Preface

Over the last two decades, advancements in computer technology have enabled accelerated research and development of Monte-Carlo computational methods. This book is a compilation of invited papers from some of the most forward-thinking statistical researchers. These authors present new developments in Monte-Carlo simulation-based statistical modeling, thereby creating an opportunity for the exchange ideas among researchers and users of statistical computing.

Our aim in creating this book is to provide a venue for timely dissemination of the research in Monte-Carlo simulation-based statistical modeling to promote further research and collaborative work in this area. In the era of big data science, this collection of innovative research not only has remarkable potential to have a substantial impact on the development of advanced Monte-Carlo methods across the spectrum of statistical data analyses but also has great promise for fostering new research and collaborations addressing the ever-changing challenges and opportunities of statistics and data science. The authors have made their data and computer programs publicly available, making it possible for readers to replicate the model development and data analysis presented in each chapter and readily apply these new methods in their own research.

The 18 chapters are organized into three parts. Part I includes six chapters that present and discuss general Monte-Carlo techniques. Part II comprises six chapters with a common focus on Monte-Carlo methods used in missing data analyses, which is an area of growing importance in public health and social sciences. Part III is composed of six chapters that address Monte-Carlo statistical modeling and their applications.

Part I: Monte-Carlo Techniques (Chapters "Joint Generation of Binary, Ordinal, Count, and Normal Data with Specified Marginal and Association Structures in Monte-Carlo Simulations"–"Quantifying the Uncertainty in Optimal Experiment Schemes via Monte-Carlo Simulations")

Chapter "Joint Generation of Binary, Ordinal, Count, and Normal Data with Specified Marginal and Association Structures in Monte-Carlo Simulations" presents a unified framework for concurrently generating data that include the four major types of distributions (i.e., binary, ordinal, count, and normal) with specified marginal and association structures. In this discussion of an important supplement to existing methods, Hakan Demirtas unifies the Monte-Carlo methods for specified types of data and presents his systematic and comprehensive investigation for mixed data generation. The proposed framework can then be readily used to simulate multivariate data of mixed types for the development of more sophisticated simulation, computation, and data analysis techniques.

In Chapter "Improving the Efficiency of the Monte-Carlo Methods Using Ranked Simulated Approach", Hani Samawi provides an overview of his development of ranked simulated sampling; a key approach for improving the efficiency of general Monte-Carlo methods. Samawi then demonstrates the capacity of this approach to provide unbiased estimation.

In Chapter "Normal and Non-normal Data Simulations for the Evaluation of Two-Sample Location Tests", Jessica Hoag and Chia-Ling Kuo discuss Monte-Carlo simulation of normal and non-normal data to evaluate two-sample location tests (i.e., statistical tests that compare means or medians of two independent populations).

Chapter "Anatomy of Correlational Magnitude Transformations in Latency and Discretization Contexts in Monte-Carlo Studies" proposes a general assessment of correlational magnitude changes in the latency and discretization contexts of Monte-Carlo studies. Further, authors Hakan Demirtas and Ceren Vardar-Acar provide a conceptual framework and computational algorithms for modeling the correlation transitions under specified distributional assumptions within the realm of discretization in the context of latency and the threshold concept. The authors illustrate the proposed algorithms with several examples and include a simulation study that demonstrates the feasibility and performance of the methods.

Chapter "Monte-Carlo Simulation of Correlated Binary Responses" discusses the Monte-Carlo simulation of correlated binary responses. Simulation studies are a well-known, highly valuable tool that allows researchers to obtain powerful conclusions for correlated or longitudinal response data. In cases where logistic modeling it used, the researcher must have appropriate methods for simulating correlated binary data along with associated predictors. In this chapter, author Trent Lalonde presents an overview of existing methods for simulating correlated binary response data and compares those methods with methods using R software.

Chapter "Quantifying the Uncertainty in Optimal Experiment Schemes via Monte-Carlo Simulations" provides a general framework for quantifying the sensitivity and uncertainty that result from the misspecification of model parameters in optimal experimental schemes. In designing life-testing experiments, it is widely accepted that the optimal experimental scheme depends on unknown model parameters, and that misspecified parameters can lead to substantial loss of efficiency in the statistical analysis. To quantify this effect, Tony Ng, Yu-Jau Lin, Tzong-Ru Tsai, Y.L. Lio, and Nan Jiang use Monte-Carlo simulations to evaluate the robustness of optimal experimental schemes.

