Biometrics 67, 1672–1679
December 2011

BOOK REVIEWS

EDITOR: GUILHERME J. M. ROSA

REILLY, C. **Statistics in Human Genetics and Molecular Biology.** Chapman & Hall/CRC, Boca Raton, Florida, 2009. Xii + 266 pp. $61.95/£39.99, ISBN 9781420072631.

"I maintain, however, that in every special doctrine of nature only so much science proper can be found as there is mathematics in it." So wrote Immanuel Kant in his *Philosophy of Material Nature* over two centuries ago. Thankfully, from its earliest origins in Mendel's gardens, the field of genetics has embraced the importance and use of mathematical and statistical thinking. However, the complexity of modern genetic research forces us far beyond the comfortable Punnett squares with which we started.

Ever-increasingly complex and massive genetic data sets are generated on unprecedented scales and offer great promise to unveil mysterious genetic secrets and crack the codes for complex traits and human disorders; however, there is also great possibility for confusion and error. Thankfully, some brave souls are willing to serve as guides to rigorous application and understanding of statistical approaches to genetically informative data. Cavan Reilly is among them.

Reilly's new book aims at introducing basic knowledge of the mathematical and statistical problems in genetics and genomics and providing an overview on a variety of popular methods that are vital to making sense of modern genomic data. The target audience of this book is, ideally, graduate students in statistically related fields. The book is self-contained and well organized, covering a substantial breadth of the core topics in genetics and genomics. Along with a nice introductory review on basic concepts in molecular biology and statistics, such as Mendel's laws and the extreme value distribution theory, it presents a diverse set of problems in genetic mapping, sequence alignment, analyses of gene expression data from microarray experiments, and sundry approaches that have been used to address these problems. Each chapter includes its own list of reference, relevant worked examples and end-of-chapter exercises to cement understanding of these statistical applications. *Repetitio est mater studiorum*, we can almost hear Professor Reilly enjoining. A few chapters also list bibliographic notes for further reading. Overall, this book is a valuable reference source for both statistics-oriented and human-genetics-oriented researchers and graduate students to learn the specialized methodology for analysis of diverse genetic data.

The book contains 15 chapters. Chapters 1 and 3 contain a brief summary of the basic concepts and terms in biology and biotechnology including DNA, RNA, protein, gene, chromosome, genetic marker, and physical map. The key concepts are clearly explained and are within reach of nonbiology major graduate students. Chapters 2 and 9 recapitulate essential statistical theory and principles such as hidden Markov models, likelihood-based statistics, Bayesian inference, and Markov chain Monte Carlo sampling. The mathematical prerequisites are moderate, assuming the reader has basic knowledge of calculus, algebra, probability, and statistics. Chapters 4 to 6 are perhaps a bit anachronistic at this point, but nevertheless effectively outline genetic mapping theory for two-point or multipoint parametric and nonparametric linkage analysis, and association study for Mendelian and complex phenotypes. The Elston–Stewart algorithm and the Lander–Green algorithm for linkage computation, and the Identity-by-Descent sharing,

the Haseman–Elston regression, and the variance components models for gene detection are delineated. Chapters 7 and 8 are devoted to pairwise (DNA, RNA, or protein) sequence alignment and chapters 10 and 11 focus on multiple sequence alignment and feature recognition. FASTA, BLAST, Dot plots, dynamic programming, and hidden Markov model approaches are illustrated in global and local alignments for understanding functional, structural, or evolutionary relationships between the sequences, identifying genes, and discovering motifs. Chapters 12 to 15 cover differential gene expression detection as well as classification and cluster analysis using gene expression data sets. Methods for quantification of expression level, normalization of microarray data, two-sample comparison of differential expression, and control of family-wise error rate or false discovery rate are elaborated. Cluster analysis is introduced as a class of techniques to find groups of genes that behave in similar ways in response to a variety of stimuli and sets of subjects that have similar gene expression profiles. Discriminate analysis, regression trees, nearest neighbor classifiers, support vector machines, random forests, and few other methods are described as classification techniques useful in gene expression measurements.

