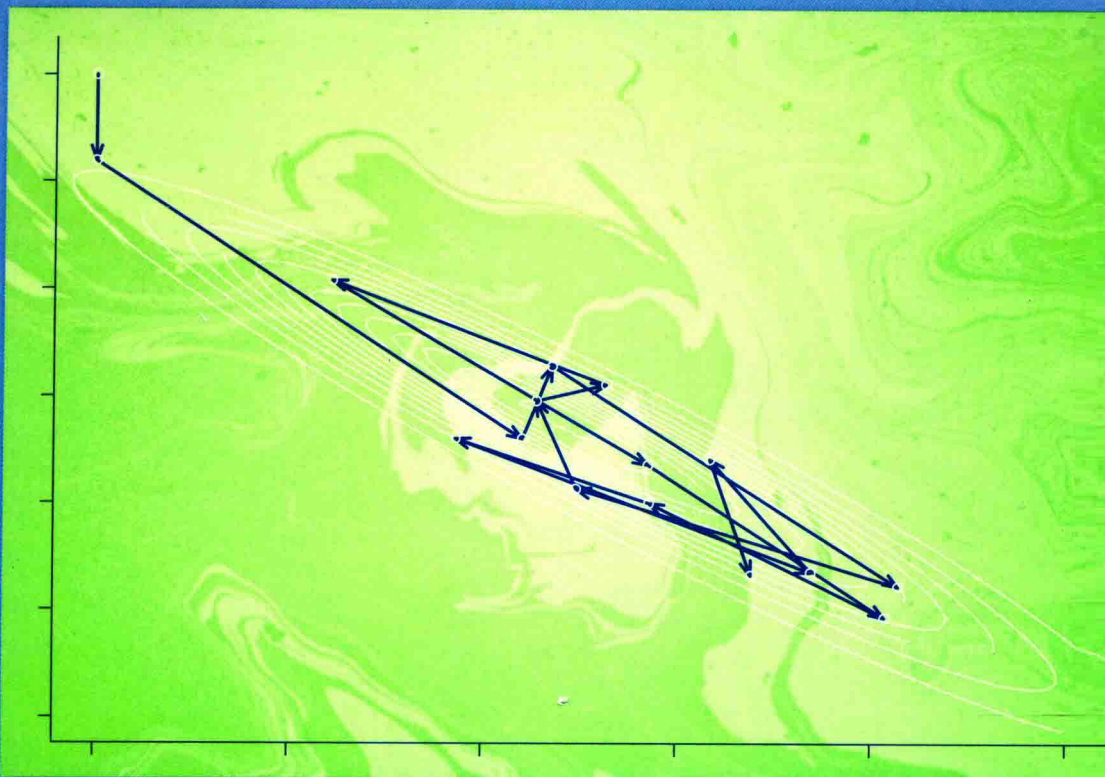


WILEY SERIES IN COMPUTATIONAL STATISTICS



Bayesian Modeling Using WinBUGS

Ioannis Ntzoufras

 **WILEY**

WWW.
LINK AVAILABLE

Bayesian Modeling Using WinBUGS

Ioannis Ntzoufras

Department of Statistics

Athens University of Economics and Business

Athens, Greece



WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2009 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data is available.

Ntzoufras, Ioannis, 1973-
Bayesian modeling using WinBUGS / Ioannis Ntzoufras.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-14114-4 (pbk.)
1. Bayesian statistical decision theory. 2. WinBUGS. I. Title.
QA279.5.N89 2009
519.5'42—dc22 2008033316

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Bayesian Modeling Using WinBUGS

WILEY SERIES IN COMPUTATIONAL STATISTICS

Consulting Editors:

Paolo Giudici

University of Pavia, Italy

Geof H. Givens

Colorado State University, USA

Bani K. Mallick

Texas A&M University, USA

Wiley Series in Computational Statistics is comprised of practical guides and cutting edge research books on new developments in computational statistics. It features quality authors with a strong applications focus. The texts in the series provide detailed coverage of statistical concepts, methods and case studies in areas at the interface of statistics, computing, and numerics.

With sound motivation and a wealth of practical examples, the books show in concrete terms how to select and to use appropriate ranges of statistical computing techniques in particular fields of study. Readers are assumed to have a basic understanding of introductory terminology.

The series concentrates on applications of computational methods in statistics to fields of bioinformatics, genomics, epidemiology, business, engineering, finance and applied statistics.

