The Statistician (1999)
48, *Part* 4, *pp.* 607–626

Book reviews

Books for review

If you would like to review a book for *The Statistician*, and thereby to retain it for your collection, please contact the Book Reviews Editor, whose details can be found by clicking on 'books currently available' in the information about *The Statistician* on the Royal Statistical Society's Web site:

`http://www.rss.org.uk/journals.html`

Reviews in this issue

 0039–0526/99/48607

nomial with dependent coefficients is presented by N. Ranganathan and M. Sambandham.

G. Sankaranarayanan and S. Sivaraman turn their attention to the $M/M_b/1$ bulk queue with randomly varying arrival and service rates, whereas N. Krishnaji looks at the problem of predicting the sign of $X_t - X_{t-1}$, where X_t represents a time series in agricultural economics, assuming that the sequence has a two-step memory. The objective of M. Ghosh is to obtain the sharpest Berry–Esseen rate of convergence to normality, $O(n^{-1/2})$, for Studentized jackknife estimators. P. K. Sen deals with estimators for the strength of a bundle of parallel filaments in a simple random sampling scheme and the jackknife versions.

J. K. Ghosh and K. Subramanyam put forward a selection procedure based on maximum likelihood for choosing an appropriate model; they contrast the asymptotic theory when deciding between a Poisson and a geometric model with that of D. R. Cox in 1962. The estimation of drift of a diffusion process is described by B. R. Bhat and the estimation of the parameters for series systems with k out of m: G subsystems is considered by A. P. Basu and N. Ebrahimi. Finally, G. J. Babu gives a short overview of literature on the bootstrap.

The book gives an interesting insight into the directions of research pursued by these probabilists. It is doubly unfortunate, both for the authors and for the reader, that the work has taken so long to put into print (the most recent references in the book are dated 1989). The editors greatly regret the unforeseen delay.

Adrienne (Freda) Kemp
University of St Andrews

Statistical Regression with Measurement Error
C.-L. CHENG AND J. W. VAN NESS, 1999
London, Arnold
262 pp., £35.00
ISBN 0 340 61451 7

This hardback book forms part of 'Kendall's library of statistics', which is part of a radical expansion of Kendall's *Advanced Theory of Statistics*, including a series of monographs covering some specialized statistical topics. This particular monograph in the series is intended to provide a comprehensive coverage of measurement error (ME) models.

Three basic types of ME models are discussed; the structural model, the functional model and the ultrastructural model. Topics include model identifiability, parameter estimation, confidence intervals, asymptotic theory, finite sample properties, modified least squares, non-normal models, robust methods, prediction and statistical calibration.

The authors state that their goal was to provide 'a comprehensive coverage of the subject that emphasizes the ideas and the practical implementation of the theory without too great an emphasis on the theorem–proof format'. This, I think, they have achieved.

The book is intended to be accessible to readers with a year's course in probability and statistics, but at quite an advanced level.

The book has eight chapters and one appendix. Each chapter concludes with a brief discussion of the material, some historical background and some comments relating the results described to other material in the literature. There are also exercises for the reader to complete, some of which are described as research (no solutions). The introductory text makes it clear that the authors feel that perhaps, too often, we ignore the true situation of errors in both variables, preferring to use ordinary regression. I certainly benefited from being made more aware of the situations under which I could safely ignore errors in both variables.

Chapter 1 provides an 'Introduction to linear measurement error models', with sections on identifiability (including conditions for identifiability) and maximum likelihood estimation, and then sections on three main model types: ME models with correlated errors, the equation error model and the Berkson model.

Chapter 2 deals with 'Properties of estimates and predictors', including asymptotic properties of the model parameter estimates, finite sample properties, implications regarding confidence regions, prediction and calibration under ME models. I found the section on confidence intervals fascinating and worrying. In such a 'simple' situation of introducing errors on both variables, we find that our well-trusted asymptotic confidence intervals are problematical.