Part II: Monte-Carlo Methods for Missing Data (Chapters "Markov Chain Monte-Carlo Methods for Missing Data Under Ignorability Assumptions"–"Application of Markov Chain Monte-Carlo Multiple Imputation Method to Deal with Missing Data from the Mechanism of MNAR in Sensitivity Analysis for a Longitudinal Clinical Trial")

Chapter "Markov Chain Monte-Carlo Methods for Missing Data Under Ignorability Assumptions" presents a fully Bayesian method for using the Markov chain Monte-Carlo technique for missing data to sample the full conditional distribution of the missing data given observed data and the other parameters. In this chapter, Haresh Rochani and Daniel Linder show how to apply these methods to real datasets with missing responses as well as missing covariates. Additionally, the authors provide simulation settings to illustrate this method.

In Chapter "A Multiple Imputation Framework for Massive Multivariate Data of Different Variable Types: A Monte-Carlo Technique", Hakan Demirtas discusses multiple imputation for massive multivariate data of variable types from planned missingness designs with the purpose to build theoretical, algorithmic, and implementation-based components of a unified, general-purpose multiple imputation framework. The planned missingness designs are highly useful and will likely increase in popularity in the future. For this reason, the proposed multiple imputation framework represents an important refinement of existing methods.

Chapter "Hybrid Monte-Carlo in Multiple Missing Data Imputations with Application to a Bone Fracture Data" introduces the Hybrid Monte-Carlo method as an efficient approach for sampling complex posterior distributions of several correlated parameters from a semi-parametric missing data model. In this chapter, Hui Xie describes a modeling approach for missing values that does not require assuming specific distributional forms. To demonstrate the method, the author provides an R program for analyzing missing data from a bone fracture study.

Chapter "Statistical Methodologies for Dealing with Incomplete Longitudinal Outcomes Due to Dropout Missing at Random" considers key methods for handling longitudinal data that are incomplete due to missing at random dropout. In this chapter, Ali Satty, Henry Mwambi, and Geert Muhlenbergs provide readers with an overview of the issues and the different methodologies for handling missing data in longitudinal datasets that result from dropout (e.g., study attrition, loss of follow-up). The authors examine the potential strengths and weaknesses of the various methods through two examples of applying these methods.

In Chapter “Applications of Simulation for Missing Data Issues in Longitudinal Clinical Trials”, Frank Liu and James Kost present simulation-based approaches for addressing missing data issues in longitudinal clinical trials, such as control-based imputation, tipping-point analysis, and a Bayesian Markov chain Monte-Carlo method. Computation programs for these methods are implemented and available in SAS.

In Chapter “Application of Markov Chain Monte-Carlo Multiple Imputation Method to Deal with Missing Data from the Mechanism of MNAR in Sensitivity Analysis for a Longitudinal Clinical Trial”, Wei Sun discusses the application of Markov chain Monte-Carlo multiple imputation for data that is missing not at random in longitudinal datasets from clinical trials. This chapter compares the patterns of missing data between study subjects who received treatment and study subjects who received a placebo.

Part III: Monte-Carlo in Statistical Modellings and Applications (Chapters “Monte-Carlo Simulation in Modeling for Hierarchical Generalized Linear Mixed Models”–“Bootstrap-Based LASSO-type Selection to Build Generalized Additive Partially Linear Models for High-Dimensional Data”)

Chapter “Monte-Carlo Simulation in Modeling for Hierarchical Generalized Linear Mixed Models” adds a discussion of Monte-Carlo simulation-based hierarchical models, taking into account the variability at each level of the hierarchy. In this chapter, Kyle Irimata and Jeffrey Wilson discuss Monte-Carlo simulations for hierarchical linear mixed-effects models to fit the hierarchical logistic regression models with random intercepts (both random intercepts and random slopes) to multilevel data.