A minor downside we see in this book is that a few issues of practical importance such as how to select Logarithm of the ODds threshold in linkage analysis and how to conduct multiple testing adjustments for correlated multiple markers in association studies are ignored; however, no book can cover all. Another downside is that some cutting-edge topics such as statistical analysis in genome-wide association studies, and analysis of copy number variation data, whole exome, and whole genome sequence data have not been thoroughly included.

In conclusion, this is a useful text book for beginners trained in applied mathematics and statistics to take in a panoramic snapshot of the very evolving field of statistical genetics and genomics.

Xiang-Yang Lou and David B. Allison
Department of Biostatistics
University of Alabama at Birmingham
Birmingham, Alabama, USA

GUERRA, R. and GOLDSTEIN, D. R. **Meta-Analysis and Combining Information in Genetics and Genomics.** Chapman & Hall/CRC, Boca Raton, Florida, 2009. xxiii + 335 pp. $99.95/£63.99. ISBN 9781584885221.

The book *Meta-Analysis and Combining Information in Genetics and Genomics* is a collection of contributed articles by various authors on the topics, edited by Rudy Guerra and Darlene R. Goldstein. Guerra and Goldstein do a fantastic job of describing the scope and contents of the collection, both in their preface and in their Chapter 1, so my review will somewhat duplicate their own review.

For someone who is interested in either metaanalysis or genomics, this book provides a great overview of both. Guerra and Goldstein have written a very nice introduction to familiarize readers first with the basics of metaanalysis, next with the basics of modern genomics, and finally with the intersection of the two. Books that are collections of papers such as this tend to be just that, a collection of papers; however, Guerra and Goldstein have done a wonderful job in introducing the material, and in organizing the collection coherently.

The book is well arranged into three major sections: part I contains four articles on combining similar data types with genotype data, including genome-wide linkage scans. (If the genetic terminology is unfamiliar, you can read Guerra and Goldstein's introductory section.) Part II also considers combining similar data types, but with gene expression data, and includes six articles. The final section, considers how to combine data of different types, which involves special challenges of comparability and scaling, and has six more papers. The book concludes with a lengthy reference list, as well as an index that I found particularly helpful.

Rather than provide a detailed review of each of the 16 articles individually, I would instead suggest that you read Guerra and Goldstein's reviews in their preface and in their Chapter 1, as mine would simply duplicate theirs. But I will say that the material is very accessible to readers of *Biometrics*, with an ample mix of theory (not too hard), data analysis, and simulation. The spectrum of methods is considered: likelihood, frequentist, Bayes', and empirical Bayes', so there is something for everyone.

As a researcher in the field of multiple comparisons and multiple tests (MCMT), I have always been intrigued by metaanalysis. Metaanalytic tests form an alternative to MCMT, in that the component test data are combined to arrive at an overall conclusion, rather than separated out to find individual conclusions as in the case of MCMT. Also, metaanalytic (composite) tests are the foundation of closed testing, one of the more powerful methods for MCMT. Finally, the two approaches have a similar, potentially fatal, flaw: how does one choose the collection of items from which to form the inferences? Thus, I was interested in how the book treated the subject of MCMT. My first reaction was negative, because they botched the reference to my book *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (Westfall and Young, 1993), joint with Stan Young—it was published in 1993, not 1999! I was further disappointed that the book did not better elucidate the MCMT/metaanalysis differences. A recent book by Brad Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction* (Efron, 2010), for example, is also mostly about genomics, but written completely from the point of view of MCMT.

As well, there are metaanalysis techniques that come closer to combining data and simultaneously providing individual inferences. I was disappointed by the lack of mention of the truncated product p-value combination test, which allows specific inferences and also solves the problem of missing (due to insignificance) p-values discussed in the book. I was also disappointed with the short shrift given to the popular method of "enrichment analysis" (or combining decisions), which was dismissed in Chapter 1 using references no later than 1978.

Further nitpicks: the book could use a little tighter editing. There are mildly contradictory statements, such as when discussing Fisher's combination test it is first stated that the hypotheses are assumed all the same (which is not a necessary assumption), and later it is assumed that "each of the k null hypotheses is true," suggesting (correctly) that the hypotheses can be different. Also, while funnel plots are mentioned repeatedly in the "Introduction," no picture is given. Finally, there are occasional awkward grammatical constructions, although they are mostly minor and typical for a book of this type.

