STATISTICS:
TEXTBOOKS and MONOGRAPHS

Bayesian Model Selection and Statistical Modeling

Tomohiro Ando

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Bayesian Model Selection and Statistical Modeling

Tomohiro Ando

Keio University

Kanagawa, Japan

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

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

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4398-3614-9 (Hardback)

Library of Congress Cataloging-in-Publication Data

Ando, Tomohiro.
Bayesian model selection and statistical modeling / Tomohiro Ando.
p. cm. -- (Statistics, textbooks, and monographs)
Includes bibliographical references and index.
ISBN 978-1-4398-3614-9 (hardcover : alk. paper)
1. Bayesian statistical decision theory. 2. Mathematical statistics. 3. Mathematical models. I. Title. II. Series.

QA279.5.A55 2010
519.5'42--dc22 2010017141

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

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

Bayesian Model Selection and Statistical Modeling

STATISTICS: Textbooks and Monographs

D. B. Owen

Founding Editor, 1972–1991

STATISTICS: Textbooks and Monographs

Recent Titles

Computer-Aided Econometrics, *edited by David E.A. Giles*

The EM Algorithm and Related Statistical Models, *edited by Michiko Watanabe and Kazunori Yamaguchi*

Multivariate Statistical Analysis, Second Edition, Revised and Expanded, *Narayan C. Giri*

Computational Methods in Statistics and Econometrics, *Hisashi Tanizaki*

Applied Sequential Methodologies: Real-World Examples with Data Analysis, *edited by Nitis Mukhopadhyay, Sujay Datta, and Saibal Chattopadhyay*

Handbook of Beta Distribution and Its Applications, *edited by Arjun K. Gupta and Saralees Nadarajah*

Item Response Theory: Parameter Estimation Techniques, Second Edition, *edited by Frank B. Baker and Seock-Ho Kim*

Statistical Methods in Computer Security, *edited by William W. S. Chen*

Elementary Statistical Quality Control, Second Edition, *John T. Burr*

Data Analysis of Asymmetric Structures, *Takayuki Saito and Hiroshi Yadohisa*

Mathematical Statistics with Applications, *Asha Seth Kapadia, Wenyaw Chan, and Lemuel Moyé*

Advances on Models, Characterizations and Applications, *N. Balakrishnan, I. G. Bairamov, and O. L. Gebizlioglu*

Survey Sampling: Theory and Methods, Second Edition, *Arijit Chaudhuri and Horst Stenger*

Statistical Design of Experiments with Engineering Applications, *Kamel Rekab and Muzaffar Shaikh*

Quality by Experimental Design, Third Edition, *Thomas B. Barker*

Handbook of Parallel Computing and Statistics, *Erricos John Kontoghiorghes*

Statistical Inference Based on Divergence Measures, *Leandro Pardo*

A Kalman Filter Primer, *Randy Eubank*

Introductory Statistical Inference, *Nitis Mukhopadhyay*

Handbook of Statistical Distributions with Applications, *K. Krishnamoorthy*

A Course on Queueing Models, *Joti Lal Jain, Sri Gopal Mohanty, and Walter Böhm*

Univariate and Multivariate General Linear Models: Theory and Applications with SAS, Second Edition, *Kevin Kim and Neil Timm*

Randomization Tests, Fourth Edition, *Eugene S. Edgington and Patrick Onghena*

Design and Analysis of Experiments: Classical and Regression Approaches with SAS, *Leonard C. Onyiah*

Analytical Methods for Risk Management: A Systems Engineering Perspective, *Paul R. Garvey*

Confidence Intervals in Generalized Regression Models, *Esa Uusipaikka*

Introduction to Spatial Econometrics, *James LeSage and R. Kelley Pace*

Acceptance Sampling in Quality Control, *Edward G. Schilling and Dean V. Neubauer*

Applied Statistical Inference with MINITAB®, *Sally A. Lesik*

Nonparametric Statistical Inference, Fifth Edition, *Jean Dickinson Gibbons and Subhabrata Chakraborti*

Bayesian Model Selection and Statistical Modeling, *Tomohiro Ando*

Preface

Bayesian model selection is a fundamental part of the Bayesian statistical modeling process. In principle, the Bayesian analysis is straightforward. Specifying the data sampling and prior distributions, a joint probability distribution is used to express the relationships between all the unknowns and the data information. Bayesian inference is implemented based on the posterior distribution, the conditional probability distribution of the unknowns given the data information. The results from the Bayesian posterior inference are then used for the decision making, forecasting, stochastic structure explorations and many other problems. However, the quality of these solutions usually depends on the quality of the constructed Bayesian models. This crucial issue has been realized by researchers and practitioners. Therefore, the Bayesian model selection problems have been extensively investigated.

