CAMBRIDGE

Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

Computational Bayesian Statistics

An Introduction

Meaningful use of advanced Bayesian methods requires a good understanding of the fundamentals. This engaging book explains the ideas that underpin the construction and analysis of Bayesian models, with particular focus on computational methods and schemes. The unique features of the text are the extensive discussion of available software packages combined with a brief but complete and mathematically rigorous introduction to Bayesian inference. The text introduces Monte Carlo methods, Markov chain Monte Carlo methods, and Bayesian software, with additional material on model validation and comparison, transdimensional MCMC, and conditionally Gaussian models. The inclusion of problems makes the book suitable as a textbook for a first graduate-level course in Bayesian computation with a focus on Monte Carlo methods. The extensive discussion of Bayesian software – R/R-INLA, OpenBUGS, JAGS, STAN, and BayesX – makes it useful also for researchers and graduate students from beyond statistics.

MARIA ANTÓNIA AMARAL TURKMAN was, until 2013, a full professor in the Department of Statistics and Operations Research, Faculty of Sciences, University of Lisbon. Though retired from the university, she is still a member of its Centre of Statistics and Applications, where she held the position of scientific coordinator until 2017. Her research interests are Bayesian statistics, medical and environmental statistics, and spatiotemporal modeling, with recent publications on computational methods in Bayesian statistics, with an emphasis on applications in health and forest fires. She has served as vice president of the Portuguese Statistical Society. She has taught courses on Bayesian statistics and computational statistics, among many others.

CARLOS DANIEL PAULINO is a senior academic researcher in the Center of Statistics and Applications and was an associate professor with habilitation in the Department of Mathematics of the Instituto Superior Técnico, both at the University of Lisbon. He has published frequently on Bayesian statistics and categorical data, with emphasis on applications in biostatistics. He has served as president of the Portuguese Statistical Society. He has taught many undergraduate and graduate-level courses, notably in mathematical statistics and Bayesian statistics.

PETER MÜLLER is a professor in the Department of Mathematics and the Department of Statistics & Data Science at the University of Texas at Austin. He has published widely on computational methods in Bayesian statistics, nonparametric Bayesian statistics, and decision problems, with emphasis on applications in biostatistics and bioinformatics. He has served as president of the International Society for Bayesian Analysis and as chair for the Section on Bayesian Statistics of the American Statistical Association. Besides many graduate-level courses, he has taught short courses on Bayesian biostatistics, Bayesian clinical trial design, nonparametric Bayesian inference, medical decision-making, and more.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

INSTITUTE OF MATHEMATICAL STATISTICS TEXTBOOKS

Editorial Board
Nancy Reid (University of Toronto)
Ramon van Handel (Princeton University)
Xuming He (University of Michigan)
Susan Holmes (Stanford University)

ISBA Editorial Representative
Peter Müller (University of Texas at Austin)

IMS Textbooks give introductory accounts of topics of current concern suitable for advanced courses at master's level, for doctoral students and for individual study. They are typically shorter than a fully developed textbook, often arising from material created for a topical course. Lengths of 100–290 pages are envisaged. The books typically contain exercises.

In collaboration with the International Society for Bayesian Analysis (ISBA), selected volumes in the IMS Textbooks series carry the "with ISBA" designation at the recommendation of the ISBA editorial representative.

Other Books in the Series (*with ISBA)

1. *Probability on Graphs*, by Geoffrey Grimmett
2. *Stochastic Networks*, by Frank Kelly and Elena Yudovina
3. *Bayesian Filtering and Smoothing*, by Simo Särkkä
4. *The Surprising Mathematics of Longest Increasing Subsequences*, by Dan Romik
5. *Noise Sensitivity of Boolean Functions and Percolation*, by Christophe Garban and Jeffrey E. Steif
6. *Core Statistics*, by Simon N. Wood
7. *Lectures on the Poisson Process*, by Günter Last and Mathew Penrose
8. *Probability on Graphs (Second Edition)*, by Geoffrey Grimmett
9. *Introduction to Malliavin Calculus*, by David Nualart and Eulàlia Nualart
10. *Applied Stochastic Differential Equations*, by Simo Särkkä and Arno Solin
11. **Computational Bayesian Statistics*, by M. Antónia Amaral Turkman, Carlos Daniel Paulino, and Peter Müller

