Biometrics 71, 274–277
March 2015

BOOK REVIEWS

EDITOR:
TAESUNG PARK

ATKINSON, A. C. and BISWAS, A. C. **Randomized Response-Adaptive Designs in Clinical Trials**. Chapman and Hall/CRC, Boca Raton, Florida.

Randomization plays a key role in clinical trials and produces treatment groups in which the distribution of both known and unknown prognostic factors are similar. Statistical consideration of efficient estimation and powerful statistical tests prefer equal allocation, so a popular method of allocation in clinical trials is a 50:50 randomization if there are only two treatments. However, in a 50:50 randomization, half of the patients will be treated with the worse treatment. This is unethical, because as many patients as possible should receive better treatment. In order to solve this problem, the randomized response-adaptive design is an option to be considered.

The authors present the fundamental ideas and methods of randomized response-adaptive design. This book is clearly written and well-structured for a graduate course as well as consulting statisticians. Chapter 2 describes randomization methods which achieve balance across the treatment and covariates. Properties of the methods such as selection bias, loss, and bias are investigated. Chapter 3 covers response-adaptive design for binary responses whose objective is to skew the sequential allocation procedure in favor of better treatment for ethical consideration. For example, Urn's designs and the Play-the-winner rule are discussed. Both Chapters 4 and 5 deal with the development of response-adaptive designs when responses are either continuous or longitudinal responses. Randomized response-adaptive designs discussed above deals with the ethical concerns, but statisticians want to have some form of optimality in the allocation process to have higher efficiency in statistical inferences. Chapters 7, 8, and 9 cover such optimal allocation designs.

The European Medicines Agency addresses the response-adaptive designs, half a page in length, in their guidelines for clinical trials in small populations. Therefore, this book is particularly useful for the development of orphan drugs.

However, many current phase III clinical trials are non-inferiority trials in which the efficacy of a new treatment and an active control are supposed to be almost similar. In such designs, usefulness of randomized response-adaptive designs may be limited. Another practical issue is how to ensure that the un-blinded analysis results of previous patients is not known to both investigators and patients. The independent third party that deals with the un-blinded analysis results of previous patients might be needed. Therefore, effort might be needed to maintain the integrity of the clinical trials.

Seung-Ho Kang
Department of Statistics
Yonsei University
Seoul, Korea
seungho@yonsei.ac.kr

ZHOU, X.-H., ZHOU, C., LUI, D., and DING, X. **Applied Missing Data Analysis in the Health Sciences.** Wiley, Hoboken, New Jersey.

In all fields of study, missing data are a common problem for any data collection process. When we are trying to answer scientific questions of interest, researchers ask how the missing values will be handled. The statistical literature to answer this question is well developed, but very technical and complicated for researchers who are not experts in statistics. Therefore, researchers may recognize the missing values, but fail to respond appropriately. Therefore, the authors attempt

© 2015, The International Biometric Society

to guide the researchers with experience in applied data analysis to understand and implement the methods described in this book.

This book is organized into eight chapters. They may be classified into four parts, Chapters 1–3, Chapters 4–6, Chapter 7, and Chapter 8. The first part corresponds to an introduction to missing data. Chapter 1 introduces concepts on patterns and mechanism of missing data and some real-work examples analyzed through the book. They are collected from a wide range of clinical studies including cross-sectional data, longitudinal data, and survival data. Chapter 2 provides an overview of the existing approaches and classifies them according to whether they remove observations with missing data, utilize all available data; in example, one can use both complete and partially observed data, or impute missing values. Chapter 3 outlines some designs and conducts strategies to avoid or remove missing data in biomedical research studies based on a National Research Council report and a report in the New England Journal of Medicine (Little et al., 2012) because many missing data arise from a poor study design and lack of careful planning. It also includes the ways to reduce the impact of missing data when missing data are unavoidable. The cross-sectional data analysis refers to the analysis that looks at the data collected on subjects at one time point. Chapter 4 provides the statistical methods for analyzing such kind of cross-sectional data with missing values. They include maximum likelihood (ML) approach, Bayesian methods, multiple imputation (MI), and inverse probability weighting (IPW) method. In addition, doubly robust estimators are discussed. At the end of Chapter 4, the set of R codes are displayed to carry out the complete-case and the EM algorithm, the Bayesian analysis, the analysis based on the mean imputation and the MI, and the IPW estimation, respectively. In contrast to the cross-sectional studies, longitudinal studies have the defining feature of repeated measures collected on individuals over time. Chapter 5 begins by introducing the modelling approaches and statistical inferences for longitudinal data without missing values, and then introduces the settings of longitudinal data with missing values and simple methods to deal with missingness. When only the response variable is subject to monotone missingness, the likelihood-based method and the IPW generalized estimating equation (IPW-GEE) approaches are suggested and IPW-GEE methods are also extended to the situation of intermittent missingness (when subjects miss intermittent visits), or some measurements in a particular visit are missing. The likelihood approach and the weighted estimating equation approach provide valid inference with the monotone missingness pattern and their extensions to the intermittent missingness are straightforward. However, if the covariates are subject to missingness too, neither approach can be extended simply. Therefore, the MI procedure and the Bayesian approach are introduced in the setting where both outcomes and covariates are subject to missingness. In survival analysis, the outcome of interest, the failure time, may be censored such that either the failure time or the censoring time may be observed. Chapter 6 provides methods for survival analysis where the covariates are possibly missing.

