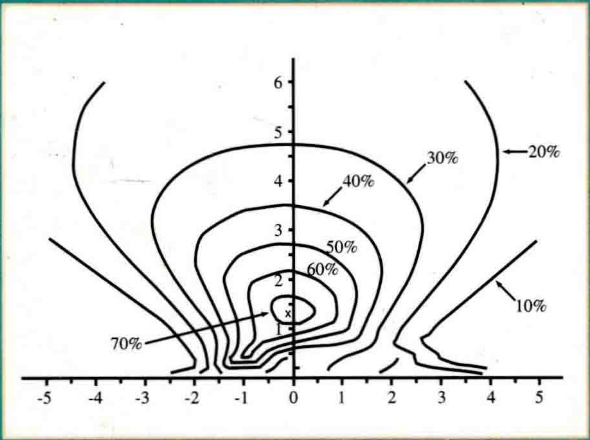
STATISTICS:
a DEKKER series of TEXTBOOKS and MONOGRAPHS

Computational Methods in Statistics and Econometrics



Hisashi Tanizaki

Dekker Media

Media Included

Computational Methods in Statistics and Econometrics

Hisashi Tanizaki
Kobe University
Kobe, Japan

MARCEL DEKKER, INC. NEW YORK • BASEL

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

ISBN: 0-8247-4804-2

This book is printed on acid-free paper.

Headquarters
Marcel Dekker, Inc., 270 Madison Avenue, New York, NY 10016, U.S.A.
tel: 212-696-9000; fax: 212-685-4540

Distribution and Customer Service
Marcel Dekker, Inc., Cimarron Road, Monticello, New York 12701, U.S.A.
tel: 800-228-1160; fax: 845-796-1772

World Wide Web
http://www.dekker.com

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/Professional Marketing at the headquarters address above.

Current printing (last digit):

10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

Computational Methods in Statistics and Econometrics

1. The Generalized Jackknife Statistic, *H. L. Gray and W. R. Schucany*
2. Multivariate Analysis, *Anant M. Kshirsagar*
3. Statistics and Society, *Walter T. Federer*
4. Multivariate Analysis: A Selected and Abstracted Bibliography, 1957–1972, *Kocherlakota Subrahmaniam and Kathleen Subrahmaniam*
5. Design of Experiments: A Realistic Approach, *Virgil L. Anderson and Robert A. McLean*
6. Statistical and Mathematical Aspects of Pollution Problems, *John W. Pratt*
7. Introduction to Probability and Statistics (in two parts), Part I: Probability; Part II: Statistics, *Narayan C. Giri*
8. Statistical Theory of the Analysis of Experimental Designs, *J. Ogawa*
9. Statistical Techniques in Simulation (in two parts), *Jack P. C. Kleijnen*
10. Data Quality Control and Editing, *Joseph I. Naus*
11. Cost of Living Index Numbers: Practice, Precision, and Theory, *Kali S. Banerjee*
12. Weighing Designs: For Chemistry, Medicine, Economics, Operations Research, Statistics, *Kali S. Banerjee*
13. The Search for Oil: Some Statistical Methods and Techniques, *edited by D. B. Owen*
14. Sample Size Choice: Charts for Experiments with Linear Models, *Robert E. Odeh and Martin Fox*
15. Statistical Methods for Engineers and Scientists, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
16. Statistical Quality Control Methods, *Irving W. Burr*
17. On the History of Statistics and Probability, *edited by D. B. Owen*
18. Econometrics, *Peter Schmidt*
19. Sufficient Statistics: Selected Contributions, *Vasant S. Huzurbazar (edited by Anant M. Kshirsagar)*
20. Handbook of Statistical Distributions, *Jagdish K. Patel, C. H. Kapadia, and D. B. Owen*
21. Case Studies in Sample Design, *A. C. Rosander*
22. Pocket Book of Statistical Tables, *compiled by R. E. Odeh, D. B. Owen, Z. W. Birnbaum, and L. Fisher*
23. The Information in Contingency Tables, *D. V. Gokhale and Solomon Kullback*
24. Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods, *Lee J. Bain*
25. Elementary Statistical Quality Control, *Irving W. Burr*
26. An Introduction to Probability and Statistics Using BASIC, *Richard A. Groeneveld*
27. Basic Applied Statistics, *B. L. Raktoe and J. J. Hubert*
28. A Primer in Probability, *Kathleen Subrahmaniam*
29. Random Processes: A First Look, *R. Syski*
30. Regression Methods: A Tool for Data Analysis, *Rudolf J. Freund and Paul D. Minton*
31. Randomization Tests, *Eugene S. Edgington*
32. Tables for Normal Tolerance Limits, Sampling Plans and Screening, *Robert E. Odeh and D. B. Owen*
33. Statistical Computing, *William J. Kennedy, Jr., and James E. Gentle*
34. Regression Analysis and Its Application: A Data-Oriented Approach, *Richard F. Gunst and Robert L. Mason*
35. Scientific Strategies to Save Your Life, *I. D. J. Bross*
36. Statistics in the Pharmaceutical Industry, *edited by C. Ralph Buncher and Jia-Yeong Tsay*
37. Sampling from a Finite Population, *J. Hajek*
38. Statistical Modeling Techniques, *S. S. Shapiro and A. J. Gross*
39. Statistical Theory and Inference in Research, *T. A. Bancroft and C.-P. Han*
40. Handbook of the Normal Distribution, *Jagdish K. Patel and Campbell B. Read*
41. Recent Advances in Regression Methods, *Hrishikesh D. Vinod and Aman Ullah*
42. Acceptance Sampling in Quality Control, *Edward G. Schilling*
43. The Randomized Clinical Trial and Therapeutic Decisions, *edited by Niels Tygstrup, John M Lachin, and Erik Juhl*

