

**STATISTICS:**  
**TEXTBOOKS and MONOGRAPHS**

# **Bayesian Model Selection and Statistical Modeling**

**Tomohiro Ando**



**CRC Press**  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

02125  
A553

# 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

© 2010 by Taylor and Francis Group, LLC  
Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

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)

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.

---

**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*

## ***Editors***

N. Balakrishnan  
*McMaster University*

William R. Schucany  
*Southern Methodist University*

## ***Editorial Board***

Thomas B. Barker  
*Rochester Institute of Technology*

Nicholas Jewell  
*University of California, Berkeley*

Paul R. Garvey  
*The MITRE Corporation*

Sastry G. Pantula  
*North Carolina State  
University*

Subir Ghosh  
*University of California, Riverside*

Daryl S. Paulson  
*Biosciences Laboratories, Inc.*

David E. A. Giles  
*University of Victoria*

Aman Ullah  
*University of California,  
Riverside*

Arjun K. Gupta  
*Bowling Green State  
University*

Brian E. White  
*The MITRE Corporation*

# 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

<b>Preface</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Statistical models . . . . .	1
1.2 Bayesian statistical modeling . . . . .	6
1.3 Book organization . . . . .	8
<b>2 Introduction to Bayesian analysis</b>	<b>13</b>
2.1 Probability and Bayes' theorem . . . . .	13
2.2 Introduction to Bayesian analysis . . . . .	15
2.3 Bayesian inference on statistical models . . . . .	17
2.4 Sampling density specification . . . . .	19
2.4.1 Probability density specification . . . . .	19
2.4.2 Econometrics: Quantifying price elasticity of demand . . . . .	20
2.4.3 Financial econometrics: Describing a stock market behavior . . . . .	21
2.4.4 Bioinformatics: Tumor classification with gene expression data . . . . .	22
2.4.5 Psychometrics: Factor analysis model . . . . .	23
2.4.6 Marketing: Survival analysis model for quantifying customer lifetime value . . . . .	24
2.4.7 Medical science: Nonlinear logistic regression models . . . . .	25
2.4.8 Under the limited computer resources . . . . .	26
2.5 Prior distribution . . . . .	26
2.5.1 Diffuse priors . . . . .	26
2.5.2 The Jeffreys' prior . . . . .	27
2.5.3 Conjugate priors . . . . .	27
2.5.4 Informative priors . . . . .	27
2.5.5 Other priors . . . . .	28
2.6 Summarizing the posterior inference . . . . .	28
2.6.1 Point estimates . . . . .	28
2.6.2 Interval estimates . . . . .	29
2.6.3 Densities . . . . .	29
2.6.4 Predictive distributions . . . . .	30
2.7 Bayesian inference on linear regression models . . . . .	30
2.8 Bayesian model selection problems . . . . .	33

2.8.1	Example: Subset variable selection problem . . . . .	33
2.8.2	Example: Smoothing parameter selection problem . . . . .	35
2.8.3	Summary . . . . .	37
<b>3</b>	<b>Asymptotic approach for Bayesian inference</b>	<b>43</b>
3.1	Asymptotic properties of the posterior distribution . . . . .	43
3.1.1	Consistency . . . . .	43
3.1.2	Asymptotic normality of the posterior mode . . . . .	44
3.1.3	Example: Asymptotic normality of the posterior mode of logistic regression . . . . .	45
3.2	Bayesian central limit theorem . . . . .	46
3.2.1	Bayesian central limit theorem . . . . .	47
3.2.2	Example: Poisson distribution with conjugate prior . . . . .	49
3.2.3	Example: Confidence intervals . . . . .	50
3.3	Laplace method . . . . .	51
3.3.1	Laplace method for integral . . . . .	51
3.3.2	Posterior expectation of a function of parameter . . . . .	53
3.3.3	Example: Bernoulli distribution with a uniform prior . . . . .	55
3.3.4	Asymptotic approximation of the Bayesian predictive distribution . . . . .	57
3.3.5	Laplace method for approximating marginal posterior distribution . . . . .	58
<b>4</b>	<b>Computational approach for Bayesian inference</b>	<b>63</b>
4.1	Monte Carlo integration . . . . .	63
4.2	Markov chain Monte Carlo methods for Bayesian inference . . . . .	64
4.2.1	Gibbs sampler . . . . .	65
4.2.2	Metropolis-Hastings sampler . . . . .	65
4.2.3	Convergence check . . . . .	67
4.2.4	Example: Gibbs sampling for seemingly unrelated regression model . . . . .	68
4.2.5	Example: Gibbs sampling for auto-correlated errors . . . . .	73
4.3	Data augmentation . . . . .	76
4.3.1	Probit model . . . . .	76
4.3.2	Generating random samples from the truncated normal density . . . . .	78
4.3.3	Ordered probit model . . . . .	79
4.4	Hierarchical modeling . . . . .	81
4.4.1	Lasso . . . . .	81
4.4.2	Gibbs sampling for Bayesian Lasso . . . . .	82
4.5	MCMC studies for the Bayesian inference on various types of models . . . . .	83
4.5.1	Volatility time series models . . . . .	83
4.5.2	Simultaneous equation model . . . . .	84

