Book Reviews

Editor: Ananda Sen

Monte Carlo Methods and Stochastic Processes
Emmanuel Gobet
Chapman and Hall/CRC, 2016, 310 pages., £44.99, hardcover

Readership: Researchers in pure and applied probability, statistics and applied mathematics, who are interested in stochastic modelling.

This book, initially originating from course notes, covers an exciting area at the intersection of stochastic processes and simulation-based methods. For an applied researcher seeking to understand a dynamic real-life system in engineering, natural sciences or biosciences, stochastic processes serve as a fundamental modelling instrument. However, the utility of the model depends on our ability to operationalise it, to equip ourselves with tools that inform the model from available data, to provide predictions of the system's behaviour and to extrapolate them into scenarios not yet observed. Often the classical apparatus of mathematics, probability theory and statistics fails to advance towards that goal as analytic, and even numerical solutions are notoriously difficult to develop for stochastic processes beyond a relatively small set of special cases. This book introduces the reader to the methodology that can push this envelope.

The book starts with an overview of the history of the Monte-Carlo method followed by laying out a modern toolbox for stochastic simulations and general modelling of dependence used to describe multivariate vector-valued outcomes (Part A). Also included in this part are basic asymptotic results and variance reduction designs.

At the core of the method, is the observation that partial differential equations emerging in the stochastic process theory are similar to the ones that describe heat or fluid dynamics in the natural sciences. This is presented in Part B via Brownian motion and related processes defined through stochastic differential equations. Monte-Carlo methods developed for solving partial differential equations are then brought to bear on the study of stochastic processes.

Part C, the final part, is devoted to non-linear processes. It features backward stochastic differential equations and associated dynamic programming method, empirical regression and elements of approximation theory, and non-linear and interacting diffusion processes.

The book is self-contained and supplies key facts from a typical advanced probability theory and statistics course in a set of appendices. While maintaining mathematical rigour, the book skillfully avoids complex mathematical issues such as measurability that often make the theory of stochastic processes inaccessible to applied researchers. The style of the book walks the fine line, which makes it an invaluable resource for scientists in pure and applied probability, statistics and applied mathematics, who are interested in stochastic modelling. Examples mostly come out of physics, but the comprehensive exposition of the method makes the book really valuable to any stochastic modeller.

Alexander Tsodikov: *tsodikov@umich.edu*
Department of Biostatistics
School of Public Health

SPH II Room M4142
University of Michigan
1415 Washington Heights
Ann Arbor, MI 48109, USA

Nonlinear Modeling of Solar Radiation and Wind Speed Time Series
Luigi Fortuna, Giuseppe Nunnari and Silvia Nunnari
Springer, 2016, xv + 98 pages, $54.99, paper cover
ISBN: 978-3-319-38764-2

Readership: Upper level undergraduate students, graduate students, practitioners looking to augment their repertoire of methods.

This volume, which is part of Springer Briefs in Energy, reviews and demonstrates various methods to statistically analyse and build forecasting models from time series of wind speed and solar irradiance. Understanding the behaviour of such time series is important for harvesting the power from these resources. Utilities and independent system operators use forecasts of these variables, translated into power, to make important decisions on how to run their systems more economically and efficiently while maintaining reliable power for their customers. Thus, this book presents real useful examples.

The book covers some of the most used methods in time series analysis, demonstrating their implementation with real data derived from the National Renewable Energy Laboratory. It is unfortunate that the data are not sampled at a higher rate to be more relevant for the shortest range forecasting problems (less than an hour) where they might be quite useful. The authors demonstrate the various techniques using MATLAB and clearly point the reader to the correct MATLAB functions in order to perform similar analyses. The book concisely reviews quite a few of the appropriate methods for such analyses and for forecasting and clustering problems.

