dard probability models and robust regression. Each of the separate chapters on regression, hierarchical linear models, generalized linear models, multivariate models, and mixture models contains a good discussion of the Bayesian approach to the specific model, with at least one example worked in detail.

Discussion of specific models in Carlin and Louis is condensed into Chapter 7 along with the problem of ensemble estimates. In hierarchical modeling, the scientific question of interest sometimes revolves around the ensemble histogram of the estimates for the individual units or possibly a ranking of the units rather than inference on the individual units. For example, in public health it may be of more interest to identify the 10 states with the highest infant mortality rates than to estimate the infant mortality rates in each state. The discussion of ensemble estimates is based on using loss functions to achieve the desired goal. It is an extension of the chapter on empirical Bayes methods and picks up the threads of decision theory from Chapter 1. The section makes use of the constrained empirical Bayes approach presented in the earlier chapter. The ideas related to ensemble estimates are not presented in Gelman et al.

The three case studies in Chapter 8 of Carlin and Louis are virtuoso, state-of-the-art applications of advanced Bayesian methods. The first is an analysis of longitudinal AIDS data. The second is a robust analysis of a clinical trial that illustrates the role of the prior elicitation process. The third illustrates spatio-temporal mapping of lung cancer rates. Because the data are not included, the examples cannot be replicated as pedagogical exercises. Carlin and Louis's appendix on currently available Bayesian software, could become, as they admit, outdated very rapidly. Hopefully, current discussion to integrate more Bayesian methods into Splus will come to fruition.

Very special in the Gelman et al. book is the presentation of the potential outcomes approach to causal inference, the role of study design in Bayesian inference, and the missing-data chapters. These ideas are crucial to the interpretation of analyses that we do every day. Statisticians can easily get caught up in estimation and computation without stopping to think about the meaning of what they are estimating. It is admirable that Gelman et al. chose to include this material in their book. In addition to presenting an overview of the generality of the observed- and missing-data paradigm, Chapter 7 challenges the standard Bayesian notion that the method of data collection is irrelevant to Bayesian analysis and discusses the role of randomization in Bayesian analysis. The chapter also summarizes the distinction between finite population and superpopulation inference. Chapter 17 contains a synopsis of concepts of missing and observed at random and the nuts and bolts of multiple imputation that provides a concise reference for missing-data problems.

Because the two books are written by active researchers with some overlapping interests, not surprisingly they contain subtle differences in attribution of who did what. These will be obvious to people reading the books who are in the know, and they may produce an occasional smile. Luckily, the differences are not obvious enough to disturb readers unfamiliar with the personalities involved. In general, Carlin and Louis do a better job of presenting a balanced view where there are controversies, such as in model checking and assessing convergence. Despite the differences in the material, both books have extensive bibliographies. Thus even when one book or the other does not cover a topic, the appropriate references are generally included.

The books differ in their characters as textbooks and references as well as in the background that they assume of the reader. Both books aspire in their introductions to be textbooks as well as reference volumes. Historically, Gelman et al. evolved as a textbook and was used as teaching material at several universities. It is, however, also a rich reference for Bayesian analysis. Carlin and Louis more nearly resembles two monographs put into one book. The one part is on empirical Bayes methods and guiding principles of Bayesian procedures and frequentist evaluation, and the other is on Bayesian methods and computation, with a grand finale of case studies at the end. Carlin and Louis requires more background of the reader than does Gelman et al. Carlin and Louis admit in the Preface that many of the details need to be filled in by going to the library. This may stand in the way of Carlin and Louis being a user-friendly introductory text, but the book does present the state of the art in Bayesian and empirical Bayes methods. It is an excellent reference book and can be used as a more advanced textbook or together with Gelman et al.