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

Computational Bayesian Statistics

An Introduction

Maria Antónia Amaral Turkman
University of Lisbon

Carlos Daniel Paulino
University of Lisbon

Peter Müller
University of Texas, Austin

CAMBRIDGE
UNIVERSITY PRESS

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi - 110025, India

103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108703741
DOI: 10.1017/9781108646185

© Maria Antónia Amaral Turkman, Carlos Daniel Paulino, and Peter Müller 2019

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2019

A catalogue record for this publication is available from the British Library

ISBN 978-1-108-48103-8 Hardback
ISBN 978-1-108-70374-1 Paperback

Additional resources for this publication at www.cambridge.org/9781108703741

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

CAMBRIDGE

Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

Contents

Preface to the English Version viii
Preface ix

1 Bayesian Inference 1
1.1 The Classical Paradigm 2
1.2 The Bayesian Paradigm 5
1.3 Bayesian Inference 8
1.3.1 Parametric Inference 8
1.3.2 Predictive Inference 12
1.4 Conclusion 13
Problems 14

2 Representation of Prior Information 17
2.1 Non-Informative Priors 18
2.2 Natural Conjugate Priors 23
Problems 26

3 Bayesian Inference in Basic Problems 28
3.1 The Binomial ∧ Beta Model 28
3.2 The Poisson ∧ Gamma Model 30
3.3 Normal (Known μ) ∧ Inverse Gamma Model 31
3.4 Normal (Unknown μ, σ^2) ∧ Jeffreys' Prior 31
3.5 Two Independent Normal Models ∧ Marginal Jeffreys' Priors 33
3.6 Two Independent Binomials ∧ Beta Distributions 34
3.7 Multinomial ∧ Dirichlet Model 36
3.8 Inference in Finite Populations 38
Problems 40

4 Inference by Monte Carlo Methods 43
4.1 Simple Monte Carlo 43
4.1.1 Posterior Probabilities 47
4.1.2 Credible Intervals 47

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

4.1.3 Marginal Posterior Distributions 48
4.1.4 Predictive Summaries 50
4.2 Monte Carlo with Importance Sampling 50
4.2.1 Credible Intervals 54
4.2.2 Bayes Factors 56
4.2.3 Marginal Posterior Densities 57
4.3 Sequential Monte Carlo 59
4.3.1 Dynamic State Space Models 59
4.3.2 Particle Filter 60
4.3.3 Adapted Particle Filter 62
4.3.4 Parameter Learning 63
Problems 64

5 Model Assessment 70
5.1 Model Criticism and Adequacy 70
5.2 Model Selection and Comparison 76
5.2.1 Measures of Predictive Performance 76
5.2.2 Selection by Posterior Predictive Performance 81
5.2.3 Model Selection Using Bayes Factors 83
5.3 Further Notes on Simulation in Model Assessment 85
5.3.1 Evaluating Posterior Predictive Distributions 85
5.3.2 Prior Predictive Density Estimation 86
5.3.3 Sampling from Predictive Distributions 87
Problems 88

6 Markov Chain Monte Carlo Methods 90
6.1 Definitions and Basic Results for Markov Chains 91
6.2 Metropolis–Hastings Algorithm 94
6.3 Gibbs Sampler 98
6.4 Slice Sampler 105
6.5 Hamiltonian Monte Carlo 107
6.5.1 Hamiltonian Dynamics 107
6.5.2 Hamiltonian Monte Carlo Transition Probabilities 111
6.6 Implementation Details 113
Problems 116