Chapter “Monte-Carlo Methods in Financial Modeling” demonstrates the use of Monte-Carlo methods in financial modeling. In this chapter, Chuanshu Ji, Tao Wang, and Leicheng Yin discuss two areas of market microstructure modeling and option pricing using Monte-Carlo dimension reduction techniques. This approach uses Bayesian Markov chain Monte-Carlo inference based on the trade and quote database from Wharton Research Data Services.

Chapter “Simulation Studies on the Effects of the Censoring Distribution Assumption in the Analysis of Interval-Censored Failure Time Data” discusses using Monte-Carlo simulations to evaluate the effect of the censoring distribution assumption for interval-censored survival data. In this chapter, Tyler Cook and

Jianguo Sun investigate the effectiveness and flexibility of two methods for regression analysis of informative case I and case II interval-censored data. The authors present extensive Monte-Carlo simulation studies that provide readers with guidelines regarding dependence of the censoring distribution.

Chapter "Robust Bayesian Hierarchical Model Using Monte-Carlo Simulation" uses Monte-Carlo simulation to demonstrate a robust Bayesian multilevel item response model. In this chapter, Geng Chen uses data from patients with Parkinson's disease, a chronic progressive disease with multidimensional impairments. Using these data, Chen illustrates applying the multilevel item response model to not only deal with the multidimensional nature of the disease but also simultaneously estimate measurement-specific parameters, covariate effects, and patient-specific characteristics of disease progression.

In Chapter "A Comparison of Bootstrap Confidence Intervals for Multi-level Longitudinal Data Using Monte-Carlo Simulation", Mark Reiser, Lanlan Yao, and Xiao Wang present a comparison of bootstrap confidence intervals for multilevel longitudinal data using Monte-Carlo simulations. Their results indicate that if the sample size at the lower level is small, then the parametric bootstrap and cluster bootstrap perform better at the higher level than the two-stage bootstrap. The authors then apply the bootstrap methods to a longitudinal study of preschool children nested within classrooms.

Chapter "Bootstrap-Based LASSO-Type Selection to Build Generalized Additive Partially Linear Models for High-Dimensional Data" presents an approach to using a bootstrap-based LASSO-type selection to build generalized additive partially linear models for high-dimensional data. In this chapter, Xiang Liu, Tian Chen, Yuanzhang Li, and Hua Liang first propose a bootstrap-based procedure to select variables with penalized regression and then apply their procedure to analyze data from a breast cancer study and an HIV study. The two examples demonstrate the procedure's flexibility and utility in practice. In addition, the authors present a simulation study that shows, when compared with the penalized regression approach, their variable selection procedure performs better.

As a general note, the references for each chapter are included immediately following the chapter text. We have organized the chapters as self-contained units so readers can more easily and readily refer to the cited sources for each chapter.

To facilitate readers' understanding of the methods presented in this book, corresponding data and computing program can be requested from the first editor by email at DrDG.Chen@gmail.com.

The editors are deeply grateful to many who have supported the creation of this book. We thank the authors of each chapter for their contributions and their generous sharing of their knowledge, time, and expertise to this book. Second, our sincere gratitude goes to Ms. Diane C. Wyant from the School of Social Work, University of North Carolina at Chapel Hill for her expert editing and comments of this book which substantially uplift the quality of this book. We gratefully acknowledge the professional support of Hannah Qiu (Springer/ICSA Book Series coordinator) and Wei Zhao (associate editor) from Springer Beijing that made publishing this book with Springer a reality.

We welcome readers' comments, including notes on typos or other errors, and look forward to receiving suggestions for improvements to future editions of this book. Please send comments and suggestions to any of the editors listed below.

October 2016

Ding-Geng (Din) Chen
University of North Carolina, Chapel Hill, USA
University of Pretoria, South Africa

John Dean Chen
Credit Suisse, New York, NY, USA

Springer

http://www.springer.com/978-981-10-3306-3

Monte-Carlo Simulation-Based Statistical Modeling

Chen, D.-G.; Chen, J.D. (Eds.)

2017, XX, 430 p. 64 illus., 33 illus. in color., Hardcover
ISBN: 978-981-10-3306-3