In teaching a half-semester course, I used the first six chapters and half of Chapter 7 of Gelman et al. I then turned to Chapter 5 of Carlin and Louis for computational methods. The more leisurely pace of the opening chapters in Gelman et al. is easier on the teacher as well as the students. It is a useful exercise to reproduce the example analyses presented in the text in both books. The examples presented in Gelman et al. were done in Splus, which makes them convenient teaching tools. The datasets used in Gelman et al. are also now available from Gelman's web site at Columbia University. I preceded the leap to the advanced level of the MCMC methods in Chapter 5 of Carlin and Louis with Casella and George's (1992) "Gibbs for Kids" and some excellent introductory material on the Metropolis algorithm from the Kass and Wasserman (1995) short-course manuscript. For teaching data augmentation and the EM algorithm, including Louis's calculation of the missing information, I use Tanner's (1993) book. Bernardo and Smith (1994) provide the most complete appendix on conjugate distributions. In the future, I would look for an alternative to Chapter 6 of Gelman et al. for discussing model checking. In a course that covered decision theory, empirical Bayes methods, or evaluation of procedures, Carlin and Louis's Chapters 1, 3, and 4 would be quite valuable.

Bayesian Data Analysis and *Bayes and Empirical Bayes Methods for Data Analysis* are important contributions to practical Bayesian and empirical Bayes data analysis. Because of the differences in level and choice of material, the two books complement each other quite well both as references and as textbooks. They will likely wind up together on many statisticians' shelves.

M. Elizabeth HALLORAN
Emory University

REFERENCES

Bernardo, J. M., and Smith, A. F. M. (1994), *Bayesian Theory*, New York: Wiley.

Casella, G., and George, E. I. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167–174.

de Finetti, B. (1974), *Theory of Probability*, New York: Wiley.

Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Sciences*, 7, 457–472.

Kass, R. E., and Wasserman, L. (1995), *A Short Course in Bayesian Data Analysis*, New York: Springer.

Tanner, M. A. (1993), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, New York: Springer-Verlag.

Neural Networks for Pattern Recognition.

C. M. BISHOP. Oxford, U.K.: Oxford University Press, 1995. xvii + 482 pp. $55.

Pattern Recognition and Neural Networks.

B. D. RIPLEY. New York: Cambridge University Press, 1996. vii + 403 pp. $52.95.

These two recent books are important additions to the rapidly growing body of literature on pattern recognition and artificial neural networks. According to current definition and usage, pattern recognition "refers to a technology that recognizes and analyzes patterns automatically by machine" (Boden 1987) whose goal is "to clarify. . .complicated mechanisms of decision-making processes and to automate these functions using computers" (Fukunaga 1990). Human beings are subjective experts in both discernment and imposition of patterns in nature; pattern recognition by machines commonly excludes gestalt theory, although figure/ground phenomena and other aspects of sensory perception are common to both. An artificial neural network is "a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use" (Aleksander and Morton 1990; adapted by Haykin 1994)—yet, as Ripley wonders, one could ask how a machine comes to have "natural" properties. Both authors pause to make such definitions and then proceed rapidly to elaborate these themes. Ripley's overture is particularly useful for readers unfamiliar with the fields, as it draws on several motivating practical areas of application that are addressed in more detail throughout the book. Both authors fail to append the necessary qualifier "artificial" to their "neural networks," and neither addresses problems surrounding anatomical and functional details of actual neural networks in

living human minds/brains, such as the kind suggested, for instance, by Wichmann and DeLong (1993) in normal and dysfunctional basal ganglia-thalamocortical "motor" circuitry in Parkinsonism, or by Arbib (1995), and thus both benefit somewhat indirectly from the currently increasing awareness of and interest in fundamental and clinical brain science across many areas of research.

In statistics and biostatistics, pattern recognition and artificial neural networks are not as well established as they are in computer science, engineering, and industrial applications, yet their status is changing as statisticians participate in continued cross-fertilization with these and other disciplines. Reasons for the relative paucity to date may involve preferences in engineering and industry for certainty over uncertainty and determinism over randomness in applications that do not seem to warrant formal statistical inference. Whereas availability of massive amounts of data may indeed favor deterministic algorithms and reduce the need for classical distributional assumptions in some cases, statistical models of uncertainty and inference play a central role in Bishop's and Ripley's expert treatments.