7 Model Selection and Trans-dimensional MCMC 129
7.1 MC Simulation over the Parameter Space 129
7.2 MC Simulation over the Model Space 131
7.3 MC Simulation over Model and Parameter Space 135
7.4 Reversible Jump MCMC 138
Problems 143

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

8 Methods Based on Analytic Approximations 150
8.1 Analytical Methods 151
8.1.1 Multivariate Normal Posterior Approximation 151
8.1.2 The Classical Laplace Method 154
8.2 Latent Gaussian Models (LGM) 159
8.3 Integrated Nested Laplace Approximation 161
8.4 Variational Bayesian Inference 164
8.4.1 Posterior Approximation 164
8.4.2 Coordinate Ascent Algorithm 165
8.4.3 Automatic Differentiation Variational Inference 168
Problems 168

9 Software 172
9.1 Application Example 173
9.2 The BUGS Project: WinBUGS and OpenBUGS 173
9.2.1 Application Example: Using R2OpenBUGS 175
9.3 JAGS 181
9.3.1 Application Example: Using R2jags 181
9.4 Stan 185
9.4.1 Application Example: Using RStan 186
9.5 BayesX 192
9.5.1 Application Example: Using R2BayesX 194
9.6 Convergence Diagnostics: the Programs CODA and BOA 198
9.6.1 Convergence Diagnostics 199
9.6.2 The CODA and BOA Packages 201
9.6.3 Application Example: CODA and BOA 203
9.7 R-INLA and the Application Example 213
9.7.1 Application Example 215
Problems 222

Appendix A. Probability Distributions 224
Appendix B. Programming Notes 229
References 232
Index 241

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

Preface to the English Version

This book is based on lecture notes for a short course that was given at the *XXII Congresso da Sociedade Portuguesa de Estatística*. In the translation from the original Portuguese text we have added some additional material on sequential Monte Carlo, Hamiltonian Monte Carlo, transdimensional Markov chain Monte Carlo (MCMC), and variational Bayes, and we have introduced problem sets. The inclusion of problems makes the book suitable as a textbook for a first graduate-level class in Bayesian computation with a focus on Monte Carlo methods. The extensive discussion of Bayesian software makes it useful also for researchers and graduate students from beyond statistics.

The core of the text lies in Chapters 4, 6, and 9 on Monte Carlo methods, MCMC methods, and Bayesian software. Chapters 5, 7, and 8 include additional material on model validation and comparison, transdimensional MCMC, and conditionally Gaussian models. Chapters 1 through 3 introduce the basics of Bayesian inference, and could be covered fairly quickly by way of introduction; these chapters are intended primarily for review and to introduce notation and terminology. For a more in-depth introduction we recommend the textbooks by Carlin and Louis (2009), Christensen et al (2011), Gelman et al (2014a) or Hoff (2009).

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

Preface

In 1975, Dennis Lindley wrote an article in *Advances in Applied Probability* titled "The future of statistics: a Bayesian 21st century," predicting for the twenty-first century the predominance of the Bayesian approach to inference in statistics. Today one can certainly say that Dennis Lindley was right in his prediction, but not exactly in the reasons he gave. He did not foresee that the critical ingredient would be great advances in computational Bayesian statistics made in the last decade of the twentieth century. The "Bayesian solution" for inference problems is highly attractive, especially with respect to interpretability of the inference results. However, in practice, the derivation of such solutions involves in particular the evaluation of integrals, in most cases multi-dimensional, that are difficult or impossible to tackle without simulation. The development of more or less sophisticated computational methods has completely changed the outlook. Today, Bayesian methods are used to solve problems in practically all areas of science, especially when the processes being modeled are extremely complex. However, Bayesian methods can not be applied blindly. Despite the existence of many software packages for Bayesian analysis, it is critical that investigators understand what these programs output and why.