My bottom line is that this book is well worth having as a reference book for those interested in metaanalysis and/or genomics. Although edited collections such as this are, in my experience, less useful than more focused texts, this one is better than most. Despite its minor flaws, Guerra and Goldstein have done an admirable job putting the collection together.

References

Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. New York: Cambridge University Press.

Westfall, P. H. and Young, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York: Wiley.

Peter H. Westfall
Rawls College of Business
Texas Tech University
Lubbock, Texas, USA

OJA, H. **Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks**. Springer, New York, 2010. Xiii + 232 pp. $79.95/€74.85, ISBN 9781441904676.

This book gives a comprehensive overview of multivariate nonparametric estimators and related hypotheses, based on spatial signs and ranks. It covers many research topics investigated by the research group of the author Prof. Hannu Oja, as can be expected from a volume within the series "Lecture Notes in Statistics." However, also many other related research results with adequate references are presented throughout all chapters.

The structure of the book is clear and well outlined in the preface. The first four chapters give an overview of the main tools that are needed to construct the nonparametric methods described in the following chapters. The introductory Chapter 1 gives a very concise introduction to the general concept of spatial sign, spatial rank, and spatial signed rank. It also provides a short overview of alternative procedures, such as those based on marginal signs and ranks, or Oja signs and ranks. Chapter 2 mainly contains definitions of the multivariate parametric and semiparametric location and scatter models used in the rest of the book, based on different assumptions of symmetry. Already in this chapter, all the material is presented in a unified setting. For example, the different types of symmetry are defined through an invariance property of certain transformations. Multivariate location and scatter functionals, and their finite-sample versions, are introduced in Chapter 3, as well as some important properties such as first and second moments, breakdown point, and influence function. Finally, Chapter 4 is a more detailed version of the first chapter. It explains how univariate signs and ranks can be generalized to the multivariate case, by avoiding a formulation that requires an ordering of the data. Also sign and rank covariance matrices are introduced here.

The next four chapters comprehend the study of location in the one-sample setting. Chapter 5 focuses on different versions of Hotelling's T^2 test, which corresponds with the identity score function and the sample mean as location estimator. The spatial sign score function similarly leads to the spatial sign test and the spatial median, and is studied in Chapter 6. Affine invariant tests are discussed as well, which leads to Tyler's scatter matrix. Next, the spatial signed-rank test and the related Hodges–Lehmann location estimator are considered in Chapter 7. In all these chapters limiting distributions under the null and the alternative hypothesis are summarized, which can, for example, be used for sample size and power computations. To conclude the location setting, a comparison of the efficiency and robustness of the tests and estimators is presented in Chapter 8.

The one-sample setting is further studied in Chapter 9, which treats different types of the spatial sign and spatial rank covariance matrices as well as their use for principal component analysis. Nonparametric tests about the independence between two sets of variables (extending the parametric Wilks' test and Pillai's trace test) are handled in Chapter 10 and lead to multivariate generalizations of Blomqvist quadrant test, Spearman's rho, and Kendall's tau test.

The book continues with several other multivariate procedures. The study of location in multiple samples is covered in Chapters 9 and 10. Tests for independent samples are defined in Chapter 9 based on the identity score function, the spatial sign score function, and the spatial rank score function, and their limiting distributions are discussed. Also differences between location centers (treatment effects) are studied. For detecting differences in treatments within a randomized block design, multivariate extensions of the parametric two-way analysis of variance test and the nonparametric Friedman test are given in Chapter 12. The setting of multivariate multiple linear regression (i.e., regression with multiple response and explanatory variables) is studied in Chapter 13, and leads to a unified treatment of least squares regression, least absolute deviation regression based on spatial signs, and least absolute deviation based on spatial ranks. The final Chapter 14 considers cluster-correlated data, such as grouped data or longitudinal data, for which parametric mixed models are typically used. Here again, nonparametric variants for the one- and two-sample location setting are presented.

The structure within each chapter is very consistent. All the definitions and the theoretical results on the estimators and the hypothesis tests are explained in a unified framework,

which essentially is based on different score functions. In this way differences and similarities between estimators, tests, and asymptotic results are emphasized. Also the procedure to construct affine equivariant estimators (or affine invariant tests) is repeatedly explained. The required model assumptions for each procedure are always clearly stated. Most chapters also contain some examples on real data, and corresponding R code from the packages MNM (multivariate nonparametric methods) and SpatialNP. This makes the whole work to be a great well-structured resource about the construction, the theory, and the use of nonparametric methods based on spatial signs and ranks.