44. Regression Analysis of Survival Data in Cancer Chemotherapy, *Walter H. Carter, Jr., Galen L. Wampler, and Donald M. Stablein*
45. A Course in Linear Models, *Anant M. Kshirsagar*
46. Clinical Trials: Issues and Approaches, *edited by Stanley H. Shapiro and Thomas H. Louis*
47. Statistical Analysis of DNA Sequence Data, *edited by B. S. Weir*
48. Nonlinear Regression Modeling: A Unified Practical Approach, *David A. Ratkowsky*
49. Attribute Sampling Plans, Tables of Tests and Confidence Limits for Proportions, *Robert E. Odeh and D. B. Owen*
50. Experimental Design, Statistical Models, and Genetic Statistics, *edited by Klaus Hinkelmann*
51. Statistical Methods for Cancer Studies, *edited by Richard G. Cornell*
52. Practical Statistical Sampling for Auditors, *Arthur J. Wilburn*
53. Statistical Methods for Cancer Studies, *edited by Edward J. Wegman and James G. Smith*
54. Self-Organizing Methods in Modeling: GMDH Type Algorithms, *edited by Stanley J. Farlow*
55. Applied Factorial and Fractional Designs, *Robert A. McLean and Virgil L. Anderson*
56. Design of Experiments: Ranking and Selection, *edited by Thomas J. Santner and Ajit C. Tamhane*
57. Statistical Methods for Engineers and Scientists: Second Edition, Revised and Expanded, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
58. Ensemble Modeling: Inference from Small-Scale Properties to Large-Scale Systems, *Alan E. Gelfand and Crayton C. Walker*
59. Computer Modeling for Business and Industry, *Bruce L. Bowerman and Richard T. O'Connell*
60. Bayesian Analysis of Linear Models, *Lyle D. Broemeling*
61. Methodological Issues for Health Care Surveys, *Brenda Cox and Steven Cohen*
62. Applied Regression Analysis and Experimental Design, *Richard J. Brook and Gregory C. Arnold*
63. Statpal: A Statistical Package for Microcomputers—PC-DOS Version for the IBM PC and Compatibles, *Bruce J. Chalmer and David G. Whitmore*
64. Statpal: A Statistical Package for Microcomputers—Apple Version for the II, II+, and IIe, *David G. Whitmore and Bruce J. Chalmer*
65. Nonparametric Statistical Inference: Second Edition, Revised and Expanded, *Jean Dickinson Gibbons*
66. Design and Analysis of Experiments, *Roger G. Petersen*
67. Statistical Methods for Pharmaceutical Research Planning, *Sten W. Bergman and John C. Gittins*
68. Goodness-of-Fit Techniques, *edited by Ralph B. D'Agostino and Michael A. Stephens*
69. Statistical Methods in Discrimination Litigation, *edited by D. H. Kaye and Mikel Aickin*
70. Truncated and Censored Samples from Normal Populations, *Helmut Schneider*
71. Robust Inference, *M. L. Tiku, W. Y. Tan, and N. Balakrishnan*
72. Statistical Image Processing and Graphics, *edited by Edward J. Wegman and Douglas J. DePriest*
73. Assignment Methods in Combinatorial Data Analysis, *Lawrence J. Hubert*
74. Econometrics and Structural Change, *Lyle D. Broemeling and Hiroki Tsurumi*
75. Multivariate Interpretation of Clinical Laboratory Data, *Adelin Albert and Eugene K. Harris*
76. Statistical Tools for Simulation Practitioners, *Jack P. C. Kleijnen*
77. Randomization Tests: Second Edition, *Eugene S. Edgington*
78. A Folio of Distributions: A Collection of Theoretical Quantile-Quantile Plots, *Edward B. Fowlkes*
79. Applied Categorical Data Analysis, *Daniel H. Freeman, Jr.*
80. Seemingly Unrelated Regression Equations Models: Estimation and Inference, *Virendra K. Srivastava and David E. A. Giles*

