SIAM Review
Issue 47, Vol. 4

chosen for the purpose, covered in logical order, and treated in some generality. There are many exercises; they tend to be the fill-in-the-details or the explore-related-topics types, though a sizeable minority are more specific. On this side of the Atlantic the treatment seems rather pure and abstract, but it is in keeping with a Continental tradition that places topics such as approximation in a general context and that views a solid grounding in rigorous functional analysis as indispensable.

A price is paid for all the generality and thoroughness: after 400 pages, one needs to go to the second and third volumes (not yet translated?) to find topics from integration on. It is not clear to me how these books would fit the usual curriculum in the United States. They might be an excellent choice for supplementary reading for very strong students in a standard course, or as the basis for a seminar that followed a standard course. They could also serve as a background reference for any user of basic analysis who wishes to have at hand an overall picture of the rigorous logical development of the subject.

RICHARD BEALS
Yale University

Measure Theory and Filtering: Introduction and Applications. *By Lakhdar Aggoun and Robert Elliott.* Cambridge University Press, Cambridge, UK, 2004. $75.00. x+258 pp., hardcover. ISBN 0-521-83803-7.

Measure theory has long been considered as an abstract setup, of interest mainly to pure mathematicians. Applied mathematicians could live with probability distributions of random variables, which, although clearly related, appear to be a more concrete mathematical tool. However, measure theory is not just a mathematical abstraction to provide a rigorous framework for probability theory. It has appeared, in the development of research, as a powerful tool to obtain significant results in applied fields, like control, filtering, and more recently finance and others. Today, Girsanov's theorem is a standard tool for engineers and economists. More widely, the technique of change of measures is extremely useful in applications. The concept of risk-neutral probability in finance, which has permitted so many developments, relies on these concepts and methods. However, in spite of this broad spectrum of applications, measure theory has remained confined to abstract probability books. This is why the book of Aggoun and Elliott is extremely timely in filling a gap between the users of applications of measure theory and the theoreticians. The main interest of this book is to combine the mathematical developments with numerous significant applications. So it will be a reference for those who want to use the potential of measure theory in concrete problems. The numerous applications treated constitute a framework for other applications and will certainly suggest further developments. They are diversified and well presented. They concern statistical problems, finance, genetics, and population modeling among the main ones. The book covers discrete as well as continuous time models, although the main emphasis is on discrete time models, which play the major role in applications and which are easier to study for practitioners. This book will be particularly useful as a textbook in applied mathematics, control, statistics, economics, and engineering. It will be quite interesting for research-oriented people, eager to use these techniques in various situations. It is well written and self-contained.

I am convinced that it will raise a lot of interest and remain a reference for a long time to come.

ALAIN BENSOUSSAN
University of Texas at Dallas

Bayesian Nonparametrics via Neural Networks. *By Herbert H. K. Lee.* SIAM, Philadelphia, 2004. $42.00. x+96 pp., softcover. ASA-SIAM Series on Statistics and Applied Probability. Vol. 13. ISBN 0-89871-563-6.

Professor Lee states that his goal in writing this book is "to put neural network models firmly into a statistical framework, treating them with the accompanying rigor normally accorded to statistical models."

Neural networks have been developed mainly by the machine learning community, as an algorithm for predictions. Statisticians have generally been dismissive of neural networks, considering them to be black boxes not based on a probability model. Lee hopes this book will shift the focus back to modeling.

In Chapter 2, Lee surveys nonparametric regression models, including local models and basis function models, and shows the relationships between them. Local methods in one dimension include simple moving window averages and locally weighed scatterplot smoothers (LOWESS). Spline smoothers partition the axis and fit low order polynomials on each region that are required to be continuous at the knots which are the boundary points of the regions. Extension of these methods to higher dimensions can be problematic. Generalized additive models (GAM) are an extension to multivariate splines that use separate univariate spline smoothers for each dimension. If the axes are rotated to improve the spline fit, it is called projection pursuit regression (PPR). Partitioning the dimensions separately into tree-based regions is the basis for classification and regression trees (CART). Basis methods use an infinite set of basis functions such as a set of functions that span the space such as polynomials, trigonometric functions, or wavelets. Any function can be approximated closely as a linear sum of the basis functions. The coefficients are more interpretable when the basis functions are orthogonal. CART can be considered to be a basis function model using step functions as the basis.

The equations of neural network regression models correspond to a linear sum of k logistic functions. The parameter of each logistic is a linear combination of the covariates. Thus they constitute a nonparametric regression method using logistic basis functions. This means an error model can be associated with a neural network. Parameter estimates and their statistical properties can be determined from the error model. The modeling perspective offers a systematic way for fitting neural networks. Lee takes the Bayesian approach to statistics, which views the parameters of the model as random variables. The prior distribution measures the subjective belief in the possible parameter values before viewing the data. The posterior distribution of the parameters combines information from their prior distribution and from the data by using Bayes' theorem.