Perhaps the simplest artificial neural network, the McCulloch–Pitts neuron (McCulloch and Pitts 1943, 1948) or single-unit perceptron has the form $y = \text{sign}(w_0 + \sum_{i=1}^{p} w_i x_i)$, where the sign function takes the value $+1$ if its argument is nonnegative and -1 otherwise. This artificial neural network is a type of linear regression of y on $x_1, \ldots, x_p$ with unknown weights (parameters) to be estimated and without any formal error term (although many can be specified). This model is not ordinary least squares regression, but rather some form of discriminant analysis with an indicator variable, y, as the output of the "neuron" modeled as a linear function of its inputs, the x's, which are themselves indicator variables. The neuron "fires" or does not fire, depending on whether the summation is positive or not. Networks are constructed by hooking up interconnected banks of artificial neurons, usually more general than these to allow functional relationships more complex than simple linear forms, with x's feeding into every such node and the outputs of these nodes feeding into similar nodes at another level, and so forth. Such an arrangement is a layered feed-forward neural network or, equivalently, a multilayer perceptron. Layers in between input and output are called hidden layers, because they are not observable directly. As their name implies, nodes at the same level are not connected to one another directly in a feed-forward net; otherwise, the construction is a recurrent net, Hopfield net, Boltzmann machine, or associative memory whose stable patterns can be identified using Markov chain Monte Carlo, for instance.

The layered feed-forward neural network is an important pattern recognizer and a powerful tool for general problems of classification and function approximation (Friedman 1994). Many forms of statistical models can be represented as artificial neural nets, and vice versa. For instance, when output y is an indicator variable and modeled as a function of inputs x that is log-linear in its parameters, we have a form of multiple logistic regression or logistic discrimination that classifies new examples according to maximum posterior probability. This is equivalent to the "softmax" approach of artificial neural networks and indeed predated it. Extensions to more than two classes lead to canonical correlation analysis and flexible discriminant analysis by optimal scoring (Hastie, Tibshirani, and Buja 1994). When regression functions are not necessarily linear members of a general feature space spanned by smooth functions of x, artificial neural networks have been shown to be equivalent to various forms of generalized additive models, projection pursuit regression, and multivariate adaptive regression splines (Friedman 1994; Ripley 1994, and in his present work). Expansions by Fourier and wavelets bases are also possible. Bishop and Ripley provide much more detail on relationships such as these between layered feed-forward nets and statistical models with which we may be more familiar, and well as radial basis functions, nonparametric regression, and high-dimensional density estimation. Additional high-quality expositions can be found in texts by, for example, Duda, Hart, and Stork (1997), Mardia, Kent, and Bibby (1979), McLachlan (1992), and Michie, Spiegelhalter, and Taylor (1994). Their demonstrations of close connections between statistical methods such as these provide readers with some idea of the depth and breadth of pattern recognition and artificial neural networks, a truly multidisciplinary field.