The first chapter introduces some general time series analysis methods, including linear de-trending, noise reduction, power spectrum analysis, autocorrelations, mutual information, computation of fractal dimension and Hurst Exponents, multifractal analysis, Lyopanov Spectrum analysis and various time series clustering methods. Chapter 2 demonstrates applications to solar radiation time series while Chapter 3 focuses on wind speed time series. Chapter 4 introduces several prediction models for solar and wind speed time series, emphasising embedding phase space methods, and demonstrates these methods in Chapters 5 (solar) and 6 (wind). Chapters 7 (solar) and 8 (wind) are devoted to demonstrating clustering methods, including transforming the raw variables into derived variables that better enable defining clusters in the new phase spaces. These chapters demonstrate the impact of choosing the number of clusters to use. These methods provide deeper insight into understanding the data.

The book includes many plots that demonstrate the techniques on real data and an analysis of the results shown in those plots. This is a great place to start thinking about how to apply statistical methods to wind and irradiance time series. Some practitioners also apply machine learning and artificial intelligence methods, of which only neural networks and neuro-fuzzy approaches are treated in this treatise.

Native English-speaking readers may find the inexact English usage and grammar, as well as the various typographical errors, a bit distracting. In one case, a figure of the spectrum plots in Chapter 2 on forecasting for Solar is labelled as Wind, leaving the reader to wonder whether it is just the label that is incorrect, or whether the wrong plot was inserted.

In summary, this book will be useful to mathematical, statistical or engineering students studying realistic examples of time series analysis and forecasting. To be most useful for the energy industry, however, the forecasts should include exogenous meteorological variables that will help improve the accuracy of the forecasts.

Sue Ellen Haupt: *haupt@ucar.edu*
National Center for Atmospheric Research
Boulder, CO 80301, USA

Modern Data Science with R
Benjamin S. Baumer, Daniel T. Kaplan and Nicholas J. Horton
Chapman & Hall/CRC, 2017, 556 pages, £51.19, hardcover
ISBN: 978-1-498-72448-7

Readership: Undergraduate and graduate students, data analysts and readers with some statistical background.

Modern Data Science with R illustrates how the statistical software R/RStudio can be used to conduct statistical programming on various forms of data. Its focus is on applying the statistical and computer science concepts to real-world problems. This book is suitable as a textbook for undergraduate students. It is also a useful reference for graduate students, data analysts and readers with some background in statistics, who may use this book to develop their data analytic skills in R in order to carry out data science projects.

Modern Data Science with R covers a wide range of topics, from technical data analysis to spatial data to the ethics of data presentation. The book begins with an introduction to data science, which covers data visualisation, good graphics practice, data manipulation and professional ethics in presenting data. The book then introduces statistics and modelling that covers statistical foundations, statistical learning and predictive analytics, unsupervised learning and simulation. Finally, the book presents specialised topics in data science, including more practical topics such as interactive data graphics, SQL, setting up databases, spatial data and network science. The detailed appendices present topics such as an introduction to R/RStudio and relevant packages therein, algorithmic thinking and regression modelling.

This book is unique because it incorporates theoretical fundamentals such as statistical learning and regression modelling with the modern, practical elements of data science, including setting up databases and debugging. There are a number of similar textbooks which use different software platforms, such as XLMiner and Python, especially those written by and/or for readers with a computer science background rather than a statistical background. There are also books which use R for readers who have familiarity with both statistical and computer science concepts. *Modern Data Science with R* is different from these books as it presents an abundance of R codes, functions and packages clearly with several useful examples. For people with a statistical background, the book covers computational topics like simulation and also includes appropriate computer science topics such as Data Wrangling, Database Querying using SQL and Text as Data. The book is well-structured and is presented in an easy-to-understand manner, making it suitable for a wide range of readers.

In addition, the book features a number of exercises for readers to practice and improve their skills. The real-world and practical examples presented throughout the textbook help readers develop a solid understanding of the concepts. The book's website offers some online exercises on data visualisation and comprehensive R resources for instructors for each chapter of the

textbook. Instructors and students will certainly find it helpful to look at these additional R examples in their courses. These resources are available at http://mdsr-book.github.io/.