A complete list of titles in this series appears at the end of the volume.

To Ioanna and our baby daughter

PREFACE

Since the mid-1980s, the development of widely accessible powerful computers and the implementation of Markov chain Monte Carlo (MCMC) methods have led to an explosion of interest in Bayesian statistics and modeling. This was followed by an extensive research for new Bayesian methodologies generating the practical application of complicated models used over a wide range of sciences. During the late 1990s, BUGS emerged in the foreground. BUGS was a free software that could fit complicated models in a relatively easy manner, using standard MCMC methods. Since 1998 or so, WinBUGS, the Windows version of BUGS, has earned great popularity among researchers of diverse scientific fields. Therefore, an increased need for an introductory book related to Bayesian models and their implementation via WinBUGS has been realized.

The objective of the present book is to offer an introduction to the principles of Bayesian modeling, with emphasis on model building and model implementation using WinBUGS. Detailed examples are provided, ranging from very simple to more advanced and realistic ones. Generalized linear models (GLMs), which are familiar to most students and researchers, are discussed. Details concerning model building, prior specification, writing the WinBUGS code and the analysis and interpretation of the WinBUGS output are also provided. Because of the introductory character of the book, I focused on elementary models, starting from the normal regression models and moving to generalized linear models. Even more advanced readers, familiar with such models, may benefit from the Bayesian implementation using WinBUGS.

Basic knowledge of probability theory and statistics is assumed. Computations that could not be performed in WinBUGS are illustrated using R. Therefore, a minimum knowledge of R is also required.

This manuscript can be used as the main textbook in a second-level course of Bayesian statistics focusing on modeling and/or computation. Alternatively, it can serve as a companion (to a main textbook) in an introductory course of a Bayesian statistics. Finally, because of its structure, postgraduate students and other researchers can complete a self-taught tutorial course on Bayesian modeling by following the material of this book.

All datasets and code used in the book are available in the book's Webpage: www.stat-athens.aueb.gr/~jbn/winbugs_book.

IOANNIS NTZOUFRAS

Athens, Greece

June 29, 2008

ACKNOWLEDGMENTS

I am indebted to the people at Wiley publications for their understanding and assistance during the preparation of the manuscript. Acknowledgments are due to the anonymous referees. Their suggestions and comments led to a substantial improvement of the present book. I would particularly like to thank Dimitris Fouskakis, colleague and good friend, for his valuable comments on an early version of chapters 1–6 and 10–11. I am also grateful to Professor Brani Vidakovic for proposing and motivating this book. Last but not least, I wish to thank my wife Ioanna for her love, support, and patience during the writing of this book as well as for her suggestions on the manuscript.

I. N.

ACRONYMS

ACF	Autocorrelation
AIC	Akaike information criterion
ANOVA	Analysis of variance
ANCOVA	Analysis of covariance
AR	Attributable risk
BF	Bayes factor
BIC	Bayes information criterion
BOA	Bayesian output analysis (R package)
BP	Bivariate Poisson
BOD	Biological oxygen demand (data variable in example 6.3)
BUGS	Bayesian inference using Gibbs (software)
CDF	Cumulative distribution function
COD	Chemical oxygen demand (data variable in example 6.3)
CODA	Convergence diagnostics and output analysis software for Gibbs sampling analysis (R package)
CPO	Conditional Predictive Ordinate
CR	corner (constraint)
CV	Cross-validation

CV-1	Leave-one-out cross-validation
DAG	Directed acyclic graph
DI	Dispersion index
DIBP	Diagonal inflated bivariate Poisson distribution
DIC	Deviance information criterion
GLM	Generalized linear model
GP	Generalized Poisson
GVS	Gibbs variable selection
ICPO	Inverse conditional predictive ordinate
i.i.d.	Independent identically distributed
LS	Logarithmic score
MAP	Maximum a posteriori
MP model	Median probability
MCMC	Markov chain Monte Carlo
MCE	Monte Carlo error
ML	Maximum likelihood
MLE	Maximum-likelihood estimate/estimator
NB	Negative binomial
OR	Odds ratio
PBF	Posterior Bayes factor
PD	Poisson difference
p.d.f.	Probability density function
PO	Posterior model odds
PPO	Posterior predictive ordinate
RJMCMC	Reversible jump Markov chain Monte Carlo
RR	Relative risk
SD	Standard deviation
SE	Standard error
SSVS	Stochastic search variable selection
STZ	sum-to-zero (constraint)
TS	Total solids(data variable in example 6.3)
TVS	Total volatile solids (data variable in example 6.3)
WinBUGS	Windows version of BUGS (software)
ZI	Zero inflated
ZID	Zero inflated distribution
ZIP	Zero inflated Poisson distribution
ZINB	Zero inflated negative binomial distribution