This book is mainly focused on the Cox proportional (PH) model. At first, the authors review the Cox PH model when covariate data are completely observed and then introduce the enhanced analyses, the weighted methods, and imputation methods. In Chapters 4–6, the focus was on the data that were missing at random (MAR), that is, the missing mechanism is explained by the observed variable only. Under the MAR assumption, the observed-data likelihood can be written as a product of the missing mechanism part and the outcome model part. Therefore, the ML estimation of the outcome model could be done separately from estimating the missing mechanism. However, the MAR assumption can be violated in the situation where the missing mechanism depends on the missing data itself, or on some other unobserved variables that are associated with missing data. In case of missing not at random (MNAR), the observed-data likelihood involves the joint distribution of the outcome and the missing indicator. Chapter 7 focuses on four types of methods to deal with MNAR data such as likelihood-based methods, Bayesian methods, MI methods, and estimating equation methods. The likelihood-based methods include pattern mixture models, selection models, shared random effect models, and mixed effects hybrid models, that is, a combination of the pattern mixture model and the shared random effects model. In Bayesian analysis, the generalized linear models with MNAR covariates are handled with depth. Two MI approaches are introduced such as the proper MI by the Bayesian analysis and the approximate Bayesian bootstrap hot deck imputation. As in Chapters 4–6, the IPW methods apply to non-ignorable missing data. The methods handling missing data can be applied to make causal inference in randomized clinical trials with noncompliance and possibly missing outcome data. In practice, two types of noncompliance are encountered such as all-or-none noncompliance, that is, either compliance or not, and partial noncompliance. For dealing with the problem of all-or-none noncompliance, four different methods are presented. They are, instrumental variable methods, moment-based methods, ML and Bayesian methods. In addition, the MI method for dealing with noncompliance and MAR missing outcome data are presented.

To conclude, this book may be very helpful for researchers who are not familiar with computer programming because all examples analyzed in each chapter are coded with R and Stata language, and all source codes are provided publicly. For those with a stronger background in statistics, it provides full understanding to the statistical literature related to the missing data problems in clinical studies.

Reference

Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., Neaton, J. D., Rotnitzky, A., Scharfstein, D., Shih, W. J., Siegel, J. P., and Stern, H. (2012). The prevention and treatment of missing data in clinical trials. *The New England Journal Medicine* **367**, 1355–1360.

Jinheum Kim
Department of Applied Statistics
University of Suwon
Gyeonggi-do, Korea
jinhkim@suwon.ac.kr

CHEN, M.-H., KUO, L., and LEWIS, P. O. **Bayesian Phylogenetics: Methods, Algorithms, and Applications**. Chapman and Hall/CRC, Boca Raton, Florida.

This book provides an extensive and concrete account of modern Bayesian phylogenetics from a perspective how they differ from typical statistical methods in both of theoretical and computational aspects. The book's content mostly provides an overview of subareas in Bayesian phylogenetics that can be understood at the graduate student level. Each chapter opens with an introduction of issues, motivations, and recent developments. Readers can learn various applications, as well as how those can be grouped within the framework of Bayesian phylogenetics. In this regard, this book will be a good starting point for entering each subarea of Bayesian phylogenetics. Along with extensive and up-to-date overview of Bayesian phylogenetics, I believe that this book can serve as a road map for both starters and those who are already in this field.