In summary, *Modern Data Science with R* presents a variety of topics with several illustrative and engaging examples in R. This book is a valuable resource to all those studying and interested in data science.

Shuangzhe Liu: *Shuangzhe.Liu@canberra.edu.au*
Faculty of Science and Technology, University of Canberra
Canberra ACT 2601, Australia

Noninferiority Testing in Clinical Trials: Issues and Challenges
Tie-Hua Ng
Chapman & Hall/CRC, 2014, 208 pages, £56.99, hardcover
ISBN: 978-1-466-56149-6

Readership: Graduate students, clinical trialists and statistics researchers.

This book is intended for statisticians and non-statisticians working in the area of drug development, and it is written to be easily understood by a broad audience without any prior knowledge of non-inferiority testing. It covers important aspects of non-inferiority testing, including one-sided and two-sided hypotheses, equivalence/non-inferiority margins, fixed-margin and the synthesis methods, controversial issues with switching between superiority and non-inferiority testing, tests with binary, continuous and time-to-event endpoints, multiple historical studies and meta-analysis, regulatory guidances, and issues and challenges in these tests.

The book is very well structured. Chapter 1 introduces the hypotheses and equivalence/non-inferiority margin and defines notation which are used consistently throughout the book. Chapter 2 discusses the choice of such margins and offers a comprehensive list of references that deals with the same topic. Chapters 3 and 4 deal with non-inferiority tests with continuous, survival and binary endpoints. Chapter 5 describes two commonly used statistical approaches. The first six chapters have been covered by short course/tutorials at various conferences. Chapters 9–11 offer regulatory guidelines and illustrates with real examples some controversial issues and challenges related to these tests.

This book targets both statisticians and non-statisticians and will promote better communication between them.

Lili Zhao: *zhaolili@umich.edu*
Department of Biostatistics
University of Michigan
1415 Washington Heights
Ann Arbor, MI 48109, USA

Financial Analytics with R

Mark J. Bennet and Dirk L. Hugen
Cambridge University Press, 2016, 377 pages, €55.99, hardcover
ISBN: 978-1-107-15075-1

Readership: Undergraduate students in economics finance and statistics.

The book is a well-narrated introduction to mining of financial data for undergraduate students. In fact, the appropriate readership seems to be second or third year students in business schools, economics departments with a course or two on statistical methods under their sleeves. An exposure to some algorithmic computational methods would also be a desirable background of a student attending a course based on this book. Such a course would be also of interest for those undergraduate students in mathematics and computer science that are interested in seeing solution of economics and financial problems through 'big data' approach. Finally, the book could serve as a first guide to machine learning and data mining methods for a self-taught financial analyst-to-be.

The focus of the text is on presenting and developing modern computational tools for performing financial analytics in the form of 'a laptop laboratory for data sciences'. This so-called laptop laboratory is built based on the statistical language R and is supported by numerous add-in packages facilitating some sophisticated and problem-specific computational tools. Additionally, in the chapter on data exploration, the RSQLite Package that bridges the R and SQL database exploration software has been utilised. Data science is a relatively new field, and the book delivers a convincing account of practical financial analytics problems that can be approached from data mining perspective and through algorithmic analysis. The latter differentiates it from a typical financial statistics textbook. The hands-on data approach is promoted throughout the book with a considerable success. It is evident that the content was previously tested in a classroom and thus it should be relatively easy to use the book as a foundation for a course on financial and business analytics.