The aim of this text, associated with a minicourse given at the *XXII Congresso da Sociedade Portuguesa de Estatística*, is to present the fundamental ideas that underlie the construction and analysis of Bayesian models, with particular focus on computational methods and schemes.

We start in Chapter 1 with a brief summary of the foundations of Bayesian inference with an emphasis on the principal differences between the classical and Bayesian paradigms. One of the main pillars of Bayesian inference, the specification of prior information, is unfortunately often ignored in applications. We review its essential aspects in Chapter 2. In Chapter 3, analytically solveable examples are used to illustrate the Bayesian solution to statistical inference problems. The "great idea" behind the development of computational Bayesian statistics is the recognition that Bayesian infer-

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

ence can be implemented by way of simulation from the posterior distribution. Classical Monte Carlo methods are presented in Chapter 4 as a first solution for computational problems. Model validation is a very important question, with its own set of concepts and issues in the Bayesian context. The most widely used methods to assess, select, and compare models are briefly reviewed in Chapter 5.

Problems that are more complex than the basic ones in Chapter 4 require the use of more sophisticated simulation methods, in particular Markov chain Monte Carlo (MCMC) methods. These are introduced in Chapter 6, starting as simply as possible. Another alternative to simulation is the use of posterior approximations, which is reviewed in Chapter 8. The chapter describes, in a generic fashion, the use of integrated nested Laplace approximation (INLA), which allows for substantial improvements in both computation times (by several factors), and in the precision of the reported inference summaries. Although applicable in a large class of problems, the method is more restrictive than stochastic simulation. Finally, Chapter 9 is dedicated to Bayesian software. The possibility of resorting to MCMC methods for posterior simulation underpins the development of the software BUGS, which allows the use of Bayesian inference in a large variety of problems across many areas of science. Rapid advances in technology in general have changed the paradigm of statistics, with the increasing need to deal with massive data sets ("Big Data"), often of spatial and temporal types. As a consequence, posterior simulation in problems with complex and high-dimensional data has become a new challenge, which gives rise to new and better computational methods and the development of software that can overcome the earlier limitations of BUGS and its successors, WinBUGS and OpenBUGS. In Chapter 9 we review other statistics packages that implement MCMC methods and variations, such as JAGS, Stan, and BayesX. This chapter also includes a brief description of the R package R-INLA, which implements INLA.

For the compilation of this text we heavily relied on the book *Estatística Bayesiana* by Paulino, A. Turkman, and Murteira, published by Fundação Calouste Gulbenkian in 2003. As all copies of this book were sold a long while ago, we also extensively used preliminary work for an upcoming second edition, as well as material that we published in the October 2013 edition of the bulletin of the Sociedade Portuguesa de Estatística (SPE).

This text would not have been completed in its current form without the valuable and unfailing support of our dear friend and colleague Giovani Silva. We owe him sincere thanks. We are also thankful to the Sociedade Portuguesa de Estatística for having proposed the wider theme of Bayesian

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-108-70374-1 — Computational Bayesian Statistics
M. Antónia Amaral Turkman, Carlos Daniel Paulino, Peter Müller
Frontmatter
More Information

statistics and for the opportunity to give a minicourse at the 22nd conference of the society. We also acknowledge the institutional support from the Universidade de Lisboa through the Centro de Estatística e Aplicações (PEst-OE/MAT/UI0006/2014, UID/MAT/00006/2013), in the Department of Statistics and Operations Research in the Faculdade de Ciências and of the Department of Mathematics in the Instituto Superior Técnico. We would like to acknowledge that the partial support by the Funda cão para a Ciência e Tecnologia through various projects over many years enabled us to build up this expertise in Bayesian statistics.

Finally, we would like to dedicate this book to Professor Bento Murteira to whom the development of Bayesian statistics in Portugal owes a lot. In fact, Chapter 1 in this book reflects in many ways the flavor of his writings.

© in this web service Cambridge University Press www.cambridge.org