81. Response Surfaces: Designs and Analyses, *Andre I. Khuri and John A. Cornell*
82. Nonlinear Parameter Estimation: An Integrated System in BASIC, *John C. Nash and Mary Walker-Smith*
83. Cancer Modeling, *edited by James R. Thompson and Barry W. Brown*
84. Mixture Models: Inference and Applications to Clustering, *Geoffrey J. McLachlan and Kaye E. Basford*
85. Randomized Response: Theory and Techniques, *Arijit Chaudhuri and Rahul Mukerjee*
86. Biopharmaceutical Statistics for Drug Development, *edited by Karl E. Peace*
87. Parts per Million Values for Estimating Quality Levels, *Robert E. Odeh and D. B. Owen*
88. Lognormal Distributions: Theory and Applications, *edited by Edwin L. Crow and Kunio Shimizu*
89. Properties of Estimators for the Gamma Distribution, *K. O. Bowman and L. R. Shenton*
90. Spline Smoothing and Nonparametric Regression, *Randall L. Eubank*
91. Linear Least Squares Computations, *R. W. Farebrother*
92. Exploring Statistics, *Damaraju Raghavarao*
93. Applied Time Series Analysis for Business and Economic Forecasting, *Sufi M. Nazem*
94. Bayesian Analysis of Time Series and Dynamic Models, *edited by James C. Spall*
95. The Inverse Gaussian Distribution: Theory, Methodology, and Applications, *Raj S. Chhikara and J. Leroy Folks*
96. Parameter Estimation in Reliability and Life Span Models, *A. Clifford Cohen and Betty Jones Whitten*
97. Pooled Cross-Sectional and Time Series Data Analysis, *Terry E. Dielman*
98. Random Processes: A First Look, Second Edition, Revised and Expanded, *R. Syski*
99. Generalized Poisson Distributions: Properties and Applications, *P. C. Consul*
100. Nonlinear L_p-Norm Estimation, *Rene Gonin and Arthur H. Money*
101. Model Discrimination for Nonlinear Regression Models, *Dale S. Borowiak*
102. Applied Regression Analysis in Econometrics, *Howard E. Doran*
103. Continued Fractions in Statistical Applications, *K. O. Bowman and L. R. Shenton*
104. Statistical Methodology in the Pharmaceutical Sciences, *Donald A. Berry*
105. Experimental Design in Biotechnology, *Perry D. Haaland*
106. Statistical Issues in Drug Research and Development, *edited by Karl E. Peace*
107. Handbook of Nonlinear Regression Models, *David A. Ratkowsky*
108. Robust Regression: Analysis and Applications, *edited by Kenneth D. Lawrence and Jeffrey L. Arthur*
109. Statistical Design and Analysis of Industrial Experiments, *edited by Subir Ghosh*
110. *U*-Statistics: Theory and Practice, *A. J. Lee*
111. A Primer in Probability: Second Edition, Revised and Expanded, *Kathleen Subrahmaniam*
112. Data Quality Control: Theory and Pragmatics, *edited by Gunar E. Liepins and V. R. R. Uppuluri*
113. Engineering Quality by Design: Interpreting the Taguchi Approach, *Thomas B. Barker*
114. Survivorship Analysis for Clinical Studies, *Eugene K. Harris and Adelin Albert*
115. Statistical Analysis of Reliability and Life-Testing Models: Second Edition, *Lee J. Bain and Max Engelhardt*
116. Stochastic Models of Carcinogenesis, *Wai-Yuan Tan*
117. Statistics and Society: Data Collection and Interpretation, Second Edition, Revised and Expanded, *Walter T. Federer*
118. Handbook of Sequential Analysis, *B. K. Ghosh and P. K. Sen*
119. Truncated and Censored Samples: Theory and Applications, *A. Clifford Cohen*
120. Survey Sampling Principles, *E. K. Foreman*
121. Applied Engineering Statistics, *Robert M. Bethea and R. Russell Rhinehart*
122. Sample Size Choice: Charts for Experiments with Linear Models: Second Edition, *Robert E. Odeh and Martin Fox*
123. Handbook of the Logistic Distribution, *edited by N. Balakrishnan*
124. Fundamentals of Biostatistical Inference, *Chap T. Le*
125. Correspondence Analysis Handbook, *J.-P. Benzécri*