In Chapter 3 Lee discusses the issues in estimating the parameters of a neural network model. First the parameter may lack interpretability, so it is hard to determine a reasonable prior belief. Lee shows situations where this lack of interpretability occurs. The rest of this chapter looks at methods for determining the prior. He shows that the use of hierarchical models for the parameters is the most satisfactory solution. Vague but proper priors for the hyperparameters allow the posterior to be mainly influenced by the data. He shows that noninformative improper priors instead of hierarchical models for neural networks lead to improper posteriors. Restricting the parameter space will give a proper posterior. Priors developed by Jeffrey's rule and independent Jeffreys priors also are improper and lead to improper posteriors. He gives a brief summary of the Markov chain Monte Carlo methods for drawing a sample from the posterior distribution. Finally, he shows that Bayesian estimators for neural networks are asymptotically consistent, either by allowing the number of parameters to increase to infinity at a slower rate than the number of observations, or by including the number of parameters as another parameter and putting a prior on it.

Chapter 4 explores the problem of choosing a suitable model from a set of possible models, either for improving prediction or for its own sake. In the neural net context this means finding the covariates to be included and the number of hidden nodes. Bayes factors for model comparison could be used; however, when improper priors are used the probability of the data given the model requires that the normalizing constant for each model be known, and this usually is an intractable integral. The Bayesian information criterion (BIC) is seen to be an easily applied approximation to the Bayes factors which avoids this difficulty. Model averaging is another approach for improving predictions. The space of possible models gets explosively large as the number of covariates and hidden nodes increases. Prediction models are weighted by their posterior probabilities. Bagging (bootstrap aggregating) is a frequentist model averaging tool which can be brought into the Bayesian framework by using the Bayesian bootstrap. Various algorithms for exploring the model space are presented, including greedy algorithms and stochastic algorithms such as Markov chain Monte Carlo model composition (MC3), Bayesian random searching (BARS), and reversible jump Markov chain Monte Carlo (RJMCMC).

The various methods for modeling using neural networks are trialed and compared using the "ozone data set" for regression and the "loans data set" for classification that are both well known in the nonparametric literature. Lee has succeeded in showing that Bayesian neural network models are a valuable tool in the statistical toolbox for performing nonparametric regression. They are flexible and can model reasonably well behaved functions and can successfully capture high-dimensional effects. However, they are complex and the parameters lack interpretability, which makes it difficult to choose priors. This book will be a useful addition to both the statistical and machine learning communities. Statisticians will see that neural nets offer an effective method of nonparametric regression, while the machine learning community will find that statistical modeling ideas can lead to improved predictions.

WILLIAM M. (BILL) BOLSTAD
University of Waikato, New Zealand

Integer Partitions. *By George E. Andrews and Kimmo Eriksson.* Cambridge University Press, Cambridge, UK, 2004. $24.99. x+141 pp., softcover. ISBN 0-521-60090-1.

Partition theory has made it into popular culture. In episode 12 of *Numb3rs*, Charlie's student Amita Ramanujan talks with Charlie about her doctoral dissertation in combinatorics in front of a board on which the Rogers–Ramanujan identities are written. What do you do with an undergraduate who's had some taste of discrete mathematics and would like to learn about these identities? Send them to this book.

Andrews and Eriksson have written a beautiful little book for an undergraduate seminar, an independent project, or that really talented high school student who would like to get a taste of what current research in combinatorics looks like. In the first chapter it introduces students to the basic problem, How many ways can an integer be written as an unordered sum of positive integers? The book proceeds patiently and carefully with clear summaries and many exercises, including exploratory exercises. The style is informal and engaging.

The Rogers–Ramanujan identities form a recurrent theme. They state that the number of partitions of an integer into parts that differ by at least 2 (or that differ by at least 2 and have no parts of size 1) is the same as the number of partitions into parts congruent to ± 1 modulo 5 (respectively, ± 2 modulo 5). The first two-thirds of the book leads students through simple bijective proofs and the use of generating functions, building to Robin Chapman's recent and elegantly simple proof of these identities. A familiarity with infinite series is all that is needed to follow this proof. The initial chapters are also sprinkled with interesting information. There is a reference to Barry McCoy's talk at the International Congress of Mathematicians on applications of the Rogers–Ramanujan identities to physics, a detailed derivation of explicit formulas for the number of partitions into one, two, three, or four parts, and a proof that $\lim_{n\to\infty} p(n)^{1/n} = 1$.

But it is in the final four chapters where this book really shines, because here the authors introduce a selection of results in partition theory culled from the past decade. Chapter 9 presents the second author's lecture hall partition theorem, which is deceptively simple to state. A lecture hall partition with N parts allows parts of size 0, and the parts satisfy the inequalities

$$\frac{\lambda_1}{1} \le \frac{\lambda_2}{2} \le \cdots \le \frac{\lambda_N}{N}.$$

In other words, the line of sight from $(0,0)$ to (j, λ_j) passes through or above the points (i, λ_i) for all $1 \le i < j$. The theorem, proven by Eriksson and Bousquet-Mélou, is