Chapter 12: Statistical Methods for Classifying Mass Spectrometry Database Search. Proteomics is among the key modern high-throughput biological experimental techniques that are described in this chapter.

In summary, the 12 chapters are a well-written collection from multiple authors that I recommend for the intended audience. Several chapters include exercises (Chapters 3, 4, 5, 6, and 7).

Font changes, inconsistent reference style, and substandard plot quality could easily be remedied in a second version.

Tom BURR
Los Alamos National Laboratory

REFERENCES

Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Boca Raton, FL: Springer.

Venables, W., and Ripley, B. (1999), *Modern Applied Statistics With SPlus*, New York: Springer.

Wahba, G. (1999), "Support Vector Machines, Reproducing Kernel Hilbert Spaces, and the Randomized GACV," in *Advances in Kernel Methods*, eds. B. Scholkopg, C. Burges, and A. Smola, Cambridge, MA: MIT Press, pp. 69–88.

Probability Models for DNA Sequences Evolution (2nd ed.), by Richard DURRETT, New York: Springer, 2008, ISBN 978-0387781686, 432 pp., $84.95.

To review the second edition of *Probability Models for DNA Sequences Evolution*, we should first compare it with the first edition. The first edition in 2002 has five chapters and 240 pages. The second edition has nine chapters and 431 pages. This means that the second edition is a major addition to the population genetics literature. The bibliography of the second edition stretches from pages 399 to 426. About half of the references are subsequent to the year 2000.

The chapter titles are Basic Models, Estimation and Hypothesis Testing, Recombination, Population Combinations, Stepping Stone Model, Natural Selection, Diffusion Processes, Multidimensional Diffusions, and Genome Rearrangement. These give some idea of the topics, but the detailed section headings give a much better idea. For example, Chapter 7 (Diffusion) has sections Infinitesimal mean and variance, Examples of diffusions, Transition probabilities, Hitting probabilities, Stationary measures, Occupation times, Green's functions, Examples, Conditioned processes, Boundary behavior, Site frequency spectrum, and Fluctuating selection. For the other chapters, there is a listing of the sections at *www.math.cornell.edu/~durrett/Gbook/Gbook.html*.

The book is written at an advanced level. For mathematicians, the book introduces the necessary biology and genetics. For biologists, the book does assume some knowledge of applied stochastic processes and some mathematical maturity. Markov processes play a major role in later parts of the book This book would be appropriate for advanced graduate students in probability and mathematical statistics with interests in genetics. The author states that the goal of the book is to present "useful analytic results in population genetics, together with their proofs."

Richard Durrett enjoys writing and his style shows it. For example, on page 249 we read "As an antidote to the mathematical skullduggery, we will give some anecdotes ..." The quotations at the beginning of each chapter are as welcome as they are fun. There are lots of figures in the book which add to the readability. The literature review is excellent and it is intertwined with the text.

The topic of the book is important, as evidenced by the fact that a second edition has been issued. The author is clearly expert in his field. The book is up-to-date as evidenced by the references. The writing style is friendly and the presentation combines both the theorem/proof development with less formal discussions.

For statisticians or probability specialists looking for a way of moving into the exciting area of population genetics, this book is excellent. The book is filled with a large variety of models with expressions for probabilities and statistical moments, and with asymptotic results. For those using the first edition, it is time to upgrade. In conclusion, I highly recommend the book.

Myron HLYNKA
University of Windsor

Bayesian Reliability, by Michael S. HAMADA, Alyson G. WILSON, C. Shane REESE, and Harry F. MARTZ, New York: Springer, 2008, ISBN 978-0387779485, 436 pp., $89.95.

Reliability, defined colloquially as a property that a thing works when we want to use it or by ISO definition as the ability of an item to perform a required function under a given environment and operating condition and for a stated period of time, is the theme of this book. The authors give the reader a thorough statistical understanding of lifetime or failure time analysis of products or systems. They use different representations to specify properties of random variables as the probability density (the probability mass function), the reliability function (the survival function), the cumulative distribution function (the unreliability function), and the hazard function (the instantaneous failure rate function).

The statistical theme of this book is the application of Bayesian statistical methods to problems in reliability and lifetime analysis. The Markov chain Monte Carlo algorithms, namely Metropolis–Hastings and Gibbs sampler algorithms, were used to implement Bayesian analysis. These algorithms make computations feasible and straightforward. The computational code uses several free downloadable statistical software packages like WinBUGS, YADAS, and R. The authors pay attention in this chapter, entitled Advanced Bayesian Modeling and Computational Methods, to Bayesian goodness-of-fit testing, model diagnostics or model validation, reliability test design, and assurance test planning.