ZIGP	Zero inflated generalized Poisson distribution
ZIBP	Zero inflated bivariate Poisson distribution

CONTENTS

Preface	xvii
Acknowledgments	xix
Acronyms	xxi
1 Introduction to Bayesian Inference	1
1.1 Introduction: Bayesian modeling in the 21st century	1
1.2 Definition of statistical models	3
1.3 Bayes theorem	3
1.4 Model-based Bayesian inference	4
1.5 Inference using conjugate prior distributions	7
1.5.1 Inference for the Poisson rate of count data	7
1.5.2 Inference for the success probability of binomial data	8
1.5.3 Inference for the mean of normal data with known variance	9
1.5.4 Inference for the mean and variance of normal data	11
1.5.5 Inference for normal regression models	12
1.5.6 Other conjugate prior distributions	14
1.5.7 Illustrative examples	14
1.6 Nonconjugate analysis	24
Problems	27
	vii

2	Markov Chain Monte Carlo Algorithms in Bayesian Inference	31
2.1	Simulation, Monte Carlo integration, and their implementation in Bayesian inference	31
2.2	Markov chain Monte Carlo methods	35
2.2.1	The algorithm	36
2.2.2	Terminology and implementation details	37
2.3	Popular MCMC algorithms	42
2.3.1	The Metropolis–Hastings algorithm	42
2.3.2	Componentwise Metropolis–Hastings	45
2.3.3	The Gibbs sampler	71
2.3.4	Metropolis within Gibbs	76
2.3.5	The slice Gibbs sampler	76
2.3.6	A simple example using the slice sampler	77
2.4	Summary and closing remarks	81
	Problems	81
3	WinBUGS Software: Introduction, Setup, and Basic Analysis	83
3.1	Introduction and historical background	83
3.2	The WinBUGS environment	84
3.2.1	Downloading and installing WinBUGS	84
3.2.2	A short description of the menus	85
3.3	Preliminaries on using WinBUGS	88
3.3.1	Code structure and type of parameters/nodes	88
3.3.2	Scalar, vector, matrix, and array nodes	89
3.4	Building Bayesian models in WinBUGS	93
3.4.1	Function description	93
3.4.2	Using the <code>for</code> syntax and array, matrix, and vector calculations	97
3.4.3	Use of parentheses, brackets and curly braces in WinBUGS	98
3.4.4	Differences between WinBUGS and R/Spplus syntax	98
3.4.5	Model specification in WinBUGS	99
3.4.6	Data and initial value specification	100
3.4.7	An example of a complete model specification	107
3.4.8	Data transformations	108
3.5	Compiling the model and simulating values	108
3.6	Basic output analysis using the <code>sample monitor</code> tool	117
3.7	Summarizing the procedure	120
3.8	Chapter summary and concluding comments	121
	Problems	121
4	WinBUGS Software: Illustration, Results, and Further Analysis	125
4.1	A complete example of running MCMC in WinBUGS for a simple model	125

4.1.1	The model	125
4.1.2	Data and initial values	127
4.1.3	Compiling and running the model	127
4.1.4	MCMC output analysis and results	129
4.2	Further output analysis using the inference menu	132
4.2.1	Comparison of nodes	133
4.2.2	Calculation of correlations	136
4.2.3	Using the summary tool	137
4.2.4	Evaluation and ranking of individuals	138
4.2.5	Calculation of deviance information criterion	140
4.3	Multiple chains	141
4.3.1	Generation of multiple chains	141
4.3.2	Output analysis	142
4.3.3	The Gelman–Rubin convergence diagnostic	143
4.4	Changing the properties of a figure	145
4.4.1	General graphical options	145
4.4.2	Special graphical options	145
4.5	Other tools and menus	148
4.5.1	The node info tool	148
4.5.2	Monitoring the acceptance rate of the Metropolis–Hastings algorithm	148
4.5.3	Saving the current state of the chain	149
4.5.4	Setting the starting seed number	149
4.5.5	Running the model as a script	149
4.6	Summary and concluding remarks	149
	Problems	150

