djurić, P. M. and Doucet, A. (2001). Model Selection by MCMC Computation. *Signal Processing* **81**, 19-37.
- Anscombe, F. J. (1963). Bayesian Inference Concerning Many Parameters with Reference to Supersaturated Designs. *Bulletin of the International Statistical Association* **40**, 733-741.
- Antoniak, C. E. (1974). Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *Annals of Statistics* **2**, 1152-1174.
- Arbuckle, J. L. (2009). *Amos 18 User's Guide*. Chicago: SPSS Inc.
- Arnold, B. C. and Press, S. J. (1983). Bayesian Inference for Pareto Populations. *Journal of Econometrics* **21**, 287-306.
- Arnold, B. C. and Press, S. J. (1989). Bayesian Estimation and Prediction for Pareto Data. *Journal of the American Statistical Association* **84**, 1079-1084.
- Asmussen, S. P., Glynn, P. and Thorisson, H. (1992). Stationarity Detection in the Initial Transient Problem. *ACM Transactions on Modeling and Computer Simulation* **2**, 130-57.
- Athreya, K. B. and Ney, P. (1978). A New Approach to the Limit Theory of Recurrent Markov Chains. *Transactions of the American Mathematical Society* **245**, 493-501.



- Athreya, K. B., Doss, H. and Sethuraman, J. (1996). On the Convergence of the Markov Chain Simulation Method. *Annals of Statistics* **24**, 69-100.
- Atkinson, A. C. (1977). An Easily Programmed Algorithm for Generating Gamma Random Variates. *Journal of the Royal Statistical Society, Series A* **140**, 232-234.
- Atkinson, A. C. (1978). Posterior Probabilities for Choosing a Regression Model. *Biometrika* **65**, 39-48.
- Atkinson, A. C. (1980). Tests of Pseudo-Random Numbers. *Applied Statistics* **29**, 164-171.
- Atkinson, A. C. and Pearce, M. C. (1976). The Computer Generation of Beta, Gamma, and Normal Random Variables. *Journal of the Royal Statistical Society, Series A* **139**, 431-461.
- Ayres, D. (1994). *Information, Entropy, and Progress*. New York: American Institute of Physics Press.
- Bakan, D. (1960). The Test of Significance in Psychological Research. *Psychological Bulletin* **66**, 423-437.
- Balakrishnan, N. and Nevzorov, V. B. (2003). *A Primer on Statistical Distributions*. New York: John Wiley & Sons.
- Baldus, D. C., Pulaski, C. and Woodworth, G. (1983). Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience. *Journal of Criminal Law and Criminology* **74**, 661-753.
- Baldus, D. C., Pulaski, C. and Woodworth, G. (1990). *Equal Justice and the Death Penalty: A Legal and Empirical Analysis*. Boston: Northeastern University Press.
- Barankin, E. W. and Maitra, A. P. (1963). Generalization of the Fisher-Darmois-Koopman-Pittman Theorem on Sufficient Statistics. *Sankhyā, Series A* **25**, 217-244.
- Barnard, G. A. (1991). Discussion of Aitkin. *Journal of the Royal Statistical Society, Series B* **53**, 128-130.
- Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. New York: John Wiley & Sons.
- Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. London: Chapman & Hall.
- Barnett, V. (1973). *Comparative Statistical Inference*. New York: John Wiley & Sons.
- Barnett, V. and Lewis, T. (1978). *Outliers in Statistical Data*. New York: Wiley & Sons.
- Bartels, L. M. (1997). Specification Uncertainty and Model Averaging. *American Journal of Political Science* **41**, 641-674.
- Bartholomew, D. J. (1965). A Comparison of Some Bayesian and Frequentist Inferences. *Biometrika* **52**, 19-35.
- Bartlett, M. S. (1937). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Proceedings of the Royal Society A: Mathematical, Physical, and Engineering Sciences* **160**, 268-278.
- Basu, S. (1994). Variations of Posterior Expectations for Symmetric Unimodal Priors in a Distribution Band. *Sankhyā, Series A* **31**, 320-334.
- Basu, S. and DasGupta, A. (1995). Robust Bayesian Analysis with Distribution Bands. *Statistics and Decisions* **13**, 333-349.
- Basu, S., Jammalamadaka, S. R. and Liu, W. (1996). Local Posterior Robustness with Parametric Priors: Maximum and Average Sensitivity. In *Maximum Entropy and Bayesian Methods*, G. Heidlbreder (ed.). Dordrecht: Kluwer, pp.97-106.
- Baum, L. E. and Eagon, J. A. (1967). An Inequality with Applications to Statistical Estimation for Probabilistic Functions of Markov Processes and to a Model for Ecology. *Bulletin of the American Mathematical Society* **73**, 360-363.
- Baum, L. E. and Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite Markov Chains. *Annals of Mathematical Statistics* **37**, 1554-1563.
- Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Annals of Mathematical Statistics* **41**, 164-171.

- Bauwens, L., Lubrano, M. and Richard, J-F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford, UK: Oxford University Press.
- Bayer, D. and Diaconis, P. (1992). Trailing the Dovetail Shuffle to its Lair. *Annals of Applied Probability* **2**, 294-313.
- Bayes, T. (1763). An Essay Towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London* **53**, 370-418.
- Beale, E. M. L. (1977). Discussion of the Paper by Professor Dempster, Professor Laird and Dr. Rubin. *Journal of the Royal Statistical Society, Series B* **139**, 22.
- Beale, E. M. L. and Little, R. J. A. (1975). Missing Values in Multivariate Analysis. *Journal of the Royal Statistical Society, Series B* **37**, 129-145.
- Beckett, L. and Diaconis, P. (1994). Spectral Analysis for Discrete Longitudinal Data. *Advances in Mathematics* **103**, 107-128.
- Bedrick, E. J., Christensen, R. and Johnson, W. (1997). Bayesian Binomial Regression: Predicting Survival at a Trauma Center. *The American Statistician* **51**, 211-218.
- Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics*. New York: John Wiley & Sons.
- Berger, J. O. (1984). The Robust Bayesian Viewpoint. In *Robustness of Bayesian Analysis*, Joseph B. Kadane (ed.). Amsterdam: North Holland, pp.63-144.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second Edition. New York: Springer-Verlag.
- Berger, J. O. (1986a). Discussion: On the Consistency of Bayes Estimates. *Annals of Statistics* **14**, 30-37.
- Berger, J. O. (1986b). Bayesian Salesmanship. In *Bayesian Inference and Decision Techniques with Applications: Essays in Honor of Bruno de Finetti*, Arnold Zellner (ed.). Amsterdam: North Holland, pp.473-488.
- Berger, J. O. (1990). Robust Bayesian Analysis: Sensitivity to the Prior. *Journal of Statistical Planning and Inference* **25**, 303-28.
- Berger, J. O. (1994). An Overview of Robust Bayesian Analysis. *Test* **3**, 5-124.
- Berger, J. O. (2001). Bayesian Analysis: A Look at Today and Thoughts of Tomorrow. In *Statistics in the 21st Century*, Adrian E. Raftery, Martin A. Tanner and Martin T. Wells (eds.). New York: Chapman & Hall, pp.275-290.
- Berger, J. O. (2003). Could Fisher, Jeffreys and Neyman Have Agreed on Testing? *Statistical Science* **18**, 1-32.
- Berger, J. and Berliner, L. M. (1986). Robust Bayes and Empirical Bayes Analysis with  $\epsilon$ -Contaminated Priors. *The Annals of Statistics* **14**, 461-486.
- Berger, J. O. and Bernardo, J. M. (1989). Estimating the Product of Means: Bayesian Analysis with Reference Priors. *Journal of the American Statistical Association* **84**, 200-207.
- Berger, J. O. and Bernardo, J. M. (1992). On the Development of the Reference Prior Method. In *Bayesian Statistics 4*, James O. Berger, José M. Bernardo, A. P. Dawid and Adrian F. M. Smith (eds.). Oxford: Oxford University Press, pp.35-49.
- Berger, J. O., Brown, L. D. and Wolpert, R. L. (1994). A Unified Conditional Frequentist and Bayesian Test for Fixed and Sequential Simple Hypothesis Testing. *Annals of Statistics* **22**, 1787-1807.
- Berger, J. O. and Delampady, M. (1987). Testing Precise Hypotheses. With Discussion. *Statistical Science* **2**, 317-352.
- Berger, J. O., Ghosh, J. K. and Mukhopadhyay, N. (2003). Approximations to the Bayes Factor In Model Selection Problems and Consistency Issues. *Journal of Statistical Planning and Inference* **112**, 241-258.
- Berger, J. O. and Mortera, J. (1999). Default Bayes Factors for Nonnested Hypothesis Testing. *Journal of the American Statistical Association* **94**, 542-554.
- Berger, J. O. and O'Hagan, A. (1988). Ranges of Posterior Probabilities for Unimodal Priors with Specified

- Quantiles. In *Bayesian Statistics 3*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.45-65.
- Berger, J. O. and Pericchi, L. R. (1996a). The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association* **91**, 109-122.
- Berger, J. O. and Pericchi, L. R. (1996b). The Intrinsic Bayes Factor for Linear Models. In *Bayesian Statistics 5*, J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.23-42.
- Berger, J. O., Pericchi, L. R. and Varshavsky, J. A. (1998). Bayes Factors and Marginal Distributions in Invariant Situations. *Sankhyā, Series A (Special Issue on Bayesian Analysis)* **60**, 307-321.
- Berger, J. O. and Sellke, T. (1987). Test of a Point Null Hypothesis: the Irreconcilability of Significance Levels and Evidence. With Discussion. *Journal of the American Statistical Association* **82**, 112-139.
- Berger, J. O. and Wolpert, R. L. (1984). *The Likelihood Principle*. Hayward, CA: Institute of Mathematical Statistics Monograph Series.
- Berger, J. O. and Wolpert, R. L. (1988). *The Likelihood Principle*. Second Edition. Hayward, CA: Institute of Mathematical Statistics Monograph Series.
- Berger, J. O. and Yang, R. (1994). Noninformative Priors and Bayesian Testing for the AR(1) Model. *Econometric Theory* **10**, 461-482.
- Berger, J. O., Bernardo, J. M. and Mendoza, M. (1989). On Priors That Maximize Expected Information. In *Recent Developments in Statistics and Their Applications*, J. Klein and J. Lee (eds.). Seoul, ROK: Freedom Academy, pp.1-20.
- Berger, J. O., Boukai, B. and Wang, Y. (1997) Unified Frequentist and Bayesian Testing of a Precise Hypothesis. *Statistical Science* **12**, 133-160.
- Berk, R. H. (1972). Consistency and Asymptotic Normality of MLE's for Exponential Models. *Annals of Mathematical Statistics* **43**, 193-204.
- Bernardo, J. M. (1979). Reference Prior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society, Series B* **41**, 113-147.
- Bernardo, J. M. (1984). Monitoring the 1982 Spanish Socialist Victory: A Bayesian Analysis. *Journal of the American Statistical Association* **79**, 510-515.
- Bernardo, J. M. (1992). Simulated Annealing in Bayesian Decision Theory. *Computational Statistics. Proceedings of 10th Symposium on Computational Statistics, Volume 1*. Heidelberg: Physica-Verlag. 547-552.
- Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: John Wiley & Sons.
- Berndt, E., Hall, B., Hall, R. and Hausman, J. (1974). Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement* **3/4**, 653-665.
- Berry, D. and Hartigan, J. A. (1993). A Bayesian Analysis for Changepoint Problems. *Journal of the American Statistical Association* **88**, 309-319.
- Berry, W. D., Rinquist, E. J., Fording, R. C. and Hanson, R. L. (1998). Measuring Citizen and Government Ideology in the American States, 1960-93. *American Journal of Political Science* **42**, 327-348.
- Bertsimas, D. and Tsitsiklis, J. N. (1993). Simulated Annealing. *Statistical Science* **8**, 10-15.
- Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society, Series B* **36**, 192-236.
- Besag, J. and Green, P. J. (1993). Spatial Statistics and Bayesian Computation. *Journal of the Royal Statistical Society, Series B* **55**, 1-52.
- Besag, J., Green, P. J., Higdon, D. M. and Mengersen, K. L. (1995). Bayesian Computation and Stochastic Systems (with discussion). *Statistical Science* **10**, 3-66.
- Best, D. J. (1979). Some Easily Programmed Pseudo-Random Normal Generators. *Australian Computing Journal* **11**, 60-62.

- Best, D. J. (1983). A Note on Gamma Variate Generators With Shape Parameter Less Than Unity. *Computing, Archives for Informatics and Numerical Computation* **30**, 185-188.
- Best, N. G., Spiegelhalter, D. J., Thomas, A. and Brayne, C. E. G. (1996). Bayesian Analysis of Realistically Complex Models. *Journal of the Royal Statistical Society, Series A* **159**, 323-342.
- Bickel, P. and Doksum, K. (1977). *Mathematical Statistics*. San Francisco: Holden-Day.
- Billera, L. J. and Diaconis, P. (2001). A Geometric Interpretation of the Metropolis-Hastings Algorithm. *Statistical Science* **16**, 335-339.
- Billingsley, P. (1995). *Probability and Measure*. Third Edition. New York: John Wiley & Sons.
- Birkes, D. and Dodge, Y. (1993). *Alternative Methods of Regression*. New York: Wiley & Sons.
- Birnbaum, A. (1962). On the Foundations of Statistical Inference. *Journal of the American Statistical Association* **57**, 269-306.
- Blalock, H. M. (1961). *Causal Inferences in Nonexperimental Research*. Chapel Hill, NC: University of North Carolina Press.
- Blight, B. J. N. (1970). Estimation from a Censored Sample for the Exponential Family. *Biometrika* **57**, 389-395.
- Blom, G., Holst, L. and Sandell, D. (1994). *Problems and Snapshots from the World of Probability*. New York: Springer-Verlag.
- Blyth, S. (1995). The Dead of the Gulag: an Experiment in Statistical Investigation. *Applied Statistics*, 307-21.
- Boehmke, F. J. (2009). Approaches to Modeling the Adoption and Modification of Policies with Multiple Components. *State Politics and Policy Quarterly* **9**, 229-252.
- Bohachevsky, I. O., Johnson, M. E. and Stein, M. L. (1986). Generalized Simulated Annealing for Function Optimization. *Technometrics* **28**, 209-217.
- Bohi, D. R., Toman, M. A. and Wells, M. A. (1996). *The Economics of Energy Security*. Boston: Kluwer Academic Publishers.
- Boole, G. (1854). *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*. London: Macmillan.
- Booth, J. G. and Hobert, J. P. (1999). Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm. *Journal of the Royal Statistical Society, Series B* **61**, 265-285.
- Boscardin, J. W. and Gelman, A. (1996). Bayesian Computation for Parametric Models of Heteroscedasticity in the Linear Model. In *Advances in Econometrics, Volume 11 (Part A)*, R. C. Hill (ed.). Connecticut: JAI Press Inc., pp.87-109.
- Bose, S. (1994a). Bayesian Robustness with Mixture Classes of Priors. *Annals of Statistics* **22**, 652-667.
- Bose, S. (1994b). Bayesian Robustness with More Than One Class of Contaminations. *Journal of Statistical and Inference* **40**, 177-187.
- Bowman, K. O. and Beauchamp, J. J. (1975). Pitfalls with Some Gamma Variate Simulation Routines. *Journal of Statistical Computation and Simulation* **4**, 141-154.
- Box, G. E. P. (1980). Sampling and Bayes' Inference in Scientific Modeling and Robustness. *Journal of the Royal Statistical Society, Series A* **143**, 383-430.
- Box, G. E. P. (1995). Discussion. *Journal of the Royal Statistical Society, Series B* **57**, 77.
- Box, G. E. P. and Müller, M. E. (1958). A Note on Generation of Normal Deviates. *Annals of Mathematical Statistics* **28**, 610-611.
- Box, G. E. P. and Tiao, G. C. (1968). A Bayesian Approach to Some Outlier Problems. *Biometrika* **55**, 119-129.
- Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: John Wiley & Sons.

- Boyles, R. A. (1983). On the Convergence of the EM Algorithm. *Journal of the Royal Statistical Society, Series B* **45**, 47-50.
- Bozdogan, H. (1987). Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions. *Psychometrika* **52**, 345-370.
- Bradlow, E. T. and Zaslavsky, A. M. (1997). Case Influence Analysis in Bayesian Inference. *Journal of Computational and Graphical Statistics* **6**, 314-331.
- Bradlow, E. T., Weiss, R. E. and Cho, M. (1998). Bayesian Identification of Outliers in Computerized Adaptive Tests. *Journal of the American Statistical Association* **93**, 910-919.
- Braithwaite, R. B. (1953). *Scientific Explanation*. New York: Harper & Brothers.
- Brandstätter, Eduard. (1999). Confidence Intervals As An Alternative To Significance Testing. *Methods of Psychological Research Online* **4.2**, 33-46.
- Bratton, M. and Van De Walle, N. (1994). Neopatrimonial Regimes and Political Transitions in Africa. *World Politics* **46**, 453-489.
- Bratton, M. and Van De Walle, N. (1997). Political Regimes and Regime Transitions in Africa, 1910-1994. *ICPSR Study Number I06996*. Ann Arbor: Inter-University Consortium for Political and Social Research.
- Braun J. V., Braun R. K. and Muller H. G. (2000). Multiple Changepoint Fitting Via Quasilikelihood, With Application to DNA Sequence Segmentation. *Biometrika* **87**, 301-314.
- Brémaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. New York: Springer-Verlag.
- Brent, R. P. (1974). A Gaussian Pseudo-Random Number Generator. *Communications of the Association for Computing Machinery* **17**, 704-706.
- Breslow, N. (1990). Tests of Hypotheses in Overdispersed Poisson Regression and Other Quasi-Likelihood Models. *Journal of the American Statistical Association* **85**, 565-571.
- Broemeling, L. D. (1985). *Bayesian Analysis of Linear Models*. New York: Marcel Dekker.
- Brooks, R. J. (1972). Decision Theory Approach to Optimal Regression Designs. *Biometrika* **59**, 563-571.
- Brooks, S. P. (1998a). Markov Chain Monte Carlo and its Applications. *The Statistician* **47**, 69-100.
- Brooks, S. P. (1998b). Quantitative Convergence Assessment for Markov Chain Monte Carlo Via Cusums. *Statistics and Computing* **8**, 267-274.
- Brooks, S. P. (2002). Discussion on the Paper by Spiegelhalter, Best, Carlin, and van der Linde. *Journal of the Royal Statistical Society, Series B* **64**, 616-618.
- Brooks, S. P., Fan, Y. and Rosenthal, J. S. (2006). Perfect Forward Simulation via Simulated Tempering. *Communications in Statistics, Simulation and Computation* **35**, 683-713.
- Brooks, S. P. and Roberts, G. O. (1999). On Quantile Estimation and Markov Chain Monte Carlo Convergence. *Biometrika* **86**, 710-717.
- Brooks, S. P. and Gelman, A. (1998a). Convergence Assessment Techniques for Markov Chain Monte Carlo. *Statistics and Computing* **8**, 319-335.
- Brooks, S. P. and Gelman, A. (1998b). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics* **7**, 434-455.
- Brooks, S. P. and Giudici, P. (2000). MCMC Convergence Assessment via Two-Way ANOVA. *Journal of Computational and Graphical Statistics* **9**, 266-285.
- Brooks, S. P. and Morgan, B. J. T. (1995). Optimization Using Simulated Annealing. *The Statistician* **44**, 241-257.
- Brooks, S. P. and Roberts, G. O. (1998). Convergence Assessment Techniques for Markov Chain Monte Carlo. *Statistics and Computing* **8**, 319-335.
- Brooks, S. P., Dellaportas, P. and Roberts, G. O. (1997). An Approach to Diagnosing Total Variation Convergence of MCMC Algorithms. *Journal of Computational and Graphical Statistics* **6**, 251-265.

- Brooks, S. and Roberts, G. (1997). On Quantile Estimation and MCMC Convergence. *Biometrika* **86**, 710-717.
- Brown, L. D. (1964). Sufficiency Statistics in the Case of Independent Random Variables. *Annals of Mathematical Statistics* **35**, 1456-1474.
- Brown, L. D. (1986). *Foundations of Exponential Families*. Hayward, CA: Institute of Mathematical Statistics Monograph Series 6.
- Brown, L. D., Cohen, A. and Strawderman, W. E. (1980). Complete Classes for Sequential Tests of Hypotheses. *Annals of Statistics* **8**, 377-398.
- Brown, L. D., Cohen, A. and Strawderman, W. E. (1989). *Correction of: Complete Classes for Sequential Tests of Hypotheses*. (1980, *Annals of Statistics* **8**, 377-398). *Annals of Statistics* **17**, 1414-1416.
- Brown, P. J., Fearn, T. and Vannucci, M. (1999). The Choice of Variables in Multivariate Regression: A Non-Conjugate Bayesian Decision Theory Approach. *Biometrika* **86**, 635-648.
- Browne, E. C., Frendreis, J. P. and Gleiber, D. W. (1986). The Process of Cabinet Dissolution: An Exponential Model of Duration and Stability in Western Democracies. *American Journal of Political Science* **30**, 628-650.
- Bryk, A. S. and Raudenbush, S. W. (1989). Toward a More Appropriate Conceptualization of Research on School Level Effects: A Three Level Hierarchical Linear Model. In *Multilevel Analysis of Educational Data*, R. Darrell Bock (ed.). San Diego: Academic Press, pp.159-204.
- Bryk, A. S. and Raudenbush, S. W. (2001). *Hierarchical Linear Models*. Second Edition. Beverly Hills: Sage.
- Buck, S. F. (1960). A Method of Estimation of Missing Values in Multivariate Data Suitable for Use with an Electronic Computer. *Journal of the Royal Statistical Society, Series B* **22**, 302-306.
- Buck, C. E., Cavanaugh, W. G. and Litton, C. D. (1996). *Bayesian Approach to Interpreting Archaeological Data*. New York: John Wiley & Sons.
- Burford, R. L. (1973). A Better Additive Congruential Random Number Generator. *Decision Sciences* **4**, 190-193.
- Burford, R. L. (1975). A Better Additive Congruential Random Number Generator? A Reply. *Decision Sciences* **6**, 199-201.
- Burkhardt, H. and Schoenfeld, A. H. (2003). Improving Educational Research: Toward a More Useful, More Influential, and Better-Funded Enterprise. *Educational Researcher*, **32**, 3-14.
- Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodal Inference: A Practical Information-Theoretical Approach*. Second Edition. New York: Springer-Verlag.
- Burstein, L., Kim, K-S. and Delandshere, G. (1989). Multilevel Investigations of Systematically Varying Slopes: Issues, Alternatives, and Consequences. In *Multilevel Analysis of Educational Data*, R. Darrell Bock (ed.). San Diego: Academic Press, pp.233-279.
- Butcher, J. C. (1961). A Partition Test for Pseudo-Random Numbers. *Mathematics of Computation* **15**, 198-199.
- Campbell, A. and Converse, P. (1999). American National Election Study, 1960. ICPSR07216-v3. Ann Arbor, MI: *Inter-university Consortium for Political and Social Research* [distributor].
- Carver, Ronald P. (1978). The Case Against Statistical Significance Testing. *Harvard Education Review* **48**, 378-99.
- Carver, Ronald P. (1993). The Case Against Statistical Significance Testing, Revisited. *Journal of Experimental Education* **61**, 287-92.
- Canes-Wrone, B., Brady, D. W. and Cogan, J. F. (2002). Out of Step, Out of Office: Electoral Accountability and House Members' Voting. *American Political Science Review* **96**, 127-140.
- Canova, F. (1994). Statistical Inference in Calibrated Models. *Journal of Applied Econometrics* **9**, S123-S144.

- Carlin, B. P. and Chib, S. (1995). Bayesian Model Choice Via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Series B* **57**, 473-484.
- Carlin, B. P. and Gelfand, A. E. (1990). Approaches for Empirical Bayes Confidence Intervals. *Journal of the American Statistical Association* **409**, 105-114.
- Carlin, B. P. and Louis, T. A. (1996). Identifying Prior Distributions That Produce Specific Decisions, With Application to Monitoring Clinical Trials. In *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, D. Berry, K. Chaloner and J. Geweke (eds.). New York: John Wiley & Sons, pp.493-503.
- Carlin, B. P. and Louis, T. A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*. Second Edition. New York: Chapman & Hall.
- Carlin, B. P. and Louis, T. A. (2009). *Bayes and Empirical Bayes Methods for Data Analysis*. Third Edition. New York: Chapman & Hall.
- Carlin, B. P. and Polson, N. G. (1991). Inference for Nonconjugate Bayesian Models Using the Gibbs Sampler. *Canadian Journal of Statistics* **19**, 399-405.
- Carlin, B. P., Chaloner, K., Church, T., Louis, T. A. and Matts, J. P. (1993). Bayesian Approaches for Monitoring Clinical Trials with an Application to Toxoplasmic Encephalitis Prophylaxis. *The Statistician* **42**, 355-367.
- Carlin, B. P., Chaloner, K., Louis, T. A. and Rhame, F. S. (1995). Elicitation, Monitoring, and Analysis for an AIDS Clinical Trial. In *Case Studies in Bayesian Statistics, Volume II*, Constantine Gatsonis, James S. Hodges, Robert E. Kass and Nozer D. Singpurwalla (eds.). New York: Springer-Verlag, pp.48-78.
- Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992). Hierarchical Bayesian Analysis of Changepoint Problems. *Applied Statistics* **41**, 389-405.
- Carlin, J. B. (1992). Meta-analysis for  $2 \times 2$  Tables: A Bayesian Approach. *Statistics in Medicine* **11**, 141-158.
- Carroll, R. J. and Ruppert, D. (1982). Robust Estimation in Heteroscedastic Linear Models. *Annals of Statistics* **10**, 429-441.
- Carroll, R. J. and Ruppert, D. (1988). *Transforming and Weighting in Regression*. New York: Chapman & Hall.
- Carter, W. J., Jr. and Myres, R. H. (1973). Maximum Likelihood Estimation from Linear Combinations of Discrete Probability Functions. *Journal of the American Statistical Association* **68**, 203-206.
- Casella, G. (1985). An Introduction to Empirical Bayes Data Analysis. *The American Statistician* **39**, 83-87.
- Casella, G. (2005). James-Stein Estimator. In *Encyclopedia of Biostatistics 4*. Peter Armitage and Theodore Colton (eds.). New York: Wiley & Sons.
- Casella, G. and Berger, R. L. (1987a). Reconciling Bayesian and Frequent Evidence in the One-Sided Testing Problem. *Journal of the American Statistical Association* **82**, 106-111.
- Casella, G. and Berger, R. L. (1987b). Rejoinder. *Journal of the American Statistical Association* **82**, 133-135.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Second Edition. Belmont, CA: Duxbury Advanced Series.
- Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician* **46**, 167-174.
- Casella, G., Lavine, M. and Robert, C. (2001). Explaining the Perfect Sampler. *The American Statistician* **55**, 299-305.
- Casella, G., Mengersen, K. L., Robert, C. P. and Titterton, D. M. (2002) Perfect Slice Samplers for Mixtures of Distributions. *Journal of the Royal Statistical Society, Series B*, **64**, 777-790.
- Casella, G. and Robert, C. P. (1996). Rao-Blackwellization of Sampling Schemes. *Biometrika* **93**, 81-94.