Chapter 3 compares model assumptions and modifying least squares. The introduction talks of the issues facing users of ME models, some of which are all too familiar in any model choice setting.

Chapter 4 presents alternative approaches to the ME model, including instrumental variables, grouping methods and methods based on ranks and higher order moments.

Chapter 5 presents the linear ME model with vector explanatory variables (or in other words multiple regression), including the equation error

model, maximum likelihood and alternative approaches to estimating parameters.

Chapter 6 presents a development to polynomial ME models; extensions to general non-linear relationships are not quite so straightforward, and indeed some questions remain unanswered.

Chapter 7 describes robust estimation in ME models, including robust orthogonal regression and computational methods. The usual ME model estimates are highly non-robust (to outliers or estimation assumptions), so this is an important area. The final chapter is a catch-all, including estimation of the true variables, obtaining identifiability assumption information, and relationships to other latent variables models including factor analysis. The final section is about terminology, of which several variants exist for this area alone.

All told, this is a useful book and one well worth dipping into for those concerned with regression modelling.

Marian Scott
Glasgow University

Regression Graphics: Ideas for Studying Regressions through Graphics
R. D. Cook, 1998
Chichester, Wiley
xviii + 350 pp., £65.00
ISBN 0 471 19365 8

This book is in Wiley's 'Probability and statistics series' and is about ideas for the graphical analysis of regression data. The motivation for the work comes from Cook's wondering whether it is possible to conduct a regression analysis by using just graphics. A strong differentiation between this book and most others that provide some coverage of the subject is that the discussion here focuses solely on graphics unless non-graphical methods are essential for progress. However, a major element of this work is the provision of a strong theoretical basis for the graphical methods suggested. As a result, unless one is familiar with the techniques used, it is not easy to read.

The first two chapters provide a relatively light introduction to graphical issues and introduces some of the ideas that are explored later. Since the book concentrates almost exclusively on regression problems, the intent is to study existing and new graphical methods which facilitate the understanding of the conditional distribution of $y|x$. A central theme of the book is based on finding a simplified version of $y|x$ by reducing the algebraic dimension of the predictor vector

(a) without losing information on the response and
(b) including minimal assumptions on the nature of the conditional distribution $y|x$.

A key concept in the search for simplification is that of the *dimension reduction subspace*. Once found, a *sufficient summary plot* can be produced, i.e. a plot of y *versus* a lower dimensional vector of predictors. These ideas are introduced in Chapter 4, where three-dimensional scatterplots are used to analyse problems with two predictors and a many valued response. Chapter 5 continues these ideas for regression with binary data. The next four chapters expand these ideas by allowing for many predictors. These ideas are expressed as propositions with justifications; a knowledge of mathematical statistics and finite dimensional vector spaces is required for the justifications.

As well as discussing numerical methods for estimating a central dimension subspace, the latter chapters cover inverse regression, ideas for studying the effects of individual predictors, methods for visualizing predictor transformations and graphics for model assessment.

Cook acknowledges that colour can facilitate the interpretation of graphical displays and the lack of colour in the book does detract from some of the plots. However, colour versions of these plots, as well as other information, are available via `http://www.stat.umn.edu/RegGraph/`.

I still need to find time to understand more fully some of the methods that Cook is proposing. Clearly there are other techniques, such as principal component analysis and partial least squares, which can also be used to obtain a dimension reduction subspace. No reference to these traditional multivariate techniques was made, and so it is difficult to evaluate the merits of the various approaches.

I strongly believe that more widespread and innovative use of data visualization methods is the key to the future of applied statistics, and applied statisticians. Unfortunately the ideas and methods such as those discussed by Chambers *et al.* (1983) are still not given adequate coverage in the statistical education of scientists, engineers or statisticians. There are still too many who see graphics as a relatively unimportant 'making the results look pretty' exercise. Cook's book attempts to put some rigour into the use of graphical methods in the regression situation. As a result it is not ideally suited to the applied statistician looking for guidance on how to use these methods. However, since this book can only bring more attention to