4.5.3	Quantile regression . . . . .	86
4.5.4	Graphical models . . . . .	88
4.5.5	Multinomial probit models . . . . .	88
4.5.6	Markov switching models . . . . .	90
4.6	Noniterative computation methods for Bayesian inference . . . . .	93
4.6.1	The direct Monte Carlo . . . . .	93
4.6.2	Importance sampling . . . . .	94
4.6.3	Rejection sampling . . . . .	95
4.6.4	Weighted bootstrap . . . . .	96
<b>5</b>	<b>Bayesian approach for model selection</b>	<b>101</b>
5.1	General framework . . . . .	101
5.2	Definition of the Bayes factor . . . . .	103
5.2.1	Example: Hypothesis testing 1 . . . . .	104
5.2.2	Example: Hypothesis testing 2 . . . . .	105
5.2.3	Example: Poisson models with conjugate priors . . . . .	106
5.3	Exact calculation of the marginal likelihood . . . . .	108
5.3.1	Example: Binomial model with conjugate prior . . . . .	108
5.3.2	Example: Normal regression model with conjugate prior and Zellner's $g$ -prior . . . . .	109
5.3.3	Example: Multi-response normal regression model . . . . .	111
5.4	Laplace's method and asymptotic approach for computing the marginal likelihood . . . . .	113
5.5	Definition of the Bayesian information criterion . . . . .	115
5.5.1	Example: Evaluation of the approximation error . . . . .	116
5.5.2	Example: Link function selection for binomial regression . . . . .	116
5.5.3	Example: Selecting the number of factors in factor analysis model . . . . .	118
5.5.4	Example: Survival analysis . . . . .	121
5.5.5	Consistency of the Bayesian information criteria . . . . .	124
5.6	Definition of the generalized Bayesian information criterion . . . . .	125
5.6.1	Example: Nonlinear regression models using basis expansion predictors . . . . .	126
5.6.2	Example: Multinomial logistic model with basis expansion predictors . . . . .	132
5.7	Bayes factor with improper prior . . . . .	141
5.7.1	Intrinsic Bayes factors . . . . .	142
5.7.2	Partial Bayes factor and fractional Bayes factor . . . . .	146
5.7.3	Posterior Bayes factors . . . . .	147
5.7.4	Pseudo Bayes factors based on cross validation . . . . .	148
5.7.4.1	Example: Bayesian linear regression model with improper prior . . . . .	148