The book consists of independent but cohesive thirteen chapters. For each chapter, established experts in the field were invited to write a chapter of their own, jointly with the editors. The book appears to be the result of a concerted effort by the entire Bayesian phylogenetics community.

The book opens with an introduction to Bayesian phylogenetics and overview of this book.

In Chapter 2, they introduce the Bayesian paradigm and various priors in Bayesian phylogenetics, and then demonstrate the application of priors on branch lengths, substitution models and tree topology.

The next two chapters focus on Bayesian model selection based on marginal likelihood. Chapter 3 covers basic ideas and formulas of estimators of marginal likelihood such as Generalized harmonic mean (GHM), inflated density ration (IDR), Path sampling (PS), Generalized path sampling (GPS) and Generalized steppingstone (GSS). Chapter 4 reviews Akaike information criterion through Markov chain Monte Carlo (AICM) as a posterior simulation-based analog of the AIC model selection criterion. In a simulation study, they assess the performance of these estimators and the maximum a posteriori (MAP) methods under uncorrelated relaxed clock model. They also demonstrate the comparison of coalescent tree prior models based on a real-world HIV-1 data set.

Topics in Chapter 5 and 6 are about the marginal likelihood estimator for the variable topology case. Chapter 5 focuses on generalizing the GSS method to accommodate varying tree topology. They propose the reference distribution for tree topologies and edge lengths and then compare their proposed method with a brute-force approach based on 6-taxon binary un-rooted tree. In Chapter 6, with extending the IDR method to varying tree topology case, they prove the statistical consistency for the varying tree topology versions of GHM, GSS and IDR.

Chapter 7 provides extensive reviews for fundamental Markov Chain Monte Carlo (MCMC) algorithms for Bayesian phylogeny inference. To overcome the difficulties of these algorithms in high-dimensional problems, they present sequential stochastic approximation Monte Carlo algorithm (SSAMC) as an efficient algorithm to overcome the curse of dimensionality.

In Chapter 8, they introduce sequential Monte Carlo (SMC, also known as particle filter methods) as one promising complement to MCMC from computational perspective. The review especially focuses on the computational gains by SMC, and provides the details of SMC algorithm.

Chapter 9 provides a framework for combining multi-locus data sets such as large scale bio-geographic or population genetic datasets. Under the assumption that data blocks are independent, they introduce the combined marginal likelihood integrates marginal likelihood estimates over all independent data blocks.

Chapter 10 covers the issues on the recombination in phylogenetics. They discuss the impact from the ignorance of recombination, and then present various inference methods in the presence of recombination. They review Ancestral Recombination Graph (ARG) and cover inference, sampling methods, and software used for inferring the ARG.

Chapter 11 focuses on a general overview of modern Bayesian nonparametric methods for inference of effective population size trajectories. After describing the Bayesian model formulation along with coalescent prior distribution for gene genealogy, the authors review priors on effective population size trajectory such as multiple change-point models, coalescent priors as a point process and Gaussian process-based nonparametric priors.

Chapter 12 opens with an observation that, in conventional likelihood-based framework, unrealistic assumptions are often imposed when the evolutionary models are too biologically rich to be computationally feasible. As an alternative, the authors present endpoint-conditioned evolutionary inference, especially focusing on both dependent and independent site models.

Finally, Chapter 13 covers Bayesian inference of species divergence time which is the product of the branch rate and the branch length. They review priors on branch rates (e.g., auto-correlated rate, local molecular clock) on node times (e.g., generic prior, branching-process priors). Priors for calibration are also discussed.

The book is concise, and the examples accompanying each topic are clear and supported by relevant illustrations. I have only a minor complaint regarding the absence of a separate chapter for basic definition and models in phylogenetics. Even if the book's focus is on the Bayesian methods and these items are briefly handled over some chapters, a chapter dedicated to the understanding of basic definitions at the beginning would help to catch the flow of the book.

In summary, Bayesian phylogenetics provides an up-to-date review of methods and application of Bayesian Phylogenetics and should be useful to researchers and graduate students.

Joungyoun Kim
Biostatistics and Clinical Epidemiology Center
Samsung Medical Center
Seoul, Korea
joungyoun@gmail.com

BANERJEE, S., CARLIN, B. P., and GELFAND, A. E. **Hierarchical Modeling and Analysis for Spatial Data. Second Edition**. CRC Press/Chapman & Hall. Monographs on Statistics and Applied Probability 135, Boca Raton, Florida, 2015. 562 pp. $ 99.95 . ISBN-13: 978-1-4398-1917-3 (Hardcover).