The information in the book is not complete, in the sense that most results are just presented without proofs. A detailed study of each method thus requires consulting the corresponding paper(s), which seems feasible as many references are provided. My only minor remark is that the description is sometimes a bit too concise and the author seems to presume that the reader is already somewhat familiar with the necessary tools. For example, the first (introductory) chapter immediately introduces the multivariate sign, rank, and signed-rank score function, whereas the comprehensive motivation for these definitions is given in Chapter 4. The first figures only show up on page 37 of Chapter 4 and the first real data example on page 47 at the start of Chapter 5.

To conclude, I consider this book as an encyclopedic work, ideal for researchers who want to have a clear overview on the development and state of the art of many nonparametric methods based on spatial signs and ranks. It is also very valuable for users of the R statistical software packages MNM and SpatialNP.

Mia Hubert
Department of Mathematics—Lstat
Katholieke Universiteit Leuven, Belgium

KING, R., MORGAN, B. J. T., GIMENEZ, O., and BROOKS, S. P. **Bayesian Analysis for Population Ecology.** Chapman & Hall/CRC, Boca Raton, Florida, 2010. Xiii + 442 pp. $82.95/£50.99, ISBN 9781439811870.

Recent concerns over human impacts on the natural environment have increased interest in the fields of ecological and environmental statistics. Bayesian methods continue to play a large role in these areas, and King, Morgan, Gimenez, and Brooks (hereafter KMGB) have been instrumental in developing Bayesian methods for analyzing ecological data—specifically data from the study of marked individuals. The primary strengths of this book are the authors' extensive practical experience in applying Bayesian methods and the advanced material on model selection and multimodel inference, particularly via reversible jump Markov chain Monte Carlo (RJMCMC). This would be a valuable reference for those already familiar with core Bayesian methods, and who are looking to learn more about ecological statistics or to implement these methods for complex ecological data.

The opening section of the book introduces both population ecology and classical, likelihood-based inference. Chapters 1 and 2 describe the field of ecological statistics and the types of studies, data, and models referenced through the remainder of the text. This is a good start for those in the statistical community who have not worked with ecological data and who are interested in learning about the field. However, most in the *Biometrics* audience will be able to skim over the third chapter, which provides an overview of frequentist methods. The chapter moves quickly and much of the content is not referenced later in the text. The most important topics for readers not yet familiar with classical inference are those on modeling and likelihood construction, which are key to the Bayesian approach in general and to understanding the examples presented in later chapters.

The four chapters in the next section cover Bayesian statistics from the basics (Bayes' theorem and prior specification) through to algorithms for sampling from posterior distributions with and without model uncertainty (specifically Markov chain Monte Carlo [MCMC] and RJMCMC). The material in these chapters moves quickly at times and would be ideal for those who already have some knowledge of Bayesian methods. This complements the other recent books by McCarthy (2007) and Link and Barker (2010) that provide gentler introductions to Bayesian inference for ecological data but do not consider the application and complexities of multimodel inference in as great depth. Note that Link and Barker (2010) do present multimodel inference from a different perspective, but the methods they discuss do not play a large part in the examples presented in the final section of their book.

The third and final section of KMGB's book provides specific examples of Bayesian applications to mark–recapture/resight/recovery data, and it is here that the authors' experience really comes through. The four chapters in this section each begin by discussing one or two problems in the study of marked individuals followed by a complete Bayesian solution including details on how to specify models, select priors, choose proposal distributions to implement efficient (RJ)MCMC sampling, and draw inferences from posterior distributions. Several fully worked examples taken mostly from the authors' own research are presented in each chapter, and these go a long way in helping to unravel some of the art of Bayesian inference. The material is well presented and will be informative both to statisticians seeking an introduction to ecological modeling and to ecologists wishing to learn about Bayesian inference.