A default framework for the Bayesian model selection is based on the Bayes factor, which provides the scientific foundations for various fields of natural sciences, social sciences and many other areas of study. From the Bayes factor, Bayesian information criterion (BIC), generalized Bayesian information criterion (GBIC), and various types of Bayesian model selection criteria have been proposed. One of the main objectives of this book is to provide comprehensive explanations of the concepts and derivations of the default framework for the Bayesian model selection, together with a wide range of practical examples of model selection criteria.

The Bayesian inference on a statistical model was previously complex. It is now possible to implement the various types of the Bayesian inference thanks to advances in computing technology and the use of new sampling methods, including Markov chain Monte Carlo (MCMC). Such developments together with the availability of statistical software have facilitated a rapid growth in the utilization of Bayesian statistical modeling through the computer simulations. Nonetheless, model selection is central to all Bayesian statistical modeling. There is a growing need for evaluating the Bayesian models constructed by the simulation methods.

Recent Bayesian model selection studies have been mainly focusing on the evaluation of Bayesian models constructed by the simulation methods. We have seen advances of theoretical development for this area of study. A secondary objective of this book is to give plenty of simulation-based Bayesian model evaluation methods with practical advice. Various types of simulation-based Bayesian model selection criteria are explained, including the numerical calculation of the Bayes factors, the Bayesian predictive information criteria and the deviance information criteria. This book also provides a theoretical basis for the analysis of these criteria.

In addition, Bayesian model averaging is applied to many problems. By averaging over many different Bayesian statistical models, it can incorporate model uncertainty into the solution of the decision problems. In the modeling process, researchers and practitioners generally face a problem: how to specify the weight to average over all models as well as how to calculate the number of models to be combined. Heavily weighting the best fitting models, Bayesian model selection criteria have also played a major role in these issues. The third purpose of book is to cover the model averaging.

R code for several worked examples that appear in the book is available. From the link http://labs.kbs.keio.ac.jp/andotomohiro/Bayesianbook.htm, readers can download the R code to run the programs.

The author would like to acknowledge the many people who contributed to the preparation and completion of this book. In particular, the author would like to acknowledge with his sincere thanks Sadanori Konishi (Kyushu University) and Arnold Zellner (University of Chicago), from whom the author has learned so much about the concepts of Bayesian statistics and statistical modeling. The author would like to thank Ruey Tsay (University of Chicago) for an opportunity to visit Booth School of Business, University of Chicago, where he gained much experience.

The author's ideas on Bayesian statistics and statistical modeling for interdisciplinary studies have been greatly influenced by: Neeraj Bharadwaj (Temple University), Pradeep Chintagunta (University of Chicago), Alan Gelfand (Duke University), John Geweke (University of Iowa), Genshiro Kitagawa (Institute of Statistical Mathematics), Takao Kobayashi (University of Tokyo), Hedibert Lopes (University of Chicago), Teruo Nakatsuma (Keio University), Yasuhiro Omori (University of Tokyo), Nicholas Polson (University of Chicago) and many scholars.

The author is grateful to four anonymous reviewers for comments and suggestions that allowed him to improve the original draft greatly. David Grubbs patiently encouraged and supported the author throughout the final preparation of this book. The author would like to express his sincere thanks to all of these people.

TOMOHIRO ANDO

Contents

Chapter 1

Introduction

1.1 Statistical models

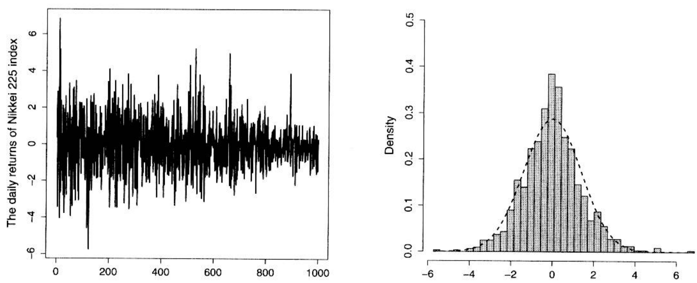
The practice of statistical modeling undergoes continual change as a result of both methodological developments and progress in the computer environment. The high-performance computers facilitated widespread advances in the development of statistical modeling theory to capture the underlying nature of a phenomenon. It is evident that the amount of information has been increasing both in size and variety thanks to recent advancement of science technology. With the advancement of computers and the information age, the challenge of understanding vast amounts of complicated data has led to the development of various types of statistical models.