The books differ in some content areas and coverage. Ripley focuses on pattern recognition and artificial neural networks through a more traditional statistical treatment, with chapters on statistical decision theory and linear and flexible discriminants and a lengthy appendix titled "Statistical Sidelines." Bishop, on the other hand, takes a more classical approach as an applied mathematician, focusing on optimization algorithms, preprocessing, and feature extraction with a statistical flavor, and includes appendixes on Lagrange multipliers and calculus of variations, to name a few topics. Ripley's page count is a bit inflated, as he has added wide margins throughout for footnotes, explanation, sidelines, and opinion; without these, the book would be approximately 80% of its current length. However, this pleasant feature, as well as beautifully presented graphics, including a color figure on page 303 in the pursuit of revealing projections of the *Leptograpsus* crab data, adds to the book's overall handsome appearance. Bishop's page content is more standard, in LaTeX and with a slightly smaller font size. Bishop's graphics are adequate, yet prior and posterior distributions are depicted incorrectly in several places (a minor quibble). For instance, the areas under (proper) prior density curves in Figures 2.4, 10.1, and 10.15 are much greater than those under posterior densities; Figure 2.5 has it about right, although all curves there depict posterior densities, and the discrepancy could mislead a naive reader who perhaps needs to be reminded that both marginal and conditional densities are, in fact, densities. Bishop's overall appearance is very similar to Hertz, Krogh, and Palmer's *Introduction to the Theory of Neural Computation* (1991); as acknowledged at the book's beginning, some of Bishop's 160 figures have been adapted from this and other sources. Indeed, one could say, as does Geoffrey Hinton in the Foreword, that Bishop is Hertz, Krogh, and Palmer without the statistical physics—a welcome subtraction for this reviewer who finds Bishop more accessible. As can be seen by comparing the tables of contents, Ripley is more comprehensive than Bishop; Ripley's synoptic coverage of the field perhaps complements Bishop's tighter focus. However, neither text deals adequately with special features involved in the analysis of time series by pattern recognition techniques or artificial neural networks (as in Weigend and Gershenfeld 1993). Bishop's emphasis from the beginning is less tied to statistical applications and more on "how to" from a numerical analytic and applied mathematical viewpoint, compared to Ripley's "what" within a theoretical and far-reaching statistical framework.

Even though both books take optical character recognition as a partially solved canonical problem (Le Cun et al. 1989), Bishop contains no data-analytic examples, whereas Ripley contains many, just as Ripley contains no exercises and Bishop contains many. For instance, Bishop has four ordered exercises that demonstrate failure of common intuition when dealing with spaces of many dimensions, teaching one in detail about the so-called "curse of dimensionality" (Chap. 1, pp. 28–30). For those interested in comparative analyses of actual data, Ripley provides five datasets throughout the book: crabs (which will be familiar to readers of Venables and Ripley 1994), Cushing's syndrome, diabetes, forensic glass, and viruses. Whereas both authors are equally deep and broad in their expositions, Ripley stresses the generality of the many concepts he introduces and demonstrates notable erudition in the process through his insights and an encyclopedic 35-page bibliography. (However, the useful review article by Cheng and Titterington [1994] is omitted.) Bishop is broad in a different sense. Hinton notes in his Foreword that Bishop exhibits his deep understanding and creativity as he explains everything from scratch, taking the reader with basic mathematical literacy on a journey to some equally alluring knowledge domains. In my opinion, Bishop may serve better as a graduate-level textbook, whereas Ripley may be better as a reference book. As an illustration, suppose that one wants to learn about VC dimension, which puts bounds on the worst-case performance of a trained artificial neural network ("VC" stands for Vapnik and Chervonenkis, who proved several important theorems about the "generalization ability" of function approximators; see Moody 1991 for an introduction to generalization and learning in artificial neural networks.) Ripley provides the reader with a scholarly treatment of the topic (8 pages including proof outlines, "omitting details of measurability") with many up-to-date results and references in probability theory. There is some difficult material here, so I would qualify the recommendation of a previous reviewer who found Ripley to be an "altogether accessible text . . . [not a] hard work to read" (Naylor 1996). Conversely, on VC dimension, Bishop takes the reader step-by-step through the basic ideas (3 pages) and leaves remaining points to outside references and exercises. Each author provides detail

in his own manner; the differences perhaps are due to their orientations (Bishop in a computer science and applied mathematics department and Ripley in a statistics department). I preferred Bishop's exposition, but at the same time appreciated knowing where to find more probabilistic analytical results in Ripley when needed. Generally, Bishop charts a detailed middle course through central concepts without real-world data yet plenty of exercises, whereas Ripley covers two extremes: sophisticated statistical theoretic treatment and comparative data analytic results. I would have liked to see some more depth with one or more of Ripley's data examples and discussion of the scientific implications of his findings and of substantive differences between results of his various models, if any.