- Celeux, G. and Diebolt, J. (1985). The SEM Algorithm: A Probabilistic Teacher Algorithm Derived from the EM Algorithm for the Mixture Problem. *Computational Statistics* **2**, 73-82.
- Celeux, G., Chauveau, D. and Diebolt, J. (1996). Stochastic Versions of the EM Algorithm: an Experimental Study in the Mixture Case. *Journal of Statistical Computation and Simulation* **55**, 287-314.
- Celeux, G., Forbes, F., Robert, C. P. and Titterton, D. M. (2006). Deviance Information Criteria for Missing Data Models. *Bayesian Analysis* **1**, 1-24.
- Celeux, G., Hurn, M. and Robert, C. P. (2000). Computational and Inferential Difficulties with Mixture Posterior Distributions. *Journal of the American Statistical Association* **95**, 957-70.
- Černý, V. (1985). A Thermodynamic Approach to the Traveling Salesman Problem: An Efficient Simulation. *Journal of Optimization Theory and Applications* **45**, 41-51.
- Chakrabarti, A. and Ghosh, J. K. (2006). A Generalization of BIC for the General Exponential Family. *Journal of Statistical Planning and Inference* **136**, 2847-2872.
- Chaloner, K. and Brant, R. (1988). A Bayesian Approach to Outlier Detection and Residual Analysis. *Biometrika* **75**, 651-659.
- Chaloner, K. and Duncan, G. T. (1983). Assessment of a Beta Prior Distribution: PM Elicitation. *The Statistician* **27**, 174-180.
- Chaloner, K. and Duncan, G. T. (1987). Some Properties of the Dirichlet-Multinomial Distribution and Its Use in Prior Elicitation. *Communications in Statistics, Part A, Theory and Methods* **16**, 511-523.
- Chan, K. S. (1993). Asymptotic Behavior of the Gibbs Sampler. *Journal of the American Statistical Association* **88**, 320-326.
- Chan, K. S. and Geyer, C. J. (1994). Discussion of "Markov Chains for Exploring Posterior Distributions." *Annals of Statistics* **22**, 1747-1758.
- Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM Estimation for Time Series Models Involving Counts. *Journal of the American Statistical Association* **90**, 242-252.
- Chang, T. and Villegas, C. (1986). On a Theorem of Stein Relating Bayesian and Classical Inferences in Group Models. *Canadian Journal of Statistics* **14**, 289-296.
- Chao, M. T. (1970). The Asymptotic Behavior of Bayes' Estimators. *Annals of Mathematical Statistics* **41**, 601-608.
- Chatfield, C. (2003). *The Analysis of Time Series*. Sixth Edition. New York: Chapman & Hall.
- Chay, S. C., Fardo, R. D. and Mazumdar, M. (1975). On Using the Box-Müller Transformation with Multiplicative Congruential Pseudo-Random Number Generators. *Applied Statistics* **24**, 132-135.
- Chen, J. and Gupta A. K. (1997). Testing and Locating Change-points With Application To Stock Prices. *Journal of the American Statistical Association* **92**, 739-747.
- Chen, J. and Ibrahim, J. G. (2006). The Relationship Between the Power Prior and Hierarchical Models. *Bayesian Analysis* **1**, 551-574.
- Chen, M-H. and Shao, Q-M. (1999). Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *Journal of Computational and Graphical Statistics* **8**, 69-92.
- Chen, M-H. and Schmeiser, B. (1993). Performance of the Gibbs, Hit-and-Run, and Metropolis Samplers. *Journal of Computational and Graphical Statistics* **2**, 251-72.
- Chen, M-H., Dey, D. and Ibrahim, J. G. (2000). Bayesian Criterion Based Model Assessment for Categorical Data. *Biometrika* **91**, 45-63.
- Chen, M-H., Shao, Q-M. and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.
- Chen, T. T. and Fienberg, S. E. (1974). Two-dimensional Contingency Tables with Both Completely and Partially Cross-Classified Data. *Biometrics* **30**, 629-642.
- Cheng, R. C. H. (1977). The Generation of Gamma Variables with Non-Integral Shape Parameter. *Applied Statistics* **26**, 71-75.



- Cheng, R. C. H. (1985). Generation of Multivariate Normal Samples with Given Mean and Covariance Matrix. *Journal of Statistical Computation and Simulation* **21**, 39-49.
- Cheng, R. C. H. and Feast, G. M. (1979). Some Simple Gamma Variate Generators. *Applied Statistics* **28**, 290-295.
- Cheng, R. C. H. and Feast, G. M. (1980). Gamma Variate Generators with Increased Shape Parameter Range. *Communications of the ACM* **23**, 389-393.
- Chiang, T-S. and Chow, Y. (1988). On Eigenvalues and Optimal Annealing Rate. *Mathematical Operations Research* **13**, 508-511.
- Chib, S. (1992). Bayes Inference in the Tobit Censored Regression Model. *Journal of Econometrics* **51**, 79-99.
- Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association* **90**, 1313-1321.
- Chib S. (1998). Estimation and Comparison of Multiple Change-point Models. *Journal of Econometrics* **86**, 221-241.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*. **49**, 327-335.
- Chib, S. and Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika* **85**, 347-361.
- Chib, S. and Jeliazkov, I. (2001). Marginal Likelihood from the Metropolis-Hastings Output. *Journal of the American Statistical Association* **96**, 270-281.
- Chib, S. and Tiwari, R. C. (1991). Robust Bayes Analysis in Normal Linear Regression With An Improper Mixture Prior. *Communications in Statistics, Part A, Theory and Methods* **20**, 807-829.
- Christiansen, C. L. and Morris, C. N. (1997). Hierarchical Poisson Regression Modeling. *Journal of the American Statistical Association* **92**, 618-632.
- Chu, C-S. J. and White, H. (1992). A Direct Test for Changing Trend. *Journal of Business and Economic Statistics* **10**, 298-299.
- Chubb, J. E. and Moe, T. (1988). Politics, Markets and the Organization of American Schools. *American Political Science Review* **82**, 1065-1089.
- Chubb, J. E. and Moe, T. (1990). *Politics, Markets and America's Schools*. Washington, DC: Brookings Institution.
- Chung, K. L. (1974). *A Course in Probability Theory*. San Diego: Academic Press.
- Cioffi-Revilla, C. and Lai, D. (1995). War and Politics in Ancient China, 2700 B.C. to 722 B.C.: Measurement and Comparative Analysis. *Journal of Conflict Resolution* **39**, 467-494.
- Cioffi-Revilla, C. and Lai, D. (2001). The Second International System: China and East Asia, 5000 B.C. to 200 B.C. Presented at the Hong Kong Meeting of the International Studies Association, July 26-28, 2001.
- Clements, Michael. (2004). Evaluating the Bank of England Density Forecasts of Inflation. *Economic Journal* **114**, 844-866.
- Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* **74**, 829-36.
- Cleveland, W. S. (1981). LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *The American Statistician* **35**, 54.
- Cleveland, W. S. (1993). *Visualizing Data*. Murray Hill, NJ: Hobart Press.
- Clogg, C. C., Petkova, E. and Haritou, A. (1995). Statistical Methods for Comparing Regression Coefficients Between Models. *American Journal of Sociology* **100**, 1261-1293.
- Clyde, M. (1999). Bayesian Model Averaging and Model Search Strategies. In *Bayesian Statistics 6*, J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.157-185.

- Clyde, M., Müller, P. and Parmigiani, G. (1995). Optimal Designs for Heart Defibrillators. *Case Studies in Bayesian Statistics II. Lecture Notes in Statistics* **105**, 278-292. New York: Springer-Verlag.
- Cocchi, D. and Mouchart, M. (1996). Quasi-Linear Bayes Estimation in Stratified Finite Populations. *Journal of the Royal Statistical Society, Series B* **58**, 293-300.
- Cohen, J. (1962). The Statistical Power of Abnormal-Social Psychological Research: A Review. *Journal of Abnormal and Social Psychology* **65**, 145-153.
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (revised edition). New York: Academic Press.
- Cohen, J. (1988). A Power Primer. *Psychological Bulletin* **112**, 155-159.
- Cohen, J. (1992). Statistical Power Analysis. *Current Directions In Psychological Science* **1**, 98-101.
- Cohen, J. (1994). The Earth is Round ( $p < .05$ ). *American Psychologist* **12**, 997-1003.
- Cohen, J., Nagin, D., Wallstrom, G. and Wasserman, L. (1998). Hierarchical Bayesian Analysis of Arrest Rates. *Journal of the American Statistical Association* **93**, 1260-1270.
- Columbo, B. (1952). Preliminary Analysis of Recent Demographic Trends in Italy. *Population Index* **18**, 265-279.
- Conell, C. and Cohn, S. (1995). Learning from Other People's Actions: Environmental Variation and Diffusion in French Coal Mining Strikes, 1890-1935. *American Journal of Sociology* **101**, 366-403.
- Congdon, P. (2001). *Bayesian Statistical Modeling*. New York: Wiley & Sons.
- Congdon, P. (2003). *Applied Bayesian Modelling*. New York: Wiley & Sons.
- Congdon, P. (2005). *Bayesian Models for Categorical Data*. New York: Wiley & Sons.
- Congdon, P. (2010). *Applied Bayesian Hierarchical Methods*. New York: Wiley & Sons.
- Conigliani, C., Castro, J. I. and O'Hagan, A. (2000). Bayesian Assessment of Goodness of Fit Against Nonparametric Alternatives. *Canadian Journal of Statistics* **28**, 327-342.
- Conquest, R. (1978). *Kolyma: The Arctic Death Camps*. New York: Viking Adult.
- Consonni, G. and Veronese, P. (1992). Conjugate Priors for Exponential Families Having Quadratic Variance Functions. *Journal of the American Statistical Association* **87**, 1123-1127.
- Conte, S. D. and de Boor, C. (1980). *Elementary Numerical Analysis: An Algorithmic Approach*. Third Edition. New York: McGraw-Hill.
- Cook, D. R. and Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman & Hall.
- Corcoran, J. N. and Tweedie, R. L. (1998). Perfect Sampling From Independent Metropolis-Hastings Chains. *Journal of Statistical Planning and Inference* **104**, 297-314.
- Cosmides, L. and Tooby, J. (1996). Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment Under Uncertainty. *Cognition* **58**, 1-73.
- Coveyou, R. R. (1960). Serial Correlation in the Generation of Pseudo-Random Numbers. *Journal of the Association for Computing Machinery* **7**, 72-74.
- Coveyou, R. R. (1970). Random Numbers Fall Mainly in the Planes (Review). *ACM Computing Reviews*, 225.
- Coveyou, R. R. and MacPherson, R. D. (1967). Fourier Analysis of Uniform Random Number Generators. *Journal of the ACM* **14**, 100-119.
- Cowles, M. K. (2002). MCMC Sampler Convergence Rates for Hierarchical Normal Linear Models: A Simulation Approach. *Statistics and Computing* **12**, 377-389.
- Cowles, M. K. and Carlin, B. P. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association* **91**, 883-904.
- Cowles, M. K., Roberts, G. O. and Rosenthal, J. S. (1999). Possible Biases Induced by MCMC Convergence Diagnostics. *Journal of Statistical Computation and Simulation* **64**, 87-104.

- Cowles, M. K. and Rosenthal, J. S. (1998). A Simulation Approach to Convergence Rates for Markov Chain Monte Carlo Algorithms. *Statistics and Computing* **8**, 115-124.
- Cox, D. R. (1961). Tests of Separate Families of Hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 105-123.
- Cox, D. R. (1983). Some Remarks on Overdispersion. *Biometrika* **70**, 269-274.
- Crain, B. R. and Morgan, R. L. (1975). Asymptotic Normality of the Posterior Distribution for Exponential Models. *Annals of Statistics* **3**, 223-227.
- Craiu, R. V. and Meng, X-L. (2011). Perfection Within Reach: Exact MCMC Sampling. In *Handbook of Markov Chain Monte Carlo*, Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li Meng (eds.). Boca Raton: Chapman & Hall/CRC, pp.199-223.
- Cramer, J. S. (1994). *Econometric Applications of Maximum Likelihood Methods*. Cambridge: Cambridge University Press.
- Creutz, M. (1979). Confinement and the Critical Dimensionality of Space-Time. *Physical Review Letters* **43**, 553-56.
- Creutz, M., Jacobs, L. and Rebbi, C. (1983). Monte Carlo Computations in Lattice Gauge Theories. *Physical Review* **95**, 201.
- Cuevas, A. and Sanz, P. (1988). On Differentiability Properties of Bayes Operators. In *Bayesian Statistics 3*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.569-577.
- Cui, L., Tanner, M. A., Sinha, D. and Hall, W. J. (1992). Comment: Monitoring Convergence of the Gibbs Sampler: Further Experience with the Gibbs Stopper. *Statistical Science* **7**, 483-486.
- Cyert, R. M. and DeGroot, M. H. (1987). *Bayesian Analysis and Uncertainty in Economic Theory*. Totowa, NJ: Rowman & Littlefield.
- Dalal, S. R. and Hall, G. J., Jr. (1980). On Approximating Parametric Bayes Models by Nonparametric Bayes Models. *Annals of Statistics* **8**, 664-772.
- Dalal, S. R. and Hall, W. J. (1983). Approximating Priors by Mixtures of Natural Conjugate Priors. *Journal of the Royal Statistical Society, Series B* **45**, 278-286.
- Dale, A. I. (1991). *A History of Inverse Probability: From Thomas Bayes to Karl Pearson*. New York: Springer-Verlag.
- Damien, P., Wakefield, J. C. and Walker, S. (1999). Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variables. *Journal of the Royal Statistical Society, Series B* **61**, 331-344.
- Dar, R., Serlin, R. C. and Omer, H. (1994). Misuse of Statistical Tests in Three Decades of Psychotherapy Research. *Journal of Consulting and Clinical Psychology* **62**, 75.
- Darmois, G. (1935). Sur les Lois de Probabilités à Estimation Exhaustive. *Académie de Sciences Paris* **200**, 1265-1266.
- Datta, G. S. and Ghosh, M. (1991). Bayesian Prediction in Linear Models: Applications to Small Area Estimation. *Annals of Statistics* **19**, 1748-1770.
- Davidian, M. and Carroll, R. J. (1987). Variance Function Estimation. *Journal of the American Statistical Association* **82**, 1079-1091.
- Davis, D. H. (1992). *Energy Politics*. New York: St. Martins Press.
- Davis, P. J. and Rabinowitz, P. (1984). *Methods of Numerical Integration*. Second Edition. San Diego: Academic Press.
- Davis, W. W. (1978). Bayesian Analysis of the Linear Model Subject to Linear Inequality Constraints. *Journal of the American Statistical Association* **73**, 573-579.
- Dawid, A. P. (1979). Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society, Series B* **41**, 1-31.

- Dawid, A. P. (1982). The Well-Calibrated Bayesian. *Journal of the American Statistical Association* **77**, 605-613.
- Dawid, A. P. (1983). Invariant Prior Distributions. In *Encyclopedia of Statistical Sciences*, S. Kotz and N. L. Johnson (eds.). New York: John Wiley & Sons, pp.228-236.
- Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization Paradoxes in Bayesian and Structural Inference. *Journal of the Royal Statistical Society, Series B* **35**, 189-233.
- Dearden, R. and Clancy, D. (2002). Particle Filters for Real-Time Fault Detection in Planetary Rovers. *Proceedings of the Thirteenth International Workshop on Principles of Diagnosis*, 1-6.
- De Bruijn, N. G. (1981). *Asymptotic Methods in Analysis*. Dover Edition. New York: Dover Publications.
- Deely, J. J. and Lindley, D. V. (1981). Bayes Empirical Bayes. *Journal of the American Statistical Association* **76**, 833-841.
- Deely, J. J. and Zimmer, W. J. (1976). Asymptotic Optimality of the Empirical Bayes Procedure. *Annals of Statistics* **4**, 576-580.
- de Finetti, B. (1930). Funzione Caratteristica di un Fenomeno Aleatorio. *Mem. Accad. Naz. Lincei* **4**, 86-133.
- de Finetti, B. (1937). La Prévision: ses Lois Logiques, ses Sources Subjectives. *Annales de l'Institut Henri Poincaré* **7**, 1-68. Translated by H. E. Kyburg Jr. in H. E. Kyburg Jr. and H. E. Smokler (ed.). *Studies in Subjective Probability* (1964). New York: Wiley, 93-158.
- de Finetti, B. (1972). *Probability, Induction, and Statistics*. New York: John Wiley & Sons.
- de Finetti, B. (1974). *Theory of Probability, Volume 1*. New York: John Wiley & Sons.
- de Finetti, B. (1975). *Theory of Probability, Volume 2*. New York: John Wiley & Sons.
- DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: John Wiley & Sons.
- DeGroot, M. H. (1973). Doing What Comes Naturally: Interpreting a Tail Area as a Posterior Probability or as a Likelihood Ratio. *Journal of the American Statistical Association* **68**, 966-969.
- DeGroot, M. H. (1986). *Probability and Statistics*. Second Edition. Reading, MA: Addison-Wesley.
- De Oliveira, V. (2013). Hierarchical Poisson Models For Spatial Count Data. *Journal of Multivariate Analysis* **122**, 393-408.
- Del Pino, G. (1989). The Unifying Role of Iterative Generalized Least Squares in Statistical Algorithms. *Statistical Science* **4**, 394-408.
- Delampady, M. and Dey, D. K. (1994). Bayesian Robustness for Multiparameter Problems. *Journal of Statistical Planning and Inference* **50**, 375-382.
- Dellaportas, P. and Smith, A. F. M. (1993). Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling. *Applied Statistics* **42**, 443-459.
- Demir, B., Bovolo, F. and Bruzzone, L. (2012). A Novel System for Classification of Image Time Series with Limited Ground Reference Data. 2012 IEEE International on Geoscience and Remote Sensing Symposium (IGARSS). Munich, Germany, 22-27 July 2012, 158-161.
- de Morgan, A. (1837). Review of Laplace's Théorie Analytique des Probabilités. *Dublin Review* **2**, 338-354, **3**, 237-248.
- de Morgan, A. (1838). *An Essay on Probabilities and their Application to Life Contingencies and Insurance Offices*. London: Longman, Orme, Brown, Green, & Longmans.
- de Morgan, A. (1847). *Formal Logic; or, the Calculus of Inference, Necessary and Probable. Formal Logic*. London: Taylor & Walton.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1-38.
- Denis, D. J. (2005) The Modern Hypothesis Testing Hybrid: R. A. Fisher's Fading Influence. *Journal de la Société Française de Statistique* **145**, 5-26.

- DeSouza, C. M. (1992). An Approximate Bivariate Bayesian Method for Analyzing Small Frequencies. *Biometrics* **48**, 1113-1130.
- Devore, J. (1999). *Probability and Statistics for Engineering and the Sciences*. Boston: PWS Publishing.
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.
- Dey, D. K. and Birmiwal, L. R. (1994). Robust Bayesian Analysis Using Entropy and Divergence Measures. *Statistical Probability Letters* **20**, 287-294.
- Dey, D. K. and Micheas, A. (2000). Ranges of Posterior Expected Losses and  $\epsilon$ -Robust Actions. In *Robust Bayesian Analysis*, David Ríos Insua and Fabrizio Ruggeri (eds.). New York: Springer-Verlag, pp.71-88.
- Dey, D. K., Ghosh, S. K. and Mallick, B. K. (2000). *Generalized Linear Models: A Bayesian Perspective*. New York: Marcel Dekker.
- Dey, D. K., Lou, K. and Bose, S. (1998). A Bayesian Approach to Loss Robustness. *Statistics and Decisions* **16**, 65-87.
- Diaconis, P. (1988). *Group Representations in Probability and Statistics*. Hayward, CA: Institute of Mathematical Statistics.
- Diaconis, P. and Freedman, D. A. (1980). Finite Exchangeable Sequences. *Annals of Probability* **8**, 745-764.
- Diaconis, P. and Freedman, D. A. (1986). On the Consistency of Bayes Estimates. *Annals of Statistics* **14**, 1-67.
- Diaconis, P. and Ylvisaker, D. (1979). Conjugate Priors for Exponential Families. *Annals of Statistics* **7**, 269-281.
- Diaconis, P. and Ylvisaker, D. (1985). Quantifying Prior Opinion. In *Bayesian Statistics 2*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.). Amsterdam: North Holland Press, pp.133-156.
- Diaconis, P. and Saloff-Coste, L. (1993). Comparison Theorems for Reversible Markov Chains. *Annals of Applied Probability* **3**, 696-730.
- Diaconis, P. and Saloff-Coste, L. (1996). Logarithmic Sobolev Inequalities for Finite Markov Chains. *Annals of Applied Probability* **6**, 695-750.
- Diaconis, P. and Stroock, D. W. (1991). Geometric Bounds for Eigenvalues of Markov Chains. *Annals of Applied Probability* **1**, 36-61.
- Dias, J. and Wedel, M. (2004). An Empirical Comparison of EM, SEM and MCMC Performance for Problematic Gaussian Mixture Likelihoods. *Statistics and Computing* **14**, 323-332.
- DiCiccio, T. J. and Stern, S. E. (1994). Frequentist and Bayesian Bartlett Correction of Test Statistics Based on Adjusted Profile Likelihood. *Journal of the Royal Statistical Society, Series B* **56**, 397-408.
- Dickey, J. M. (1974). Bayesian Alternatives to the F-test and Least-Squares Estimator in the Normal Linear Model. In *Studies in Bayesian Econometrics and Statistics*, S. Fienberg and A. Zellner (eds.). Amsterdam: North Holland, pp.515-554.
- Diebolt J. and Ip, E. H. S. (1996). Stochastic EM: Method and application. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.). New York: Chapman & Hall, pp.259-273.
- Diebolt, J. and Robert, C. P. (1994). Estimation of Finite Mixture Distributions through Bayesian Sampling. *Journal of the Royal Statistical Society, Series B* **56**, 363-375.
- Dieter, U. (1975). Statistical Interdependence of Pseudo-Random Numbers Generated by the Linear Congruential Method. *Applications of Number Theory to Numerical Analysis*, S. K. Zaremba (ed.). San Diego: Academic Press, 287-318.
- Dieter, U. and Ahrens, J. H. (1971). An Exact Determination of Serial Correlations of Pseudo-Random Numbers. *Numerische Mathematik* **17**, 101-123.

- Dieter, U. and Ahrens, J. H. (1973). A Combinatorial Method for Generation of Normally Distributed Random Variables. *Computing* **11**, 137-146.
- Dobson, A. J. (1990). *An Introduction to Generalized Linear Models*. New York: Chapman & Hall.
- Doebelin, W. (1940). Éléments d'une Théorie Générale des Chaînes Simples Constantes de Markoff. *Annales Scientifiques de l'Ecole Normale Supérieure* **57**, 61-111.
- Doob, J. L. (1990). *Stochastic Processes*. New York: Wiley & Sons.
- Doob, J. L. (1996). The Development of Rigor in Mathematical Probability (1900-1950). *American Mathematical Monthly* **103**, 586-595.
- Downham, D. Y. (1970). The Runs Up and Test. *Applied Statistics* **19**, 190-192.
- Downham, D. Y. and Roberts, F. D. K. (1967). Multiplicative Congruential Pseudo-Random Number Generators. *Computer Journal* **10**, 74-77.
- Draper, D. (1995). Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society, Series B* **57**, 45-97.
- Draper, D., Hodges, J. S., Mallows, C. L. and Pregibon, D. (1993). Exchangeability and Data Analysis. *Journal of the Royal Statistical Society, Series A* **156**, 9-37.
- Duane, S., Kennedy, A. D., Pendleton, B. J. and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B* **195**, 216-222.
- Dudewicz, E. J. (1975). Random Numbers: The Need, the History, the Generators. In *A Modern Course on Statistical Distributions in Scientific Work*, Volume 2, G. P. Patil, S. Kotz and J. K. Ord (eds.). Boston: D. Reidel, pp.25-36.
- Dudewicz, E. J. (1976). Speed and Quality of Random Numbers for Simulation. *Journal of Quality Technology* **8**, 171-178.
- Dudley, R. M. and Haughton, D. (1997). Information Criteria for Multiple Data Sets and Restricted Parameters. *Statistica Sinica* **7**, 265-284.
- Dufflo, M. (1996). *Random Iterative Models*. In *Series Applications of Mathematics*, I. Karatzas and M. Yor (eds.). New York: Springer-Verlag, Volume 34.
- DuMouchel, W. and Normand, S. (2000). Computer-Modeling and Graphical Strategies for Meta-Analysis. In *Meta-Analysis in Medicine and Health Policy*, D. K. Stangl and D. A. Berry (eds.). New York: Marcel Dekker, pp.127-78.
- Earman, J. (1992). *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.
- Eaves, D. M. (1985). On Maximizing Missing Information about a Hypothesis. *Journal of the Royal Statistical Society, Series B* **47**, 263-266.
- Ebrahimi, N., Habibullah, M. and Soofi, E. S. (1992). Testing Exponentiality Based on Kullback-Leibler Information. *Journal of the Royal Statistical Society, Series B* **54**, 739-748.
- Edgeworth, F. Y. (1892a). Correlated Averages. *Philosophical Magazine*, 5th Series, **34**, 190-204.
- Edgeworth, F. Y. (1892b). The Law of Error and Correlated Averages. *Philosophical Magazine*, 5th Series, **34**, 429-438 & 518-526.
- Edgeworth, F. Y. (1893a). Exercises in the Calculation of Errors. *Philosophical Magazine*, 5th Series, **36**, 98-111.
- Edgeworth, F. Y. (1893b). Note on the Calculation of Correlation Between Organs. *Philosophical Magazine*, 5th Series, **36**, 350-351.
- Edgeworth, F. Y. (1921). Molecular Statistics. *Journal of the Royal Statistical Society* **84**, 71-89.
- Edwards, A. W. F. (1992). *Likelihood*. Baltimore: Johns Hopkins University Press.
- Edwards, R. G. and Sokal, A. D. (1988). Generalization of the Fortuin-Kasteleyn-Swendsen-Wang Representation and Monte Carlo Algorithm. *Physical Review Letters* **38**, 2009-2012.