5.8	Expected predictive likelihood approach for Bayesian model selection . . . . .	149
5.8.1	Predictive likelihood for model selection . . . . .	150
5.8.2	Example: Normal model with conjugate prior . . . . .	152
5.8.3	Example: Bayesian spatial modeling . . . . .	152
5.9	Other related topics . . . . .	155
5.9.1	Bayes factors when model dimension grows . . . . .	155
5.9.2	Bayesian $p$ -values . . . . .	156
5.9.3	Bayesian sensitivity analysis . . . . .	157
5.9.3.1	Example: Sensitivity analysis of Value at Risk . . . . .	158
5.9.3.2	Example: Bayesian change point analysis . . . . .	160
<b>6</b>	<b>Simulation approach for computing the marginal likelihood</b>	<b>169</b>
6.1	Laplace-Metropolis approximation . . . . .	169
6.1.1	Example: Multinomial probit models . . . . .	170
6.2	Gelfand-Day's approximation and the harmonic mean estimator . . . . .	172
6.2.1	Example: Bayesian analysis of the ordered probit model . . . . .	172
6.3	Chib's estimator from Gibb's sampling . . . . .	174
6.3.1	Example: Seemingly unrelated regression model with informative prior . . . . .	176
6.3.1.1	Calculation of the marginal likelihood . . . . .	177
6.4	Chib's estimator from MH sampling . . . . .	179
6.5	Bridge sampling methods . . . . .	181
6.6	The Savage-Dickey density ratio approach . . . . .	182
6.6.1	Example: Bayesian linear regression model . . . . .	182
6.7	Kernel density approach . . . . .	185
6.7.1	Example: Bayesian analysis of the probit model . . . . .	185
6.8	Direct computation of the posterior model probabilities . . . . .	187
6.8.1	Reversible jump MCMC . . . . .	187
6.8.2	Example: Reversible jump MCMC for seemingly unrelated regression model with informative prior . . . . .	188
6.8.3	Product space search and metropolized product space search . . . . .	190
6.8.4	Bayesian variable selection for large model space . . . . .	192
<b>7</b>	<b>Various Bayesian model selection criteria</b>	<b>199</b>
7.1	Bayesian predictive information criterion . . . . .	199
7.1.1	The posterior mean of the log-likelihood and the expected log-likelihood . . . . .	199
7.1.2	Bias correction for the posterior mean of the log-likelihood . . . . .	201
7.1.3	Definition of the Bayesian predictive information criterion . . . . .	201
7.1.4	Example: Bayesian generalized state space modeling . . . . .	204

7.2	Deviance information criterion . . . . .	214
7.2.1	Example: Hierarchical Bayesian modeling for logistic regression . . . . .	215
7.3	A minimum posterior predictive loss approach . . . . .	216
7.4	Modified Bayesian information criterion . . . . .	218
7.4.1	Example: $P$ -spline regression model with Gaussian noise . . . . .	220
7.4.2	Example: $P$ -spline logistic regression . . . . .	221
7.5	Generalized information criterion . . . . .	222
7.5.1	Example: Heterogeneous error model for the analysis motorcycle impact data . . . . .	226
7.5.2	Example: Microarray data analysis . . . . .	227
<b>8</b>	<b>Theoretical development and comparisons</b>	<b>235</b>
8.1	Derivation of Bayesian information criteria . . . . .	235
8.2	Derivation of generalized Bayesian information criteria . . . . .	237
8.3	Derivation of Bayesian predictive information criterion . . . . .	238
8.3.1	Derivation of BPIC . . . . .	239
8.3.2	Further simplification of BPIC . . . . .	243
8.4	Derivation of generalized information criterion . . . . .	245
8.4.1	Information theoretic approach . . . . .	245
8.4.2	Derivation of GIC . . . . .	248
8.5	Comparison of various Bayesian model selection criteria . . . . .	250
8.5.1	Utility function . . . . .	250
8.5.2	Robustness to the improper prior . . . . .	252
8.5.3	Computational cost . . . . .	252
8.5.4	Estimation methods . . . . .	253
8.5.5	Misspecified models . . . . .	253
8.5.6	Consistency . . . . .	253
<b>9</b>	<b>Bayesian model averaging</b>	<b>257</b>
9.1	Definition of Bayesian model averaging . . . . .	257
9.2	Occam's window method . . . . .	259
9.3	Bayesian model averaging for linear regression models . . . . .	260
9.4	Other model averaging methods . . . . .	261
9.4.1	Model averaging with AIC . . . . .	262
9.4.2	Model averaging with predictive likelihood . . . . .	262
	<b>Bibliography</b>	<b>265</b>
	<b>Index</b>	<b>285</b>