As mentioned previously, many of the ideas and principles of artificial neural networks are common to the statistical sciences, although sometimes under different names. There is a resurgence of interest in artificial neural networks and pattern recognition in various statistical and biostatistical communities in the United States and abroad, yet, as pointed out by Poggio and Girosi (1992), such "epidemics" (their term) have a wavelength of a few decades or so, and

> there is nothing special or particularly new about neural networks, since they can be regarded as graphical notation for specific regression and classification techniques. The enthusiasm around Neural Networks has taken the form of real intellectual epidemics with too much hype and exaggerations . . . the previous peaks of the epidemics had names such as Gestalt in the 20's, Cybernetics in the 40's, Perceptrons in the 60's . . . Until recently, however, even the best techniques lacked in solid theoretical basis and a good understanding of their working principles. It is possible that, years from now, the main legacy of the Neural Networks epidemics of today will be a renewed interest in a variety of statistical and function approximation techniques and the recognition that the latter may be relevant to the problem of intelligence and the brain.

There is no neural network hype in Bishop and Ripley. On the contrary, these two books add to the solid theoretical basis and understanding of working principles of artificial neural networks to which Poggio and Girosi refer. Breiman (1994), commenting on a review article on artificial neural networks by Cheng and Titterington (1994), seems to have had a similar (although not entirely favorable opinion) on the current status of statistical thinking in artificial neural network research. Cherkassky, Friedman, and Weschler (1994), in an edited collection of papers with the provocative title *From Statistics to Neural Networks*, noted that statistics is one of the oldest fields to study pattern recognition problems and has perhaps seen the greatest duplication of its approaches in other fields. In this same work, Friedman (1994) quoted Efron (paraphrasing Santayana): "Statistics has been the most successful information science, and those who ignore it are condemned to reinvent it." This insight, whether pertaining to statistics, history, memory of life events, or the kind of prediction of the future and understanding of the past (e.g., Weigend and Gershenfeld 1993) that statisticians provide, applies to many current activities in artificial neural network research and its uses. But although it is indeed true that statistical methods are beautiful and elegant and have a large measure of generality and practical importance, the field as a whole does not appear to generate the levels of scientific inspiration and enthusiasm enjoyed by other fields that may, ultimately, rely on statistical inferences to establish their conclusions. Statisticians have much to gain by paying attention to the exciting activity in artificial neural networks research. It makes sense, then, for interested statisticians to read up on the topics covered by Bishop and Ripley, if for no other reason than to observe how many of the techniques that we use are being applied, fine-tuned, and generalized. Tibshirani rightly advocated that we statisticians should worry less about statistical optimality and more about finding methods that work, tackle difficult real data problems, and, in general, sell ourselves better (Tibshirani 1994).

More information on both books is available via the World Wide Web; for Bishop, go to http://www.ncrg.aston.ac.uk/NNPR/, and for Ripley, go to http://www.stats.ox.ac.uk/ripley/PRbook/ and to http://www.stats.ox.ac.uk/pub/PRNN. All of Ripley's datasets are available by anonymous ftp from statlib and the U.K. mirror site at Oxford, or from markov.stats.ox.ac.uk, IP address 163.1.20.1 as noted in the book. I was favorably impressed with the "Complements" updates Ripley has set up there, including notes, errata, references for up-to-date changes (a 6-month lag when I visited) and display mathematics. (These add-ons, not all of which are on Bishop's WWW book page, may be possible because Ripley, not the publisher, holds copyright.) Interested readers can also visit ftp://ftp.sas.com/pub/neural/FAQ4.html#questions, a site not related directly to either book, for FAQ's (frequently asked questions) and answers on artificial neural networks, including many references and reviews for beginning, intermediate, and advanced readers. In making some of their material available on the Web, both Bishop and Ripley show that they are well aware of the current revolution in electronic publishing and seem to have concluded that if there is to be a hard copy version of their work, that version should be attractive: Ripley seems to have been more keen on appearance than Bishop, yet content has not been compromised in either work.