- Edwards, W., Lindman, H. and Savage, L. J. (1963). Bayesian Statistical Inference for Psychological Research. *Psychological Research* **70**, 193-242.
- Efron, B. (1978). The Geometry of Exponential Families. *Annals of Statistics* **6**, 362-376.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7**, 1-26.
- Efron, B. (1986). Why Isn't Everyone a Bayesian? *The American Statistician* **40**, 1-11.
- Efron, B. (1998). R. A. Fisher in the 21st Century. *Statistical Science* **13**, 95-122.
- Efron, B. and Hinkley, D. V. (1978). Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information. *Biometrika* **65**, 457-482.
- Efron, B. and Morris, C. N. (1971). Limiting the Risk of Bayes and Empirical Bayes Estimators, Part I: The Bayes Case. *Journal of the American Statistical Association* **66**, 807-815.
- Efron, B. and Morris, C. N. (1972a). Limiting the Risk of Bayes and Empirical Bayes Estimators, Part II: The Empirical Bayes Case. *Journal of the American Statistical Association* **67**, 130-139.
- Efron, B. and Morris, C. N. (1972b). Empirical Bayes on Vector Observations: An Extension of Stein's Method. *Biometrika* **59**, 335-347.
- Efron, B. and Morris, C. N. (1973). Stein's Estimation Rule and Its Competitors: An Empirical Bayes Approach. *Journal of the American Statistical Association* **68**, 117-130.
- Efron, B. and Morris, C. N. (1975). Data Analysis Using Stein's Estimator and Its Generalizations. *Journal of the American Statistical Association* **70**, 311-319.
- Efron, B. and Morris, C. N. (1976). Multivariate Empirical Bayes and Estimation of Covariance Matrices. *Annals of Statistics* **4**, 22-32.
- Efron, B. and Morris, C. N. (1977). Stein's Paradox In Statistics. *Scientific American* **May**, 119-127.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- Eglese, R. W. (1990). Simulated Annealing: A Tool for Operational Research. *European Journal of Operational Research* **46**, 271-281.
- Eichenauer-Herrmann, J. (1996). Equidistribution Properties of Inversive Congruential Pseudorandom Numbers With Power of Two Modulus. *Metrika* **44**, 199-205.
- Eliason, S. R. (1983). *Maximum Likelihood Estimation: Logic and Practice*. Thousand Oaks, CA: Sage.
- Emerson, J. D. and Hoaglin, D. C. (1983). Resistant Lines for y Versus x. In *Understanding Robust and Exploratory Data Analysis*, D. C. Hoaglin, Frederick Mosteller and John Tukey (eds.). New York: Wiley & Sons, pp.129-165.
- Escobar, M. D. (1994). Estimating Normal Means with a Dirichlet Process Prior. *Journal of the American Statistical Association* **89**, 268-277.
- Escobar, M. D. and West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association* **90**, 577-588.
- Espelid, T. O. and Genz, A. (1992). *Numerical Integration: Recent Developments, Software, and Applications*. Dordrecht, NL: Kluwer.
- Estlin, T., Volpe, R., Nesnas, I. A. D., Mutz, D., Fisher, F., Englehardt, B. and Chien, S. (2001). Decision-Making in a Robotic Architecture for Autonomy. *The 6th International Symposium on Artificial Intelligence, Robotics, and Automation in Space*, Montreal Canada, 18-21.
- Evans, M. J., Fraser, D. A. S. and Monette, G. (1986). On Principles and Arguments to Likelihood. *Canadian Journal of Statistics* **14**, 181-199.
- Evans, M. J., Hastings, N. and Peacock, B. (2000). *Statistical Distributions*. New York: John Wiley & Sons.
- Evans, M. J. and Swartz, T. (1995). Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems. *Statistical Science* **10**, 254-272.
- Evans, S. J. W. (1994). Discussion of the Paper by Spiegelhalter, Freedman, and Parmar. *Journal of the Royal Statistical Society, Series A* **157**, 395.

- Faber, J. (1989). *Annual Data on Nine Economic and Military Characteristics of 78 Nations (SIRE NATDAT), 1948-1983*. Ann Arbor: Inter-University Consortium for Political and Social Research and Amsterdam, and Amsterdam, the Netherlands: Europa Institute, Steinmetz Archive.
- Fabius, J. (1964). Asymptotic Behavior of Bayes' Estimates. *Annals of Mathematical Statistics* **35**, 846-856.
- Fahrmeir, L. and Kaufmann, H. (1985). Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. *The Annals of Statistics* **13**, 342-368.
- Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Second Edition. New York: Springer.
- Falk, M. (1999). A Simple Approach to the Generation of Uniformly Distributed Random Variables With Prescribed Correlations. *Communications in Statistics, Part B, Simulation and Computation* **28**, 785-791.
- Falk, R. and C. W. Greenbaum. (1995). Significance Tests Die Hard. *Theory and Psychology* **5**, 396-400.
- Falkenrath, R. (2001). Analytical Models and Policy Prescription: Understanding Recent Innovation in U.S. Counterterrorism. *Studies in Conflict and Terrorism* **24**, 159-181.
- Fang, K.-T., Hickernell, F. J. and Niederreiter, H. (2002). *Monte Carlo and Quasi-Monte Carlo Methods*. New York: Springer-Verlag.
- Fang, K.-T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. New York: Chapman & Hall.
- Farnum, N. R. and Stanton, L. W. (1987). Some Results Concerning the Estimation of Beta Distribution Parameters in PERT. *Journal of the Operational Research Society* **38**, 287-90.
- Fearnhead, P. (2006). Exact and Efficient Bayesian Inference for Multiple Changepoint Problems. *Statistics and Computing* **16**, 203-213.
- Feller, W. (1990). *An Introduction to Probability Theory and its Applications*. Volume 1. New York: John Wiley & Sons.
- Feller, W. (1990). *An Introduction to Probability Theory and its Applications*. Volume 2. New York: John Wiley & Sons.
- Ferguson, T. S. (1967). *Mathematical Statistics: A Decision-Theoretic Approach*. San Diego: Academic Press.
- Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *Annals of Statistics* **1**, 209-230.
- Ferguson, T. S. (1983). Bayesian Density Estimation by Mixtures of Normal Distributions. In *Recent Advances in Statistics*, H. Rizvi and J. Rustagi (eds.). San Diego: Academic Press, pp.287-302.
- Ferguson, T. S. and Phadia, E. G. (1979). Bayesian Nonparametric Estimation Based on Censored Data. *The Annals of Statistics* **7**, 163-186.
- Ferrari, P. A., Frigessi, A. and Schonmann, R. J. (1993). Convergence of Some Partially Parallel Gibbs Samplers with Annealing. *Annals of Applied Probability* **3**, 137-153.
- Ferreira, M. A. R. and Gamerman, D. (2000). Dynamic Generalized Linear Models. In *Generalized Linear Models: A Bayesian Perspective*, Dipak K. Dey, Sujit K. Ghosh and Bani K. Mallick (eds.). New York: Marcel Dekker, pp.41-53.
- Fienberg, S. E. (2006). When Did Bayesian Inference Become Bayesian? *Bayesian Analysis* **1**, 1-40.
- Fill, J. A. (1991). Eigenvalue Bounds on Convergence to Stationarity for Nonreversible Markov Chains with an Application to the Exclusion Process. *Annals of Applied Probability* **1**, 62-67.
- Fill, J. A. (1998). An Interruptible Algorithm for Perfect Sampling Via Markov Chains. *Annals of Applied Probability* **8**, 131-162.
- Fill, J. A., Machida, M., Murdoch, D. J., Rosenthal, J. S. (1999). Extensions of Fill's Perfect Rejection Sampling Algorithm to General Chains. *Random Structures and Algorithms* **17**, 290-316.
- Firth, D. (1987). On the Efficiency of Quasi-Likelihood Estimation. *Biometrika* **74**, 233-245.



- Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Statistical Society of London A* **222**, 309-360.
- Fisher, R. A. (1925a). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1925b). Theory of Statistical Estimation. *Proceedings of the Cambridge Philosophical Society* **22**, 700-725.
- Fisher, R. A. (1930). Inverse Probability. *Proceedings of the Cambridge Philosophical Society* **26**, 528-535.
- Fisher, R. A. (1934). *The Design of Experiments*. First Edition. Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1935). The Fiducial Argument in Statistical Inference. *Annals of Eugenics* **6**, 391-398.
- Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*. Second Edition. New York: Hafner.
- Fishman, G. S. and Moore, L. R. (1982). A Statistical Evaluation of Multiplicative Congruential Random Number Generators With Modulus  $2^{31}-1$ . *Journal of the American Statistical Association* **77**, 129-136.
- Fishman, G. (2003). *Monte Carlo*. New York: Springer-Verlag.
- Flegal, J. M. (2012). Applicability of Subsampling Bootstrap Methods In Markov Chain Monte Carlo. In *Monte Carlo and Quasi-Monte Carlo Methods 2010*, Leszek Plaskota and Henryk Wozniakowski (eds.). Springer Berlin Heidelberg, pp.363-372.
- Fleischer, M. A. (1995). Simulated Annealing: Past, Present, and Future. *Proceedings of the 27th Winter Simulation Conference* **00**, 155-161.
- Flournoy, N. and Tsutakawa, R. K. (eds.). (1991). *Statistical Multiple Integration*. Providence: American Mathematical Society.
- Fomby, T. B., Hill, R. C. and Johnson, S. R. (1980). *Advanced Econometric Methods*. New York: Springer-Verlag.
- Fort, G., Moulines, E. and Priouret, P. (2011). Convergence of Adaptive and Interacting Markov Chain Monte Carlo Algorithms. *The Annals of Statistics* **39**, 3262-3289.
- Fortini, S. and Ruggeri, F. (2000). On the Use of the Concentration Function in Bayesian Robustness. In *Robust Bayesian Analysis*, David Ríos Insua and Fabrizio Ruggeri (eds.). New York: Springer-Verlag, pp.109-126.
- Fox, L. (1971). How to Get Meaningless Answers in Scientific Computation and What to Do About It. *IMA Bulletin* **7**, 296-302.
- Fraser, D. A. S. (1963). On Sufficiency and the Exponential Family. *Journal of the Royal Statistical Society, Series B* **25**, 115-123.
- Fredenhagen, K. and Marcu, M. (1987). A Modified Heat-Bath Method Suitable for Monte Carlo Simulations on Vector and Parallel Processors. *Physical Letters, B* **193**, 486-488.
- Freedman, D. A. (1963). On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case. *Annals of Mathematical Statistics* **34**, 1386-1403.
- Freedman, D. A. (1965). On the Asymptotic Behavior of Bayes Estimates in the Discrete Case II. *Annals of Mathematical Statistics* **36**, 454-456.
- Freedman, D. A. and Diaconis, P. (1982). de Finetti's Theorem for Symmetric Location Families. *Annals of Statistics* **10**, 184-189.
- Freedman, L. S. and Spiegelhalter, D. J. (1983). The Assessment of Subjective Opinion and Its Use in Relation to Stopping Rules for Clinical Trials. *Statistician* **32**, 153-160.
- Freeman, D. (1983). *Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth*. Cambridge, MA: Harvard University Press.
- Frieze, A., Kannan, R. and Polson, N. G. (1994). Sampling from Log-Concave Distributions. *Annals of Applied Probability* **4**, 812-837.
- Frigessi, A., Hwang, C.-R., Di Stefano, P. and Sheu, S.-J. (1993). Convergence Rates of the Gibbs Sampler, the Metropolis Algorithm, and Other Single-Site Updating Dynamics. *Journal of the Royal Statistical Society, Series B* **55**, 205-220.

- Fuller, A. T. (1976). The Period of Pseudo-Random Numbers Generated by Lehmer's Congruential Method. *Computer Journal* **19**, 173-177.
- Fulman, J. and Wilmer, E. L. (1999). Comparing Eigenvalue Bounds for Markov Chains: When Does Poincare Beat Cheeger? *Annals of Applied Probability* **9**, 1-13.
- Galant, D. (1969). Gauss Quadrature Rules for the Evaluation of  $2\pi^{-1/2} \int_0^\infty \exp(-x^2)f(x)dx$ . *Mathematics of Computation* **23**, 674.
- Galton, F. (1869). *Heredity Genius: An Inquiry into its Laws and Consequences*. Second Edition. London: Macmillan.
- Galton, F. (1875). Statistics by Intercomparison, with Remarks on the Law of Frequency of Error. *Philosophical Magazine*, 4th Series (**49**), 33-46.
- Galton, F. (1886). Regression Towards Mediocrity in Hereditary Stature. *Journal of the Anthropological Institute* **15**, 246-263.
- Galton, F. (1892). *Finger Prints*. London: Macmillan.
- Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo*. Second Edition. New York: Chapman & Hall.
- Garip, F. and Western, B. (2011). Model Comparison and Simulation for Hierarchical Models: Analyzing Rural-Urban Migration in Thailand. In *Handbook of Markov Chain Monte Carlo*, Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li Meng (eds.), 563-574. Boca Raton: Chapman & Hall/CRC. pp.563-574.
- Garthwaite, P. H. and Dickey, J. M. (1988). Quantifying Expert Opinion in Linear Regression Problems. *Journal of the Royal Statistical Society, Series B* **50**, 462-474.
- Garthwaite, P. H. and Dickey, J. M. (1992). Elicitation of Prior Distributions for Variable Selection Problems in Regression. *Annals of Statistics* **20**, 1697-1719.
- Gates, C. E. (1978). On Generating Random Normal Deviates Using the Butler Algorithm. *Proceedings of the Statistical Computing Section*. Alexandria, VA: American Statistical Association, pp.111-114.
- Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium*. Hamburg: Perthes et Besser.
- Gauss, C. F. (1823). *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae*. Göttingen: Königlichen Gesellschaft der Wissenschaften.
- Gauss, C. F. (1855). *Méthode des Moindres Carrés. Mémoires sur la Combination des Observations*. Translated by J. Bertrand. Paris: Mallet-Bachelier.
- Gavasakar, U. (1988). A Comparison of Two Elicitation Methods for a Prior Distribution for a Binomial Parameter. *Management Science* **34**, 784-790.
- Gawande, K. (1998). Comparing Theories of Endogenous Protection: Bayesian Comparison of Tobit Models Using Gibbs Sampling Output. *Review of Economics and Statistics* **80**, 128-140.
- Gebhardt, F. (1964). On the Risk of Some Strategies for Outlying Observations. *The Annals of Mathematical Statistics* **35**, 1524-1536.
- Geisser, S. and Eddy, W. F. (1979). A Predictive Approach to Model Selection. *Journal of the American Statistical Association* **74**, 153-160.
- Gelfand, A. E. and Dey, D. K. (1991). On Bayesian Robustness of Contaminated Classes of Priors. *Statistical Decisions* **9**, 63-80.
- Gelfand, A. E. and Mitter, S. K. (1993). Metropolis-type Annealing Algorithms for Global Optimization in  $\mathbb{R}^d$ . *SIAM Journal of Control and Optimization* **31**, 111-131.
- Gelfand, A. E. and Sahu, S. K. (1994). On Markov Chain Monte Carlo Acceleration. *Journal of Computational and Graphical Statistics* **3**, 261-276.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* **85**, 398-409.

- Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1995). Efficient Parameterizations For Normal Linear Mixed Models. *Biometrika* **82**, 479-488.
- Gelman, A. (1992). Iterative and Non-Iterative Simulation Algorithms. *Computing Science and Statistics* **24**, 433-438.
- Gelman, A. (1996). Inference and Monitoring Convergence. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.). New York: Chapman & Hall, pp.131-144.
- Gelman, A. (2005). Analysis of Variance: Why It Is More Important Than Ever. *Annals of statistics* **33**, 1-53.
- Gelman, A. (2006). Prior Distributions for Variance Parameters in Hierarchical Models. *Bayesian Analysis* **1**, 515-533.
- Gelman, A., Gilks, W. R. and Roberts, G. O. (1996). Efficient Metropolis Jumping Rules. In *Bayesian Statistics 5*, J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.599-608.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Gelman, A., Meng, X.-L. and Stern, H. S. (1996). Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies. *Statistica Sinica* **6**, 733-807.
- Gelman, A. and Rubin, D. B. (1992a). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **7**, 457-472.
- Gelman, A. and Rubin, D. B. (1992b). A Single Sequence from the Gibbs Sampler Gives a False Sense of Security. In *Bayesian Statistics 5*, J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.223-253.
- Gelman, A. and Rubin, D. B. (1995). Avoiding Model Selection in Bayesian Social Research *Sociological Methodology* **25**, 165-173.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003). *Bayesian Data Analysis*. Second Edition. New York: Chapman & Hall.
- Gelman, A., Huang, Z., van Dyk, D. A. and Boscardin, J. W. (2008). Using Redundant Parameterizations to Fit Hierarchical Models. *Journal of Computational and Graphical Statistics* **17**, 95-122.
- Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.
- Geman, S. and Hwang, C.-R. (1986). Diffusions for Global Optimization. *SIAM Journal of Control and Optimization* **24**, 1031-1043.
- Gentle, J. E. (1990). Computer Implementation of Random Number Generators. *Journal of Computational and Applied Mathematics* **31**, 119-125.
- Gentle, J. E. (1998). *Random Number Generation and Monte Carlo Methods*. New York: Springer-Verlag.
- George, E. I. and McCulloch, R. E. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association* **85**, 398-409.
- George, E. I., Makov, U. E. and Smith, A. F. M. (1993). Conjugate Likelihood Distributions. *Scandinavian Journal of Statistics* **20**, 147-156.
- George, R. (1963). Normal Random Variables. *Communications of the Association for Computing Machinery* **6**, 444.
- Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica* **57**, 1317-1339.
- Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In *Bayesian Statistics 4*, J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (eds.). Oxford: Oxford University Press, pp.169-193.

- Geweke, J. (1993). Bayesian Treatment of the Independent Student-t Linear Model. *Journal of Applied Econometrics* **8**, S19-S40.
- Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. New York: Wiley & Sons.
- Geweke, J. and Petrella, L. (1998). Prior Density-ratio Class Robustness in Econometrics. *Journal of Business and Economic Statistics* **16**, 469-478.
- Geyer, C. J. (1991). Markov Chain Monte Carlo Maximum Likelihood. *Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface*, 156-63.
- Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science* **7**, 473-511.
- Geyer, C. J. (1995). Conditioning in Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics* **4**, 148-154.
- Geyer, C. J. (2011). Importance Sampling, Simulated Temporing, and Umbrella Sampling. In *Handbook of Markov Chain Monte Carlo*, Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li Meng (eds.). Boca Raton: Chapman & Hall/CRC, pp.3-47.
- Geyer, C. J. and Thompson, E. A. (1995). Annealing Markov Chain Monte Carlo With Applications to Ancestral Inference. *Journal of the American Statistical Association* **90**, 909-920.
- Ghosh, J. K. (1969). Only Linear Transformations Preserve Normality. *Sankhyā, Series A* **31**, 309-312.
- Ghosh, M. and Meeden, G. (1997). *Bayesian Methods for Finite Population Sampling*. New York: Chapman & Hall.
- Ghosh, M. and Mukerjee, R. (1992). Hierarchical and Empirical Bayes Multivariate Estimation. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, M. Ghosh and P. K. Pathak (eds.). Hayward, CA: Institute of Mathematical Statistics Monograph Series 17, pp.1-12.
- Giakoumatos, S. G., Delaportas, P. and Politis, D. N. (2005). Bayesian Analysis of the Unobserved ARCH model. *Statistics and Computing* **15**, 103-111.
- Gidas, B. (1985). Nonstationary Markov Chains and Convergence of the Annealing Algorithm. *Journal of Statistical Physics* **39**, 73-131.
- Gigerenzer, G. (1987). Probabilistic Thinking and the Fight Against Subjectivity. In Krüger, Lorenz, Gerd Gigerenzer and Mary Morgan, eds. *The Probabilistic Revolution*. Volume 2. Cambridge, MA: MIT Press.
- Gigerenzer, G. (1998a). We Need Statistical Thinking, Not Statistical Rituals. *Behavioral and Brain Sciences* **21**, 199-200.
- Gigerenzer, G. (1998b). The Superego, the Ego, and the Id in Statistical Reasoning. In G. Keren and C. Lewis, eds. *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gigerenzer, G. (2004). Mindless Statistics. *The Journal of Socio-Economics* **33**, 587-606.
- Gigerenzer, G. and D. J. Murray. (1987). *Cognition as Intuitive Statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gilks, W. R. (1992). Derivative-Free Adaptive Rejection Sampling for Gibbs Sampling. In *Bayesian Statistics 4*, J. O. Berger, J. M. Bernardo, A. P. Dawid and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.641-649.
- Gilks, W. R. (1996). Full Conditional Distributions. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (ed.). New York: Chapman & Hall, pp.75-88.
- Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995). Adaptive Rejection Metropolis Sampling Within Gibbs Sampling. *Applied Statistics* **44**, 455-472.
- Gilks, W. R. and Roberts, G. O. (1996). Strategies for Improving MCMC. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.). New York: Chapman & Hall, pp.89-114.

- Gilks, W. R., Roberts, G. O. and Sahu, S. K. (1998). Adaptive Markov Chain Monte Carlo Through Regeneration. *Journal of the American Statistical Association* **93**, 1045-1054.
- Gilks, W. R., Thomas, A. and Spiegelhalter, D. J. (1994). A Language and Program for Complex Bayesian Modelling. *The Statistician* **43**, 169-177.
- Gilks, W. R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics* **41**, 337-348.
- Gill, J. (1999). The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly* **52**, 647-674.
- Gill, J. (2000). *Generalized Linear Models: A Unified Approach*. Thousand Oaks, CA: Sage.
- Gill, J. (2006). *Essential Mathematics for Political and Social Research*. Cambridge, England: Cambridge University Press.
- Gill, J. (2010). Critical Differences in Bayesian and Non-Bayesian Inference. In *Current Methodological Developments of Statistics in the Social Sciences*. Stanislav Kolenikov, Lori Thombs and Douglas Steinley (eds.). 135-158. New York: John Wiley & Sons, pp.383-408.
- Gill, J. and Casella, G. (2007). Nonparametric Priors For Ordinal Bayesian Social Science Models: Specification and Estimation. *Journal of the American Statistical Association* **104**, 453-464
- Gill, J. and Casella, G. (2004). Dynamic Tempered Transitions for Exploring Multimodal Posterior Distributions. *Political Analysis*, **12**, 425-433.
- Gill, J. and Freeman, J. (2013). Dynamic Elicited Priors for Updating Covert Networks. *Network Science*, **1**, 68-94.
- Gill, J. and King, G. (2004). What to Do When Your Hessian Is Not Invertible: Alternatives to Model Respecification in Nonlinear Estimation. *Sociological Methods and Research* **33**, 54-87.
- Gill, J. and Meier, K. J. (2000). Public Administration Research and Practice: A Methodological Manifesto. *Journal of Public Administration Research and Theory*, **10**, 157-200.
- Gill, J. and Walker, L. (2005). Elicited Priors for Bayesian Model Specifications in Political Science Research. *Journal of Politics* **67**, 841-872.
- Gill, J. and Witko C. (2013). Bayesian Analytical Methods: A Methodological Prescription for Public Administration. *Journal of Public Administration Research and Theory* **23**, 457-494.
- Gill, J. and Womack, A. (2013). The Multilevel Model Framework. In *The Sage Handbook of Multilevel Modeling*. Marc A. Scott, Jeffrey S. Simonoff and Brian D. Marx (eds.). Thousand Oaks, CA: Sage Publications, pp.3-20.
- Gill, P., Murray, W. and Wright, M. H. (1981). *Practical Optimization*. San Diego: Academic Press.
- Gliner, J. A., Leech, N. L. and Morgan, G. A. (2002). Problems With Null Hypothesis Significance Testing (NHST): What Do the Textbooks Say? *The Journal of Experimental Education* **71**, 83-92.
- Goel, P. K. (1983). Information Measures and Bayesian Hierarchical Models. *Journal of the American Statistical Association* **78**, 408-410.
- Goel, P. K. and DeGroot, M. H. (1979). Comparison of Experiments and Information Measures. *Annals of Statistics* **7**, 1066-1077.
- Goel, P. K. and DeGroot, M. H. (1980). Only Normal Distributions Have Linear Posterior Expectations in Linear Regression. *Journal of the American Statistical Association* **75**, 895-900.
- Goel, P. K. and DeGroot, M. H. (1981). Information About Hyperparameters in Hierarchical Models. *Journal of the American Statistical Association* **76**, 140-147.
- Goldberger, A. S. (1964). *Econometric Theory*. New York: Wiley & Sons.
- Golder, E. R. (1976). The Spectral Test for the Evaluation of Congruential Pseudo-Random Generators. *Applied Statistics* **25**, 173-180.
- Golder, E. R. and Settle, J. G. (1976). The Box-Müller Method for Generating Pseudo-Random Normal Deviates. *Applied Statistics* **25**, 12-20.