126. Quadratic Forms in Random Variables: Theory and Applications, *A. M. Mathai and Serge B. Provost*
127. Confidence Intervals on Variance Components, *Richard K. Burdick and Franklin A. Graybill*
128. Biopharmaceutical Sequential Statistical Applications, *edited by Karl E. Peace*
129. Item Response Theory: Parameter Estimation Techniques, *Frank B. Baker*
130. Survey Sampling: Theory and Methods, *Arijit Chaudhuri and Horst Stenger*
131. Nonparametric Statistical Inference: Third Edition, Revised and Expanded, *Jean Dickinson Gibbons and Subhabrata Chakraborti*
132. Bivariate Discrete Distribution, *Subrahmaniam Kocherlakota and Kathleen Kocherlakota*
133. Design and Analysis of Bioavailability and Bioequivalence Studies, *Shein-Chung Chow and Jen-pei Liu*
134. Multiple Comparisons, Selection, and Applications in Biometry, *edited by Fred M. Hoppe*
135. Cross-Over Experiments: Design, Analysis, and Application, *David A. Ratkowsky, Marc A. Evans, and J. Richard Alldredge*
136. Introduction to Probability and Statistics: Second Edition, Revised and Expanded, *Narayan C. Giri*
137. Applied Analysis of Variance in Behavioral Science, *edited by Lynne K. Edwards*
138. Drug Safety Assessment in Clinical Trials, *edited by Gene S. Gilbert*
139. Design of Experiments: A No-Name Approach, *Thomas J. Lorenzen and Virgil L. Anderson*
140. Statistics in the Pharmaceutical Industry: Second Edition, Revised and Expanded, *edited by C. Ralph Buncher and Jia-Yeong Tsay*
141. Advanced Linear Models: Theory and Applications, *Song-Gui Wang and Shein-Chung Chow*
142. Multistage Selection and Ranking Procedures: Second-Order Asymptotics, *Nitis Mukhopadhyay and Tumulesh K. S. Solanky*
143. Statistical Design and Analysis in Pharmaceutical Science: Validation, Process Controls, and Stability, *Shein-Chung Chow and Jen-pei Liu*
144. Statistical Methods for Engineers and Scientists: Third Edition, Revised and Expanded, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
145. Growth Curves, *Anant M. Kshirsagar and William Boyce Smith*
146. Statistical Bases of Reference Values in Laboratory Medicine, *Eugene K. Harris and James C. Boyd*
147. Randomization Tests: Third Edition, Revised and Expanded, *Eugene S. Edgington*
148. Practical Sampling Techniques: Second Edition, Revised and Expanded, *Ranjan K. Som*
149. Multivariate Statistical Analysis, *Narayan C. Giri*
150. Handbook of the Normal Distribution: Second Edition, Revised and Expanded, *Jagdish K. Patel and Campbell B. Read*
151. Bayesian Biostatistics, *edited by Donald A. Berry and Dalene K. Stangl*
152. Response Surfaces: Designs and Analyses, Second Edition, Revised and Expanded, *André I. Khuri and John A. Cornell*
153. Statistics of Quality, *edited by Subir Ghosh, William R. Schucany, and William B. Smith*
154. Linear and Nonlinear Models for the Analysis of Repeated Measurements, *Edward F. Vonesh and Vernon M. Chinchilli*
155. Handbook of Applied Economic Statistics, *Aman Ullah and David E. A. Giles*
156. Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators, *Marvin H. J. Gruber*
157. Nonparametric Regression and Spline Smoothing: Second Edition, *Randall L. Eubank*
158. Asymptotics, Nonparametrics, and Time Series, *edited by Subir Ghosh*
159. Multivariate Analysis, Design of Experiments, and Survey Sampling, *edited by Subir Ghosh*