The text generally does a good job at describing multimodel inference and RJMCMC, though one or two of the advanced topics could have been expanded. In particular, there is only a brief discussion of convergence diagnostics for RJMCMC, and the ideas are not implemented in any examples. KMGB do cover convergence diagnostics for standard MCMC algorithms and repeatedly stress the importance of sensitivity analysis, presenting clear examples to show how the prior may affect the posterior distribution both with and without model uncertainty. Though diagnostics for RJMCMC are still being actively developed, assessing convergence is crucial for any iterative method and further direction is needed for anyone who intends to implement these methods.

For those hoping to try the methods themselves, Chapter 7 provides sample computer code in R and WinBUGS to implement MCMC and RJMCMC algorithms for two problems and electronic versions of the code are also available online. The code is clearly written with many comments and the accompanying text provides a good discussion of the implementation choices and resulting output. Introductions to R and WinBUGS are included in the appendices for those who may not have used the software. Readers may wish for further code for all of the examples in Section III and the authors do note on their website that "more [code] will be added in due course." This would be a great addition to the text.

References

Link, W. A. and Barker, R. J. (2010). *Bayesian Inference with Ecological Applications*. San Diego, California: Academic Press.

McCarthy, M. A. (2007). *Bayesian Methods for Ecology*. New York: Cambridge University Press.

Simon Bonner
Department of Statistics
University of Kentucky
Lexington, Kentucky, USA

FERRATY, F. and ROMAIN, Y. (eds) **The Oxford Handbook of Functional Data Analysis.** Oxford University Press, New York, 2011. xviii + 494 pp. $152.95/£85.00, ISBN 9780199568444.

Functional data analysis (FDA) is a subject with a rapidly growing popularity. One could call it statistics for curves and surfaces. It covers many problems that come with modern high-intensity measurements such as time series, spectra, and images. The observations are discrete but they are obtained by (regular or irregular) sampling of continuous phenomena. One typical problem is alignment (warping) to maximize the similarity between a master curve or surface, the template, and the other ones. Another is regression on curves or surfaces to predict values of a dependent variable.

This book is a volume in the growing series of *Oxford Handbooks in Mathematics*. The printing and binding is of very high quality, the layout is well balanced, and the choice of font types is excellent. This is a beautiful book.

The editors explain in the preface that the book grew out of the activities of the working group STAPH at The Institute of Mathematics at The University of Toulouse. Apparently, its members are strongly mathematically oriented and that determined the spirit of the book. Only two of the chapters (one by Jim Ramsay, the other by Garett James) seriously deal with real data analysis, all the others are (very) heavy on theory. Hence, the title of the book is misleading: "data analysis" should be "statistics."

The theoretical character is also illustrated by the low number of graphs, only 26 in total. Of them 20 are in the two applied chapters I mentioned above. Of the other six, only one more or less relates to FDA (it shows time-varying coefficients). Compare that to the well-known book by Ramsay and Silverman, which contains over a hundred graphs. Because curves allow us to see so much, in my view graphs are essential in this field.

Also remarkable is the very low number of references to the chemometric literature. It probably is the area where practical interest in regression on curves (mainly spectra) started and still is actively being investigated.

See the web page of the book (search with the ISBN on the OUP website, www.oup.co.uk) for the complete table of contents and the authors of the chapters. There are three parts. The first part, regression modeling for FDA, presents foundations. Part II, benchmark methods for FDA, contains chapters on resampling, functional principal component analysis, curve registration (alignment), classification, and sparse FDA. Here we find the chapters on real data analysis. In Part III, towards stochastic background in infinite-dimensional spaces, the mathematicians go all out. I am not convinced of the relevance of that material to data analysis. Even fans of the *Annals of Statistics* would have a hard time there.

Smoothing is an essential tool for FDA. I am very much in favor of regression bases and penalty methods, because I have used them successfully in a number of papers about regression on curves and surfaces (with Brian Marx). I am surely biased, but I was surprised to see that in this book kernel smoothers are used almost exclusively. I consider that rather old-fashioned.

To summarize: this is not a real handbook on FDA, because its scope is too limited. The blurb on the back cover says: "As a consequence, this book should appeal to a wide audience of engineers, practitioners and graduate students, as well as academic researchers, not only in statistics and probability but also in the numerous related application areas." That is only true for a small part of the book.