I came to *Neural Networks for Pattern Recognition and Pattern Recognition and Neural Networks* with interests in the development and application of these methodologies to the study of human brain function by medical imaging (i.e., using artificial neural networks to study actual large-scale neural networks), and both books are helpful in this regard. Together with Hertz et al. (1991) if one has an affinity for statistical physics and Arbib (1995) for biologists and neuroscientists, both books should be in the library of any student, teacher, or researcher with a keen interest in modern statistical methods, a large volume of meaningful data to analyze (including simulations), and a fast workstation with good numerical and graphical capabilities.

Nicholas LANGE
McLean Hospital, Harvard Medical School

REFERENCES

Aleksander, I., and Morton, H. (1990), *An Introduction to Neural Computing*, London: Chapman and Hall.

Arbib, M. A. (ed.) (1995), *The Handbook of Brain Theory and Neural Networks*, Cambridge, MA: MIT Press.

Boden, M. A. (1987), "Pattern Recognition," in *The Oxford Companion to the Mind*, ed. R. L. Gregory, Oxford, U.K.: Oxford University Press, pp. 48–50.

Breiman, L. (1994), Comment on "Neural Networks: A Review from a Statistical Perspective," by B. Cheng and D. M. Titterington, *Statistical Science*, 9, 38–42.

Cheng, B., and Titterington, D. M. (1994), "Neural Networks: A Review From a Statistical Perspective" (with comments), *Statistical Science*, 9, 2–54.

Cherkassky, V., Friedman, J. H., and Weschler, H. (eds.) (1994), *From Statistics to Neural Networks, Theory and Pattern Recognition Applications*. Series F: Computer and Systems Sciences, Vol. 136, NATO-PCO Database, Berlin: Springer-Verlag.

Duda, R. O., Hart, P. E., and Stork, D. G. (1997), *Pattern Classification and Scene Analysis* (2nd ed.) New York: Wiley.

Friedman, J. H. (1994), "An Overview of Predictive Learning and Function Approximation," in *From Statistics to Neural Networks, Theory and Pattern Recognition Applications*, Series F: Computer and Systems Sciences, Vol 136. NATO-PCO Database, eds. V. Cherkassky, J. H. Friedman, and H. Weschler, Berlin: Springer-Verlag.

Fukunaga, K. (1990), *Introduction to Statistical Pattern Recognition* (2nd ed.), New York: Academic Press.

Hastie, T., Tibshirani, R., and Buja, A. (1994), "Flexible Discriminant Analysis by Optimal Scoring," *Journal of the American Statistical Association*, 89, 1255–1270.

Haykin, S. (1994), *Neural Networks: A Comprehensive Foundation*, New York: Macmillian.

Hertz, J., Krogh, A., and Palmer, R. G. (1991), *Introduction to the Theory of Neural Computation*. Reading, MA: Addison-Wesley.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989), "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, 1, 541–551.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, London: Academic Press.

McCulloch, W. S., and Pitts, W. (1943), "A Logical Calculus of Ideas Immanent in Neuronal Activity," *Bulletin of Biophysics*, 5, 115–133. Reprinted in Anderson, J. A. and Rosenfeld, E. (eds.) (1988), *Neurocomputing: Foundations of Research*, Cambridge, MA: MIT Press.