160. Statistical Process Monitoring and Control, *edited by Sung H. Park and G. Geoffrey Vining*
161. Statistics for the 21st Century: Methodologies for Applications of the Future, *edited by C. R. Rao and Gábor J. Székely*
162. Probability and Statistical Inference, *Nitis Mukhopadhyay*
163. Handbook of Stochastic Analysis and Applications, *edited by D. Kannan and V. Lakshmikantham*
164. Testing for Normality, *Henry C. Thode, Jr.*
165. Handbook of Applied Econometrics and Statistical Inference, *edited by Aman Ullah, Alan T. K. Wan, and Anoop Chaturvedi*
166. Visualizing Statistical Models and Concepts, *R. W. Farebrother*
167. Financial and Actuarial Statistics: An Introduction, *Dale S. Borowiak*
168. Nonparametric Statistical Inference: Fourth Edition, Revised and Expanded, *Jean Dickinson Gibbons and Subhabrata Chakraborti*
169. Computer-Aided Econometrics, *edited by David E. A. Giles*
170. The EM Algorithm and Related Statistical Models, *edited by Michiko Watanabe and Kazunori Yamaguchi*
171. Multivariate Statistical Analysis: Second Edition, Revised and Expanded, *Narayan C. Giri*
172. Computational Methods in Statistics and Econometrics, *Hisashi Tanizaki*
173. Applied Sequential Methodologies, *edited by Nitis Mukhopadhyay, Sujay Datta, and Saibal Chattopadhyay*

Additional Volumes in Preparation

To My Family,
Miyuki, Toshifumi and Yasuyuki

Preface

As the personal computer progresses day by day, computer-intensive procedures have been developed in the field of statistics and econometrics. The computational procedures in statistics and econometrics include both Monte Carlo methods and nonparametric methods (or distribution-free methods). In the first half of this book, the Monte Carlo methods are discussed. That is, some representative random number generation methods and their applications are shown. The second half of this book is related to computer-intensive statistical techniques other than Monte Carlo methods and simulations, where the nonparametric methods are introduced.

Chapter 1 is an introduction to statistics and econometrics, which corresponds to my lecture notes in mathematical statistics course (about 15 lectures, each 90 minutes long) for first-year graduate students. Based on Chapter 1, the Monte Carlo and nonparametric methods are discussed in Chapters 2 – 8.

In the Monte Carlo methods, we discuss how to generate various random draws. Almost all the random draws are based on the uniform random draws. Therefore, it is one of the most important tasks to investigate uniform random number generation. Transforming the variable from the uniform random variable, various random draws are generated in Chapter 2, e.g., Bernoulli random draws, binomial random draws, normal random draws, χ^2 random draws, t random draws, F random draws, exponential random draws, gamma random draws, beta random draws, Cauchy random draws, logistic random draws and others.

Importance resampling, rejection sampling and the Metropolis-Hastings algorithm are the methods to generate random draws from any distribution, which are useful tools for random number generation even when it is not easy to generate the random draws. Three sampling methods are discussed in Chapter 3. Thus, in the Monte Carlo methods, random number generation is very important. Once we have the random draws, simply the arithmetic average of the random draws indicates the estimate of mean. The arithmetic average is approximately distributed with a normal random variable by the central limit theorem. Therefore, the statistical inference also becomes quite easy, using the random draws.

As some applications in the Monte Carlo methods, Baysian inference (Chapter 4), bias correction of the ordinary least squares (OLS) estimator in the autoregressive models (Chapter 5) and nonlinear non-Gaussian state space modeling (Chapter 6) are shown.