Paul H. C. Eilers
Erasmus University Medical Center
Rotterdam, The Netherlands

LAWSON, J. **Design and Analysis of Experiments with SAS.** Chapman & Hall/CRC, Boca Raton, Florida, 2010. Xiii + 582 pp. $99.99/£63.99, ISBN 9781420060607.

The book was written for first-year graduate students who intend to conduct experiments and consequently need to know something about experimental designs. The author writes that to be fully understood, a student using this book should have had previous courses in calculus, introductory statistics, basic statistical theory, applied linear models, and have had some familiarity with SAS. This reviewer feels that this is likely more than adequate. Although basic matrix notation for analysis of linear models is used throughout the book, the material is well within the reach of students from applied sciences or engineering.

This book deserves to be seriously considered as an external reference for a beginning, one semester course in experimental design. It is a handy, reference showing how to get things done quickly using SAS. It deals with many of the problems of generating designs that can be used for the research problem

at hand, the construction and implementation of the design, and the proper statistical analysis of the resulting data. Unfortunately, the theory underlying the constructions is largely ignored. For students who really want to learn about experimental designs, the instructor would have to supplement the material considerably. This book provides the SAS code for the various constructions and analyses required.

The book begins with an introductory chapter dealing with data collection, purpose of experimental design, and types of designs. Chapter 2, on completely randomized designs with one factor discusses topics such as replication and randomization. Matrix notation is used to illustrate normal equations, and concepts such as estimable functions, quadratic forms, and preplanned and unplanned tests of hypotheses. Chapter 3 treats two level factorial experiments in the completely randomized designs in considerable detail. Chapter 4 extends these concepts to the randomized block designs. Chapter 5 begins with designs to study variances and ends with designs for fixed and random factors. Chapter 6 begins with fractional factorials in the 2^k system, discusses the Plackett–Burman designs, and proceeds to the mixed level factorials and orthogonal arrays. An expanded discussion of the relative merits of the various Plackett–Burman and modifications thereof would be welcome.

Chapter 7 deals with incomplete block designs, full and partial confounding in the two, three, p level and mixed systems. Chapter 8 introduces the split-plot designs for two factor and more general split plot designs for the $2^{(k-p)}$ system. Chapters 9, 10, and 11 discuss crossover and repeated measures designs, response surface designs, and mixture experiments. Chapter 12 introduces the product array parameter designs. Unfortunately, relative merits of product array designs and fractional factorials in single arrays are not examined.

The book addresses a long list of topics. Unfortunately very little time is spent comparing the relative merits of the proposed designs. For students really interested in the subject, the instructor would have to supply these comparisons. Unfortunately the work is also marred by a number of errors. The definition of partially balanced incomplete block designs implied on page 261 is incorrect. (See section 27.2 in Kempthorne, 1952.) The definition of a Youden square design on page 265 is incorrect and the example on page 266 is not a Youden square. (See page 54 in Kempthorne, 1952.) Also the code, page 133, for randomizing a Latin square produces a selected square from only a subset of the possible squares. This subset consists of only squares derivable from one transformation set. (See page 186 in Kempthorne, 1952.)

Reference

Kempthorne, O. (1952). *The Design and Analysis of Experiments*. New York: John Wiley & Sons.

Francis Giesbrecht
Emeritus Professor of Statistics
North Carolina State University
Raleigh, North Carolina, USA

GAETAN, C. and GUYON, X. **Spatial Statistics and Modeling.** Springer, New York, 2010. Xiv + 302 pp. $89.95/€85.55, ISBN 9780387922560.

Spatial statistics is now a mature subject with a range of developments and book level publications. It is somehow conventional to subdivide the topic in three categories: random attributes with (1) geostatistical (spatially continuous, point referenced data) and (2) lattice/network (spatially discrete, area referenced data) structure, and (3) random geometry, in particular under the most common format of point processes. Literature in the area ranges from publications focused on a particular category to more comprehensive ones covering some or all of them. This book falls in the latter case. The preface sets the scene pretty much in a format of an introductory chapter with examples of data types, anticipating basic terminology and context. However, as a distinctive aspect, the five chapters are not developed around such a subdivision. Instead, the authors have chosen to present characterization of spatial processes (Chapters 1 to 3), followed by a chapter on simulations (Chapter 4) and another on inference (Chapter 5). The appendices cover essentials for simulation of random variables (A), background on limit theorems for random fields (B), elements of asymptotics (C) and a very brief account on related software.