——— (1948), "The Statistical Organization of Nervous Activity," *Biometrics*, 4, 91–99.
McLachlan, G. J. (1992), *Discriminant Analysis and Statistical Pattern Recognition*, New York: Wiley.
Michie, D., Spiegelhalter, D. J., and Taylor, C. C. (eds.) (1994), *Machine Learning, Neural and Statistical Classification*, New York: Ellis Horwood.
Moody, J. E. (1991), "Note on Generalization, Regularization and Architecture Selection in Non-Linear Learning Systems," in *First IEEE-SP Workshop on Neural Networks in Signal Processing*, Los Alamitos, CA: IEEE Computer Society Press, pp. 1–10.
Naylor, P. (1996), "Easy Learning (Review of Ripley), *Nature*, 381, 206.
Poggio, T., and Girosi, F. (1992), "Learning Algorithms and Network Architectures," in *Exploring Brain Functions: Models in Neuroscience, Proceedings of the Dahlem Conference*, eds. T. Poggio and D. Glaser, New York: Wiley.
Ripley, B. D. (1994), "Neural Networks and Related Methods for Classification," *Journal of the Royal Statistical Society*, Ser. B, 56, 409–456.
Tibshirani, R. (1994), Comment on "Neural Networks: A Review from a Statistical Perspective," by B. Cheng and D. M. Titterington, *Statistical Science*, 9, 48–49.
Venables, W. N., and Ripley, B. D. (1994), *Modern Applied Statistics With S-Plus*, New York: Springer-Verlag.
Weigend, A. S., and Gershenfeld, H. A. (eds.) (1993), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Reading, MA: Addison-Wesley.
Wickmann, T., and DeLong, M. R. (1993), "Pathophysiology of Parkinsonian Motor Abnormalities," in *Advances in Neurology, Parkinsons Disease: From Basic Research to Treatment*, eds. H. Narabayashi et al., New York: Raven Press.

Markov Chain Monte Carlo in Practice.

W. R. GILKS, S. RICHARDSON, and D. J. SPIEGELHALTER (Eds.). London: Chapman and Hall, 1996. xvii + 486 pp. $54.95.

Gibbs sampling and, more generally, Markov chain Monte Carlo (MCMC) were surely among the hottest topics in statistical research during the first half of the 1990s. The primary bibliographical landmark for this spurt of activity was the 1990 *JASA* article by Gelfand and Smith, which was näive in some respects yet succeeded tremendously well in pointing out and attracting attention to the wide applicability of Gibbs sampling to problems of Bayesian inference. Following the public dissemination of the basic ideas in the year or so preceding publication of the Gefand and Smith article, there was a period of some confusion among many who began working with Gibbs sampling. Then in 1991, at a conference on posterior simulation held at Ohio State University, after vigorous discussion and debate, the major issues began to be clarified. Of particular note was a presentation by Luke Tierney in which he illustrated the use of the Metropolis algorithm to simulate from posterior distributions, pointed out its close relationship to Gibbs sampling, and reviewed how both these and other methods could be framed within the theory of continuous state-space Markov chains. This signaled the advent of what has within statistics come to be called MCMC.

While many Bayesian statisticians were suddenly busy solving a large variety of practical data analysis problems that previously had appeared intractable, others were pointing to connections with the existing scientific literature. To some commentators, it has seemed ironic that MCMC made such a splash in the early 1990s when it had been around so long: the Metropolis algorithm was a standard methodology in statistical physics, having been first published in 1951, and was known to many in statistics (e.g., Besag 1986), partly from a basic paper on the subject by Hastings published in *Biometrika* in 1970. But this observation overlooks the essential driving force behind the emergence in statistics of this simulation technology—namely, the computer. I believe that MCMC became important in statistics when it did for the simple reason that in the early 1990s, large numbers of researchers could implement it on their desktops for interesting, nontrivial problems. An analogous phenomenon occurred a little over a decade earlier, when much attention was given to generalized linear models because faculty and students could at that time experience relatively short delays (typically well under a minute) in obtaining results for their analyses. Without meaning to diminish the important intellectual contributions of the people who helped bring these methods into statistical practice, I would suggest that the timing of their growth in popularity is explained primarily by computing technology.