In the nonparametric methods, nonparametric statistical tests are discussed in Chapters 7 and 8. Nonparametric tests of difference between two sample means include score tests (e.g., Wilcoxon rank sum test, normal score test, logistic score test, Cauchy score test and so on) and Fisher's randomization test (or Fisher's permutation test). The nonparametric tests of difference between two sample means are discussed in Chapter 7. One of the features of nonparametric tests is that we do not have to impose any assumption on the underlying distribution. From no restriction on the distribution, it could be expected that nonparametric tests would be less powerful than the conventional parametric tests such as the t test. However, it is shown that the Wilcoxon rank sum test is as powerful as the t test under the location-shift alternatives and moreover that the Wilcoxon test is sometimes much more powerful than the t test. Especially, the remarkable fact about the Wilcoxon test is that it is about 95 per cent as powerful as the t test for normal data. It is known that Pitman's asymptotic relative efficiency of the normal score test relative to the t test is greater than one under the location-shift alternatives. This implies that the power of the normal score test is always larger than that of the t test. It is known that the normal score test is less powerful than the Wilcoxon test if the tails of the underlying distributions are diffuse. Since in general the nonparametric tests require a large computational burden, however, there are few studies on small sample properties although asymptotic properties from various aspects were studied in the past.

In addition to testing difference between two sample means, in Chapter 8 we also consider testing independence between two samples, which corresponds to testing correlation coefficient and regression coefficient. Small sample properties are discussed in the nonparametric statistical tests part of the book.

Thus, some selected representative computer-intensive methods are treated in this book, where the source codes are shown by Fortran 77 and sometimes C languages for the purpose of practical understanding. For this book, I used seven personal computers, i.e.,

- Athlon 1.4 GHz CPU, Windows 2000 Operating System
- Pentium III 1.4 GHz Dual CPU, and Windows 2000 Operating System
- Pentium III 1.0 GHz Dual CPU, and Windows 2000 Operating System
- Athlon 2000+ Dual CPU, and Windows 2000 Operating System
- Athlon 2000+ Dual CPU, and Linux (Slackware 8.0) Operating System (see `http://www.slackware.com` for Slackware)
- Pentium III 1.4 GHz Dual CPU, and Linux (Plamo Linux 2.2.5, which is equivalent to Slackware+Japanese) Operating System (see `http://www.linet.jp/Plamo` or `http://plamo-linux.jp` for Plamo Linux, which is a Japanese site)
- Pentium III 1.0 GHz Dual CPU, and Linux (Plamo Linux 2.2.5) Operating System

For almost two years, my personal computers have been running all the time to prepare this book. Under the Windows 2000 Operating System, the following Fortran and C compilers are used for computation.

- Open WATCOM C/C++ and Fortran 1.0 (`http://www.openwatcom.org`)
- Cygwin (`http://www.cygwin.com`)
- DJGPP (`http://www.delorie.com/djgpp`)

In addition to the free compilers shown above, in Section 7.5, Tables 7.4 and 7.5, the IMSL library (`http://www.vni.com/products/imsl`) with Microsoft Fortran PowerStation Version 4.00 is used to obtain the percent points of the normal, t and F distributions.

I am indebted to many people for assistance with this book. All cannot be mentioned in this short space, but I would like to acknowledge the Acquisitions Editor Taisuke Soda (Marcel Dekker, Inc.), who suggested that I write a book on statistical computing. I presented some chapters at the Econometrics workshop, Graduate School of Economics, Kobe University. I would like to thank the participants at the workshop for valuable comments. Furthermore, as mentioned above, Chapter 1 is based on my lecture notes in mathematical statistics course. Graduate students found some errors in the lecture notes. I am grateful to them for that. This research was partially supported by Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C)(2) #14530033 and Grants-in-Aid for the 21st Century COE Program #15COE500.

I used LaTeX to write this book. I did all of the typing, programming and so on. Therefore, all mistakes are mine. In programming, I use Fortran 77 in the first half of the book (Chapters 2 – 6) and C in the last half (Chapters 7 and 8). The *pointer*, which is often used in C language, is not familiar to Fortran users. Therefore, in Chapters 7 and 8, I did not use the pointer. Instead, I declared the external variables in the source codes, which are similar to the `common` sentence in Fortran 77. Finally, all the source codes used in this book are available in the CD-ROM attached with this book (therefore, readers do not have to type the source codes from the beginning).

Hisashi Tanizaki

Contents