- Goldstein, H. (1985). *Multilevel Statistical Models*. New York: Halstead Press.
- Goldstein, H. (1986). Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares. *Biometrika* **73**, 43-56.
- Goldstein, H. (1987). *Multilevel Models in Education and Social Research*. Oxford: Oxford University Press.
- Golstein, H. (2003). *Multilevel Statistical Models*. Third Edition. Oxford: Oxford University Press, Kendall Library of Statistics 3.
- Goldstein, H. and Rasbash, J. (1996). Improved Approximations for Multilevel Models with Binary Responses. *Journal of the Royal Statistical Society, Series A* **159**, (1996), 505-513.
- Goldstein, M. (1991). Belief Transforms and the Comparison of Hypotheses. *The Annals of Statistics* **19**, 2067-2089.
- Golomb, S. W. (1967). *Shift Register Sequences*. San Francisco: Holden-Day.
- Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations*. Third Edition. Baltimore: Johns Hopkins University Press.
- Golub, G. H. and Welsch, J. H. (1969). Calculation of Gauss Quadrature Rules. *Mathematics of Computation* **23**, 221-230.
- Gönen, M., Johnson, W. O., Yonggang, L. and Westfall, P. H. (2005). The Bayesian Two-Sample  $t$  Test. *The American Statistician* **59**, 252-257.
- Good, I. J. (1950). *Probability and the Weighting of Evidence*. London: Griffin.
- Good, I. J. (1957). On the Serial Test for Random Sequences. *Annals of Mathematical Statistics* **28**, 262-264.
- Good, I. J. (1972). Statistics and Today's Problems. *American Statistician* **26**, 11-19.
- Good, I. J. (1976). The Bayesian Influence, or How to Sweep Subjectivism Under the Carpet. In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science II*, William L. Harper and Clifford A. Hooker (eds.). Dordrecht: D. Reidel, pp.125-174.
- Good, I. J. (1980a). Some History of the Hierarchical Bayes Methodology. In *Bayesian Statistics 2*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (ed.). Amsterdam: North Holland, pp.489-515.
- Good, I. J. (1980b). The Contributions of Jeffreys to Bayesian Statistics. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*. Arnold Zellner (ed.). Amsterdam: North Holland.
- Good, I. J. (1983a). The Bayes/Non-Bayes Compromise or Synthesis and Box's Current Philosophy. *Journal of Statistical Computation and Simulation* **18**, 234-236.
- Good, I. J. (1983b). *Good Thinking: The Foundations of Probability and Its Applications*. Minneapolis: University of Minnesota Press.
- Good, I. J. (1985). Weight of Evidence: A Brief Survey. In *Bayesian Statistics 2*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.). Amsterdam: North Holland Press, pp.249-270.
- Good, I. J. (1992). The Bayes/Non-Bayes Compromise: A Brief Review. *Journal of the American Statistical Association* **87**, 597-606.
- Good, I. J. and Crook, J. F. (1974). The Bayes/Non-Bayes Compromise and the Multinomial Distribution. *Journal of the American Statistical Association* **69**, 711-720.
- Goodman, J. and Sokal, A. D. (1989). Multigrid Monte Carlo Method. Conceptual Foundations. *Physics Review D* **40**, 2035-2071.
- Goodman, S. N. (1993). P values, Hypothesis Tests, and Likelihood: Implications for Epidemiology of a Neglected Historical Debate. *American Journal of Epidemiology* **137**, 485-496.
- Goodman, S. N. (1999). Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy. *Annals of Internal Medicine* **130**, 995-1004.

- Gorenstein, S. (1967). Testing a Random Number Generator. *Communications of the Association for Computing Machinery* **10**, 111-118.
- Gossett, W. S. (1908a). The Probable Error of a Mean. *Biometrika* **6**, 1-25.
- Gossett, W. S. (1908b). Probable Error of a Correlation Coefficient. *Biometrika* **6**, 302-309.
- Goutis, C. and Casella, G. (1999). Explaining the Saddlepoint Approximation. *The American Statistician* **53**, 216-224.
- Granger, C. W. J. (1999). Empirical Modeling in Economics: Specification and Evaluation. Cambridge: Cambridge University Press.
- Granger, G. W. and Hatanaka, M. (1964). *Spectral Analysis of Economic Time Series*. Princeton: Princeton University Press.
- Gray, H. L. (1988). On Unification of Bias Reduction and Numerical Approximation. In *Probability and Statistics Essays in Honor of Franklin A. Graybill*, J. N. Srivastava (ed.). Amsterdam: North Holland, pp.105-116.
- Gray, V. and Lowery, D. (1996). *The Population Ecology of Interest Representation: Lobbying Communities in the American States*. Ann Arbor, Michigan: University of Michigan Press.
- Gray, V. and Lowery, D. (2001). The institutionalization of state communities of organized interests. *Political Research Quarterly* **54**, 265-284.
- Green, P. J. (1984). Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and Some Robust and Resistant Alternatives. *Journal of the Royal Statistical Society, Series B* **46**, 149-192.
- Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* **82**, 711-732.
- Green, P. J. and Murdoch, D. J. (1998). Exact Sampling for Bayesian Inference: Towards General Purpose Algorithms. *Bayesian Statistics* **6**, 301-321.
- Greene, W. (2011). *Econometric Analysis*. Seventh Edition. Upper Saddle River, NJ: Prentice Hall.
- Greenwald, A. G. (1975). Consequences of Prejudice against the Null Hypothesis. *Psychological Bulletin* **82**, 1-20.
- Greenwald, A. G., Gonzalez, R., Harris, R. and Guthrie, D. (1996). Effect Sizes and p-values: What Should Be Reported and What Should Be Replicated? *Psychophysiology* **33**, 175-183.
- Greenwood, J. A. (1974). A Fast Generator for Gamma-Distributed Random Variables. In *Compstat 1974: Proceedings in Computational Statistics*, G. Bruckman, F. Ferschl and L. Schmetterer (eds.). Vienna: Physica Verlag. pp.19-27.
- Grenander, U. (1983). Tutorial in Pattern Theory. Unpublished Manuscript, Division of Applied Mathematics. Providence, RI: Brown University.
- Grimmett, G. R. and Stirzaker, D. R. (1992). *Probability and Random Processes*. Oxford: Oxford University Press.
- Guan, Y., Fleißner, R. and Joyce, P. (2006). Markov Chain Monte Carlo in Small Worlds. *Statistics and Computing* **16**, 193-202.
- Guihenneuc-Jouyaux, C. and Robert, C. P. (1998). Valid Discretization via Renewal Theory. In *Discretization and MCMC Convergence Assessment, Lecture Notes in Statistics*, **135**. Christian P. Robert (ed.). New York: Springer-Verlag, pp. 67-98.
- Guo, S. W. and Thompson, E. A. (1991). Monte Carlo Estimation of Variance Component Models for Large Complex Pedigrees. *IMA Journal of Mathematics Applied in Medicine and Biology* **8**, 171-189.
- Guo, S. W. and Thompson, E. A. (1994). Monte Carlo Estimation of Mixed Models for Large Complex Pedigrees. *Biometrics* **50**, 417-432.
- Gupta, S. S. and Berger, J. O. (1982). *Statistical Decision Theory and Related Topics III*. San Diego: Academic Press.

- Gustafson, P. (1996). Local Sensitivity of Inferences to Prior Marginals. *Journal of the American Statistical Association* **91**, 774-781.
- Gustafson, P. (2000). Local Robustness in Bayesian Analysis. In *Robust Bayesian Analysis*, David Ríos Insua and Fabrizio Ruggeri (eds.). New York: Springer-Verlag, pp.71-88.
- Gustafson, P. and Wasserman, L. (1995). Local Sensitivity Diagnostics for Bayesian Inference. *The Annals of Statistics* **23**, 2153-2167.
- Gutiérrez-Peña, E. and Smith, A. F. M. (1995). Conjugate Parameterizations for Natural Exponential Families. *Journal of the American Statistical Association* **90**, 1347-1356.
- Guttman, I. (1973). Care and Handling of Univariate or Multivariate Outliers in Detecting Spuriousity-A Bayesian Approach. *Technometrics* **15**, 723-738.
- Guttman, I., Dutter, R. and Freeman, P. R. (1978). Care and Handling of Univariate Outliers in the General Linear Model to Detect Spuriousity. *Technometrics* **20**, 187-193.
- Haag, J. (1924). Sur un Probleme Deneral de Probabilités et Ses Diverses Applications. *Proceedings of the International Congress of Mathematics*. Toronto, 659-674.
- Haber, S. (1970). Numerical Integration of Multiple Integrals. *SIAM Review* **12**, 481-526.
- Haberman, S. J. (1974a). *The Analysis of Frequency Data*. Chicago: University of Chicago Press.
- Haberman, S. J. (1974b). Log-linear Models for Frequency Tables Derived by Indirect Observations: Maximum Likelihood Equations. *Annals of Statistics* **2**, 911-924.
- Haberman, S. J. (1977). Product Models for Frequency Tables Involving Indirect Observations. *Annals of Statistics* **5**, 1124-1147.
- Hadjicostas, P. and Berry, S. M. (1999). Improper and Proper Posteriors with Improper Priors in a Poisson-Gamma Hierarchical Model. *Test* **8**, 147-166.
- Häggström, Olle. (2002). *Finite Markov Chains and Algorithmic Applications*. London: London Mathematical Society (Student Texts **52**).
- Hàjek, B. (1985). A Tutorial Survey of Theory and Applications of Simulated Annealing. In *Proceedings of the 24th IEEE Conference on Decision and Control*, New York, IEEE, pp.55-760.
- Hàjek, B. (1988). Cooling Schedules for Optimal Annealing. *Mathematical Operations Research* **13**, 311-329.
- Haldane, J. B. S. (1938). The Estimation of the Frequency of Recessive Conditions in Man. *Annals of Eugenics* **7**, 255-262.
- Haller, H. and Krauss, S. (2002). Misinterpretations of Significance: A Problem Students Share With Their Teachers. *Methods of Psychological Research* **7**, 1-20.
- Halpern, E. F. (1973). Polynomial Regression from a Bayesian Approach. *Journal of the American Statistical Association* **68**, 137-143.
- Halton, J. H. (1970). A Retrospective and Prospective Survey of the Monte Carlo Method. *SIAM Review* **12**, 1-63.
- Hamilton, L. C. (1992). *Regression with Graphics*. Pacific Grove, CA: Brooks/Cole Publishing Company.
- Hammersley, J. M. and Morton, J. M. (1956). A New Monte Carlo Technique: Antithetical Variables. *Proceedings of the Cambridge Philosophical Society* **52**, 449-475.
- Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo Methods*. London: Methuen.
- Hampel, F. R. (1974). The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association* **69**, 383-393.
- Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley & Sons.
- Han, C. and Carlin, B. P. (2001). MCMC Methods for Computing Bayes Factors: A Comparative Review. *Journal of the American Statistical Society* **96**, 1122-1132.
- Hannan, M. T., Carroll, G. R., Dundon, E. A. and Torres, J. C. (1995). Organizational Evolution in



- a Multinational Context: Entries of Automobile Manufacturers in Belgium, Britain, France, Germany, and Italy. *American Sociological Review* **60**, 509-528.
- Hanushek, E. A. and Jackson, J. E. (1977). *Statistical Methods for Social Scientists*. San Diego: Academic Press.
- Haque, M. M., Chin, H. C. and Huang, H. (2010). Applying Bayesian Hierarchical Models To Examine Motorcycle Crashes At Signalized Intersections. *Accident Analysis & Prevention* **42**, 203-212.
- Harlow, L. L. E., Mulaik, S. A. and Steiger, J. H. (1997). What If There Were No Significance Tests? Lawrence Erlbaum Associates Publishers.
- Haro-López, R., Mallick, B. K. and Smith, A. F. M. (2000). Binary Regression Using Data Adaptive Robust Link Functions. In *Generalized Linear Models: A Bayesian Perspective*, Dipak K. Dey, Sujit K. Ghosh and Bani K. Mallick (eds.). New York: Marcel Dekker, pp.243-253.
- Harris, T. E. (1956). The Existence of Stationary Measures for Certain Markov Processes. In *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, Volume II*. Berkeley and Los Angeles: University of California Press, pp.113-124.
- Hartigan, J. A. (1964). Invariant Prior Distributions. *Annals of Mathematical Statistics* **35**, 836-845.
- Hartigan, J. A. (1969). Linear Bayesian Models. *Journal of the Royal Statistical Society, Series B* **31**, 446-454.
- Hartigan, J. A. (1983). *Bayes Theory*. New York: Springer-Verlag.
- Harrison, M. (2006) Bombers and Bystanders in Suicide Attacks in Israel, 2000 to 2003. *Studies in Conflict & Terrorism*, **29**, 187-206.
- Hartley, H. O. (1958). Maximum Likelihood Estimation From Incomplete Data. *Biometrics* **14**, 174-194.
- Hartley, H. O. and Hocking, R. R. (1971). The Analysis of Incomplete Data. *Biometrics* **27**, 783-808.
- Harvey, A. C. 1976. Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrica* **44**, 461-5.
- Hastie, D. (2005). Towards Automatic Reversible Jump Markov Chain Monte Carlo. Un-published doctoral thesis, University of Bristol, Bristol, UK.
- Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**, 97-109.
- Healy, M. J. R. and Westmacott, M. (1956). Missing Values in Experiments Analyzed on Automatic Computers. *Applied Statistics* **5**, 203-206.
- Heath, D. and Sudderth, W. (1976). de Finetti's Theorem on Exchangeable Variables. *The American Statistician* **30**, 188-189.
- Heidelberger, P. and Welch, P. D. (1981a). Adaptive Spectral Methods for Simulation Output Analysis. *IBM Journal of Research and Development* **25**, 860-876.
- Heidelberger, P. and Welch, P. D. (1981b). A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations. *Communications of the Association for Computing Machinery* **24**, 233-245.
- Heidelberger, P. and Welch, P. D. (1983). Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research* **31**, 1109-1144.
- Hellekalek, P. (1998). Good Random Number Generators Are (not so) Easy to Find. *Mathematics and Computers in Simulation* **46**, 485-505.
- Hernandez, F. and Johnson, R. A. (1980). The Large-Sample Behavior of Transformations to Normality. *Journal of the American Statistical Association* **75**, 855-861.
- Heumann, C. (2011). James-Stein Estimator. In *International Encyclopedia of Statistical Science*. Lovric, Miodrag (ed.). New York: Springer-Verlag, pp.699-701.
- Heyde, C. C. and Morton, R. (1996). Quasi-Likelihood and Generalizing the EM Algorithm. *Journal of the Royal Statistical Society, Series B* **58**, 317-327.

- Higdon, D. M. (1998). Auxiliary Variable Methods for Markov Chain Monte Carlo With Applications. *Journal of the American Statistical Association* **93**, 585-595.
- Higham, N. J. (1996). *Accuracy and Stability of Numerical Algorithms*. Philadelphia: Society for Industrial and Applied Mathematics.
- Hill, B. (1974). On Coherence, Inadmissibility, and Inference About Many Parameters in the Theory of Least Squares. In *Studies in Bayesian Econometrics and Statistics: in Honor of Leonard J. Savage*, S. Fienberg and A. Zellner (eds.). Amsterdam: North Holland, pp.555-584.
- Hill, J. R. and Tsai, C-L. (1988). Calculating the Efficiency of Maximum Quasilikelihood Estimation. *Applied Statistics* **37**, 219-230.
- Hill, J. and Kriesi, H. (2001). Classification by Opinion Changing Behavior: A Mixture Model Approach. *Political Analysis* **9**, 301-324.
- Hills, S. E. and Smith, A. F. M. (1992). Parameterization Issues in Bayesian Inference. In *Bayesian Statistics 4*, J. O. Berger, J. M. Bernardo, A. P. Dawid and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.227-246.
- Hinkley, D. V. (1987). Comment. *Journal of the American Statistical Association* **82**, 128-129.
- Hipp, C. (1974). Sufficient Statistics and Exponential Families. *Annals of Statistics* **2**, 1283-1292.
- Hjort, N. L. (1996). Bayesian Approaches to Non- and Semiparametric Density Estimation. In *Bayesian Statistics 5*, J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.223-253.
- Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (1983). *Understanding Robust and Exploratory Data Analysis*. New York: Wiley & Sons.
- Hobert, J. P. and Casella, G. (1996). The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. *Journal of the American Statistical Association* **91**, 1461-1473.
- Hobert, J. P. and Casella, G. (1998). Functional Compatibility, Markov Chains, and Gibbs Sampling with Improper Posteriors. *Journal of Computational and Graphical Statistics* **7**, 42-60.
- Hobert, J. P., Robert, C. P. and Titterton, D. M. (1999). On Perfect Simulation for Some Mixtures of Distributions. *Journal of Statistics and Computing* **9**, 287-298.
- Hodges, J. S. and Sargent, D. J. (2001). Counting Degrees of Freedom in Hierarchical and Other Richly Parameterized Models. *Biometrika* **88**, 367-379.
- Hoel, P. G., Port, S. C. and Stone, C. J. (1987). *An Introduction to Stochastic Processes*. Prospect Heights, IL: Waveland Press.
- Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science* **14**, 382-417.
- Hoff, P. D. (2009). Multiplicative Latent Factor Models for Description and Prediction of Social Networks. *Computational and Mathematical Organization Theory* **15**, 261-272.
- Hoff, P. D. (2005). Bilinear Mixed-effects Models for Dyadic Data. *Journal of the American Statistical Association* **100**, 286-295.
- Hogarth, R. M. (1975). Cognitive Processes and the Assessment of Subjective Probability Distribution. *Journal of the American Statistical Association* **70**, 271-289.
- Hogg, R. V. and Craig, A. T. (1978). *Introduction to Mathematical Statistics*. New York: Macmillan Publishing Co.
- Holley, R. A., Kusuoka, S. and Stroock, D. W. (1989). Asymptotics of the Spectral Gap with Applications to the Theory of Simulated Annealing. *Journal of Functional Analysis* **83**, 333-347.
- Hong, Y. and Lee, Y-J. (2013). A Loss Function Approach To Model Specification Testing and Its Relative Efficiency. *Annals of Statistics* **41**, 1166-1203.
- Hora, S. C., Hora, J. A. and Dodd, N. G. (1992). Assessment of Probability Distributions for Continuous

- Random Variables: A Comparison of the Bisection and Fixed-value Methods. *Organizational Behavior and Human Decision Processes* **51**, 133-55.
- Howson, C. and Urbach, P. (1993). *Scientific Reasoning: The Bayesian Approach*. Second Edition. Chicago: Open Court.
- Hsu, J. S. J. (1995). Generalized Laplacian Approximations in Bayesian Inference. *Canadian Journal of Statistics* **23**, 399-410.
- Huang, D. S. (1970). *Regression and Econometric Methods*. New York: Wiley & Sons.
- Huber, P. J. (1972). Robust Statistics: A Review. *Annals of Mathematical Statistics* **43**, 1041-1067.
- Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures, and Monte Carlo. *Annals of Statistics* **1**, 799-821.
- Huber, P. J. (1981). *Robust Statistics*. New York: Wiley & Sons.
- Hull, T. E. and Dobell, A. R. (1964). Mixed Congruential Random Number Generators for Binary Machines. *Journal of the Association for Computing Machinery* **11**, 31-40.
- Hunter, J. E. (1997). Needed: A Ban on the Significance Test. *Psychological Science* **8**, 3-7.
- Hunter, J. E. and Schmidt, F. L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Beverly Hills: Sage.
- Hurd, W. J. (1974). Efficient Generation of Statistically Good Pseudonoise by Linearly Interconnected Shift Registers. *IEEE Transactions on Computers* **C-23**, 146-152.
- Hurn, M. (1997). Difficulties in the Use of Auxiliary Variables in Markov Chain Monte Carlo Methods. *Statistics and Computing* **7**, 35-44.
- Hurst, R. L. and Knop, R. E. (1972). Generation of Random Normal Correlated Variables: Algorithm 425. *Communications of the Association for Computing Machinery* **15**, 355-357.
- Hurvich, C. M. and Tsai, C-L. (1991). Bias of the Corrected AIC Criterion for Underfitted Regression and Time Series Models. *Biometrika* **78**, 499-509.
- Hurvich, C. M., Shumway, R. and Tsai, C-L. (1990). Improved Estimators of Kullback-Leibler Information for Autoregressive Model Selection in Small Samples. *Biometrika* **77**, 709-719.
- Hwang, J. T., Casella, G., Robert, C. P., Wells, M. T. and Farrell, R. H. (1992). Estimation of Accuracy in Testing. *Annals of Statistics* **20**, 490-509.
- Hyndman, R. J. (1996). Computing and Graphing Highest Density Regions. *The American Statistician* **50**, 120-126.
- Ibrahim, J. G. and Chen, M-H. (2000a). Power Prior Distributions for Regression Models. *Statistical Science* **15**, 46-60.
- Ibrahim, J. G. and Chen, M-H. (2000b). Prior Elicitation and Variable Selection for Generalized Linear Mixed Models. In *Generalized Linear Models: A Bayesian Perspective*, Dipak K. Dey, Sujit K. Ghosh and Bani K. Mallick (eds.). New York: Marcel Dekker, pp.41-53.
- Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian Analysis of Generalized Linear Models Using Jeffreys Prior. *Journal of the American Statistical Association* **86**, 981-986.
- Ibrahim, J. G., Chen, M-H. and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer-Verlag.
- Ibrahim, J. G., Chen, M-H. and Sinha, D. (2003). On Optimality of the Power Prior. *Journal of the American Statistical Association* **98**, 204-213.
- Imai, K. and van Dyk, D. (2004). Causal Inference with General Treatment Regimes: Generalizing the Propensity Score. *Journal of the American Statistical Association* **99**, 854-866.
- Imai, K. and van Dyk, D. A. (2005). A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation. *Journal of Econometrics* **124**, 311-334.
- Ingber, L. (1992). Genetic Algorithms and Very Fast Simulated Reannealing: A Comparison. *Mathematical Computation and Modeling* **16**, 87-100.

- Ingrassia, S. (1994). On the Rate of Convergence of the Metropolis Algorithm and Gibbs Sampler by Geometric Bounds. *Annals of Applied Probability* **4**, 347-389.
- Iosifescu, M. (1980). *Finite Markov Processes and their Applications*. New York: John Wiley & Sons.
- Irony, T. Z. and Singpurwalla, N. D. (1996). Noninformative Priors Do Not Exist: A Discussion with José M. Bernard. *Journal of Statistical Planning and Inference* **65**, 159-177.
- Jackman, S. (2000). Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo. *American Journal of Political Science* **44**, 375-404.
- James, W. and Stein, C. (1961). Estimation With Quadratic Loss. *Proceedings of the Fourth Symposium 1* J. Neyman and E. L. Scott (eds.). Berkeley: University of California Press, 361-380.
- Jagerman, D. L. (1965). Some Theorems Concerning Pseudo-Random Numbers. *Mathematics of Computation* **19**, 418-426.
- Jamshidian, M. and Jennrich, R. I. (1993). Conjugate Gradient Acceleration of the EM Algorithm. *Journal of the American Statistical Association* **88**, 221-228.
- Janssen, A. (1986). Asymptotic Properties of Neyman-Pearson Tests for Infinite Kullback-Leibler Information. *Annals of Statistics* **14**, 1068-1079.
- Jansson, B. (1966). *Random Number Generators*. Stockholm: Victor Pettersons.
- Jarner, S. F. and Roberts, G. O. (2002). Polynomial Convergence Rates of Markov Chains. *Annals of Applied Probability* **12**, 224-247.
- Jarvis, Edward. (1858). Distribution of Lunatic Reports. *American Journal of Psychiatry* **14**, 248-253.
- Jaynes, E. T. (1968). Prior Probabilities. *IEEE Transactions on Systems Science and Cybernetics* SSC-4, 227-241.
- Jaynes, E. T. (1976). Confidence Intervals vs. Bayesian Intervals. In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science II*, William L. Harper and Clifford A. Hooker (eds.). Dordrecht: D. Reidel, pp.175-257.
- Jaynes, E. T. (1980). Marginalization and Prior Probabilities. In *Bayesian Analysis in Econometrics and Statistics*, A. Zellner (ed.). North Holland: Amsterdam, pp.43-78.
- Jaynes, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics*, R. D. Rosencrantz (ed.). Dordrecht: Reidel.
- Jeffreys, H. (1961). *The Theory of Probability*. Third Edition. Oxford, England: Oxford University Press.
- Jeng, F.-C. and Woods, J. W. (1990). Simulated Annealing in Compound Gaussian Random Fields. *IEEE Transactions on Information Theory* **36**, 94-107.
- Jöhnk, M. D. (1964). Erzeugung von Betaverteilter und Gammaverteilter Zufallszahlen. *Metrika* **8**, 5-15.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 1*. New York: John Wiley & Sons.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Volume 2*. New York: John Wiley & Sons.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. New York: John Wiley & Sons.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (2000). *Continuous Multivariate Distributions, Models and Applications, Volume 1*. New York: John Wiley & Sons.
- Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). *Univariate Discrete Distributions*. New York: John Wiley & Sons.
- Johnson, R. A. (1984). The Analysis of Transformed Data: Comment. *Journal of the American Statistical Association* **79**, 314-315.
- Johnson, V. E. (1998). A Coupling-Regeneration Scheme for Diagnosing Convergence in Markov Chain Monte Carlo Algorithms. *Journal of the American Statistical Association* **93**, 238-248.
- Johnson, V. E. and Albert, J. H. (1999). *Ordinal Data Modeling*. New York: Springer-Verlag.

- Jones, G. L. and Hobert, J. P. (2001). Honest Exploration of Intractable Probability Distributions via Markov Chain Monte Carlo. *Statistical Science* **16**, 312-34.
- Jones, G. L. and Hobert, J. P. (2004). Sufficient Burn-in for Gibbs Samplers for a Hierarchical Random Effects Model. *Annals of Statistics* **32**, 784-817.
- Jones, M. C. and Lunn, A. D. (1996). Transformations and Random Variate Generation: Generalized Ratio-of-Uniforms Methods. *Journal of Statistical Computation and Simulation* **55**, 49-55.
- Jørgensen, B. (1983). Maximum Likelihood Estimation and Large-Sample Inference for Generalized Linear and Nonlinear Regression Models. *Biometrics* **70**, 19-28.
- Kackar, R. N. and Harville, D. A. (1984). Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association* **79**, 853-862.
- Kadane, J. B. (1986). Progress Toward a More Ethical Method Clinical Trials. *Journal of Medical Philosophy* **11**, 385-405.
- Kadane, J. B. (1980). Predictive and Structural Methods for Eliciting Prior Distributions. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, Arnold Zellner (ed.). Amsterdam: North Holland, pp.89-109.
- Kadane, J. B. and Chuang, D. T. (1978). Stable Decision Problems. *Annals of Statistics* **6**, 1095-1110.
- Kadane, J. B. and Srinivasan, C. (1996). Bayesian Robustness and Stability. In *Bayesian Robustness*, J. O. Berger, B. Betr , E. Moreno, L. R. Pericchi, F. Ruggeri, G. Salinetti and L. Wasserman (eds.). Hayward, CA: Institute of Mathematical Statistics Monograph Series 29, pp.139-156.
- Kadane, J. B. and Winkler, R. L. (1988). Separating Probability Elicitation From Utilities. *Journal of the American Statistical Association* **83**, 357-363.
- Kadane, J. B. and Wolfson, L. J. (1996). Priors for Design and Analysis of Clinical Trials. In *Bayesian Biostatistics*, D. A. Berry and D. K. Stangl (eds.). New York: Chapman & Hall/CRC, pp.157-186.
- Kadane, J. B. and Wolfson, L. J. (1998). Experiences in Elicitation. *Journal of The Royal Statistical Society, Series D* **47**, 3-19.
- Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S. and Peters, S. C. (1980). Interactive Elicitation of Opinion for a Normal Linear Model. *Journal of the American Statistical Association* **75**, 845-854.
- Kahn, H. (1949). Stochastic (Monte Carlo) Attenuation Analysis. No. RAND-P-88 (rev). Defense Technical Information Center.
- Kalyanam, K. (1996). Pricing Decisions under Demand Uncertainty: A Bayesian Mixture Model Approach. *Marketing Science* **15**, 207-221.
- Kaplan, A. (1964). *Conduct of Inquiry*. San Francisco: Chandler Publishing Company.
- Kaplan, H. B., Johnson, R. J., Bailey, C. A. and Simon, W. (1987). The Sociological Study of AIDS: A Critical Review of the Literature and Suggested Research Agenda. *Journal of Health and Social Behavior* **28**, 140-157.
- Karlin, S. and Taylor, H. M. (1981). *A Second Course in Stochastic Processes*. San Diego: Academic Press.
- Karlin, S. and Taylor, H. M. (1990). *A First Course in Stochastic Processes*. San Diego: Academic Press.
- Kass, R. E. (1989). The Geometry of Asymptotic Inference. *Statistical Science* **4**, 188-219.
- Kass, R. E. (1993). Bayes Factors in Practice. *The Statistician* **42**, 551-560.
- Kass, R. E. and Greenhouse, J. B. (1989). Comments on the paper by J. H. Ware. *Statistical Science* **4**, 310-317.
- Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association* **90**, 773-795.
- Kass, R. E. and Steffey, D. (1989). Approximate Bayesian Inference in Conditionally Independent Hierarchical Models. *Journal of the American Statistical Association* **84**, 717-726.