The authors' approach is to present second-order characterization and models for fixed geometry. Geostatistical and area data structures are covered in Chapter 1 with the formalism of stochastic processes with theoretical characterization of process in terms of second-order properties and modelling by covariance functions, convolutions, spatial and spatio-temporal autoregressive, spatial regression. The chapter closes with spatial prediction by kriging. Transition from Chapter 1 to 2 can be described as going from global characteristics to conditional distributions and properties by means of Gibbs–Markov random fields on networks. Typical topics such as Ising and Potts model, Hammersley–Clifford, Besag auto models, and Markov dynamics are covered. The following chapter closes the characterization of spatial processes moving to point processes. In comparison with previous chapters the topic is developed with lesser degree of formalism in favour of a more heuristic presentation of a relatively large range of topics including general characterization, Cox processes, density, and nearest neighbour and Markov processes. Marked point processes are also mentioned, very briefly, however.

The chapter on simulation is motivated by problems for which direct or acceptance/rejection methods are impractical leading to Markov chains and Markov chain Monte Carlo based simulation strategies. The text includes the necessary background on Markov chains. General descriptions of Gibbs and Metropolis–Hastings schemes are followed by description of specific algorithms for networks and point processes. The chapter includes descriptions of tools for assessing performance and convergence and coupling from the past methodology for perfect simulation. Geostatistical processes are covered by spectral and turning bands methods.

The book closes with Chapter 5 dealing with inference for the different kinds of spatial processes, with references to asymptotics. A diversity of methods is described, including the ones based on second-order characteristics, likelihood and pseudo-likelihood, simulation based, and tests for presence of spatial patterns. Bayesian methods are briefly mentioned toward the end of the chapter for spatial hierarchical geostatistical with examples for geostatistical and area data. A slight but relevant detail is that some figures mapping data or predictions within study areas are presented without paying attention to the axis aspect ratio and distorting patterns.

The text is mostly devoted to univariate spatial data. The emphasis is on the formal and theoretical description of spatial processes. Data are shown throughout the text mostly as illustrative examples. All chapters end with a nice selection of exercises and most of them extend and complete the theory with specific details.

The book is a translated version of a previous published material. The authors succeed in presenting core concepts of spatial statistics on a compact version. The preface is written as a technical introduction and does not state the authors view on the target audience. Very technical notation is used from the beginning and contents require a specialised reader with solid mathematics and statistics background. The book provides a rather general and at the same time concise coverage of main elements of spatial processes. The book can be used as support material for a general and advanced level course on spatial statistics. Primary readership of graduate students and researchers with special interest in spatial processes with emphasis in theoretical aspects can find a broad overview and pointers to more specialized accounts of specific kinds of processes. The book can attract general and biometrics audience with necessary background and interest on a general and concise account of the field with focus on the theoretical elements whereas readership with focus on applications may favour alternatives covering in greater depth the specific subarea.

Computations use R, BUGS and AntsInFileds software. A very useful web supplement (http://www.dst.unive.it/∼gaetan/ModStatSpat/) provides links to software pages, datasets, and R code.

Paulo Justiniano Ribeiro Jr.
Universidade Federal do Paraná
Curitiba, Brazil

FOULKES, A. S. **Applied Statistical Genetics with R: For Population-based Association Studies.** Springer, New York, 2009. Xxiii + 252 pp. $59.95/€48.10, ISBN 9780387895536.

Despite their increasing importance and popularity, research areas in statistical genetics are rapidly evolved due to recent technological advancements (e.g., microarrays, single nucleotide polymorphism (SNP) arrays, and next-generation sequencing) and making them inaccessible to interdisciplinary students and researchers who have little exposure to this filed. This book, *Applied Statistical Genetics with R for Population-Based Association Studies* intends to fill this gap by introducing the statistical foundations and practical tools for the analysis of population-based genetic association studies. As a typical book in *Springer Use R! Series*, it provides examples from several publicly available data sets and corresponding R codes for illustrations. The book has been reviewed several times since its publication (Lananne, 2009; Sabatti, 2009; Nagarajan, 2010).