Statistical model, a researchers' view to a phenomenon in our world, provides an useful tool for the description of stochastic system, the prediction, the information extraction, the casual inference, the decision making and so on. Simply speaking, we can regard a statistical model as a simplification of a complex reality. Statistical models are used not only in the social sciences: economics, finance marketing, psychology, sociology and political science, but are also employed in the natural sciences and engineering. Researchers and practitioners in various study fields have been using statistical models extensively.

Mathematically, a statistical model is defined as a set of probability distributions on the sample space (Cox and Hinkley (1974)). We usually consider a parametric family of distributions with densities $\{f(\boldsymbol{x}|\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$. In this case, a statistical model $f(\boldsymbol{x}|\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta}$ is thus a parameter set Θ together with a probability distribution function from the sample space to the set of all probability distributions on the sample space (McCullagh (2002)). In developing statistical models, we need to specify the components of a statistical model $f(\boldsymbol{x}|\boldsymbol{\theta})$, a probability distribution function and a parameter value. The next example illustrates a development of statistical models.

Example

Figure 1.1 shows the daily returns of Nikkei 225 index from August 28, 2001 to September 22, 2005 on which the market was open leading to a set of 1,000 samples. The vertical axis is the differences in the logarithm of the daily closing value of Nikkei 225 index and the horizontal axis is the time. The returns y_t are defined as the differences in the logarithm of the daily closing value of Nikkei 225 index $x_t = \{\log(y_t) - \log(y_{t-1})\} \times 100$, where y_t is the closing price on day t. The basic statistics, the sample mean, and the sample standard deviation, are given as $\hat{\mu} = -0.018$, $\hat{\sigma} = 1.930$, respectively.



(a) (b)

FIGURE 1.1: (a): Time series plot for Nikkei 225 index return data with sample period from August 28, 2001 to September 22, 2005. (b): The fitted normal density function with the mean $\hat{\mu} = -0.018$ and the standard deviation $\hat{\sigma} = 1.93$. Histograms of Nikkei 225 index return data are also shown.

In analyzing asset return data, however, the summary statistics may not have enough information. To present asset return data, one of the most common ways is to estimate the underlying true structure by fitting a parametric family of statistical models. Here, we consider fitting the normal distribution:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

After we specify the model, we have to determine unknown parameter values, i.e., μ and σ^2. Although there are various approaches to determine these parameter values, let us simply use the sample mean $\hat{\mu}$ and the sample standard deviation $\hat{\sigma}^2$. Figure 1.1 (b) shows the fitted normal density function $f(x|\hat{\mu}, \hat{\sigma}^2)$ with the mean $\hat{\mu} = -0.018$ and the standard deviation $\hat{\sigma} = 1.930$, respectively. Histograms of data are also shown in Figure 1.1 (b). We can see that the fitted normal density provides a rough approximation of the observations.

In summary, to describe a stochastic system of Nikkei 225 index return data, we employed the normal distribution $f(x|\mu, \sigma^2)$. The constructed statistical model $f(x|\hat{\mu}, \hat{\sigma}^2)$ might allow us to perform a forecasting of future Nikkei 225 index returns.

As shown in the above example, we first specified the probability distribution and then determined the parameter values within the specified model.

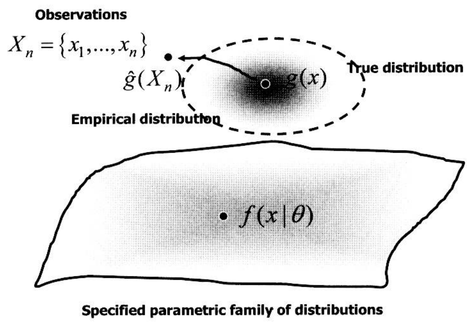


FIGURE 1.2: The relationship between the true model $g(x)$ and the empirical distribution $\hat{g}(\boldsymbol{X}_n)$ constructed by the observed data $\boldsymbol{X}_n = \{x_1, ..., x_n\}$. Once we specify the parameter values $\boldsymbol{\theta}$ of the specified model $f(x|\boldsymbol{\theta})$, the model will be fixed at a point on the surface. Therefore, the problem reduces to the parameter estimation problem.