- Kass, R. E., Tierney, L. and Kadane, J. B. (1989). Approximate Methods for Assessing Influence and Sensitivity in Bayesian Analysis. *Biometrika* **76**, 663-674.
- Kass, R. E. and Vaidyanathan, S. K. (1992). Approximate Bayes Factors and Orthogonal Parameters, with Application to Testing Equality of Two Binomial Proportions. *Journal of the Royal Statistical Society, Series B* **54**, 129-144.
- Kass, R. E. and Wasserman, L. (1995). A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion. *Journal of the American Statistical Association* **90**, 928-934.
- Kass, R. E. and Wasserman, L. (1996). The Selection of Prior Distributions by Formal Rules. *Journal of the American Statistical Association* **91**, 1343-1370.
- Kass, R. E., Tierney, L. and Kadane, J. B. (1989). Approximate Methods for Assessing Influence and Sensitivity in Bayesian Analysis. *Biometrika* **76**, 663-674.
- Kaufmann, H. (1988). On the Existence and Uniqueness of Maximum Likelihood Estimates in Quantal and Ordinal Response Models. *Metrika* **35**, 291-313.
- Keating, J. P. and Mason, R. L. (1988). James-Stein Estimation from an Alternative Perspective. *The American Statistician* **42**, 160-164.
- Keith, J., K., Kroese, D. P. and Bryant, D. (2004). A Generalized Markov Sampler. *Methodology and Computing in Applied Probability* **6**, 29-53.
- Kelleher, C. A. and Yackee, S. W. (2009). A Political Consequence of Contracting: Organized Interests and State Agency Decision-making. *The Journal of Public Administration Research and Theory* **19**, 579-602.
- Kempthorne, P. J. (1986). Decision-Theoretic Measures of Influence in Regression. *Journal of the Royal Statistical Society, Series B* **48**, 370-378.
- Kendall, M. G. (1949). On Reconciliation of the Theories of Probability. *Biometrika* **36**, 101-116.
- Kennedy, W. J. and Gentle, J. E. (1980). *Statistical Computing*. New York: Marcel Dekker.
- Kéry, M. (2010). *Introduction to WinBUGS for Ecologists: A Bayesian Approach to Regression, ANOVA and Related Analyses*. Burlington, MA: Academic Press.
- Kéry, M. and Schaub, M. (2011). *Bayesian Population Analysis Using WinBUGS: A Hierarchical Perspective*. Burlington, MA: Academic Press.
- Keynes, J. M. (1921). *A Treatise on Probability*. London: MacMillan.
- Kim, D. (1991). A Bayesian Significance Test of the Stationarity of Regression Parameters. *Biometrika* **78**, 667-675.
- Kinderman, A. J. and Monahan, J. F. (1980). New Methods for Generating Student's t and Gamma Variables. *Computing* **25**, 369-377.
- Kinderman, A. J. and Ramage, J. G. (1976). Computer Generation of Normal Random Variables. *Journal of the American Statistical Association* **71**, 893-896.
- Kinderman, A. J., Monahan, J. F. and Ramage, J. G. (1975). Computer Generation of Random Variables with Normal and Student's Distributions. *Proceedings of the Statistical Computing Section*. Alexandria, VA: American Statistical Association. 128-131.
- King, G. (1989). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Cambridge: Cambridge University Press.
- Kirk, R. E. (1996). Practical Significance: A Concept Whose Time Has Come. *Educational and Psychological Measurement* **56**, 746-759.
- Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983). Kirkpatrick, S. Optimization by Simulated Annealing. *Science* **220**, 671-680.
- Kitagawa, G. and Akaike, H. (1982). A Quasi Bayesian Approach to Outlier Detection. *Annals of the Institute of Statistical Mathematics* **34B**, 389-398.

- Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*. New York: Springer-Verlag.
- Kleibergen, F. and van Dijk, H. K. (1993). Efficient Computer Generation of Matrix-Variate  $t$  Drawings with an Application to Bayesian Estimation of Simple Market Models. In *Computer Intensive Methods in Statistics*, W. Hardle and L. Simar (eds.). Heidelberg: Physica-Verlag, pp.30-46.
- Kloek, T. and van Dijk, H. K. (1978). Bayesian Estimates of Equation System Parameters; An Application Integration by Monte Carlo. *Econometrica* **46**, 1-19.
- Knott, M., Albanese, T. M. and Galbraith, J. (1990). Scoring Attitudes to Abortion. *The Statistician* **40**, 217-223.
- Knuth, D. E. (1981). *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Second Edition. Menlo Park, CA: Addison-Wesley.
- Kollman, K., Miller, J. H. and Page, S. E. (1997). Political Institutions and Sorting in a Tiebout Model. *American Economic Review* **87**, 977-992.
- Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Julius Springer.
- Kong, A., Liu, J. S. and Wong, W. H. (1994). Sequential Imputations and Bayesian Missing Data Problems. *Journal of the American Statistical Association* **89**, 278-288.
- Koop, G. (1992). "Objective" Bayesian Unit Root Tests. *Journal of Applied Econometrics* **7**, 65-82.
- Koopman, L. H. (1936). On Distributions Admitting a Sufficient Statistic. *Transactions of the American Mathematical Society* **39**, 399-409.
- Koppel, J. G. S. (1999). The Challenge of Administration by Regulation: Preliminary Findings Regarding the U.S Government's Venture Capital Funds. *Journal of Public Administration Research and Theory* **9**, 641-666.
- Korwar, R. M. and Hollander, M. (1976). Empirical Bayes Estimation of a Distribution Function. *The Annals of Statistics* **4**, 581-588.
- Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. Singapore: World Scientific Publications.
- Kowalski, J., Tu, X. M., Day, R. S. and Mendoza-Blanco, J. R. (1997). On the Rate of Convergence of the ECME Algorithm for Multiple Regression Models With  $t$ -distributed Errors. *Biometrika* **84**, 269-281.
- Krawczyk, H. (1992). How to Predict Congruential Generators. *Journal of Algorithms* **13**, 527-545.
- Kreft, I. G. G. (1993). Using Multilevel Analysis to Assess School Effectiveness: A Study of Dutch Secondary Schools. *Sociology of Education* **66**, 104-129.
- Kreft, I. G. G. and De Leeuw, J. (1988). The Seesaw Effect: A Multilevel Problem? *Quality and Quantity* **22**, 127-137.
- Kreft, I. G. G. and De Leeuw, J. (1998). *Introducing Multilevel Modeling*. Thousand Oaks, CA: Sage.
- Kreuzenkamp, H. A. and McAleer, M. (1995). Simplicity, Scientific Inference, and Econometric Modelling. *The Economic Journal* **105**, 1-21.
- Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. New York: Chapman & Hall/CRC.
- Krommer, A. R. and Ueberhuber, C. W. (1998). *Computational Integration*. Philadelphia: Society for Industrial and Applied Mathematics.
- Kronmal, R. (1964). The Evaluation of a Pseudorandom Normal Number Generator. *Journal of the Association for Computing Machinery* **11**, 357-263.
- Kronmal, R. and Peterson, A. V., Jr. (1981). A Variant of the Acceptance-Rejection Method for Computer Generation of Random Variables. *Journal of the American Statistical Association* **76**, 446-451.
- Kronrod, A. S. (1965). *Nodes and Weights of Quadrature Formulas*. New York: Consultants Bureau.
- Krueger J. (2001). Null Hypothesis Significance Testing: On the Survival of a Flawed Method. *American Psychologist* **56**, 16-26.

- Kruschke, J. (2010). *Doing Bayesian Data Analysis: A Tutorial Introduction with R*. Burlington, MA: Academic Press.
- Krzanowski, W. J. (1988). *Principles of Multivariate Analysis*. Oxford: Oxford University Press.
- Kullback, S. (1968). *Information Theory and Statistics*. New York: Wiley & Sons.
- Kurganov, I. A. (1973). Speeches and Writings, 1945-1976. Palo Alto, CA: Online Archive of California, <http://www.oac.cdlib.org>
- Kyung, M., Gill, J. and Casella, G. (2011). New Findings from Terrorism Data: Dirichlet Process Random Effects Models for Latent Groups. *Journal of the Royal Statistical Society, Series C* **60**, 701-721.
- Kyung, M., Gill, J. and Casella, G.. (2012). Sampling Schemes for Generalized Linear Dirichlet Process Random Effects Models. With Discussion and Rejoinder. *Statistical Methods and Applications* **20**, 259-290.
- Laird, N. M. (1978). Nonparametric Maximum Likelihood Estimation of a Mixing Distribution. *Journal of the American Statistical Association* **73**, 805-811.
- Laird, N. M. and Louis, T. A. (1982). Approximate Posterior Distributions for Incomplete Data Problems. *Journal of the Royal Statistical Society, Series B* **44**, 190-200.
- Laird, N. M. and Louis, T. A. (1987). Empirical Bayes Confidence Intervals Based on Bootstrap Samples. *Journal of the American Statistical Association* **82**, 739-750.
- Laird, N. M., Lange, N. and Stram, D. O. (1987). Maximum Likelihood Computations with Repeated Measures: Applications of the EM Algorithm. *Journal of the American Statistical Association* **82**, 97-105.
- Lange, K. L. (2000). *Numerical Analysis for Statisticians*. New York: Springer-Verlag.
- Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust Statistical Modeling Using the  $t$  Distribution. *Journal of the American Statistical Association* **84**, 881-896.
- Laplace, P. S. (1774). Mémoire sur la Probabilité des Causes par le Évènements. *Mémoires de l'Académie Royale des Sciences Présentés par Divers Savans* **6**, 621-656.
- Laplace, P. S. (1781). Mémoire sur la Probabilités. *Mémoires de l'Académie Royale des Sciences de Paris* **1778**, 227-332.
- Laplace, P. S. (1811). Mémoire sur les Integrales Définies et leur Application aux Probabilités, et Specialement à Recherche du Milieu Qu'il Faut Choisir Entre les Resultats des Observations. *Mémoires de l'Académie des Sciences de Paris*, 279-347.
- Laplace, P. S. (1814). *Essai Philosophique sur les la Probabilités*. Paris: V<sup>e</sup> Courcier.
- Laud, P. W. and Ibrahim, J. G. (1995). Predictive Model Selection. *Journal of the Royal Statistical Society, Series B* **57**, 247-262.
- Lauritzen, S. and Spiegelhalter, D. J. (1988). Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems. *Journal of the Royal Statistical Society, Series B* **50**, 157-194.
- Lavine, M. (1991a). Sensitivity in Bayesian Statistics: The Prior and the Likelihood. *Journal of the American Statistical Association* **86**, 396-399.
- Lavine, M. (1991b). An Approach to Robust Bayesian Analysis for Multidimensional Parameter Spaces. *Journal of the American Statistical Association* **86**, 400-403.
- Lavine, A. M. and Schervish, M. J. (1999). Bayes Factors: What They Are and What They Are Not. *The American Statistician* **53**, 119-122.
- Law, A. M. and Kelton, W. D. (1982). *Simulation Modeling and Analysis*. New York: McGraw-Hill.
- Lawler, G. F. and Sokal, A. D. (1988). Bounds on the  $L^2$  Spectrum for Markov Chains and Their Applications. *Transactions of the American Mathematical Society* **309**, 557-580.
- Lawson, A. B. (2013). Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. Boca Raton: Chapman & Hall/CRC.



- Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag.
- Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: John Wiley & Sons.
- Leamer, E. E. (1979). Information Criteria for Choice of Regression Models: A Comment. *Econometrica* **47**, 507-510.
- Leamer, E. E. (1983). Let's Take the Con Out of Econometrics. *American Economic Review* **73**, 31-43.
- Leamer, E. E. (1984). Global Sensitivity Results for Generalized Least Squares Estimates. *Journal of the American Statistical Association* **79**, 867-870.
- Leamer, E. E. (1985). Sensitivity Analysis Would Help. *The American Economic Review* **75**, 308-313.
- Leamer, E. E. (1992). Bayesian Elicitation Diagnostics. *Econometrica* **80**, 919-942.
- Learmonth, G. P. and Lewis, P. A. W. (1973). Some Widely Used and Recently Proposed Uniform Random Number Generators. *Proceedings of Computer Science and Statistics: Seventh Annual Symposium on the Interface*, W. J. Kennedy (ed.). Ames, IA: Iowa State University. 163-171.
- L'Ecuyer, P. (1998). Uniform Random Number Generators. *1998 Winter Simulation Conference Proceedings, Society for Computer Simulation*. 97-104.
- Lee, P. M. (2004). *Bayesian Statistics: An Introduction*. Second Edition. New York: Hodder Arnold.
- Lee, S. (1998). On the Quantile Process Based on the Autoregressive Residuals. *Journal of Statistical Planning and Inference* **67**, 17-28.
- Lee, V. E. and Bryk, A. S. (1989). Multilevel Model of the Social Distribution of High School Achievement. *Sociology of Education* **62**, 172-192.
- Legendre, A. M. (1805). *Nouvelles Méthodes Pour la Détermination des Orbites des Comètes*. Paris: Courcier.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. New York: Springer-Verlag.
- Lehmann, E. L. (1993). The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two? *Journal of the American Statistical Association* **88**, 1242-1249.
- Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. New York: Springer-Verlag.
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Second Edition. New York: Springer-Verlag.
- Lehmer, D. H. (1951). Mathematical Models in Large-scale Computing Units. *Proceedings of the Second Symposium on Large Scale Digital Computing Machinery*. Cambridge: Harvard University Press. 141-146.
- Li, L. and Choe, M. K. (1997). A Mixture Model for Duration Data: Analysis of Second Births in China. *Demography*. **34**, 189-197.
- Leimkuhler, B. and Reich, S. (2005). *Simulating Hamiltonian Dynamics*. Cambridge: Cambridge University Press.
- Lempers, F. B. (1971). *Posterior Probabilities of Alternative Linear Models*. Rotterdam: Rotterdam University Press.
- Leonard, T. (1975). A Bayesian Approach to the Linear Model with Unequal Variances. *Technometrics* **17**, 95-102.
- Leonard, T. and Hsu, J. S. J. (1999). *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge, England: Cambridge University Press.
- Leonard, T., Hsu, J. S. J. and Tsui, K-W. (1989). Bayesian Marginal Inference. *Journal of the American Statistical Association* **84**, 1051-1058.
- Lesaffre, E. and Kaufmann, H. (1992). Existence and Uniqueness of the Maximum Likelihood Estimator for a Multivariate Probit Model. *Journal of the American Statistical Association* **87**, 805-811.
- Levine, R. (1996). Post-Processing Random Variables. Ph.D. Thesis, Biometrics Unit, Cornell University.

- Lewis, C. and Thayer, D. T. (2009). Bayesian Decision Theory for Multiple Comparisons. In IMS Lecture Notes Monograph Series, Optimality: the Therd Erich L. Lehmann Symposium **57**, pp.326-332.
- Lewis, J. A. (1994). Discussion of the Paper by Spiegelhalter, Freedman, and Parmar. *Journal of the Royal Statistical Society, Series A* **157**, 392.
- Lewis, S. M. and Raftery, A. E. (1997). Estimating Bayes Factors Via Posterior Simulation With the Laplace-Metropolis Estimator. *Journal of the American Statistical Association* **92**, 648-655.
- Lewis, T. G. and Payne, W. H. (1973). Generalized Feedback Shift Register Pseudorandom Number Algorithm. *Journal of the Association for Computing Machinery* **20**, 456-468.
- Li, H-L. (2004). The Sampling/Importance Resampling Algorithm. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, Andrew Gelman and Xiao-Li Meng (eds.). New York: John Wiley & Sons, pp.265-276.
- Li, L. and Choe, M. K. (1997). A Mixture Model for Duration Data: Analysis of Second Births in China. *Demography* **34**, 189-197.
- Lindley, D. V. (1957). A Statistical Paradox. *Biometrika* **44**, 187-192.
- Lindley, D. V. (1958). Fiducial Distributions and Bayes' Theory. *Journal of the Royal Statistical Society, Series B* **20**, 102-107.
- Lindley, D. V. (1961). The Use of Prior Probability Distributions in Statistical Inference and Decision. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press. 453-468.
- Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint, Parts 1 and 2*. Cambridge, England: Cambridge University Press.
- Lindley, D. V. (1968). The Choice of Variables in Multiple Regression. *Journal of the Royal Statistical Society, Series B* **30**, 31-66.
- Lindley, D. V. (1969). Discussion of *Compound Decisions and Empirical Bayes*, J. B. Copas. *Journal of the Royal Statistical Society, Series B* **31**, 397-425.
- Lindley, D. V. (1972). *Bayesian Statistics: A Review*. Philadelphia: Society for Industrial and Applied Mathematics.
- Lindley, D. V. (1983). Reconciliation of Probability Distributions. *Operations Research, Journal of Operations Research Society of America* **31**, 866-880.
- Lindley, D. V. (1985). *Making Decisions*. Second Edition. New York: John Wiley & Sons.
- Lindley, D. V. (1986). Comment. *The American Statistician* **40**, 6-7.
- Lindley, D. V. (1991). Subjective Probability, Decision Analysis and Their Legal Consequences. *Journal of the Royal Statistical Society, Series A* **154**, 83-92.
- Lindley, D. V. and Novick, M. R. (1978). The Use of More Realistic Utility Functions in Educational Applications. *Journal of Educational Measurement* **15**, 181-191.
- Lindley, D. V. and Novick, M. R. (1981). The Role of Exchangeability in Inference. *Annals of Statistics* **9**, 45-58.
- Lindley, D. V. and Singpurwalla, N. D. (1991). On the Evidence Needed to Reach Agreed Action Between Adversaries, with Application to Acceptance Sampling. *Journal of the American Statistical Association* **86**, 933-937.
- Lindley, D. V. and Smith, A. F. M. (1972). Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society, Series B* **34**, 1-41.
- Lindley, D. V., Tversky, A. and Brown, R. V. (1979). On the Reconciliation of Probability Assessments. *Journal of the Royal Statistical Society, Series A* **142**, 146-180.
- Lindsay, R. M. (1995). Reconsidering the Status of Tests of Significance: An Alternative Criterion of Adequacy. *Accounting, Organizations and Society* **20**, 35-53.
- Lindsey, J. K. (1997). *Applying Generalized Linear Models*. New York: Springer-Verlag.

- Linick, T. W., Jull, A. J. T., Toolin, L. J. and Donahue, D. J. (1986). Operation of the NSF-Arizona Accelerator Facility For Radioscope Analysis and Results From Selected Collaborative Research Projects. *Radiocarbon* **28**, 522-533.
- Liniger, W. (1961). On a Method by D. H. Lehmer for the Generation of Pseudo-Random Numbers. *Numerische Mathematik* **3**, 265-270.
- Liseo, B., Petrella, L. and Salinetti, G. (1996). Bayesian Robustness: an Interactive Approach. In *Bayesian Statistics 5*, J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley, (eds.). Oxford: Oxford University Press, pp.223-253.
- Little, R. J. A. and Rubin, D. B. (1983). On Jointly Estimating Parameters and Missing Data by Maximizing the Complete-Data Likelihood. *The American Statistician* **37**, 218-220.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Second Edition. New York: John Wiley & Sons.
- Liu, C. and Rubin, D. B. (1994). The ECME Algorithm: A Simple Extension of EM and ECM With Faster Monotone Convergence. *Biometrika* **81**, 633-648.
- Liu, C. and Rubin, D. B. (1995). ML Estimation of the  $t$  Distribution Using EM and Its Extensions, ECM and ECME. *Statistica Sinica* **5**, 19-39.
- Liu, C. and Rubin, D. B. (1998). Maximum Likelihood Estimation of Factor Analysis Using the ECME Algorithm With Complete and Incomplete Data. *Statistica Sinica* **8**, 729-748.
- Liu, C., Liu, J. S. and Rubin, D. B. (1992). A Variational Control Variable for Assessing the Convergence of the Gibbs Sampler. *Proceedings of the American Statistical Association, Statistical Computing Section*. 74-78.
- Liu, J. S. (1994). The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene Regulation Problem. *Journal of the American Statistical Association* **89**, 958-966.
- Liu, J. S. (1996a). Metropolized Independent Sampling with Comparisons to Rejection Sampling and Importance Sampling. *Statistics and Computing* **6**, 113-119.
- Liu, J. S. (1996b). Nonparametric Hierarchical Bayes via Sequential Imputations. *Annals of Statistics* **24**, 911-930.
- Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer-Verlag.
- Liu, J. S., Liang, F. and Wong, W. H. (2000). The Multiple-Try Method and Local Optimization in Metropolis Sampling *Journal of the American Statistical Association* **95**, 121-134.
- Liu, J. S. and Sabatti, C. (1999). Simulated Sintering: Markov Chain Monte Carlo with Spaces Varying Dimension. In *Bayesian Statistics*, J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (eds.). Oxford: Oxford University Press, pp.389-414.
- Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes. *Biometrika* **81**, 27-40.
- Liu, J. S., Wong, W. H. and Kong, A. (1995). Correlation Structure and Convergence Rates of the Gibbs Sampler with Various Scans. *Journal of the Royal Statistical Society, Series B* **57**, 157-169.
- Locatelli, M. (2000). Simulated Annealing Algorithms for Continuous Global Optimization: Convergence Conditions. *Journal of Optimization Theory and Applications* **104**, 121-133.
- Loftus, G. R. (1991). On the Tyranny of Hypothesis Testing In the Social Sciences. *Contemporary Psychology* **36**, 102-105.
- Loftus, G. R. (1993). A Picture Is Worth a Thousand rho Values: On the Irrelevance of Hypothesis Testing In the Microcomputer Age. *Behavioral Research Methods, Instruments, and Computers* **25**, 250-256.
- Louis, T. A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society, Series B* **44**, 226-233.
- Lu, M. (2002). Enhancing Project Evaluation and Review Technique Simulation through Artificial Neural Network-based Input Modeling. *Journal of Construction Engineering and Management* September/October, 438-45.

- Lundy, M. (1985). Applications of the Annealing Algorithm to Combinatorial Problems in Statistics. *Biometrika* **72**, 191-198.
- Lundy, M. and Mees, A. (1986). Convergence of an Annealing Algorithm. *Mathematical Programming* **34**, 111-124.
- Lunn, D. J., Thomas, A., Best, N. G. and Spiegelhalter, D. (2000). WinBUGS—A Bayesian Modelling Framework: Concepts, Structure, and Extensibility. *Statistics and Computing* **10**, 325–337.
- Lunn, D. J., Jackson, C., Spiegelhalter, D. J., Best, N. G. and Thomas, A. (2012). *The BUGS book: A Practical Introduction to Bayesian Analysis*. Boca Raton: Chapman & Hall/CRC.
- Lwin, T. (1972). Estimation of the Tail Pareto Law. *Scandinavian Actuarial Journal* **55**, 170-178.
- Lynn, R. and Vanhanen, T. (2001). National IQ and Economic Development. *Mankind Quarterly* **LXI**, 415-437.
- Macdonald, R. R. (1997). On Statistical Testing in Psychology. *British Journal of Psychology* **88**, No. 2 (May), 333-49.
- MacEachern, S. N. and Berliner, L. M. (1994). Subsampling the Gibbs Sampler. *The American Statistician* **48**, 188-190.
- MacLaren, M. D. and Marsaglia, G. (1965). Uniform Random Number Generators. *Journal of the Association for Computing Machinery* **12**, 83-89.
- Madras, N. and Sezer, D. (2010). Quantitative Bounds for Markov Chain Convergence: Wasserstein and Total Variation Distances. *Bernoulli* **16**, 882-908.
- Makelainen, T., Schmidt, K. and Styan, G. P. H. (1981). On the Existence and Uniqueness of the Maximum Likelihood Estimate of a Vector-Valued Parameter in Fixed-Size Samples. *Annals of Statistics* **9**, 758-767.
- Makov, U. E., Smith, A. F. M. and Liu, Y-H. (1996). Bayesian Methods in Actuarial Science. *The Statistician* **45**, 503-515.
- Malov, S. V. (1998). Random Variables Generated by Ranks in Dependent Schemes. *Metrika* **48**, 61-67.
- Maltzman, F. and Wahlbeck, P. J. (1996). Strategic Policy Considerations and Voting Fluidity on the Burger Court. *American Political Science Review* **90**, 581-92.
- Manski, C. F. (1995). *Identification Problems in the Social Sciences*. Cambridge: Harvard University Press.
- Marden, J. I. (2000). Hypothesis Testing: From  $p$  Values to Bayes Factors. *Journal of the American Statistical Association* **95**, 1316-1320.
- Marín, J. M. (2000). A Robust Version of the Dynamic Linear Model with an Economic Application. In *Robust Bayesian Analysis*, David Ríos Insua and Fabrizio Ruggeri (eds.). New York: Springer-Verlag, pp.373-383.
- Marinari, E. and Parisi, G. (1992). Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters* **19**, 451-458.
- Maritz, J. S. (1970). *Empirical Bayes Methods*. London: Methuen.
- Maritz, J. S. and Lwin, T. (1989). *Empirical Bayes Methods*. Second Edition. New York: Chapman & Hall.
- Marsaglia, G. (1961a). Expressing a Random Variable in Terms of Uniform Random Variables. *Annals of Mathematical Statistics* **32**, 894-898.
- Marsaglia, G. (1961b). Generating Exponential Random Variables. *Annals of Mathematical Statistics* **32**, 899-900.
- Marsaglia, G. (1964). Generating a Variable from the Tail of the Normal Distribution. *Technometrics* **6**, 101-102.
- Marsaglia, G. (1968). Random Numbers Fall Mainly in the Planes. *Proceedings of the National Academy of Sciences* **61**. 25-28.