The book consists of seven chapters to provide the relevant theoretical aspects of the analysis of genetic data and an appendix to offer a basic introduction of R. The basic statistical and genetic concepts and association studies are introduced in the first three chapters. Chapter 1 describes the types of population-based association studies and features of genetic data. Several data sets as well as corresponding R codes are also provided for illustrations. In Chapter 2, the author reviews the standard statistical concepts and tools related to the association analysis and points out several related analytic challenges due to high dimension, missing and unobservable data, and population stratification. The author also discusses several important epidemiological concepts including confounding, effect mediation, interaction, and conditional association and their subtle differences, which I find to be particularly useful even for readers with statistical background. Chapter 3 introduces two important concepts: linkage disequilibrium and Hardy–Weinberg equilibrium and their applications to population-based association studies. Two methods, principal components analysis (PCA) and multidimensional scaling (MDS), are described to identify population structures. Several packages, including SNPasso and snpMatrix, that are designed to analyze SNP array data and genome wide association studies are mentioned but without any detailed discussion.

Several more advanced research topics are discussed through Chapter 4 to Chapter 7. Chapter 4 provides a nice summary of methods to correct multiple testing problems with a large number of genotypes. Bonferroni correction, false discovery rate, as well as resampling-based methods are discussed in details with several examples. Chapter 5 offers valuable materials for haplotype reconstruction from genotype data and haplotype–trait association. A detailed note for the expectation–maximization algorithm for haplotype reconstruction and the corresponding R codes are also supplied, which I find to be especially helpful because not many books present this type of contents. Chapter 6 and Chapter 7 introduce several statistical and computational methods to model the association between traits and a large number of genotypes. Chapter 6 introduces classification and regression tree methods and Chapter 7 describes several additional approaches including random forests, logic regression, adaptive regression splines, and Bayesian variable selection. All methods are accompanied with examples and R codes.

As the author has pointed out in the preface, the book aims to (1) provide a comprehensive coverage of genetic concepts as well as fundamental statistical theory and tools used in the analysis of population-based association studies; and (2) offer integrated examples using R to facilitate readers to gain deeper understanding of statistical concepts and tools presented in the book and apply these tools to their studies. The author achieves these two goals well. First, the use

of population-based association studies avoids several complications from family-based association studies, thus allowing the author to focus on fundamentally important genetic and statistical concepts and statistical tools for the analysis of association studies. Second, several data sets are described in detail in Chapter 1 and are used as examples throughout the book. Many R packages, such as coin (Chapter 2), genetics (Chapters 2 and 3), qvalue (Chapter 4), and haplo.stats (Chapter 5), are described. Readers can readily download the data sets and corresponding R codes from the companion website and play these examples to gain immediate experience. Actually, I find that the examples and R codes are extremely helpful and have used some of these packages in several collaborative projects. However, many other important research areas in statistical genetics, such as family-based association studies, population genetics, and microarray analysis, as acknowledged by the author, are not covered. Readers are encouraged to refer other books for the coverage of these areas (e.g., Thomas, 2004; Gentleman et al., 2005). I also find that some recent methodological advancements in the analysis of population-based association studies are not or only lightly covered. As an example, a more detailed discussion of population stratification and/or admixture mapping (Chapter 4) and dependent tests from adjacent SNPs can be particularly helpful for advanced readers.

In summary, this book provides a nice and fundamental introduction of statistical genetics problems to readers with limited experience in the analysis of data from population-based genetic association studies. Readers can particularly benefit from examples and corresponding R codes provided by this book.

References

Gentleman, R, Carey, V., Huber, W., Irizarry, R., and Dudoit S. (2005). *Bioinformatics and Computational Biology Solutions Using R and Biconductos*. New York: Springer.

Lananne, C. (2009). Book reviews. *Journal of Statistical Software* **31,** 1–5.

Nagarajan, R. (2010). Reviews of books and teaching materials. *The American Statistician* **64,** 268–268.

Sabatti, C. (2009). Book Reviews. *Journal of Statistical Software* **31,** 1–5.

Thomas, D.C. (2004). *Statistical Methods in Genetic Epidemiology*. New York: Oxford University Press.

Kui Zhang
Department of Biostatistics
University of Alabama at Birmingham
Birmingham, Alabama, USA