The process determining the parameter values is called the estimation process. This estimation process adjusts the parameter values so that the specified model matches to the observed dataset. Figures 1.2–1.4 show a general image of the modeling process.

First we observe a set of data $\boldsymbol{X}_n = \{x_1, \ldots, x_n\}$. Depending on the problem, we might assume that the "true" model is contained within a set of models under our consideration. This is the M-closed framework (Bernardo and Smith (1994)). If the true model $g(x)$ belongs to the specified parametric family of distributions $f(x|\theta)$, then the true model $g(x)$ is on the surface of the specified model. In contrast to this assumption, one might follow Box (1976) in believing that "all models are wrong, but some are useful," or, none of the models under consideration is the true model. This view may be often realistic and called the M-open framework. Therefore, we usually assume that the data $\boldsymbol{X}_n$ are generated from unknown true distribution. Figure 1.2 gives an image of the relationship between the true model $g(x)$ and the empirical distribution $\hat{g}(\boldsymbol{X}_n)$ constructed by the observed data $\boldsymbol{X}_n$. Notice that we usually don't know the true model $g(x)$, but just observe the data $\boldsymbol{X}_n$. Once we specify the parameter values $\boldsymbol{\theta}$ of the specified model, the model will be fixed at a point on the model surface.

Of course, as shown in Figure 1.3, we can employ various types of probability models. In the above example, we specified the mode $f(x|\boldsymbol{\theta})$ as a normal distribution. However, we can also use Student-t distribution, Cauchy distribution and other distributions. In the model selection process, we usually

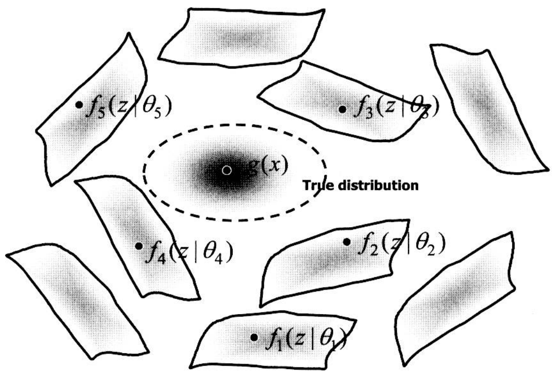


FIGURE 1.3: We can prepare various types of statistical models. For example, we can specify the probability distribution as $f_1(z|\boldsymbol{\theta}_1)$ normal distribution, $f_2(z|\boldsymbol{\theta}_2)$ Student-t distribution, $f_3(z|\boldsymbol{\theta}_3)$ Cauchy distribution, and so on.

pick one of the best models among these candidate models, or average over the set of probability models. In the context of Bayesian model selection, we usually select the model with the highest probability in the sense of the posterior model probabilities. Or, we use the Bayesian model averaging method, which averages over the set of probability models by using the posterior model probabilities.

Once we specify the probability model, the problem reduces to the parameter estimation problem. In other words, we want to specify the parameter values $\boldsymbol{\theta}$ of the specified model $f(x|\boldsymbol{\theta})$ so that the parameter value matches the true model $g(x)$. There are many approaches to estimate the unknown parameter values; the maximum likelihood method, the penalized maximum likelihood method, the generalized method of moments, robust estimation method, Bayesian estimation and so on. Figure 1.4 shows an image of the maximum likelihood method. The parameter values are determined to minimize the distance between the empirical distribution

$$\hat{g}\left(x; \boldsymbol{X}_n\right) = \frac{1}{n}\sum_{\alpha=1}^{n} I(x = x_\alpha),$$

and the specified model $f(x|\boldsymbol{\theta})$. Here $I(\cdot)$ is the indicator function that takes 1 if the relational expression in the blanket is true and 0 otherwise. The distance can be measured by using the minus of the likelihood function. Thus, the maximization of the likelihood function equals the minimized distance between the empirical distribution and the specified model $f(x|\boldsymbol{\theta})$.

Note, however, that we have to assess the closeness of the constructed statistical model $f(x|\hat{\boldsymbol{\theta}})$ not only to the true model $g(x)$, but also to the empirical