

---

## Book reviews

---

Eidhammer I, Janassen I and Taylor WR 2004: *Protein Bioinformatics: an algorithmic approach to sequence and structure analysis*. Chichester, Wiley. 355 pp. \$85 (HB), ISBN 0-470-84839-1

*Protein Bioinformatics* is based on Informatics courses taught at the University of Bergen, and is intended as a textbook and reference book for molecular biologists and computer scientists who wish to have a greater understanding of the programs and algorithms 'that are used to tackle the central problems in molecular biology'. It has three parts: (1) sequence analysis, (2) structure analysis and (3) sequence–structure analysis. Each section consists of a short explanation of the method followed by a series of detailed algorithms and examples for each type of analysis. The first section includes a short chapter on statistical methods.

As a statistician and computer scientist with an enthusiasm for dynamic programming, I was hoping that this book would increase my very rudimentary understanding and knowledge of bioinformatics. I was disappointed. My main criticism of this book is that the basic principles and motivations for using the various methods are mostly very brief and not properly explained. For example, a single line is devoted to explaining and defining dynamic programming, an approach used in most of the algorithms in this book, with no reference to Richard Bellman who was the originator of dynamic programming and who wrote several excellent books on the subject. The section on statistical methods was difficult to understand and since the explanations of the methods are so brief, it is necessary to invest a large amount of time learning the notation and working through each algorithm in turn in order to understand what is going on.

I would not recommend this book for a more general reader who wished to understand more about bioinformatics. Nevertheless, this book would probably be a useful text for those studying on a course on bioinformatics algorithms or those engaged in day-to-day bioinformatics research who require a book of

'recipes' and already know what they want to achieve.

*Reviewed by Rosemary Tate, Institute of Psychiatry, London, UK.*

Skrondal A and Rabe-Hesketh S 2004: *Generalized latent variable modelling: multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC. 508 pp £50 (HB), ISBN 1-58488-000-7

Latent variables are a common feature in statistical models representing, for example, measurement error, hypothetical constructs, unobserved heterogeneity and missing data. This book is a consolidation of the authors' research over several years into unifying and extending latent variable modelling techniques, leading to the development of the generalized linear latent and mixed models (GLLAMM) framework and associated software. This framework subsumes multilevel, factor and structural equation models for mixtures of categorical responses, counts, durations and comparative responses (rankings).

The book is organized in two parts: Methodology (I) and Applications (II). The principal advantage of this split for readers is that those primarily interested by the methods are not distracted by lengthy examples, while the applications go beyond simple illustrations of the methods to more detailed discussion of the substantive question, and issues concerning model specification, fit and interpretation. Conversely, applied readers may benefit from having descriptions of methods interspersed with examples to help them reinforce their understanding of the basic models before moving to generalizations. Some sections in Part I are technically demanding to the extent that it would have been helpful to have included some brief examples along with more cross-referencing to relevant applications in Part II. This would also make the book more accessible to readers whose first discipline is not statistics.

One valuable aim of the book is to synthesize the approaches adopted in different disciplines, including biostatistics, psychometrics and econometrics. Historically, there has been a lack of cross-disciplinary communication, which has led to some cases of methods being developed independently. Consequently, until now it has been difficult for researchers to read research from other disciplines, and recognize crossovers because of terminological differences. This was regrettable because, although statisticians working in different disciplines are faced with ostensibly different challenges, the models developed to address them often have a similar structure. The book succeeds very well in drawing out these similarities and presenting them in a unified framework with a common terminology. For instance, there is a detailed comparison of random coefficient models and factor models in Chapter 4. The authors demonstrate how what appear to be two quite different models with different applications and interpretations have essentially the same form. The GLLAMM framework includes both models as special cases, as well as syntheses of the two (e.g., multilevel factor analysis) and extensions to different response types.

Part II (Applications) includes a range of well-chosen cross-disciplinary examples. The analyses presented are thorough, with comparisons of different approaches applied to the same data set, and careful interpretation and assessment of model fit. It would have been useful, however, to have included more on model identification. The issue is of great importance to applied research but some readers will find the technical discussion of identification in Part I (Methodology) difficult to follow, and would have benefited from further examples,

particularly on the implementation of empirical methods.

One small niggle is that the authors do not state the book's intended audience or give any prerequisites. A solid grounding in probability theory and regression modelling is assumed, although some of the more technical sections of Part I can be skipped. In order to present the generalized framework for latent variable modelling of different response types, the book's scope is necessarily broad and inevitably some topics are covered very briefly. For example, Chapters 2 and 3 (on modelling different response processes and classical latent variable models) are really refreshers rather than introductions for the uninitiated. I would suggest that, to get the most out of the book, readers without at least some familiarity with multilevel modelling and factor analysis would be advised to read more introductory and in-depth texts on these techniques first; some suggestions for further reading are given at the end of each chapter. Researchers with more applied interests may prefer to start with the authors' companion text (*Multilevel and Longitudinal Modeling using Stata*, 2005, Stata Press).

Overall, the book is an extremely useful resource for statisticians working in medical and biological sciences and social sciences such as economics and psychology. Most statisticians apply some form of latent variable modelling in their research, and this book presents the latest developments in the field in a clear and engaging way.

*Reviewed by Fiona Steele, Centre for Multilevel Modelling, University of Bristol, UK.*