In any case, it is very fitting that a book on MCMC appeared in 1996. And it is in many ways an excellent book.

MCMC in Practice is aimed primarily at statisticians looking for a summary of available techniques that may be used in applications. A collection of 25 short chapters by many contributors, the book presents basic theory, issues in implementation, and applications. As may be expected of any such collection, the chapters are somewhat variable in quality. Furthermore, because they represented cutting-edge work and could not be subjected to the test of time, some chapters will ultimately be seen as less valuable than others. However, speaking as someone with fairly extensive editorial experience, I admire the book's overall quality. The topics are well-chosen and the style is very appropriate; the articles generally provide the right amount of detail for someone wishing an overview.

Chapter 1 (by Gilks, Richardson, and Spiegelhalter) is a very nice introduction, which I would recommend to anyone venturing into the world of MCMC. MCMC methods simulate observations from a distribution P by creating a Markov chain with P as its stationary distribution. This is useful because once the chain is run to convergence, subsequent states $X_1, X_2, \ldots$ visited by the chain may be considered random draws from P. Although these draws will generally not be independent, sample means of the form $(1/n)\sum_{i=1}^{n} f(X_i)$ will converge to their distributional counterparts $E_P(f(X))$ (where $X \sim P$), provided that the chain satisfies some reasonably mild regularity conditions. In Bayesian inference, P is a posterior distribution, the X_i's become draws from the posterior, and the quantities $E_P(f(X))$ may be posterior means, variances, or interval probabilities.

The Metropolis algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953) is a miraculously simple way to construct a Markov chain with P as its stationary distribution. Beginning with a provisional Markov chain C_1 satisfying only some minimal requirements, it turns out to be very easy to modify C_1, making it into another chain C_2 having the desired stationary distribution: roughly speaking, given that the chain C_2 is in a state x, one observes the next state x' from C_1; x' is accepted as the next state of C_2 with certainty if $p(x') > p(x)$, where $p(x)$ is the density of P, and with probability $p(x')/p(x)$ otherwise. If x' is not accepted, then x itself becomes the next state of C_2. In principle, from an arbitrary initial state x this chain C_2 will eventually evolve to its stationary distribution P. Furthermore, in principle, only a couple of lines of code are required to implement the method in a high-level language such as S-PLUS.

In practice, of course, the choice of C_1 matters greatly, and there are numerous important subtleties—thus the need for the book. The major concerns are reviewed in the introductory first chapter. The Metropolis algorithm may be considered a special case of the more general Metropolis–Hastings algorithm (named for the 1970 work of Hastings, though in physics the generalization is subsumed by the first eponym), which provides the basis of MCMC methods. Chapter 1 only briefly mentions another special case of MCMC—Gibbs sampling—and does not emphasize its importance. The authors may have been conscious of critical remarks to the effect that Gibbs sampling per se had received too much attention. Gibbs sampling generates observations from "full conditional" distributions—that is, distributions of one component of the vector x conditionally on all other components—and thus is limited to cases in which these conditional distributions are tractable, or at least well behaved. But Gibbs sampling turns out to be applicable to a wide variety of problems, especially those that simplify when missing data are filled in or artificial latent variables are introduced, which then become part of the simulation. Gibbs sampling thus applies in roughly the same situations as the EM algorithm. In addition, Gibbs sampling can be remarkably efficient when implemented carefully. Chapter 2 (by Spiegelhalter, Best, Gilks, and Inskip) presents a worked Gibbs sampling example—a longitudinal "growth model" for Hepatitis B titers among infants. It also illustrates the use of directed acyclic graphs in model specification, which many analysts find intuitive, and includes a brief appendix of code from BUGS, freely available software that is playing a major role in bringing Bayesian analysis into statistical practice.

Chapter 3 (by Roberts) introduces relevant Markov chain theory, concentrating mainly on discrete state spaces. Although posterior distributions