- Marsaglia, G. (1972). The Structure of Linear Congruential Sequences. In *Applications of Number Theory to Numerical Analysis*, S. K. Zaremba (ed.). San Diego: Academic Press, pp.249-286.
- Marsaglia, G. (1977). The Squeeze Method for Generating Gamma Variates. *Computers and Mathematics with Applications* **3**, 321-326.
- Marsaglia, G. (1985). A Current View of Random Number Generators. In *Computer Science and Statistics: 16th Symposium on the Interface*, L. Billard (ed.). Amsterdam: North Holland, pp.3-10.
- Marsaglia, G. and Bray, T. A. (1964). A Convenient Method for Generating Normal Variables. *SIAM Review* **6**, 260-264.
- Marsaglia, G., MacLaren, M. D. and Bray, T. A. (1964). A Fast Procedure for Generating Normal Random Variables. *Communications of the Association for Computing Machinery* **7**, 4-10.
- Marsaglia, G., Tsang, W. W. and Wang, J. (2003). Evaluating Kolmogorov's Distribution. *Journal of Statistical Software* **8**, 1-4.
- Marshall, A. W. (1956). The Use of Multi-Stage Sampling Schemes in Monte Carlo Computations. In *Symposium on Monte Carlo Methods*, M. Meyer (ed.). New York: John Wiley & Sons, pp.123-140.
- Marshall, A. W. and Olkin, I. (1967). A Multivariate Exponential Distribution. *Journal of the American Statistical Association* **62**, 30-44.
- Martikainen, P., Martelin, T., Nihtilä, E., Majamaa, K. and Seppo, K. (2005). Differences in Mortality by Marital Status in Finland from 1976 to 2000: Analyses of Changes in Marital-Status Distributions, Socio-Demographic and Household Composition, and Cause of Death. *Population Studies* **59**, 99-115.
- Martin, A. D. and Quinn, K. (2007). Assessing Preference Change on the US Supreme Court. *Journal of Law, Economics, and Organization* **23**, 365-385.
- Martín, J., Ríos Insua, D. and Ruggeri, F. (1996). Local Sensitivity Analysis in Bayesian Decision Theory. *Lecture Notes-Monograph Series* **29** (Bayesian Robustness), 119-135.
- Martín, J., Ríos Insua, D. and Ruggeri, F. (1998). Issues in Bayesian Loss Robustness. *Sankhyā, Series A* **60**, 405-417.
- Martins, A. C. R. (2009). Bayesian Updating Rules In Continuous Opinion Dynamics Models. *Journal of Statistical Mechanics: Theory and Experiment* **2**, P02017.
- Mason, R. L. and Lurie, D. (1973). Systematic Simulators of Joint Order Uniform Variates. In *Proceedings of Computer Science and Statistics: Seventh Annual Symposium on the Interface*, W. J. Kennedy (ed.). Ames, IA: Iowa State University, pp.156-162.
- Mason, W. M., Wong, G. Y. and Entwistle, B. (1983). Contextual Analysis Through the Multilevel Linear Model. In *Sociological Methodology 1983-1984*, S. Leinhardt (ed.). Oxford: Blackwell, pp.72-103.
- Matthews, P. (1993). A Slowly Mixing Markov Chain With Implications for Gibbs Sampling. *Statistics & Probability Letters* **17**, 231-236.
- Mayer, M. (1951). Report on a Monte Carlo Calculation Performed with the Eniac. In *Monte Carlo Method*, A. S. Householder, G. E. Forsyth and H. H. Germond (eds.). Applied Mathematics Series 12, Washington: National Bureau of Standards, pp.19-20.
- McArdle, J. J. (1976). Empirical Test of Multivariate Generators. In *Proceedings of the Ninth Annual Symposium on the Interface of Computer Science and Statistics*, D. C. Hoaglin and R. Welsch (eds.). Boston: Prindle, Weber, and Schmidt, pp.263-267.
- McCloskey, D. N. and Ziliak, S. T. (1996). The Standard Error of Regressions. *Journal of Economic Literature* **34**, 97-114.
- McCullagh, P. (1983). Quasi-Likelihood Functions. *Annals of Statistics* **11**, 59-67.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Second Edition. New York: Chapman & Hall.
- McCulloch, C. E. (1994). Maximum Likelihood Variance Components Estimation for Binary Data. *Journal of the American Statistical Association* **89**, 330-335.

- McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons.
- McCulloch, R. E. and Rossi, P. E. (1992). Bayes Factors for Nonlinear Hypotheses and Likelihood Distributions. *Biometrika* **79**, 663-676.
- McCullough, B. D. (1998). Assessing the Reliability of Statistical Software: Part I. *The American Statistician* **52**, 358-366.
- McCullough, B. D. (1999). Assessing the Reliability of Statistical Software: Part II. *The American Statistician* **53**, 149-159.
- McCullough, B. D. and Wilson, B. (1999). On the Accuracy of Statistical Procedures in Microsoft Excel 97. *Computational Statistics and Data Analysis* **31**, 27-37.
- McDonald, G. C. (1999). Letter to the Editor. *The American Statistician* **53**, 393.
- McGrath, K. and Waterton, J. (1986). British Social Attitudes, 1983-86 Panel Survey. London, Social and Community Planning Research.
- McGuire, J. W. (1996). Strikes in Argentina: Data Sources and Recent Trends. *Latin American Research Review* **31**, 127-150.
- McKendrick, A. G. (1926). Applications of Mathematics to Medical Problems. *Proceedings of the Edinburgh Mathematical Society* **44**, 98-130.
- McLachlan, G. J. and Basford, K. E. (1988). *Mixture Models: Inference and Application to Clustering*. New York: Marcel Dekker.
- McLachlan, G. J. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: John Wiley & Sons.
- Mead, M. (1973). *Coming of Age in Samoa*. New York: Morrow.
- Mebane, W. R., Jr. (1994). Fiscal Constraints and Electoral Manipulation in American Social Welfare. *American Political Science Review* **88**, 77-94.
- Medvedev, R. (1989). *Let History Judge: The Origins and Consequences of Stalinism*. New York: Columbia University Press.
- Meehl, P. E. (1978). Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology. *Journal of Counseling and Clinical Psychology* **46**, 806-834.
- Meehl, Paul E. (1990). Why Summaries of Research on Psychological Theories Are Often Uninterpretable. *Psychological Reports* **66**, 195-244.
- Meehl, P. E. (1997). The Problem Is Epistemology, Not Statistics: Replace Significance Tests By Confidence Intervals and Quantify Accuracy of Risky Numerical Predictions. In *What If There Were No Significance Tests*, L. L. E. Harlow, S. A. Mulaik and J. H. Steiger (eds.) Lawrence Erlbaum Associates Publishers, pp.393-425.
- Meeus, W., van de Schoot, R., Keijsers, L., Schwartz, S. J. and Branje, S. (2010). On the Progression and Stability of Adolescent Identity Formation: A Five-Wave Longitudinal Study in Early-to-Middle and Middle-to-Late Adolescence. *Child Development* **81**, 1565-1581.
- Meier, K. J. and Gill, J. (2000). *What Works: A New Approach to Program and Policy Analysis*. Boulder, CO: Westview Press.
- Meier, K. J. and Keiser, L. R. (1996). Public Administration as a Science of the Artificial: A Methodology for Prescription. *Public Administration Review* **56**, 459-466.
- Meier, K. J., Polinard, J. L. and Wrinkle, R. (2000). Bureaucracy and Organizational Performance: Causality Arguments about Public Schools. *American Journal of Political Science* **44**, 590-602.
- Meier, K. J. and Smith, K. B. (1994). Representative Democracy and Representative Bureaucracy. *Social Science Quarterly* **75**, 798-803.
- Meilijson, I. (1989). A Fast Improvement to the EM Algorithm on Its Own Terms. *Journal of the Royal Statistical Society, Series B* **51**, 127-138.

- Meng, X-L. (1994a). Posterior Predictive  $p$ -Values. *Annals of Statistics* **22**, 1142-1160.
- Meng, X-L. (1994b). On the Rate of Convergence of the ECM Algorithm. *Annals of Statistics* **22**, 326-339.
- Meng, X-L. (2000). Towards a More General Propp-Wilson Algorithm: Multistage Backward Coupling. *Monte Carlo Methods-Fields Institute Communications* **26**, 85-93.
- Meng, X-L. and Pedlow, S. (1992). EM: A Bibliographic Review With Missing Articles. *Proceedings of the Statistical Computing Section, American Statistical Association*. Alexandria, VA: American Statistical Association. 24-27.
- Meng, X-L. and Rubin, D. B. (1991). Using EM to Obtain Asymptotic Variance Covariance Matrices: the SEM Algorithm. *Journal of the American Statistical Association* **86**, 899-909.
- Meng, X-L. and Rubin, D. B. (1993). Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika* **80**, 267-278.
- Meng, X-L. and Schilling, S. (1996). Fitting Full-Information Item Factor Models and an Empirical Investigation of Bridge Sampling. *Journal of the American Statistical Association* **91**, 1254-1267.
- Meng, X-L. and van Dyk, D. A. (1999). Seeking Efficient Data Augmentation Schemes Via Conditional and Marginal Augmentation. *Biometrika* **86**, 301-320.
- Mengersen, K. L. and Robert, C. P. (1996). Testing for Mixtures: A Bayesian Entropic Approach. In *Bayesian Statistics 5*, J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.255-276.
- Mengersen, K. L. and Tweedie, R. L. (1996). Rates of Convergence of the Hastings and Metropolis Algorithms. *Annals of Statistics* **24**, 101-121.
- Mengersen, K. L., Robert, C. P. and Guihenneuc-Jouyaux, C. (1999). MCMC Convergence Diagnostics: A Review. In *Bayesian Statistics 6*, J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.415-440.
- Menzefricke, U. (1981). A Bayesian Analysis of a Change in the Precision of a Sequence of Independent Normal Random Variables at an Unknown Time Point. *Applied Statistics* **30**, 141-146.
- Metropolis, N. and Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association* **44**, 335-341.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* **21**, 1087-1091.
- Meyer, K. R. and Hall, G. R. Jr. (1992). *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*. New York: Springer-Verlag.
- Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. New York: Springer-Verlag.
- Meyn, S. P. and Tweedie, R. L. (1994). Computable Bounds for Convergence Rates of Markov Chains. *Annals of Applied Probability* **4**, 981-1011.
- Mihram, G. A. and Mihram, D. (1997). A Review and Update on Pseudo-Random Number Generation, on Seeding, and on a Source of Seeds. *ASA Proceedings of the Statistical Computing Section*. Alexandria, VA: American Statistical Association. 115-119.
- Miller, A. (2002). *Subset Selection in Regression*. Second Edition. Boca Raton: Chapman & Hall/CRC.
- Mira, A. and Tierney, L. (2001a). On the Use of Auxiliary Variables in Markov Chain Monte Carlo Sampling. *Scandinavian Journal of Statistics* **29**, 1-12.
- Mira, A. and Tierney, L. (2001b). Efficiency and Convergence Properties of Slice Samplers. *Scandinavian Journal of Statistics* **29**, 1035-53.
- Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association* **83**, 1023-1032.
- Mitra, D., Romeo, F. and Sangiovanni-Vincentelli, A. L. (1986). Convergence and Finite-Time Behavior of Simulated Annealing. *Advances in Applied Probability* **18**, 747-771.

- Mkhadri, A. (1998). On the Rate of Convergence of the ECME Algorithm. *Statistics & Probability Letters* **37**, 81-87.
- Møller, J. and Schladitz, K. (1999). Extensions of Fill's Algorithm for Perfect Simulation. *Journal of the Royal Statistical Society, Series B* **61**, 955-969.
- Monahan, J. F. and Genz, A. (1996). A Comparison of Omnibus Methods for Bayesian Computation. *Computing Science and Statistics* **27**, 471-480.
- Monahan, J. F. (2001). *Numerical Methods of Statistics*. Cambridge: Cambridge University Press.
- Montgomery, J. M. and Nyhan, B. (2010). Bayesian Model Averaging: Theoretical Developments and Practical Applications. *Political Analysis* **18**, 245-270.
- Mooney, C. Z. (1997). *Monte Carlo Simulation*. Thousand Oaks, CA: Sage.
- Moore, D. S. and McCabe, G. P. (1989). *Introduction to the Practice of Statistics*. WH Freeman/Times Books/Henry Holt & Company.
- Moran, J. L. and Solomon, P. J. (2004). A Farewell To p-values? *Critical Care and Resuscitation Journal* **6**, 130.
- Moreno, E. (1997). Bayes Factors for Intrinsic and Fractional Priors in Nested Models. In *L<sub>1</sub>-Statistical Procedures and Related Topics*. Y. Dodge (ed.). Hayward, CA: Institute of Mathematical Statistics Monograph Series, pp.257-270.
- Moreno, E. (2000). Global Bayesian Robustness for Some Classes of Prior Distributions. In *Robust Bayesian Analysis*, David Ríos Insua and Fabrizio Ruggeri (eds.). New York: Springer-Verlag, pp.45-70.
- Moreno, E. (2005). Objective Bayesian Methods for One-Sided Testing. *Test* **14**, 181-198.
- Moreno, E. and Cano, J. A. (1991). Robust Bayesian Analysis with  $\epsilon$ -Contaminations Partially Known. *Journal of the Royal Statistical Society, Series B* **53**, 143-155.
- Moreno, E. and González, A. (1990). Empirical Bayes Analysis of  $\epsilon$ -Contaminated Classes of Prior Distributions. *Brazilian Journal of Probability and Statistics* **4**, 177-200.
- Moreno, E. and Pericchi, L. R. (1991). Robust Bayesian Analysis for  $\epsilon$ -Contaminations with Shape and Quantile Restrictions. In *Proceedings of the Fifth International Symposium on Applied Stochastic Models*, R. Gutiérrez and M. Valderrama (eds.). Singapore: World Scientific, pp.454-470.
- Moreno, E. and Pericchi, L. R. (1993). Bayesian Robustness for Hierarchical,  $\epsilon$ -Contamination Models. *Journal of Statistical Planning and Inference* **37**, 159-168.
- Moreno, E., Martínez, C. and Cano, J. A. (1996). Local Robustness and Influences for Contamination Classes of Prior Distributions. In *Bayesian Robustness*, J. O. Berger, B. Betr , E. Moreno, L. R. Pericchi, F. Ruggeri, G. Salinetti and L. Wasserman (eds.). Hayward, CA: Institute of Mathematical Statistics Monograph Series 29, pp.139-156.
- Morey, R. D., Rouder, J. N., Pratte, M. S. and Speckman, P. L. (2011). Using MCMC Chain Outputs to Efficiently Estimate Bayes Factors. *Journal of Mathematical Psychology* **55**, 368-378.
- Morgan, B. J. T. (1984). *Elements of Simulation*. London: Chapman & Hall.
- Morris, C. N. (1982). Natural Exponential Families with Quadratic Variance Functions. *Annals of Statistics* **10**, 65-80.
- Morris, C. N. (1983a). Natural Exponential Families with Quadratic Variance Functions: Statistical Theory. *Annals of Statistics* **11**, 515-529.
- Morris, C. N. (1983b). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association* **78**, 47-65.
- Morrison, D. E. and Henkel, R. E. (1969). Statistical Tests Reconsidered. *The American Sociologist* **4**, 131-140.
- Morrison, D. E. and Henkel, R. E. (1970). *The Significance Test Controversy—A Reader*. Chicago: Aldine.
- Mouchart, M. and Simar, L. (1984). A Note on Least-Squares Approximation in the Bayesian Analysis of Regression Models. *Journal of the Royal Statistical Society, Series B* **46**, 124-133.



- Müller, M. E. (1958). An Inverse Method for the Generation of Random Normal Deviates on Large Scale Computers. *Mathematical Tables and Other Aids to Computation* **12**, 167-174.
- Müller, M. E. (1959a). A Comparison of Methods for Generating Normal Deviates on Digital Computers. *Journal of the Association for Computing Machinery* **6**, 376-383.
- Müller, M. E. (1959b). A Note on a Method for Generating Points Uniformly on N-dimensional Spheres. *Communications of the Association for Computing Machinery* **2**, 19-20.
- Murata, N., Yoshizawa, S. and Amari, S. (1994). Network Information Criterion-Determining the Number of Hidden Units for Artificial Neural Network Models. *IEEE Transactions on Neural Networks* **5**, 865-872.
- Murdoch, D. J. and Green, P. J. (1998). Exact Sampling From a Continuous State Space. *Scandinavian Journal of Statistics* **25**, 483-502.
- Murdoch, D. J. and Rosenthal, J. S. (1998). An Extension of Fill's Exact Sampling Algorithm to Non-Monotone Chains. Technical Report, <http://www.probability.ca/jeff/ftpdir/fill.pdf>.
- Murdoch, D. J. and Takahara, G. (2006). Perfect Sampling for Queues and Network Models. *ACM Transactions on Modeling and Computer Simulation* **16**, 76-92.
- Murray, G. D. (1977). Contribution to Discussion of Paper by A. P. Dempster, N. M. Laird and D. B. Rubin. *Journal of the Royal Statistical Society, Series B* **39**, 27-28.
- Muthén, B. and Satorra, A. (1995). Complex Sample Data in Structural Equation Modeling. *Sociological Methodology* **25**, 87-99.
- Mykland, P., Tierney, L. and Yu, B. (1995). Regeneration in Markov Chain Samplers. *Journal of the American Statistical Association* **90**, 233-241.
- Nance, R. E. and Overstreet, C. (1972). A Bibliography on Random Number Generation. *ACM Computing Reviews* **13**, 495-508.
- Nandram, B. and Chen, M.-H. (1996). Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence. *Journal of Statistical Computing and Simulation* **54**, 129-144.
- Natarajan, R. and Kass, R. E. (2000). Reference Bayesian Methods for Generalized Linear Mixed Models. *Journal of the American Statistical Association* **95**, 227-237.
- Natarajan, R. and McCulloch, C. E. (1995). A Note on the Existence of the Posterior Distribution for a Class of Mixed Models for Binomial Responses. *Biometrika* **82**, 639-643.
- Naylor, J. C. and Smith, A. F. M. (1982). Applications of a Method for the Efficient Computation of Posterior Distributions. *Applied Statistics* **31**, 214-225.
- Naylor, J. C. and Smith, A. F. M. (1983). A Contamination Model in Clinical Chemistry: An Illustration of a Method for the Efficient Computation of Posterior Distributions. *The Statistician* **32**, 82-87.
- Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo*, Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li Meng (eds.). Boca Raton: Chapman & Hall/CRC, pp.113-162.
- Neal, R. M. (1996). Sampling from Multimodal Distributions Using Tempered Transitions. *Statistics and Computing* **4**, 353-66.
- Neal, R. M. (2001). Annealed Importance Sampling. *Statistics and Computing* **11**, 125-139.
- Neal, R. M. (2003). Slice Sampling. *Annals of Statistics* **31**, 705-767.
- Neal, R. M. (1993). "Bayesian Training of Backpropagation Networks by the Hybrid Monte Carlo Method." Technical Report CRG-TR-92-1, Department of Computer Science, University of Toronto.
- Neave, H. R. (1973). On Using the Box-Müller Transformation with Multiplicative Congruential Pseudo-Random Number Generators. *Applied Statistics* **22**, 92-97.
- Neftçi, S. N. (1982). Specification of Economic Time Series Models Using Akaike's Criterion. *Journal of the American Statistical Association* **77**, 537-540.

- Nelder, J. A. (1977). Comments on "Maximum Likelihood From Incomplete Data Via the EM Algorithm." *Journal of the Royal Statistical Society, Series B* **39**, 23-24.
- Nelder, J. A. (1985). Quasi-likelihood and GLIM. *Lecture Notes in Statistics* **32**, 120-127.
- Nelder, J. A. and Lee, Y. (1992). Likelihood, Quasi-likelihood and Pseudo-likelihood: Some Comparisons. *Journal of the Royal Statistical Society, Series B* **54**, 273-284.
- Nelder, J. A. and Pregibon, D. (1987). An Extended Quasi-Likelihood Function. *Biometrika* **74**, 221-232.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society, Series A* **135**, 370-385.
- Nevzorov, V. B. and Zhukova, E. E. (1996). Wiener Process and Order Statistics. *Journal of Applied Statistical Science* **3**, 317-323.
- Newcomb, S. (1886). A Generalized Theory of the Combination of Observations So As to Obtain the Best Results. *American Journal of Mathematics* **8**, 343-366.
- Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society, Series B* **56**, 3-48.
- Neyens, T., Faes, C. and Molenberghs, G. (2012). A Generalized Poisson-Gamma Model For Spatially Overdispersed Data. *Spatial and Spatio-Temporal Epidemiology*, **3** 185-194.
- Neyman, J. and Pearson, E. S. (1928a). On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. Part I. *Biometrika* **20A**, 175-240.
- Neyman, J. and Pearson, E. S. (1928b). On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. Part II. *Biometrika* **20A**, 263-294.
- Neyman, J. and Pearson, E. S. (1933a). On the Problem of the Most Efficient Test of Statistical Hypotheses. *Philosophical Transactions of the Royal Statistical Society, Series A* **231**, 289-337.
- Neyman, J. and Pearson, E. S. (1933b). The Testing of Statistical Hypotheses in Relation to Probabilities. *Proceedings of the Cambridge Philosophical Society* **24**, 492-510.
- Neyman, J. and Pearson, E. S. (1936a). Contributions to the Theory of Testing Statistical Hypotheses. *Statistical Research Memorandum* **1**, 1-37.
- Neyman, J. and Pearson, E. S. (1936b). Sufficient Statistics and Uniformly Most Powerful Tests of Statistical Hypotheses. *Statistical Research Memorandum* **1**, 113-137.
- Ni, S. and Sun, D. (2003). Noninformative Priors and Frequentist Risks of Bayesian Estimators in Vector Autoregressive Models. *Journal of Econometrics* **115**, 159-197.
- Nickerson, R.S. (2000). Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy. *Psychological Methods* **5**, 241-301.
- Nicolaou, A. (1993). Bayesian Intervals With Good Frequentist Behaviour in the Presence of Nuisance Parameters. *Journal of the Royal Statistical Society, Series B* **55**, 377-390.
- Niederreiter, H. (1972). On the Distribution of Pseudo-Random Numbers Generated by the Linear Congruential Method. *Mathematics of Computation* **26**, 793-795.
- Niederreiter, H. (1974). On the Distribution of Pseudo-Random Numbers Generated by the Linear Congruential Method, II. *Mathematics of Computation* **28**, 1117-1132.
- Niederreiter, H. (1976). On the Distribution of Pseudo-Random Numbers Generated by the Linear Congruential Method, III. *Mathematics of Computation* **30**, 571-597.
- Nigm, A. M. and Handy, H. I. (1987). Bayesian Prediction Bounds for the Pareto Lifetime Model. *Communications in Statistics* **16**, 1761-1722.
- Norrander, B. (2000). The Multi-Layered Impact of Public Opinion on Capital Punishment Implementation in the American States. *Political Research Quarterly* **53**, 771-793.
- Norris, J. R. (1997). *Markov Chains*. Cambridge: Cambridge University Press.
- Novick, M. R. (1969). Multiparameter Bayesian Indifference Procedures. *Journal of the Royal Statistical Society, Series B* **31**, 29-64.