As opposed to other attempts in delivering data analytics to beginners, the book balances well between the elements of programming and the theoretical foundations that underpin the methodology without trivialising either of them. It starts with a gentle introduction to R that is followed by a chapter on fundamentals of financial statistics. In the appendices, the reader will find complementary material covering statistical theory and basic probability distributions. As a result, the book is, to a great extent, self-contained, allowing a reader who has gaps in assumed prerequisite knowledge (which, as mentioned above, ideally would include introduction to basic algorithmic programming and a first course in statistics) to fill them without a need for additional texts or supplementary material. Fourth chapter onwards, the book focuses on important topics for financial and business practitioners. Among them, we find models for financial securities, elements of risk analysis, time series analysis, investment efficiency, portfolio theory, predicting stock valuation, gauging the market sentiment, trading strategy, option pricing and some others. The accurate selection of data sets and examples motivates and illustrates a fairly wide range of methods that are essential for financial and business strategists. However, the text aiming at beginners must limit its methodological contents either in the range or in the depth. Seemingly, the authors have decided to follow the former path by giving a quite wide range of topics that are explained in a compact fashion and are well supported by algorithmic methods. In some parts of the text, the discussion of the code and algorithm dominates which may somewhat limit insight to analytical aspects of the problem at hand. Nevertheless, it provides important insight into how complex computational tools are built to address context

specific problems. This is an important part of education in analytical thinking with the use of software engineering and thus do not affect value of the book for a new researcher adept in the field of financial analytics.

From a general point of view, the material covered by book seems to be very adequate for the first course on big data and data mining methods for financial applications. The chosen topics are diverse and represent well the subject along with its practical implications. Each chapter is followed by a list of exercises, which occasionally appear to be somewhat short. Sometimes, the focus of exercises is more on some minor technical details in programs or methods rather than on the 'general' and universal aspects of the methodology. However, considering that the book discusses many examples and leads a reader through data analyses in a rather detailed way, this does not impact dramatically the usefulness of the book. Nevertheless, for an instructor using the book in a classroom, there will be a necessity to extend the lists of exercises. The book could also benefit from a list of projects that would test a student's acquired abilities of utilising the lessons learned in financial practice. There seem to be no internet resources for the book, which is slightly disappointing because it would give opportunity to update the book both through upgraded computational tools and relevant practical data sets. For example, a file with the R codes used throughout the book would save typing effort that currently is needed if a reader intends to access the computational tools. This is a simple example of enhancement that certainly would be appreciated by students if available on a dedicated website.

Overall, the book is a valuable addition to the growing introductory educational literature on data analytics methods in finance. It can be utilised as an in-classroom textbook and is highly recommended for this purpose.

Krzysztof Podgorski: *Krzysztof.Podgorski@stat.lu.se*
Department of Statistics
Lund University, Sweden

Observation & Experiment: An Introduction to Causal Inference
Paul R. Rosenbaum
Harvard University Press, 2017, 374 pages, £25.95, hardcover
ISBN: 978-0-674-97557-6

Readership: Graduate students and applied researchers in all fields dealing with observational data as well as randomised designs.

As the author states in his preface, the purpose of 'this book is to present the concepts of causal inference clearly, with reasonable precision, but with a minimum of technical material' (page viii). He fulfils his purpose by having most chapters (or groups of chapters) begin with an introduction to a commonly used research design followed by definitions of statistical terms necessary to analyse data using that design. He then gives a short understandable summary of a fairly simple case study which used that design. He follows up with a discussion of the conditions under which the research design can lead to valid causal inference, even though it is not a completely randomised experiment. Other chapters, also with clear examples, deal with important topics when discussing non-randomised research designs, such as bias, matching and sensitivity.

The style of the book is a bit unusual, but effective. Most of the main text of the book is written without formulas or how to implement them in R. These technical details are, however, given in 59 pages of footnotes at the end of the book, which also contain references for further study.

While the book is a very valuable contribution, I have a few quibbles. At times, the author uses idiosyncratic language. For example, he uses the word 'device' rather 'design' throughout the book and defines an 'instrument' as 'a random push toward receiving one treatment rather than another' (Page 258). Such language can be potentially confusing to readers uninitiated in the terminology of causal inference.

In addition, while implied throughout the book, more emphasis should be given to the necessity for researchers using non-randomised designs to make sure that they, and those reading the results of their research, clearly distinguish between causal and correlational results.