5 Introduction to Bayesian Models: Normal Models 151

5.1	General modeling principles	151
5.2	Model specification in normal regression models	152
5.2.1	Specifying the likelihood	153
5.2.2	Specifying a simple independent prior distribution	154
5.2.3	Interpretation of the regression coefficients	154
5.2.4	A regression example using WinBUGS	157
5.3	Using vectors and multivariate priors in normal regression models	161
5.3.1	Defining the model using matrices	161
5.3.2	Prior distributions for normal regression models	162
5.3.3	Multivariate normal priors in WinBUGS	163
5.3.4	Continuation of Example 5.1	164
5.4	Analysis of variance models	167
5.4.1	The one-way ANOVA model	167
5.4.2	Parametrization and parameter interpretation	168

5.4.3	One-way ANOVA model in WinBUGS	169
5.4.4	A one-way ANOVA example using WinBUGS	171
5.4.5	Two-way ANOVA models	173
5.4.6	Multifactor analysis of variance	184
	Problems	184

6 Incorporating Categorical Variables in Normal Models and Further Modeling Issues 189

6.1	Analysis of variance models using dummy variables	191
6.2	Analysis of covariance models	195
6.2.1	Models using one quantitative variable and one qualitative variable	197
6.2.2	The parallel lines model	197
6.2.3	The separate lines model	201
6.3	A bioassay example	203
6.3.1	Parallel lines analysis	204
6.3.2	Slope ratio analysis: Models with common intercept and different slope	212
6.3.3	Comparison of the two approaches	217
6.4	Further modeling issues	218
6.4.1	Extending the simple ANCOVA model	218
6.4.2	Using binary indicators to specify models in multiple regression	219
6.4.3	Selection of variables using the deviance information criterion (DIC)	219
6.5	Closing remarks	226
	Problems	226

7 Introduction to Generalized Linear Models: Binomial and Poisson Data 229

7.1	Introduction	229
7.1.1	The exponential family	230
7.1.2	Common distributions as members of the exponential family	231
7.1.3	Link functions	234
7.1.4	Common generalized linear models	236
7.1.5	Interpretation of GLM coefficients	238
7.2	Prior distributions	239
7.3	Posterior inference	241
7.3.1	The posterior distribution of a generalized linear model	241
7.3.2	GLM specification in WinBUGS	242
7.4	Poisson regression models	242
7.4.1	Interpretation of Poisson log-linear parameters	242
7.4.2	A simple Poisson regression example	245

7.4.3	A Poisson regression model for modeling football data	249
7.5	binomial response models	255
7.5.1	Interpretation of model parameters in binomial response models	257
7.5.2	A simple example	263
7.6	Models for contingency tables	269
	Problems	270
8	Models for Positive Continuous Data, Count Data, and Other GLM-Based Extensions	275
8.1	Models with nonstandard distributions	275
8.1.1	Specification of arbitrary likelihood using the zeros–ones trick	276
8.1.2	The inverse Gaussian model	277
8.2	Models for positive continuous response variables	279
8.2.1	The gamma model	279
8.2.2	Other models	280
8.2.3	An example	281
8.3	Additional models for count data	282
8.3.1	The negative binomial model	283
8.3.2	The generalized Poisson model	286
8.3.3	Zero inflated models	288
8.3.4	The bivariate Poisson model	291
8.3.5	The Poisson difference model	293
8.4	Further GLM-based models and extensions	296
8.4.1	Survival analysis models	297
8.4.2	Multinomial models	298
8.4.3	Additional models and further reading	300
	Problems	301
9	Bayesian Hierarchical Models	305
9.1	Introduction	305
9.1.1	A simple motivating example	306
9.1.2	Why use a hierarchical model?	307
9.1.3	Other advantages and characteristics	308
9.2	Some simple examples	308
9.2.1	Repeated measures data	308
9.2.2	Introducing random effects in performance parameters	313
9.2.3	Poisson mixture models for count data	315
9.2.4	The use of hierarchical models in meta-analysis	318
9.3	The generalized linear mixed model formulation	320
9.3.1	A hierarchical normal model: A simple crossover trial	321
9.3.2	Logit GLMM for correlated binary responses	325