- Novick, M. R., Isaacs, G. L. and DeKeyrel, D. F. (1976). CADA User's Manual-1976. Iowa City: Iowa Testing Programs, The University of Iowa.
- Novick, M. R. and Hall, W. J. (1965). A Bayesian Indifference Procedure. *Journal of the American Statistical Association* **60**, 1104-1117.
- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. New York: John Wiley & Sons.
- Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge: Cambridge University Press.
- Oakes, M. (1986). *Statistical Inference: A Commentary for the Social and Behavioral Sciences*. New York: John Wiley & Sons.
- O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics: Volume 2B, Bayesian Inference*. London: Arnold.
- O'Hagan, A. (1995). Fractional Bayes Factors for Model Comparison. *Journal of the Royal Statistical Society, Series B* **57**, 99-138.
- O'Hagan, A. (1998). Eliciting Expert Beliefs in Substantial Practical Applications. *The Statistician* **47**, 21-35.
- O'Hagan, A. and Berger, J. O. (1988). Ranges of Posterior Probabilities for Quasimodal Priors With Specified Quantiles. *Journal of the American Statistical Association* **83**, 503-508.
- Oh, M.-S. and Berger, J. O. (1992). Adaptive Importance Sampling in Monte Carlo Integration. *Journal of Statistical Computation and Simulation* **41**, 143-168.
- Olivera, S. and Gill, J. (2011). Parallel Gibbs Sampling with `snowfall`. *The Political Methodologist* **19**, 4-7.
- Orans, M. (1996). *Not Even Wrong: Margaret Mead, Derek Freeman, and the Samoans*. Novato, CA: Chandler & Sharp.
- Orchard, T. and Woodbury, M. A. (1972). A Missing Information Principle: Theory and Applications. *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability* **1**, 697-715.
- Orey, S. (1961). Strong Ratio Limit Property. *Bulletin of the American Mathematical Society* **67**, 571-574.
- Orton, C. (1997). Testing Significance or Testing Credulity? *Oxford Journal of Archaeology* **16**, 219.
- Pagan, A. (1987). Three Econometric Methodologies: A Critical Appraisal. *Journal of Economic Surveys* **1**, 3-24.
- Pang, X. (2010). Modeling Heterogeneity and Serial Correlation in Binary Time-Series Cross-sectional Data: A Bayesian Multilevel Model with AR(p) Errors. *Political Analysis* **18**, 470-498.
- Pang, X. and Gill, J. (2012). Spike and Slab Prior Distributions for Simultaneous Bayesian Hypothesis Testing, Model Selection, and Prediction, of Nonlinear Outcomes. Technical Report, <http://jgill.wustl.edu/research/current.html>.
- Pasta, J. and Ulam, S. (1953). Heuristic Studies in Problems of Mathematical Physics on High Speed Computing Machines. Technical Report, Los Alamos Scientific Lab.
- Patterson, R. L. and Richardson, W. (1963). A Decision Theoretic Model for Determining Verification Requirements. *The Journal of Conflict Resolution* **7**, 603-607.
- Patterson, T. N. L. (1968). The Optimum Addition of Points to Quadrature Formulae. *Mathematics of Computation* **23**, 847-856.
- Payne, W. H. (1977). Normal Random Numbers: Using Machine Analysis to Choose the Best Algorithm. *ACM Transactions on Mathematical Software* **3**, 346-358.
- Peach, C. (1997). Postwar Migration to Europe: Reflux, Influx, Refuge. *Social Science Quarterly* **78**, 269-283.
- Pearson, K. (1892). *The Grammar of Science*. London: Walter Scott.
- Pearson, K. (1900). On the Criterion that a Given System of Deviations from the Probable in the Case of

- a Correlated System of Variables is Such That It Can Reasonably be Supposed to Have Arisen From Random Sampling. *Philosophical Magazine*, 5th Series, **50**, 157-175.
- Pearson, K. (1907). On the Influence of Past Experience on Future Expectation. *Philosophical Magazine*, 6th Series, **13**, 365-378.
- Pearson, K. (1914). On the Probability that Two Independent Distributions of Frequency are Really Samples of the Same Population, with Special Reference to Recent Work on the Identity of Trypanosome Strains. *Biometrika* **10**, 85-143.
- Pearson, K. (1920). The Fundamental Problem of Practical Statistics. *Biometrika* **13**, 1-16.
- Pennington, R. H. (1970). *Introductory Computer Methods and Numerical Analysis*. Second Edition. London: Collier-MacMillan.
- Pericchi, L. R. and Nazaret, W. (1988). On Being Imprecise at the Higher Levels of a Hierarchical Linear Model. In *Bayesian Statistics 3*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (ed.). Oxford: Oxford University Press, pp.569-577.
- Perkins, W. C. and Menzefricke, U. (1975). A Better Additive Congruential Random Number Generator? *Decision Sciences* **6**, 194-198.
- Perks, W. (1947). Some Observations on Inverse Probability Including A New Indifference Rule. *Journal of the Institute of Actuaries* **73**, 285-312.
- Peskun, P. H. (1973). Optimum Monte Carlo Sampling Using Markov Chains. *Biometrika* **60**, 607-612.
- Petrone, S. and Raftery, A. E. (1997). A Note on the Dirichlet Process Prior in Bayesian Nonparametric Inference With Partial Exchangeability. *Statistics & Probability Letters* **36**, 69-83.
- Pettit, L. I. (1992). Bayes Factors for Outlier Models Using the Device of Imaginary Observations. *Journal of the American Statistical Association* **87**, 541-545.
- Pettit, L. I. and Smith, A. F. M. (1985). Outliers and Influential Observations in Linear Models. In *Bayesian Statistics 2*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.). Amsterdam: North Holland Press, pp.473-494.
- Pettitt, A. N., Tran, T. T., Haynes, M. A. and Hay, J. L. (2006). A Bayesian Hierarchical Model for Categorical Longitudinal Data From a Social Survey of Immigrants. *Journal of the Royal Statistical Society, Series A* **169**, 97-114.
- Phillips, D. B. and Smith, A. F. M. (1996). Bayesian Model Comparison Via Jump Diffusions. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.). New York: Chapman & Hall, pp.214-240.
- Phillips, D. T. and Beightler, C. S. (1972). Procedure for Generating Gamma Variates with Non-Integer Parameter Sets. *Journal of Statistical Computation and Simulation* **1**, 197-208.
- Phillips, P. C. B. (1991). To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends. *Journal of Applied Econometrics* **6**, 333-364.
- Phillips, P. C. B. (1995). Bayesian model selection and prediction with empirical applications. *Journal of Econometrics* **69**, 289-331.
- Pitman, E. J. G. (1936). Sufficient Statistics and Intrinsic Accuracy. *Proceedings of the Cambridge Philosophical Society* **32**, 567-579.
- Pizer, S. M. (1975). *Numerical Computing and Mathematical Analysis*. Chicago: Science Research Associates.
- Plackett, R. L. (1966). Current Trends in Statistical Inference. *Journal of the Royal Statistical Society, Series A* **129**, 249-267.
- Poirer, D. J. (1988). Frequentist and Subjectivist Perspectives on the Problems of Model Building in Economics. *Journal of Economic Perspectives* **2**, 121-144.
- Poirer, D. J. (1994). Jeffreys Prior for Logit Models. *Journal of Econometrics* **63**, 327-339.

- Polasek, W. (1984). Regression Diagnostics for General Linear Regression Models. *Journal of the American Statistical Association* **79**, 336-340.
- Polasek, W. (1987). Bounds on Rounding Errors in Linear Regression Models. *The Statistician* **36**, 221-227.
- Pole, A., West, M. and Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. New York: Chapman & Hall.
- Pollard, W. E. (1986). *Bayesian Statistics for Evaluation Research*. Thousand Oaks, CA: Sage.
- Pollard, P. (1993). How Significant is 'Significance'? In *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*, G. Keren and C. Lewis (eds.). Hillsdale, NJ: Lawrence Erlbaum Associates, pp.448-460.
- Pollard, P. and Richardson, J. T. E. (1987). On the Probability of Making Type One Errors. *Psychological Bulletin* **102**, 159-163.
- Polson, N. G. (1996). Convergence of Markov Chain Monte Carlo Algorithms. In *Bayesian Statistics 5*, J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press.
- Polson, N. G. and Scott, J. G. (2012). On the Half-Cauchy Prior For a Global Scale Parameter. *Bayesian Analysis* **7**, 887-902.
- Popper, K. (1968). *The Logic of Scientific Discovery*. New York: Harper and Row.
- Pratt, J. W. (1965). Bayesian Interpretation of Standard Inference Statements. With Discussion. *Journal of the Royal Statistical Society, Series B* **27**, 169-203.
- Prentice, M. J. and Miller, J. C. P. (1968). Additive Congruential Pseudo-Random Number Generators. *Computer Journal* **11**, 341-346.
- Press, S. J. (1989). *Bayesian Statistics: Principles, Models, and Applications*. New York: John Wiley & Sons.
- Press, S. J. and Tanur, J. M. (2001). *The Subjectivity of Scientists and the Bayesian Approach*. New York: John Wiley & Sons.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986). *Numerical Recipes: The Art of Scientific Computing*. Cambridge: Cambridge University Press.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series: Volumes I and II*. San Diego: Academic Press.
- Propp, J. G. and Wilson, D. B. (1996). Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics. *Random Structures and Algorithms* **9**, 223-252.
- Quinn, K. M., Martin, A. and Whitford, A. B. (1999). Voter Choice in Multi-Party Democracies: A Test of Competing Theories and Models. *American Journal of Political Science* **43**, 1231-1247.
- Racine, A., Grieve, A. P., Flühler, H. and Smith, A. F. M. (1986). Bayesian Methods in Practice: Experiences in the Pharmaceutical Industry. *Applied Statistics* **45**, 275-309.
- Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology* **25**, 111-164.
- Raftery, A. E. (1996). Hypothesis Testing and Model Selection. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.). New York: Chapman & Hall, pp.163-188.
- Raftery, A. E. (1999). Bayesian Model Selection in Social Research. *Sociological Methodology* **25**, 111-163.
- Raftery, A. E. and Adnan, V. E. (1986). Bayesian Analysis of a Poisson Process with a Change-Point. *Biometrika* **73**, 85-89.
- Raftery, A. E. and Banfield, J. D. (1991). Stopping the Gibbs Sampler, the Use of Morphology, and Other Issues in Spatial Statistics. *Annals of the Institute of Statistical Mathematics* **43**, 32-43.
- Raftery, A. E. and Lewis, S. M. (1992). How Many Iterations in the Gibbs Sampler? In *Bayesian Statistics 4*, J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (eds.). Oxford: Oxford University Press, pp.763-773.
- Raftery, A. E. and Lewis, S. M. (1996). Implementing MCMC. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.). New York: Chapman & Hall, pp.115-130.

- Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Cambridge: Harvard School of Business Administration.
- Ramsey, J. O. and Novick, M. R. (1980). PLU Robust Bayesian Decision Theory: Point Estimation. *Journal of the American Statistical Association* **75**, 901-907.
- Rao, C. R. and Toutenburg, H. (1995). *Linear Models: Least Squares and Alternatives*. New York: Springer-Verlag.
- Rathbun, S. L. and Black, B. (2006). Modeling and Spatial Prediction of Pre-Settlement Patterns of Forest Distribution Using Witness Tree Data. *Environmental and Ecological Statistics* **13**, 427-448.
- Raudenbush, S. and Bryk, A. S. (1986). A Hierarchical Model for Studying School Effects. *Sociology of Education* **59**, 1-17.
- Ravishanker, N. and Dey, D. K. (2002). *A First Course In Linear Model Theory*. New York: Chapman & Hall/CRC.
- Regazzini, E. (1992). Concentration Comparisons Between Probability Measures. *Sankhyā, Series B* **54**, 129-149.
- Revuz, D. (1975). *Markov Chains*. Amsterdam: North-Holland.
- Richardson, S. and Green, P. J. (1997). On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society, Series B* **59**, 731-732.
- Richey, M. (2010). The Evolution of Markov Chain Monte Carlo Methods. *The American Mathematical Monthly* **117**, 383-413.
- Ripley, B. D. (1979). Algorithm AS 137: Simulating Spatial Patterns: Dependent Samples from a Multivariate Density. *Applied Statistics* **28**, 109-112.
- Ripley, B. D. (1983). Computer Generation of Random Variables-A Tutorial. *International Statistical Review* **51**, 301-319.
- Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Association* **87**, 861-868.
- Robbins, H. (1955). An Empirical Bayes Approach to Statistics. In *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability* **1**, Berkeley: University of California Press. pp.157-164.
- Robbins, H. (1964). The Empirical Bayes Approach to Statistical Decision Problems. *Annals of Mathematical Statistics* **35**, 1-20.
- Robbins, H. (1983). Some Thoughts on Empirical Bayes Estimation. *Annals of Statistics* **1**, 713-723.
- Robert, C. P. (1995). Convergence Control Methods for Markov Chain Monte Carlo Algorithms. *Statistical Science* **10**, 231-253.
- Robert, C. P. (1996). Mixtures of Distributions: Inference and Estimation. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.). New York: Chapman & Hall, pp.441-464.
- Robert, C. P. (1997). Discussion of Richardson and Green's Paper. *Journal of the Royal Statistical Society, Series B* **59**, 758-764.
- Robert, C. P. (2001). *The Bayesian Choice: A Decision Theoretic Motivation*. Second Edition. New York: Springer-Verlag.
- Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. First Edition. New York: Springer-Verlag.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Second Edition. New York: Springer-Verlag.
- Robert, C. P. and Casella, G. (2011). A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statistical Science* **26**, 102-115.
- Robert, C. P. and Mengersen, K. L. (1999). Reparameterization Issues in Mixture Estimation and Their Bearings on the Gibbs Sampler. *Computational Statistics and Data Analysis* **29**, 325-343.

- Robert, C. P. and Cellier, D. (1998). Convergence Control of MCMC Algorithms. In *Discretization and MCMC Convergence Assessment, Lecture Notes in Statistics*, **135**. Christian P. Robert (ed.). New York: Springer-Verlag, pp.27-46.
- Robert, C. P. and Richardson, S. (1998). Markov Chain Monte Carlo Methods. In *Discretization and MCMC Convergence Assessment, Lecture Notes in Statistics*, **135**. Christian P. Robert (ed.). New York: Springer-Verlag, pp.1-25.
- Roberts, G. O. (1992). Convergence Diagnostics of the Gibbs Sampler. In *Bayesian Statistics 4*, J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (eds.). Oxford: Oxford University Press, pp.775-782.
- Roberts, G. O. (1994). Methods for Estimating  $L^2$  Convergence of Markov Chain Monte Carlo. In *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, D. Berry, K. Chaloner and J. Geweke (eds.). New York: John Wiley & Sons, pp.373-384.
- Roberts, G. O. and Polson, N. G. (1994). On the Geometric Convergence of the Gibbs Sampler. *Journal of the Royal Statistical Society, Series B* **56**, 377-384.
- Roberts, G. O. and Rosenthal, J. S. (1998). Markov Chain Monte Carlo: Some Practical Implications of Theoretical Results. *Canadian Journal of Statistics* **26**, 5-32.
- Roberts, G. O. and Rosenthal, J. S. (1999). Convergence of the Slice Sampler Markov Chains. *Journal of the Royal Statistical Society, Series B* **61**, 643-60.
- Roberts, G. O. and Rosenthal, J. S. (2011). Quantitative Non-Geometric Convergence Bounds for Independence Samplers. *Methodology and Computing in Applied Probability* **13**, 391-403.
- Roberts, G. O. and Sahu, S. K. (1997). Updating Schemes, Correlation Structure, Blocking and Parameterization for the Gibbs Sampler. *Journal of the Royal Statistical Society, Series B* **59**, 291-307.
- Roberts, G. O. and Smith, A. F. M. (1994). Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms. *Stochastic Processes and their Applications* **49**, 207-216.
- Roberts, G. O. and Tweedie, R. L. (1996). Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and Metropolis Algorithms. *Biometrika* **83**, 95-110.
- Roberts, G. O. and Rosenthal, J. S. (1998b). Two Convergence Properties of Hybrid Samplers. *Annals of Applied Probability* **8**, 397-407.
- Robinson, D. H. and Levin, J. R. (1997). Research News and Comment: Reflections On Statistical and Substantive Significance, With a Slice of Replication. *Educational Researcher* **26**, 21-26.
- Robinson, P. M. (1991). Consistent Nonparametric Entropy-Based Testing. *The Review of Economic Studies* **58**, 437-453.
- Roethlisberger, F. and Dickson, W. (1939). *Management and the Worker*. Cambridge: Cambridge University Press.
- Romney, A. K. (1999). Culture Consensus as a Statistical Model. *Current Anthropology* **40** (Supplement), S103-S115.
- Rosay, A. B. and Herz, C. D. (2000). Differences in the Validity of Self-Reported Drug Use Across Five Factors in Indianapolis, Fort Lauderdale, Phoenix, and Dallas, 1994. ICPSR02706-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].
- Rosenblatt, M. (1971). *Markov Processes. Structure and Asymptotic Behavior*. New York: Springer-Verlag.
- Rosenkranz, R. D. (1977). *Inference, Method, and Decision. Towards a Bayesian Philosophy of Science*. Dordrecht: Reidel.
- Rosenthal, J. S. (1993). Rates of Convergence for Data Augmentation on Finite Sample Spaces. *Annals of Applied Probability* **3**, 819-839.
- Rosenthal, J. S. (1995a). Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *Journal of the American Statistical Association* **90**, 558-566.
- Rosenthal, J. S. (1995b). Rates of Convergence for Gibbs Sampling for Variance Components Models. *Annals of Statistics* **23**, 740-61.

- Rosenthal, J. S. (1995c). Convergence Rates for Markov Chains. *SIAM Review* **37**, 387-405.
- Rosenthal, J. S. (1996). Analysis of the Gibbs Sampler for a Model Related to James-Stein Estimators. *Statistics and Computing* **6**, 269-75.
- Rosnow, R. L. and Rosenthal, J. S.. (1989). Statistical Procedures and the Justification of Knowledge in Psychological Science. *American Psychologist* **44**, 1276-84.
- Ross, S. (1996). *Stochastic Processes*. New York: Wiley & Sons.
- Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.
- Rozeboom, W. W. (1960). The Fallacy of the Null Hypothesis Significance Test. *Psychological Bulletin*. **57**, 416-428.
- Rubin, D. B. (1980). Using Empirical Bayes Techniques in the Law School Validity Studies. *Journal of the American Statistical Association* **75**, 801-827.
- Rubin, D. B. (1981). The Bayesian Bootstrap. *Annals of Statistics* **9**, 130-134.
- Rubin, D. B. (1984). Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *Annals of Statistics* **12**, 1151-1172.
- Rubin, D. B. (1987a). A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations When Fractions of Missing Information Are Modest: the SIR Algorithm. Discussion of Tanner & Wong (1987). *Journal of the American Statistical Society* **82**, 543-546.
- Rubin, D. B. (1987b). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- Rubin, D. B. (1988). Using the SIR Algorithm to Simulate Posterior Distributions. In *Bayesian Statistics 3*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.395-402.
- Rubin, D. B. (1991). EM and Beyond. *Psychometrika* **56**, 241-254.
- Rubin, H. (1977). Robust Bayesian Estimation. In *Statistical Decision Theory and Related Topics II*, S. S. Gupta and D. Moore (eds.). San Diego: Academic Press, pp.351-356.
- Rubin, H. (1987). A Weak System of Axioms for "Rational" Behavior and the Nonseparability of Utility From Prior. *Statistical Decisions* **5**, 47-58.
- Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*. New York: John Wiley & Sons.
- Ruggeri, F. (1990). Posterior Ranges of Functions of Parameters under Priors with Specified Quantiles. *Communications in Statistics, Part A, Theory and Methods* **19**, 127-144.
- Ruggeri, F. and Wasserman, L. (1993). Infinitesimal Sensitivity of Posterior Distributions. *Canadian Journal of Statistics* **21**, 195-203.
- Russell, B. (1929). *Mysticism and Logic and Other Essays*. New York: W.W. Norton & Company.
- Sagan, A. (2013). Market Research and Preference Data. In *The Sage Handbook of Multilevel Modeling*. Marc A. Scott, Jeffrey S. Simonoff and Brian D. Marx (eds.). Thousand Oaks, CA: Sage Publications, pp.581-598.
- Sakia, R. M. (1992). The Box-Cox Transformation Technique: A Review. *The Statistician* **41**, 169-178.
- Sala-I-Martin, X. (1997). I Just Ran Two Million Regressions. *The American Economic Review* **87**, 178-183.
- Samaniego, F. J. and Reneau, D. M. (1994). Toward a Reconciliation of the Bayesian and Frequentist Approach to Point Estimation. *Journal of the American Statistical Association* **89**, 947-957.
- Sampson, R. J. and Raudenbush, S. W. (1999). Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods. *American Journal of Sociology* **105**, 603-651.
- Sansó, B. and Pericchi, L. R. (1992). Near Ignorance Classes of Log-Concave Priors for the Location Model. *TEST*, **1**, 39-46.



- Satten, G. A. and Longini, I. M. (1996). Markov Chains With Measurement Errors: Estimating the True Course of a Marker of the Progression of HIV Disease. *Applied Statistics* **45**, 275-309.
- Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.
- Savage, L. J. (1962). *The Foundations of Statistical Inference*. London: Methuen.
- Savage, L. J. (1971). Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association* **66**, 783-801.
- Savage, L. J. (1972). *The Foundations of Statistics*. New York: Dover Publications.
- Sawa, T. (1978). Information Criteria for Discriminating among Alternative Regression Models. *Econometrica* **46**, 1273-1291.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.
- Schervish, M. J. (1995). *Theory of Statistics*. New York: Springer-Verlag.
- Schervish, M. J. (1996). P values: What They Are and What They Are Not. *The American Statistician* **50**, 203-206.
- Schervish, M. J. and Carlin, B. P. (1992). On the Convergence of Successive Substitution Sampling. *Journal of Computational and Graphical Statistics* **1**, 111-127.
- Scheuer, E. M. and Stoller, D. S. (1962). On the Generation of Normal Random Vectors. *Technometrics* **4**, 278-281.
- Schmidt, F. L. (1996). Statistical Significance Testing and Cumulative Knowledge in Psychology: Implications for the Training of Researchers. *Psychological Methods* **1**, 115-129.
- Schmidt, F. L. and Hunter, J. E. (1977). Development of a General Solution to the Problem of Validity Generalization. *Journal of Applied Psychology* **62**, 529-40.
- Schruben, L. W. (1982). Detecting Initialization Bias in Simulation Output. *Operations Research* **30**, 569-590.
- Schruben, L. W., Singh, H. and Tierney, L. (1983). Optimal Tests for Initialization Bias in Simulation Output. *Operations Research* **31**, 1167-1178.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics* **6**, 461-464.
- Scollnik, D. P. M. (2001). Actuarial Modeling with MCMC and BUGS. *North American Actuarial Journal* **5**, 95-124.
- Scott, David W. (1985). Average Shifted Histograms: Effect Nonparametric Density Estimators in Several Dimensions. *Annals of Statistics* **13**, 1024-1040.
- Sedlmeier, P. and Gigerenzer, G. (1989). Do Studies of Statistical Power Have an Effect on the Power of Studies? *Psychological Bulletin* **105**, 309-15.
- Seewald, W. (1992). Discussion of Hills and Smith (1992). In *Bayesian Statistics 4*, J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (eds.). Oxford: Oxford University Press, pp.241-243.
- Seidenfeld, T. (1985). Calibration, Coherence, and Scoring Rules. *Philosophy of Science* **52**, 274-294.
- Seidenfeld, T., Schervish, M. J. and Kadane, J. B. (1995). A Representation of Partially Ordered Preferences. *The Annals of Statistics* **23**, 2168-2217.
- Seltzer, M. H., Wong, W. H. and Bryk, A. S. (1996). Bayesian Analysis in Applications of Hierarchical Models: Issues and Methods. *Journal of Educational and Behavioral Statistics* **21**, 131-167.
- Selvin, S. (1975). On the Monty Hall problem. (Letter To the Editor.) *The American Statistician* **29**, 134.
- Sened, I. and Schofield, N. (2005). Multiparty Competition in Israel, 1988-96 *British Journal of Political Science* **35**, 635-663.
- Serlin, R. C. and Lapsley, D. K. (1993). Rational Appraisal of Psychological Research and the Good-Enough Principle. In *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*, G. Keren and C. Lewis (eds.). Hillsdale, NJ: Lawrence Erlbaum Associates, pp.199-228.
- Severini, T. A. (1991). On the Relationship Between Bayesian and Non-Bayesian Interval Estimates. *Journal of the Royal Statistical Society, Series B* **53**, 611-618.

- Severini, T. A. (1993). Bayesian Interval Estimates Which Are Also Confidence Intervals. *Journal of the Royal Statistical Society, Series B* **55**, 533-540.
- Shafer, G. (1982). Lindley's Paradox. *Journal of the American Statistical Association* **77**, 325-351.
- Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technology Journal* **27**, 379-423, and 623-56.
- Shao, J. (1989). Monte Carlo Approximations in Bayesian Decision Theory. *Journal of the American Statistical Association* **84**, 727-732.
- Shao, J. (2005). *Mathematical Statistics*. Second Edition. New York: Springer-Verlag.
- Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.
- Sharp, A. M., Register, C. A. and Grimes, P. W. (1999). *Economics of Social Issues*. Fifteenth Edition. New York: McGraw-Hill.
- Shaw, J. E. H. (1988). A Quasirandom Approach To Integration in Bayesian Statistics. *Annals of Statistics* **16**, 895-914.
- Sheynin, O. B. (1977). Laplace's Theory of Errors. *Archive for History of Exact Sciences* **17**, 1-61.
- Shyamalkumar, N. D. (2000). Likelihood Robustness. In *Robust Bayesian Analysis*, David Ríos Insua and Fabrizio Ruggeri (eds.). New York: Springer-Verlag, pp.127-144.
- Sibuya, M. (1961). Exponential and Other Variable Generators. *Annals of the Institute of Statistical Mathematics* **13**, 231-237.
- Siegmund, D. (1976). Importance Sampling in the Monte Carlo Study of Sequential Tests. *Annals of Statistics* **4**, 673-684.
- Silvapulle, M. J. (1981). On the Existence of Maximum Likelihood Estimates for the Binomial Response Models. *Journal of the Royal Statistical Society, Series B* **43**, 310-313.
- Sims, C. A. (1988). Bayesian Skepticism on Unit Root Econometrics. *Journal of Economic Dynamics and Control* **12**, 463-474.
- Sims, C. A. and Uhlig, H. (1991). Understanding Unit Rooters: A Helicopter Tour. *Econometrica* **59**, 1591-1599.
- Sinclair, A. J. and Jerrum, M. R. (1988). Conductance and the Rapid Mixing Property for Markov Chains: The Approximation of the Permanent Resolved. *Proceedings of the 20th Annual ACM Symposium on the Theory of Computing*, 235-44.
- Sinclair, A. J. and Jerrum, M. R. (1989). Approximate Counting, Uniform Generation and Rapidly Mixing Markov Chains. *Information and Computation* **82**, 93-133.
- Singleton, R., Jr. and Straight, B. C. (2004). *Approaches to Social Research*. Fourth Edition. New York: Oxford University Press.
- Sivaganesan, S. (1993). Robust Bayesian Diagnostics. *Journal of Statistical Planning and Inference* **35**, 171-188.
- Sivaganesan, S. (2000). Global and Local Robustness Approaches: Uses and Limitations. In *Robust Bayesian Analysis*, David Ríos Insua and Fabrizio Ruggeri (eds.). New York: Springer-Verlag, pp.89-108.
- Sivaganesan, S. and Berger, J. O. (1989). Ranges of Posterior Measures for Priors with Unimodal Contaminations. *Annals of Statistics* **17**, 868-889.
- Skates, S. J., Pauler, D. K. and Jacobs, I. J. (2001). Screening Based on the Risk of Cancer Calculation from Bayesian Hierarchical Change-point and Mixture Models of Longitudinal Markers. *Journal of the American Statistical Association* **96**, 429-439.
- Skene, A. M. (1983). Computing Marginal Distributions for the Dispersion Parameters of Analysis of Variance Models. *The Statistician* **32**, 99-108.
- Skene, A. M., Shaw, E. H. and Lee, T. D. (1986). Bayesian Modeling and Sensitivity Analysis. *The Statistician* **35**, 281-288.
- Skinner, B. F. (1953). *Science and Human Behavior*. Toronto: Macmillan.