Further, since the book uses fairly simple examples, the reader may miss some of the big and important ideas related to certain topics. For example, in the chapter 'Between Observational Studies and Experiments', the author introduces the topic of Simpson's Paradox with an example that only records the group (treatment versus control), gender and a nominal outcome (alive versus dead). A more detailed example with three or more independent variables, such as the one given in Bickel, Hammel, & O'Connell (1975) (which the author mentions in a footnote) or the one in Novick, M.R. (1982) would have helped the reader understand this concept much better.

The book is accessible to undergraduates and master's degree students who have completed three or more statistics courses. It is highly recommended for advanced master's degree statistics students and teachers of secondary school statistics and should be a required reading for doctoral students in any area of theoretical or applied statistics.

References

Bickel, P. J., Hammel, E. A. & O'Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science*, **187**, 398–404.
Novick, M. R. (1982). Educational testing: Inferences in relevant subpopulations. *Educational Researcher*, **11**, 4–10.

Carol Joyce Blumberg: *cblumberg@gmail.com*
Winona State University Professor Emerita and
Retired from U.S. Energy Information Administration
1201 East-West Highway, Unit 422
Silver Spring, MD 20910-6293, USA

Errors, Blunders, and Lies: How to Tell the Difference
David S. Salsburg
ASA-CRC Series on Statistical Reasoning in Science and Society. Chapman & Hall/CRC, 2017, 154 pages, £77.00, hardcover
ISBN: 9781138726987

Readership: General public interested in science and statistics with basic understanding of high school level algebra.

Many people may consider statistics as complicated. Even if it may be true, it should be possible to explain many of the ideas in a way that is easily understood. That is what David Salsburg

aims to do in this book. The title of the book somehow reminds the old and annoying saying about lies, bigger lies, statistics, but fortunately this book offers something totally different.

In Chapter 1, Salsburg tells an interesting history related to determining the distance from the Earth to the Sun. In that story, there is random errors in measurement (it was difficult to make exact observations) blunders that were not inherent in the act of measuring (the measurers did not know that they had determined their geographical position erroneously) and lies that were due to fraud in claimed scientific activity (one of the measurers had been caught lying before). These three issues (errors, blunders and lies) will each then form a section in the book. There is also a summary at the end of each chapter that repeats the key points presented in a comprehensive way.

The errors section (Chapters 2–8) is actually an introduction to statistical modelling. Salsburg deals with probability and likelihood, measurement, multilinear models, correlation and causation, and big data. Everything is written very clearly incorporating interesting (historical) stories in appropriate places. Even though a few details feel like a slight oversimplification from a statistician's point of view, the decision to keep things straightforward is certainly justified and results in an excellent overview of the complex techniques and ideas.

Section considering blunders (Chapters 9–12) contains discussion about contaminated distributions and robustness where blunders are a nuisance and cases in which blunders may actually be useful. Again, the presented (historical) examples are easy to read and describe several interesting ideas.

The rest of the book (Chapters 13–17) deal with lies. There are examples on how to evaluate if data are falsified. The cases include data on the reigns of kings, whether Davy Crockett had written all 'his' books, how valid census counts presented in the bible appear to be, how to indirectly estimate how many tanks did Rommel have and how many representatives for statistical agencies fake the interviews they should have performed. We learn that faked data usually lack the variability of real data. These chapters offer interesting stories and anecdotes on using statistical techniques in a clever way. Perhaps a section for lies could have included more recent examples of fraud in science.

In conclusion, this book is very interesting and nice to read. The general idea that observation equals truth + error remains at the centrefold through the whole book and there are many smart examples on how to apply statistical techniques to solve various problems. Chapters are generally quite short, and each take a few minutes to read. I would suggest that a reader would read one chapter a time and then take some time to think about the discourse. That would maximise the joy from reading the book.

Reijo Sund: *reijo.sund@helsinki.fi*
Centre for Research Methods, Faculty of Social Sciences
University of Helsinki, P.O. Box 18, FI-00014, Finland
&
Kuopio Musculoskeletal Research Unit (KMRU), Institute of Clinical Medicine
University of Eastern Finland (UEF), Kuopio, Finland