The first applications start in Chapter 4 dedicated to component-level data that, in later chapters, extends the model to system level (considered are models for independent and dependent component failures). Component reliability data consists in the book of sampling and prior distributions. For more complex problems with several components hierarchical models were used. The authors discuss diagnostics for addressing model fit and describe hierarchical models that facilitate the joint analyses of data collected from similar components. Subchapter 4.6 introduces the model selection methodology according to three criteria BIC (Bayesian information criterion), DIC (deviance information criterion), and AIC (Akaike information criterion). After establishing logical relationships between components and determining their functioning in a system, the authors develop probability models with dependent as well as independent components and multilevel data (in Chapter 5).

Chapter 6 considers the reliability of multiple-time-use systems that are repaired when they fail. The effectiveness of repairs varies from restoring a system to a brand new state to restoring it to the reliability just before the system last failed. Several models for failure count and failure time data collected on repairable systems allow for different degrees of repair effectiveness. The models considered include renewal processes, homogeneous and nonhomogenous Poisson processes, modulated power law processes, and a piecewise exponential model. Considered are also alternative models with real data examples with highly parallel supercomputers. This chapter also addresses quality-of-fit and evaluates current reliability and other performance criteria, which characterize the reliability of repairable systems.

Bayesian estimation methods for the standard regression models are considered in Chapter 7. The authors consider linear, nonlinear, logistic, and Poisson data for regression models. Presented are Bayesian methods for accelerated life testing models as well as reliability improvement experiments. The authors also discuss how covariates arise in accelerated life testing and in experiments to improve reliability. Under Models With Common Variables are presented the Arrhenius and Eyring models. A new statistical concept of Design of Experiments is introduced in Subchapter 7.8, Reliability Improvement Experiments. It details Taguchi's robust approach to designing as well as orthogonal polynomials for evaluation with covariate variables, factors, and interactions.

The subsequent chapter extends Bayesian methods to degradation data models. In addition to a general model for degradation data, models including both continuous and discrete covariates are considered. The authors compare reliability estimates based on degradation data with lifetime data. They also consider models for destructive degradation data as well as an alternative stochastic process-based degradation model.

Chapter 9 presents methods for collecting data for optimal design of reliability experiments. These designs attempt to allocate resources in the most efficient way to meet specified experimental goals. These goals usually involve the quality of the inferences that can be made using experimental data.

In the last chapter the authors apply obtained theory in design tests with the aim to assure that a reliability-related quantity exceeds a specified requirement with given confidence. The authors of the reviewed book have derived test plans of the Bayesian hierarchical models for binomial, Poisson, and Weibull sampling distributions.

This book brings to the reader a collection of modern Bayesian statistical methods for use in reliability and lifetime analysis for practitioners, but can serve also as a textbook for an undergraduate or graduate course in reliability. In the book are 70 illustrative examples with an additional 165 exercises downloadable from a reference website and are accompanied by a solution manual.

Adriana HORNÍKOVÁ
University of Economics in Bratislava

Applied Econometrics With R, by Christian KLIEBER and Achim ZEILEIS, New York: Springer, 2008, ISBN 978-0-387-77316-2, x + 221 pp., $54.95.

The authors have at least three main themes in this volume. One is that R, together with various packages, is sufficient for all the various statistical methods used in econometrics and has the additional advantage of being free and readily available software, including free access to the source codes. They note the availability of the software on multiple platforms and cite their own experience in that respect. A second point is the importance of reproducible research aided by various packages in R. This topic has recently received a lot attention; Leisch and Rossini (2003), Leisch (2005), Baiocchi (2007), Gentleman and Lang (2007), Peng and Eckel (2007), Peng (2007a, 2007b), Peng and Dominici (2008). A third is to demonstrate the use of a large number of R packages with data taken from various textbooks, journal articles, and government publications. The data is made available in a package called `AER`. Data from this package is used throughout the book to illustrate the use of various R packages. While not exactly a textbook in the usual sense, it does include exercises in every chapter except the first one. It is assumed that the reader will at least have had a first course in economics, but little if any prior experience with R. A good background in statistics, especially methods commonly used in econometrics, may not be essential, but certainly will be very helpful since there is little explanation of the statistical background. As a part of the emphasis on reproducible research the authors have made extensive use of the `sweave` package as well as the convenient link between R and *Latex*. Although all of the analyses in the book are reproducible from the `AER` package, the reader will want to actually run the various snippets of code that appear throughout since a good grasp of what is in the book depends on doing that. The thrust of the presentation is that the reader will learn best by actually using R and the various packages.

Introduction. This chapter provides a very quick overview and introduction to R including several elementary applications using linear regression as well as the `quantreg` and `kersmooth` packages. There is some general background information on R and its history. Data are extracted from the `AER` package for the examples.

Basic.s. This might have been titled "Much of what a new user needs to know about R." This includes the basics on handling vectors, matrices, and programming in R. There is an overview on importing and exporting data including the use of the `foreign` package. "Object orientation" is explained as in the different versions provided in R. Basic exploratory analysis in R is demonstrated as well as the use of the various graphics functions in R as well as tools for exporting graphics.