# 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.

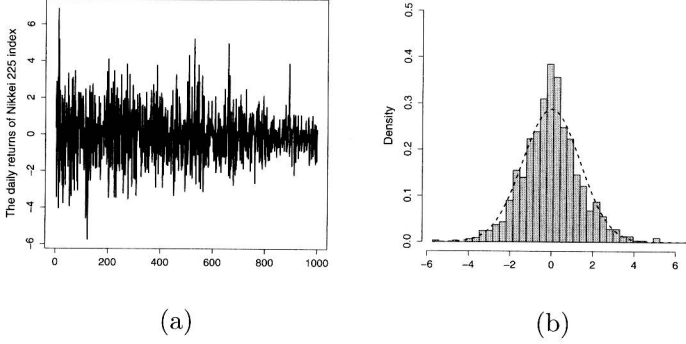
Statistical model, a researchers' view to a phenomenon in our world, provides an useful tool for the description of stochastic system, the prediction, the information extraction, the casual inference, the decision making and so on. Simply speaking, we can regard a statistical model as a simplification 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  $\{f(\mathbf{x}|\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$ . In this case, a statistical model  $f(\mathbf{x}|\boldsymbol{\theta})$  parameterized by  $\boldsymbol{\theta}$  is thus a parameter set  $\Theta$  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  $f(\mathbf{x}|\boldsymbol{\theta})$ , 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.



**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.

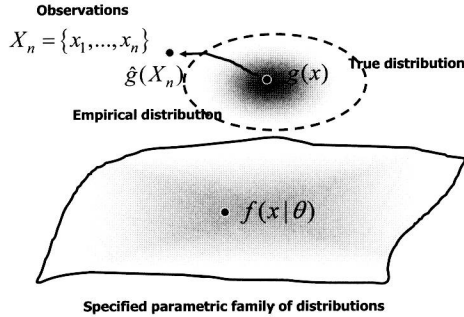
In analyzing asset return data, however, the summary statistics may not have enough information. To present asset return data, one of the most common ways is to estimate the underlying true structure by fitting a parametric family of statistical models. Here, we consider fitting the normal distribution:

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

After we specify the model, we have to determine unknown parameter values, i.e.,  $\mu$  and  $\sigma^2$ . 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.



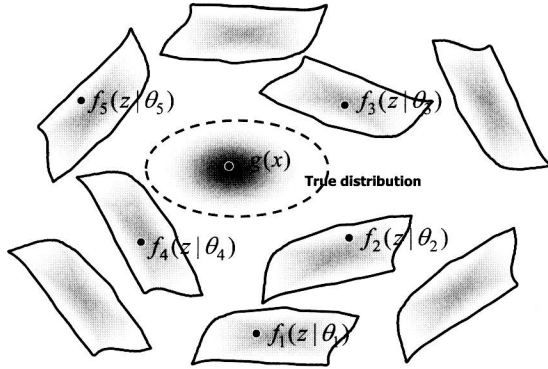
**FIGURE 1.2:** The relationship between the true model  $g(x)$  and the empirical distribution  $\hat{g}(\mathbf{X}_n)$  constructed by the observed data  $\mathbf{X}_n = \{x_1, \dots, x_n\}$ . Once we specify the parameter values  $\theta$  of the specified model  $f(x|\theta)$ , the model will be 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  $\mathbf{X}_n = \{x_1, \dots, x_n\}$ . 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  $\mathbf{X}_n$  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  $\hat{g}(\mathbf{X}_n)$  constructed by the observed data  $\mathbf{X}_n$ . Notice that we usually don’t know the true model  $g(x)$ , but just observe the data  $\mathbf{X}_n$ . Once we specify the parameter values  $\theta$  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  $f(x|\theta)$  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  $f_1(z|\theta_1)$  normal distribution,  $f_2(z|\theta_2)$  Student- $t$  distribution,  $f_3(z|\theta_3)$  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  $\theta$  of the specified model  $f(x|\theta)$  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

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

and the specified model  $f(x|\theta)$ . 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  $f(x|\theta)$ .

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