This is a very welcome second edition of a nice and very successful book written by three experts in the field of spatial and spatio-temporal hierarchical Bayesian modeling. After 10 years of publishing the first edition, the authors have completed this very updated and comprehensive monograph. To have a rough idea of the changes in this new edition, the first edition had ten chapters and this new edition has fifteen. For example, the chapter on basics of point-referenced data has been completed with a new chapter about the theory underlying point-referenced data models. The topic on modeling and analysis of point patterns that was missing in the first edition has now a full new chapter including marked point processes, space–time point patterns, and sections on measurement errors, scan statistics, and preferential sampling. The first edition chapter on multivariate spatial modeling has been split into two new chapters: one on multivariate spatial modeling for point-reference data and the other on multivariate areal data. Both chapters deserved attention by themselves. The chapter on spatio-temporal modeling has been also expanded. Two new interesting sections have been added. One on fitting dynamic spatio-temporal models using `spBayes` (Finley and Barnejee, 2013), an `R` package that fits univariate and multivariate spatio-temporal models with Markov chain Monte Carlo (MCMC), and an additional section on geostatistical modeling in the context of stochastic differential equations, in particular using stochastic diffusion processes. Two new chapter additions are also more than welcome: a complete chapter on modeling large spatial and spatio-temporal data sets and another one on spatial gradients and wombling boundaries. The last chapter in the first edition on special topics in spatial process modeling has been updated with new sections on data assimilation and space-time modeling for extremes. As a researcher interesting in disease mapping I would like to add a comment on the contents of Chapter 6 (Section 6.4). The authors recommend to use an intrinsic conditional autoregressive (intrinsic CAR) prior for the spatial component in a Poisson-normal model for areal data instead of the classical parametrization proposed by Besag, York, and Mollié (1991). Some recent papers have argued however that this prior might produce negative correlations for regions that are located further apart. I think that the parametrization proposed by Leroux, Lei, and Breslow (1999) and presented also in the book (there is a typo by the way in the reference on p. 159) or the one proposed by Dean, Ugarte, and Militino (2001) could be more appropriate.

I have no doubts that this updated text will continue being a compulsory reference for those graduate students and researchers interested in understanding and applying any of the three areas of spatial statistics: geostatistics (point-reference data), lattice (areal) data and point patterns, under a fully Bayesian context whenever possible. The literature in space and space–time modeling is growing in the last years and this book represents a more gentle and updated introduction to the field than the classical books by (8) or (2), apart from containing many exercises, good examples, and comments on computational tools in some of the chapters. The new version has avoided all the references to the `S-PLUS` commercial software, restricting its attention to `R` (R Core Team, 2014) and `WinBUGS` (Lunn et al., 2000). In addition, electronic versions of most of the data sets are maintained by Brad Carlin at the url `http://www.biostat.umn.edu/ brad/data2.html`. The license free software and the data availability facilitate also the use of this book for teaching.

The monograph is printed in color and this helps to see better some of the graphical representations that were printed in black and white in the previous edition. It is not too pleasant however, the glossy paper used in the book for two reasons: it sometimes reflects the light making the reading a bit annoying, and it results in a heavier text. Both minor inconveniences can easily be overcome using the electronic version if necessary and do not affect at all to the contents of this excellent book that I highly recommend for anyone interested in the fascinating field of space and space–time modeling. This is definitely one of those second edition books that is worthwhile having it. Many thanks to the authors for their effort.

Reference

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1–20.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. New York: Wiley

Dean, C. B., Ugarte, M. D., and Militino, A. F. (2001). Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics* **57**, 197–202.

Finley, A. O. and Banerjee, S. (2013). spBayes: Univariate and Multivariate Spatial–Temporal Modeling. R Package Version 0.3-8. http://CRAN.R-project.org/package=spBayes.

Leroux, B. G., Lei, X., and Breslow, N. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment and Clinical Trials*, M. Halloran and D. Berry (eds), 135–178. New York: Springer-Verlag.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Ripley, B. D. (1981). *Spatial Statistics*. New York: Wiley.

Maria Dolores Ugarte
Statistics and O. R. Department
Institute for Advanced Materials, INAMAT
Public University of Navarre
Pamplona, Spain
lola@unavarra.es