Linear Regression. Again it assumed that the reader will have some basic knowledge, i.e., basic knowledge of linear regression, but this moves quickly into more complicated forms, e.g., semiparametric models (called partially linear models), systems of regression equations (using the `plm` and `systemfit` packages), and regression with time series data (using the `dynlm` package). There is a short section on the analysis of *panel* data including the use of dynamic models and one on factors, interactions, and weights.

Diagnostics and Alternative Methods of Regression. Three different approaches are considered; *regression diagnostics*, i.e., jackknifing the data and comparing various statistics obtained from the full dataset versus those obtained by systematically omitting one data point, the second is diagnostic tests with alternative hypotheses such as heteroscedasticity, autocorrelation and misspecification of the functional form, the third is "robust" covariance matrix estimators. In the first the goal is to find data points that are not fitted well or have undue influence on the model fitting. The tools here are the standard ones found in most statistical packages. The second approach focuses on the tools in the `lmtest` package. For the third approach tools found in the `sandwich` package as well as some in the `MASS` package are used. The former is automatically loaded with the `AER` package whereas the latter is part of the base R code. In contrast to ordinary regression, which estimates the conditional mean, the `quantreg` package provides tools for estimating the median or one of the quartiles.

Models of Microeconometric.s. The third chapter considered ordinary linear models, this chapter extends those to more general distributions and also to count data. The authors note that many econometric packages also incorporate these, but require separate software packages as contrasted with R which subsumes them under the heading of generalized linear models. Examples are given using logit and probit regression then applications of the visualization and diagnostic tools in the `RCOCR` package are presented. The `AER` package not only provides the data used in this book, it also includes various functions, some of which are demonstrated in this chapter. The use of the package `pscl` for fitting Poisson, geometric, and negative binomial distributions in regressions with count data are shown here. The hurdle model has two parts; one is binary (is the "hurdle" crossed?) and the second a count part. This model is used to deal with a large number of zeros in the data. The `pscl` package has a function for this model. The `survival` package, well known in biostatistics, is used herein for modeling censored dependent variables. Other packages cited as useful for microeconometrics include `np`, `nnet`, `gam`, `lme4`, `mgcv`, `micEcon`, `mlogit`, `robustbase`, and `sampleSelection`.

Time Series. This chapter begins with a discussion of R data object formats. The most common being `data.frame`, but for time series there are at least two others that are more useful, `ts` and `zoo`. Examples of filtering, decomposition, and modeling are given using functions in the `stats`, `tseries`, or `forecast` packages. Microeconomics time series are often nonstationary. Examples are given to illustrate the difference between unit-root nonstationarity and deterministic trend nonstationarity using tools in the `tseries` package. This example also illustrates the possibility of two series having a common nonstationarity component. To test for cointegration, a function from the package `urca` is used. The methods discusssed in Chapter 3 for fitting dynamic regression models are extended in this chapter using functions from the `dynlm` package. In some time series, the model parameters are not constant with time. The `strucchange` package implements two general classes of tests to detect structural change or parameter instability; fluctuation tests and tests based on F tests. Other packages cited as being useful for time series analysis include `dse`, `FinTS`, `fracdiff`, `longmemo`, `mFilter`, `Rmetrics`, `tsdyn`, and `vars`. `Rmetrics` is actually a suite of some 20 packages for financial engineering and computational finance.

Programming Your Own Analysis. While R has some "point and click" features (also see the `RCommander` package), in general it uses a command line interface. From the perspective of reproducible research this is an advantage because this allows (1) constructing new functions built from those in existing packages, (2) producing a "history," and (3) documenting results in a word processor or even in a mark-up language such as HTML or *Latex*. However, R also has the *sweave()* function in the utils package that makes it easy to "weave" R code into a *Latex* document. The authors make extensive use of the latter in producing the book. To illustrate (1) three different typical tasks in econometrics are considered; simulation of power curves, bootstrapping a regression and maximizing a likelihood. In each example they go beyond on-the-shelf software. The simulation example involves generating data, setting the evaluation for one scenario, setting a loop to iterate over multiple scenarios, and then organizing the outputs into a suitable format. The bootstrapping example uses data from the `AER` package, but now the objective is to compute bootstrap standard errors and confidence intervals using case based sampling. Using functions from the boot package, a refit function is constructed to generate the regression coefficients and these are compared with those from *lm()* and finally the confidence intervals are compared. The likelihood example begins with a generalized Cobb–Douglas model which is nonlinear in the parameters. The log-likelihood function is given in analytic form, but must be maximized numerically, an iterative search is used with the *optim()* function. The last part of the Chapter presents an example of the use of the *sweave()* function and in particular how to export output from R into *Latex*, e.g., tables.