Monitor

Book Reviews

Chemometrics: from basics to wavelet transform.
Foo-tim Chau, Yi-zeng Liang, Junbin Gao, Xue-guang Shao

John Wiley and Sons, Hoboken, 2004
316 pp, ISBN 0-471-20242-8
Hardback GBP 58.95

This book is by a team of well-respected Chinese chemometricians and illustrates the present dynamic state of the subject in Asia. Reading the subtitle "from basics to wavelet transform", I wondered if they had tried to offer an A to Z but had not quite come up with the extremes (from ANOVA to Z-scores perhaps?). However, I recall that the authors are known for their extensive work on wavelets and the subject matter of the book is heavily weighted to this technique. In fact, I would have thought that a more descriptive title would have emphasized signal processing and wavelets, and not suggested a wider rendering of the subject. Chemometrics has evolved over the years into a widespread, even fuzzy, compendium of interacting tools and methods that I believe no single volume can now do justice.

The book opens with an introductory chapter on the place of chemometrics in modern analytical chemistry. This is brief, but contains interesting statistics on the popularity of chemometrics and data analysis books. Also described is the 'white/black/gray' classification of Liang, Kvalheim and Mann for multicomponent systems. Signal processing and wavelets are then introduced. The chapter concludes with a very extensive bibliography, reference list, online resources and mathematics software. This is an excellent and most useful list of resources.

One-dimensional signal processing is the subject of chapter two, with good descriptions of various filters (Savitsky-Golay, Kalman and splines), together with full MATLAB source code for Savitky-Golay and spline. (It would have been even more useful to include these more extensively throughout the book, although many are available on the publisher's website or from the authors). There follows transformations, differentiation and compression algorithms. Chapter three extends the concepts to two-dimensional signal processing, focusing on principal-component analysis and factor analysis methods. This is the strength of Professor Liang, who has developed heuristic evolving latent projections (HELP) with Olaf Kvalheim.

We are now 100 pages into the book, and the remaining text is given to a description of the wavelet transform in two chapters, an introduction to the fundamental mathematical treatment followed by applications in chemistry. Wavelet templates are introduced in relation to the more familiar sine and cosines of the Fourier transform, and their particular localized properties are emphasized. The Haar wavelet is used to describe the principles of the method, and Meyer, B-Spline and Daubechies wavelets are then given as examples. The fast wavelet transform and inverse are described along with the Biorthogonal wavelet transform and the two-dimensional wavelet transform.

Examples in chemistry may be found easily, the subject abounds with signals varying with time, wavelength, mass and so forth. Data compression was of importance when the limitations of instrumental computer systems led to difficulties in managing the burgeoning amount of data. However, the problem was temporarily solved with modern gigabyte computer memories, only to be overtaken once again by multiple-dimension methods which produce even greater demands on storage and processing. Fast wavelet transforms can be an effective solution to data compression, as described with an example of NMR. As with the FFT,

doi:10.1016/j.chemolab.2004.08.002

wavelets may also be used to denoise and smooth data, correct for baselines and remove background signals and to enhance noisy signals. Numerous and well-referenced examples in analytical chemistry include flow injection analysis, many spectroscopies, chromatography and electrochemistry. Less familiar may be uses in quantum mechanics and chemical physics.
The text concludes with appendices covering basic matrix algebra and the manipulation of matrices in MATLAB.

Overall, this is a well-written and clearly explained text on signal processing with particular focus on wavelets.

D. Brynn Hibbert
School of Chemical Sciences, University of New South Wales Sydney, Australia

Introduction to Bayesian statistics
William M. Bolstad

John Wiley and Sons, Hoboken, 2004, 354 pp, ISBN 0-471-27020-2
Hardback UKL 50.50

Given that there are, arguably, two primary approaches to statistics, frequentist (or classical) and Bayesian, most of the textbooks are written by frequentist statisticians. However, more and more people have realized the advantages of using Bayesian statistics and they have appeared increasingly in forensics and other areas of analysis. Bolstad's book focuses on Bayesian statistics as a textbook for undergraduate students as well as useful material for anyone with an interest in the subject.

Compared to frequentist statistics, the Bayesian approach provides a natural way to predict the probability of a coming event (the posterior probability) based on all the knowledge the investigator has. Frequentists only use information observed from the sample itself and do not believe that the probability that the next event will happen can be predicted. In real life, when we look at anything that we try to identify, we not only observe the object but also use the knowledge we already have which we believe can help the identification. The quality of the identification will be improved when we make a more careful observation and have more prior information about the object. Bayes' theorem tells us how to use the observation to update prior knowledge about the object. Because Bayesian inference uses prior knowledge, it gives a more complete result than the classical method. The Bayesian approach to an inference naturally considers the population parameters to be random variables and does not need to consider a hypothetical infinite number of samples. These ideas are what the author wishes to teach.

As a general statistics textbook, some essential statistical knowledge and basic probability theories are organized well in different chapters. Different sampling methods and randomized designs are introduced in Chapter 2. Statistical descriptions and graphical displays of data are given in Chapter 3. In order to help students to understand random variables, two separate chapters—Chapters 5 and 7—introduce distributions for describing discrete and continuous random variables. The calculations of expected value, variance and mean of discrete or continuous random variables are included in these two chapters. The distributions introduced in these chapters are related to choosing the prior probability of the hypothesis (so-called 'prior') which is a major step in Bayesian inference. These topics are not distinct from what may be calculated using classical statistics.

Bayes' theorem is introduced after introduction of basic probability rules in Chapter 4. Definitions of conditional and marginal probabilities, which are important to Bayes' theorem, can be found in this chapter. Applications of Bayes' theorem are covered in Chapters 6 and 8 to 14. Chapter 6 uses a discrete random variable with a discrete prior and shows the basic steps to perform a Bayesian inference. Using different distributions as priors to work out the posterior distributions under different circumstances are discussed in Chapters 8 and 10. It is interesting that the author uses Chapters 9 and 11 to compare the differences between Bayesian and frequentist inferences for proportions and means. These will help readers to understand the advantages of using Bayesian inference. Besides basic parameter inference, Bayesian inference for simple linear regression is given in Chapter 13.

A big argument against using Bayesian methods is the difficulty of choosing a suitable prior. Chapter 14 gives a short discussion of what a good prior should be and in what situations the inferences will be wrong. Mixtures of conjugate priors can be used to prevent a misspecification of the prior leading to an incorrect posterior. In this case, the posterior will be close to the likelihood function. Although Bayesian methods are criticised for reliance on prior probabilities, this approach recognizes that when the prior and likelihood are in conflict, the likelihood should take precedent because it is based on the most recent data.

The book also has five appendixes which give extra reading material when it is needed. As a textbook, each chapter has plenty of exercises and the answers of odd numbered exercises are given in Appendix E. Also provided are useful macros for Minitab users and R functions for R users. The usage of these macros and functions can be found in Appendixes C and D. Because R

can be downloaded without charge, it is more suitable to people who do not have Minitab.

There are only 15 references listed at the end of the book, but this is a textbook and does attempt to completely cover the subject within its pages. The lack of explicitly chemical examples gives a little regret to a chemist, but this does not prevent the book for being a good primary Bayesian statistics textbook that can be used by all scientists. The examples with detailed explanation of the calculations will help the reader to understand Bayesian inference and then be able to do her/his own calculations.

Jianfeng Li
School of Chemistry, University of New South Wales Sydney, Australia