- Smith, A. F. M. (1973). A General Bayesian Linear Model. *Journal of the Royal Statistical Society, Series B* **35**, 61-75.
- Smith, A. F. M. (1975). A Bayesian Approach to Inference About a Change-point in a Sequence of Random Variables. *Biometrika* **62**, 407-416.
- Smith, A. F. M. (1984). Present Position and Potential Developments: Some Personal Views: Bayesian Statistics. *Journal of the Royal Statistical Society, Series A* **147**, 245-259.
- Smith, A. F. M. (1986). Some Bayesian Thoughts on Modelling and Model Choice. *The Statistician* **35**, 97-101.
- Smith, A. F. M. and Roberts, G. O. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Series B* **55**, 3-24.
- Smith, A. F. M. and Spiegelhalter, D. J. (1980). Bayes Factors and Choice Criteria for Linear Models. *Journal of the Royal Statistical Society, Series B* **42**, 213-220.
- Smith, C. A. B. (1965). Personal Probability and Statistical Analysis. *Journal of the Royal Statistical Society, Series A* **128**, 469-499.
- Smith, Kevin B. and Meier, Kenneth J. (1995). *The Case Against School Choice: Politics, Markets, and Fools*. Armonk, NY: M.E. Sharpe.
- Smith, R. L. (1996). The Hit-And-Run Sampler: A Globally Reaching Markov Chain Sampler for Generating Arbitrary Multivariate Distributions. *Proceedings of the 1996 Winter Simulation Conference* J. M. Charnes, D. J. Morrice, D. T. Brunner and J. J. Swain (eds.). 260-264. Coronado, CA.
- Smith, W. B. and Hocking, R. R. (1972). Wishart Variate Generator. Algorithm AS53. *Applied Statistics* **21**, 241-245.
- Smyth, G. K. (1989). Generalized Linear Models with Varying Dispersion. *Journal of the Royal Statistical Society, Series B* **51**, 47-60.
- Sobol, I. M. (1994). *A Primer for the Monte Carlo Method*. New York: Chapman & Hall.
- Solzhenitsyn, A. I. (1997). *The Gulag Archipelago. Volume I*. Thomas P. Whitney and H. Willetts, translators. Boulder, CO: Westview Press.
- Sowey, E. R. (1972). A Chronological and Classified Bibliography on Random Number Generation and Testing. *International Statistical Review* **40**, 355-371.
- Sowey, E. R. (1978). A Second Classified Bibliography on Random Number Generation and Testing. *International Statistical Review* **46**, 89-102.
- Spall, J. C. and Hill, S. D. (1990). Least-Informative Bayesian Prior Distributions for Finite Samples Based on Information Theory. *IEEE Transactions on Automatic Control* **35**, 580-583.
- Spetzler, C. S. and Staël von Holstein, C. S. (1975). Probability Encoding in Decision Analysis. *Management Science* **22**, 340-358.
- Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. New York: John Wiley & Sons.
- Spiegelhalter, D. J., Freedman, L. S. and Parmar, M. K. B. (1994). Bayesian Approaches to Randomized Trials. *Journal of the Royal Statistical Society, Series A* **157**, 357-416.
- Spiegelhalter, D. J. and Smith, A. F. M. (1982). Bayes Factors for Linear and Log-linear Models with Vague Prior Information. *Journal of the Royal Statistical Society, Series B* **44**, 377-387.
- Spiegelhalter, D. J., Thomas, A., Best, N. G. and Gilks, W. R. (1996a). *BUGS 0.5\*Examples: Volume 1 (version i)*. MRC Biostatistics Unit:  
[http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS\\_Vol1.pdf](http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS_Vol1.pdf).
- Spiegelhalter, D. J., Thomas, A., Best, N. G. and Gilks, W. R. (1996b). *BUGS 0.5\*Examples: Volume 2 (version i)*. MRC Biostatistics Unit:  
[http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS\\_Vol2.pdf](http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS_Vol2.pdf).  
[http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS\\_Vol3.pdf](http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS_Vol3.pdf)

- Spiegelhalter, D. J., Thomas, A., Best, N. G. and Lunn, D. J. (2000). *WinBUGS User Manual Version 1.4* MRC Biostatistics Unit:  
<http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf>.
- Spiegelhalter, D. J., Thomas, A., Best, N. G. and Lunn, D. J. (2012). *BUGS 0.5\*Examples: Volume 3 (version i)*. MRC Biostatistics Unit:  
[http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS\\_Vol3.pdf](http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS_Vol3.pdf).
- Spiegelhalter, D., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society, Series B* **64**, 583-640.
- Stangl, D. K. (1995). Prediction and Decision Making Using Bayesian Hierarchical Models. *Statistics in Medicine* **14**, 2173-2190.
- Steen, N. M., Byrne, G. D. and Gelbard, E. M. (1969). Gaussian Quadrature for the Integrals  $\int_0^\infty f(x)dx$  and  $\int_0^b \exp(-x^2)f(x)dx$ . *Mathematics of Computation* **23**, 661-671.
- Steffey, D. (1992). Hierarchical Bayesian Modeling With Elicited Prior Information. *Communications in Statistics* **21**, 799-821.
- Stein, C. (1955). Inadmissability of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. **1** Berkeley: University of California Press, pp.197-206.
- Stein, C. (1965). Approximation of Improper Prior Measures by Prior Probability Measures. In *Bernoulli-Bayes-Laplace Anniversary Volume: Proceedings of an International Research Seminar Statistical Laboratory*, J. Neyman and L. M. Le Cam (eds.). New York: Springer-Verlag, pp.217-240.
- Stein, C. (1981). Estimation of the Mean of a Multivariate Normal Distribution. *Annals of Statistics* **9**, 1135-1151.
- Stephens, D. A. (1994). Bayesian Retrospective Multiple-Changepoint Identification. *Applied Statistics* **43**, 159-178.
- Stephens, M. (2000). Bayesian Analysis of Mixture Models with an Unknown Number of Components-An Alternative to Reversible Jump Methods. *Annals of Statistics* **28**, 40-74.
- Stern, S. E. (1997). Simulation-Based Estimation. *Journal of Economic Literature* **XXXV**, 2006-2039.
- Stern, H. S. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models *Statistics in Medicine* **19**, 2377-2397.
- Stewart, L. (1983). Bayesian Analysis Using Monte Carlo Integration-a Powerful Methodology for Handling Some Difficult Problems. *Statistician* **32**, 195-200.
- Stewart, L. and Davis, W. W. (1986). Bayesian Posterior Distributions Over Sets of Possible Models With Inferences Computed by Monte Carlo Integration. *The Statistician* **35**, 175-182.
- Stigler, S. M. (1982). Thomas Bayes' Bayesian Inference. *Journal of the Royal Statistical Society, Series A* **145**, 250-258.
- Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- Stigler, S. M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Cambridge, MA: Harvard University Press.
- Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society, Series B* **36**, 111-147.
- Stone, M. (1977a). An Asymptotic Equivalence of Choice Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Association, Series B* **39**, 44-47.
- Stone, M. (1977b). Comments On Model Selection Criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society, Series B*, **41**, 276-278.
- Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals*. Englewood Cliffs, NJ: Prentice-Hall.

- Stroud, A. H. and Secrest, D. H. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, NJ: Prentice-Hall.
- Stuart, A. and Ord, J. K. (1994). *Kendall's Advanced Theory of Statistics: Volume I, Distribution Theory*. Sixth Edition. London: Edward Arnold.
- Student, A. (1908a). On the Probable Error of a Mean. *Biometrika* **6**, 1.
- Student, A. (1908b). On the Probable Error of a Correlation Coefficient. *Biometrika* **6**, 302.
- Suchard, M. A., Weiss, R. E., Dorman, K. S. and Sinsheimer, J. S. (2003). Inferring Spatial Phylogenetic Variation along Nucleotide Sequences: A Multiple Changepoint Model. *Journal of the American Statistical Association* **98**, 427-437.
- Sundberg, R. (1974). Maximum Likelihood Theory for Incomplete Data from an Exponential Family. *Scandinavian Journal of Statistics* **1**, 49-58.
- Sundberg, R. (1976). An Iterative Method for Solution of the Likelihood Equations for Incomplete Data From Exponential Families. *Communications in Statistics, Part B, Simulation and Computation* **5**, 55-64.
- Suppes, P. (1974). The Measurement of Belief. *Journal of the Royal Statistical Society, Series B* **36**, 160-191.
- Swendsen, R. H. and Wang, J. S. (1987). Nonuniversal Critical Dynamics in Monte Carlo Simulations. *Physical Review Letters* **58**, 86-88.
- Szu, H. and Hartley, R. (1987). Fast Simulated Annealing. *Physics Letters A* **122**, 157-162.
- Tadikamalla, P. R. (1978). Computer Generation of Gamma Random Variables. *Communications of the Association for Computing Machinery* **21**, 419-422.
- Tadikamalla, P. R. and Ramberg, J. S. (1975). An Approximate Method for Generating Gamma and Other Variates. *Journal of Statistical Computation and Simulation* **3**, 275-282.
- Takeuchi, K. (1976). Distribution of Informational Statistics and A Criterion of Model Fitting. *Suri-Kagaku (Mathematical Statistics)* **153**, 12-18.
- Tanner, M. A. (1996). *Tools for Statistic Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York: Springer.
- Tanner, M. A. and Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Society* **82**, 528-550.
- Taralsden, G. and Lindqvist, B. H. (2010). Improper Priors Are Not Improper. *The American Statistician* **64**, 154-158.
- Thatcher, A. R. (1964). Relationships Between Bayesian and Confidence Limits for Predictions. *Journal of the Royal Statistical Society, Series B* **26**, 176-210.
- Theil, H. (1963). On the Use of Incomplete Prior Information in Regression Analysis. *Journal of the American Statistical Association* **58**, 401-414.
- Theil, H. (1970). On the Estimation of Relationships Involving Qualitative Variables. *American Journal of Sociology* **76**, 103-154.
- Theil, H. and Goldberger, A. S. (1961). On Pure and Mixed Statistical Estimation in Economics. *International Economic Review* **2**, 65-78.
- Thisted, R. (1988). *Elements of Statistical Computing*. New York: Chapman & Hall.
- Thompson, B. (1996). AERA Editorial Policies Regarding Statistical Significance Testing: Three Suggested Reforms. *Educational Researcher* **25**, 26-30.
- Thompson, B. (1997). Editorial Policies Regarding Statistical Significance Testing: Further Comments. *Educational Researcher* **26**, 29-32.
- Thompson, B. (2002a). Statistical, Practical, and Clinical: How Many Kinds of Significance Do Councilors Need to Consider? *Journal of Counseling and Development* **80**, 64-71.

- Thompson, B. (2002b). What Future Quantitative Social Science Research Could Look Like: Confidence Intervals for Effect Sizes. *Educational Researcher* **31**, 24-31.
- Thompson, B. (2004). The “Significance” Crisis in Psychology and Education. *Journal of Socio-Economics* **33**, 607-613.
- Thompson J. R., Palmer, T. M. and Moreno, S. (2006). Bayesian Analysis In Stata Using WinBUGS. *The Stata Journal* **6**, pp.530-549.
- Tiao, G. C. and Zellner, A. (1964a). Bayes’s Theorem and the Use of Prior Knowledge in Regression Analysis. *Biometrika* **51**, 219-230.
- Tiao, G. C. and Zellner, A. (1964b). On the Bayesian Estimation of Multivariate Regression. *Journal of the Royal Statistical Society, Series B* **26**, 277-285.
- Tierney, L. (1991). Exploring Posterior Distributions Using Markov Chains. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*. E. M. Keramidas (ed.). Fairfax Station, VA: Interface Foundation, pp.563-570.
- Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics* **22**, 1701-1728.
- Tierney, L. (1996). Introduction to General State-Space Markov Chain Theory. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.). New York: Chapman & Hall, pp.59-74.
- Tierney, L. and Kadane, J. B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association* **81**, 82-86.
- Tierney, L., Kass, R. E. and Kadane, J. B. (1989a). Fully Exponential Laplace Approximations to Expectations and Variances of Nonsensitive Functions. *Journal of the American Statistical Association* **84**, 710-716.
- Tierney, L., Kass, R. E. and Kadane, J. B. (1989b). Approximate Marginal Densities of Nonlinear Functions. *Biometrika* **76**, 425-433.
- Tierney, L. and Mira, A. (1999). Some Adaptive Monte Carlo Methods for Bayesian Inference. *Statistics in Medicine* **18**, 2507-2515.
- Titterton, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: John Wiley & Sons.
- Tobin, James. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica* **26**, 24-36.
- Tong, Y. L. (1990). *The Multivariate Normal Distribution*. New York: Springer-Verlag.
- Toothill, J. P. R., Robinson, W. D. and Adams, A. G. (1971). The Runs Up and Down Performance of Tausworthe Pseudo-Random Number Generators. *Journal of the Association for Computing Machinery* **18**, 381-399.
- Torrie, G. M. and Valleau, R. L. (1977). Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *Journal of Computational Physics* **23**, 187-199.
- Tsai, T-H. and Gill, J. (2012). **superdiag**: A Comprehensive Test Suite for Markov Chain Non- Convergence. *The Political Methodologist*. **19**, 12-18.
- Tsitsiklis, J. N. (1988). A Survey of Large Time Asymptotics of Simulated Annealing Algorithms. In *Stochastic Differential Systems, Stochastic Control Theory and Applications*, W. Fleming and P. L. Lions (eds.). New York: Springer, pp.583-599.
- Turkov, P., Krasotkina, O. and Mottl, V. (2012). The Bayesian Logistic Regression In Pattern Recognition Problems Under Concept Drift. 2012 21st International Conference on Pattern Recognition (ICPR). Tsukuba, Japan, 11-15 Nov. 2012, 2976-2979.
- Turnbull, B. W. (1976). The Empirical Distribution with Arbitrary Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society, Series B* **38**, 290-295.
- Tversky, A. (1974). Assessing Uncertainty. *Journal of the Royal Statistical Society, Series B* **36**, 148-159.

- Tversky, A. and Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science* **185**, 1124-1131.
- Tweedie, R. L. (1975). Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on a General State Space. *Stochastic Processes Applications* **3**, 385-403.
- Ulam, S. M. (1961). On Some Statistical Properties of Dynamical Systems. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability* **3**, 315-320.
- van Dijk, H. K. and Kloek, T. (1982). Monte Carlo Analysis of Skew Posterior Distributions: An Illustrative Econometric Example. *The Statistician* **32**, 216-223.
- van Dyk, D. and Meng, X-L. (2001). The Art of Data Augmentation. *Journal of Computational and Graphical Statistics* **10**, 1-50.
- van Houwelingen, H. C. (1977). Monotonizing Empirical Bayes Estimators for a Class of Discrete Distributions with Monotone Likelihood Ratio. *Statistica Neerlandica* **31**, 95-104.
- van Houwelingen, H. C. and Stijnen, T. (1993). Monotone Empirical Bayes Estimators Based on More Informative Samples. *Journal of the American Statistical Association* **88**, 1438-1443.
- van Laarhoven, P. J. M. and Aarts, E. H. L. (1987). *Simulated Annealing: Theory and Applications*. Reidel: Dordrecht.
- van Laarhoven, P. J. M., Boender, P., Aarts, E. H. L. and Rinnooy, K. A. (1989). A Bayesian Approach to Simulated Annealing. *Probability in the Engineering and Informational Sciences* **3**, 453-475.
- Vanpaemel, W. (2010). Prior Sensitivity In Theory Testing: An Apologia For the Bayes Factor. *Journal of Mathematical Psychology* **54**, 491-498.
- Varian, H. R. (1975). A Bayesian Approach to Real Estate Assessment. In *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, S.E. Fienberg and A. Zellner (eds.). North-Holland, Amsterdam, pp.195-208.
- Venables, W. N. and Ripley, B. D. (1999). *Modern Applied Statistics with S-Plus*, Third Edition. New York: Springer-Verlag.
- Venn, J. (1866). *The Logic of Chance*. London: Macmillan.
- Verma, V., Gordon, G., Simmons, R. and Thrun, S. (2004). Real-time Fault Diagnosis. *Robotics & Automation Magazine, IEEE* **11**, 56-66.
- Verma, V., Langford, J. and Simmons, R. (2001). Non-Parametric Fault Identification for Space Rovers. *International Symposium on Artificial Intelligence and Robotics in Space*.
- Verma, V., Thrun, S. and Simmons, R. (2003). Variable Resolution Particle Filter. *Proceedings of the International Joint Conference on Artificial Intelligence, AAAI*, August.
- Vidakovic, B. (1999). Linear Versus Nonlinear Rules for Mixture Normal Priors. *Annals of the Institute of Statistical Mathematics* **51**, 111-124.
- Villegas, C. (1977). On the Representation of Ignorance. *Journal of the American Statistical Association* **72**, 651-654.
- Vines, S. K., Gilks, W. R. and Wild, P. (1996). Fitting Bayesian multiple random effects models. *Statistics and Computing* **6**, 337-346.
- Volpe, R., Nesnas, I. A. D., Estlin, T., Mutz, D., Petras, R. and Das, H. (2001). The CLARAty Architecture for Robotic Autonomy. *Proceedings of the 2001 IEEE Aerospace Conference*, Big Sky, Montana, March 10-17.
- Volpe, R. and Peters, S. (2003). Rover Technology Development and Infusion for the 2009 Mars Science Laboratory Mission. *Proceedings of 7th International Symposium on Artificial Intelligence, Robotics, and Automation in Space*, Nara, Japan, May 19-23.
- Von Mises, R. (1957). *Probability, Statistics and Truth*. Second Revised English Edition prepared by H. Geiringer. London: George Allen and Unwin.

- von Neumann, J. (1951). Various Techniques Used in Connection with Random Digits, "Monte Carlo Method." *U.S. National Bureau of Standards Applied Mathematics Series* **12**, 36-38.
- Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* **57**, 307-333.
- Waagepetersen, R. and Sorensen, D. (2001). A Tutorial on Reversible Jump MCMC with a View toward Applications in QTL-Mapping. *International Statistical Review/Revue Internationale de Statistique* **69**, 49-61.
- Wagner, K. and Gill, J. (2005). Bayesian Inference in Public Administration Research: Substantive Differences from Somewhat Different Assumptions. *International Journal of Public Administration* **28**, 5-35.
- Wakefield, J. C., Gelfand, A. E. and Smith, A. F. M. (1991). Efficient Generation of Random Variates via the Ratio-of-Uniforms Method. *Statistics and Computing* **1**, 129-133.
- Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. and Gelfand, A. E. (1994). Bayesian Analysis of Linear and Non-linear Population Models by using the Gibbs Sampler. *Applied Statistics* **43**, 201-221.
- Walker, A. J. (1974). Fast Generation of Uniformly Distributed Pseudorandom Numbers with Floating-Point Representation. *Electronics Letters* **10**, 533-534.
- Wallace, N. D. (1974). Computer Generation of Gamma Random Variables with Non-Integral Shape Parameters. *Communications of the Association for Computing Machinery* **17**, 691-695.
- Walley, P., Gurrin, L. and Burton, P. (1996). Analysis of Clinical Data Using Imprecise Prior Probabilities. *Statistician* **45**, 457-485.
- Wasserman, L. (1992). Recent Methodological Advances in Robust Bayesian Inference. In *Bayesian Statistics 4*, J. O. Berger, J. M. Bernardo, A. P. Dawid and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.763-773.
- Wedderburn, R. W. M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika* **61**, 439-447.
- Wedderburn, R. W. M. (1976). On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models. *Biometrika* **63**, 27-32.
- Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithm. *Journal of the American Statistical Association* **85**, 699-704.
- Weiss, L. (1961). *Statistical Decision Theory*. New York: McGraw-Hill.
- Weiss, R. E. (1996). An Approach to Bayesian Sensitivity Analysis. *Journal of the Royal Statistical Society, Series B* **58**, 739-750.
- Wenzel, W. and Hamacher, K. (1999). A Stochastic Tunneling Approach for Global Minimization. *Physical Review Letters* **82**, 3003-3007.
- West, M. (1984). Outlier Models and Prior Distributions in Bayesian Linear Regression. *Journal of the Royal Statistical Society, Series B* **46**, 431-439.
- West, M. (1992). Modelling With Mixtures. In *Bayesian Statistics 4*, J. O. Berger, J. M. Bernardo, A. P. Dawid and A. F. M. Smith (eds.). Oxford: Oxford University Press, pp.227-246.
- West, M. (1993). Approximating Posterior Distributions by Mixtures. *Journal of the Royal Statistical Society, Series B* **55**, 409-422.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.
- West, M. and Turner, D. A. (1994). Deconvolution of Mixtures in Analysis of Neural Synaptic Transmission. *Statistician* **43**, 31-43.
- Western, B. (1995). A Comparative Study of Working-Class Disorganization: Union Decline in Eighteen Advanced Capitalist Countries. *American Sociological Review* **60**, 179-201.
- Western, B. (1998). Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach. *American Journal of Political Science* **42**, 1233-1259.



- Western, B. (1999). Bayesian Methods for Sociologists: An Introduction. *Sociological Methods & Research* **28**, 7-34.
- Westlake, W. J. (1967). A Uniform Random Number Generator Based on the Combination of Two Congruential Generators. *Journal of the Association for Computing Machinery* **14**, 337-340.
- Wetzels, R., Raaijmakers, J. G. W., Jakab, E. and Wagenmakers, E.-J. (2009). How to Quantify Support For and Against the Null Hypothesis: A Flexible WinBUGS Implementation of a Default Bayesian t-Test. *Psychonomic Bulletin & Review* **16**, 752-760.
- Wheeler, D. J. (1974). Simulation of Arbitrary Gamma Distributions. *IEEE Transactions* **6**, 167-169.
- Wheeler, D. J. (1975). An Approximation for Simulation of Gamma Distributions. *Journal of Statistical Computation and Simulation* **3**, 225-232.
- White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica* **50**, 1-26.
- White, H. (1996). *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.
- Whittaker, J. (1974). Generating Gamma and Beta Random Variables with Non-Integral Shape Parameters. *Applied Statistics* **23**, 210-214.
- Whittaker, J. (1990). *Graphical Models in Applied Mathematical Multivariate Analysis*. New York: John Wiley & Sons.
- Whittlesey, J. R. B. (1969). On the Multidimensional Uniformity of Pseudo-Random Generators. *Communications of the Association for Computing Machinery* **12**, 247.
- Wiles, P. (1965). Rationalizing the Russians. *New York Review of Books* (October, 28) **5**, 37-38.
- Wilkinson, G. N. (1977). On Resolving the Controversy in Statistical Inference. *Journal of the Royal Statistical Society, Series B* **39**, 119-171.
- Winkler, R. L. (1967). The Assessment of Prior Distributions in Bayesian Analysis. *Journal of the American Statistical Association* **62**, 776-800.
- Wolfson, L. J., Kadane, J. B. and Small, M. J. (1996). Bayesian Environmental Policy Decisions: Two Case Studies. *Ecological Applications* **6**, 1056-1066.
- Wong, G. Y. and Mason, W. M. (1985). The Hierarchical Logistic Regression Model for Multilevel Analysis. *Journal of the American Statistical Association* **80**, 513-524.
- Wong, G. Y. and Mason, W. M. (1991). Contextually Specific Effects and Other Generalizations of the Hierarchical Linear Model for Comparative Analysis. *Journal of the American Statistical Association* **86**, 487-503.
- Wong, W. H. and Li, B. (1992). Laplace Expansion for Posterior Densities of Nonlinear Functions of Parameters. *Biometrika* **79**, 393-398.
- Woodward, Phil. (2011). *Bayesian Analysis Made Simple: An Excel GUI for WinBUGS*. Boca Raton: Chapman & Hall/CRC.
- Wright, G. and Ayton, P. (1994). *Subjective Probability*. New York: John Wiley & Sons.
- Wu, C. F. J. (1983). On the Convergence Properties of the EM Algorithm. *Annals of Statistics* **11**, 95-103.
- Yang, T. Y. and Kuo, L. (2001). Bayesian Binary Segmentation Procedure For a Poisson Process With Multiple Changepoints. *Journal of Computational and Graphical Statistics* **10**, 772-785.
- Young, C. (2009). Model Uncertainty in Sociological Research: An Application to Religion and Economic Growth. *American Sociological Review* **74**, 380-397.
- Yu, B. and Mykland, P. (1998). Looking at Markov Samplers Through CUSUM Path Plots: A Simple Diagnostic Idea. *Statistics and Computing* **8**, 275-286.
- Zabell, S. (1989). R. A. Fisher on the History of Inverse Probability. *Statistical Science* **4**, 247-263.
- Zangwill, W. I. (1969). *Nonlinear Programming: A Unified Approach*. Englewood Cliffs, NJ: Prentice-Hall.
- Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley & Sons.
- Zellner, A. (1975). Bayesian Analysis of Regression Error Terms. *Journal of the American Statistical Association* **70**, 138-144.

- Zellner, A. (1976). Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms. *Journal of the American Statistical Association* **71**, 400-405.
- Zellner, A. (1985). Bayesian Econometrics. *Econometrica* **53**, 253-269.
- Zellner, A. and Chetty, V. K. (1965). Prediction and Decision Problems in Regression Models from the Bayesian Point of View. *Journal of the American Statistical Association* **60**, 608-616.
- Zellner, A. and Highfield, R. A. (1988). Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions. *Journal of Econometrics* **37**, 195-209.
- Zellner, A. and Min, C-K. (1995). Gibbs Sampler Convergence Criteria. *Journal of the American Statistical Association* **90**, 921-927.
- Zellner, A. and Moulton, B. R. (1985). Bayesian Regression Diagnostics with Applications to International Consumption and Income Data. *Journal of Econometrics* **29**, 187-211.
- Zellner, A. and Rossi, P. E. (1984). Bayesian Analysis of Dichotomous Quantal Response Models. *Journal of Econometrics* **25**, 365-393.
- Zellner, A. and Siow, A. (1980). Posterior Odds Ratios for Selected Regression Hypotheses. In *Bayesian Statistics*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.). Valencia: Valencia University Press, pp.586-603.
- Zellner, A. and Tiao, G. C. (1964). Bayesian Analysis of the Regression Model With Autocorrelated Errors. *Journal of the American Statistical Association* **59**, 763-778.
- Zhang, W. and Luck, S. J. (2011). The Number and Quality of Representations in Working Memory *Psychological Science* **22**, 1434-1441.
- Zhang, Z., Chan, K. L., Wu, Y. and Chen, C. (2004). Learning a Multivariate Gaussian Mixture Model With the Reversible Jump MCMC Algorithm. *Statistics and Computing* **14**, 343-355.
- Zhao, X. and Chu, P. S. (2006). Bayesian Multiple Changepoint Analysis of Hurricane Activity in the Eastern North Pacific: A Markov Chain Monte Carlo Approach. *Journal of Climate*. **19**, 4893-4901.
- Zhenting, H. and Qingfeng, G. (1978). *Homogeneous Denumerable Markov Chains*. Berlin: Springer-Verlag Science Press.
- Zhou, X. and Reiter, J. P. (2010). A Note on Bayesian Inference After Multiple Imputation. *The American Statistician* **64**, 159-163.
- Ziliak, S. T. and McCloskey, D. N. (2007). *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor: University of Michigan Press.
- Zuur, G., Garthwaite, P. H. and Fryer, R. J. (2002). Practical Use of MCMC Methods: Lessons from a Case Study. *Biometrical Journal* **44**, 433-455.