Chapter 8, on chi-squared tests, is one of the book's better chapters. It includes discussion of the goodness of fit, one- and two-way tables, case-control studies, and the odds ratio and contains proofs at the conclusion. The set of exercises is extensive.

Chapters 9 and 10 deal with linear regression and correlation. Many regression issues are addressed, and even curvilinear and logistic regression are given short introductions. The latter chapter includes a good discussion of spurious correlation and is noteworthy for the wealth of proofs that it contains.

Matrix algebra appears in Chapter 11, which pertains to multiple regression and correlation. The proofs associated with this chapter do not involve matrices, however.

Chapters 12–16 cover a wide range of ANOVA topics: one-way layout, fixed-effects two-way layout, multiple comparisons, random effects, mixed models, expected mean squares, blocking, Latin squares, efficiency of designs, and analysis of covariance. A number of worked examples are taken from SAT scoring. The use of matrix algebra is avoided, yet the proofs and exercise sets are substantial.

The final chapter addresses several nonparametric techniques, including the sign test, Wilcoxon tests, and Kruskal–Wallis and Friedman tests. In many cases, data are analyzed using more than one technique.

Exercises appear at the end of each chapter, after the proofs. In front of the index, answers to selected problems are supplied, organized by chapter; usually, these are the odd-numbered exercises. However, fewer answers are offered for the chapters toward the end of the text.

As described previously, this text is suitable at a rather introductory level and thus does not include intermediate-level topics such as survival analysis, time series, and the Bayesian statistical framework. Given that excellent books already exist that cover each of these topics in substantial depth, the omission is appropriate. In fact, covering all of Chiang's chapters in one semester would be a stretch. My assessment is that the instructor would be well advised to omit some or all of Chapters 13–15 (on experimental design and mixed effects) and pick up this material in a second semester. Of course, the precise selection of topics will depend on the composition of a particular class.

In parallel fashion, this text does not include sample coding and output associated with any software packages. Therefore, the instructor is free to select her or his preferred package and to present the essentials for running and interpreting results using that software. The chapters include numerous worked examples using small datasets appearing in tabular form. Although the author may be free from statistical software bias, it may have been more helpful to include a CD containing a dozen or more data files in text format to facilitate analyses using an instructor's chosen package and to permit analysis of somewhat larger datasets.

Philip F. RUST
Medical University of South Carolina

Causality: Models, Reasoning, and Inference.

Judea PEARL. New York: Cambridge University Press, 2000. ISBN 0-5217-7362-8. xvi + 384 pp. $39.95.

Those readers familiar with causality will already know the earlier work of Pearl (1988), a text on probabilistic reasoning in the framework of artificial intelligence that helped define a broad research agenda for more than a decade. Pearl's pursuit of many aspects of that agenda has led to numerous new developments, most of which have already been presented in various journals and forums. This book synthesizes and unifies the developments that treat causality, and links them with related progress by statisticians, philosophers, computer scientists, psychologists, and others. We now have a single volume that brings much of this corpus together, to reflect on where we are at and to make it easier to bring this material into classes and seminars. Of course, Pearl has his own particular approach, and others have somewhat different methods; however, the debates among them are exactly what the field needs as it matures and sorts itself out. Some of these can be followed on a web forum set up by the author at *http://bayes.cs.ucla.edu/BOOK-2K/discussion.html*.

This review is directed more toward a second and broader audience, namely those readers of this journal who were raised hearing the constant admonition "don't deduce causality" and who have been passing it on mantra-like to the next generation in many statistical contexts. What we usually want in investigating the connections between physical, social, political, or economic phenomena is precisely an understanding of cause and effect, but we have been so brainwashed that we do not revolt when our teachers or colleagues tell us to be satisfied with standard methods that offer considerably less. Is causality a mirage in the desert, or is it the paydirt that we should be out looking for? Why do some social scientists vigorously pursue it while many statisticians and mathematicians still shrink into the shadows in timidity or embarrassment?

The answer to this last question may help to explain this state of affairs. We had been warned by such titans as Karl Pearson and Bertrand Russell to keep our hands off this murky stuff. For example, the former described causality as an arcane fetish of then-modern science, much inferior to the notion of correlation. Consider the issue as it arises in connection with Simpson's paradox, the familiar situation where a treatment is found to be favored over nontreatment in a given population, but if the population is partitioned by means of another variable, then treatment is inferior to nontreatment in each of the two subpopulations. The only paradox is when one brings in notions of causality; otherwise the situation is easily explained in terms of typical probability calculus that we regularly assign our students to work out. Unfortunately, in many applications, like the causes or treatments of disease, causality is the real issue. Faced with the situation just described, should one prescribe the treatment or not? If our existing probability theory does not get at this issue, then we should demand more.

According to Pearl, early in Russell's career (circa 1913), Russell characterized the "law of causality" as "a relic of a bygone age, surviving only because it is believed to do no harm." Later, however (circa 1948), Russell wrote extensively on the subject; the difference may be largely due to the definition of terms, because many notions of causality have been debated since the time of Aristotle or even Heraclitus.

These warnings notwithstanding, one area where classical statistics is clearly up to the task of providing strong evidence of causality is in the experimental realm, where randomization in design is used to control for confounding variables. The situation is less rosy in fields that depend on observational data or, even worse, where counterfactual arguments are the main thrust. For the former, I think of epidemiology as a typical application, where unfortunately every textbook writer can easily provide many examples of conflicting studies, faulty designs, overlooked factors, and sources of bias. One thus appreciates the accomplishment in studies that persist in being accepted as definitive. For fields depending on counterfactual arguments (I think of many social, political, and even economic studies carried out to support policy decisions), we would be hard pressed to assume that they fare any better, even though it might be much less likely that they will be found out.

What Pearl and his coworkers have brought to this universe of uncontrolled and often unobserved variables is a more rigorous logical superstructure that can be used to sort out things to a level that previously was rarely achievable. It is not surprising that these developments would emanate from investigators with a strong interest in artificial intelligence. In attempting to model cognitive processes, one must take a microscope to the laws of inductive reasoning. The basic components of these developments are the following: a directed, acyclic graph theoretical framework (Bayesian networks built on causal relationships), the set of Markovian parents of variables represented by nodes on such a graph, the concept of interventions in such networks [using the so-called "do(x)" operator], and the development of an axiomatic calculus for such networks that allows the algorithmic reduction of one form to another to investigate the various aspects of causality, of which there are many.

With respect to the all-important "do(x)" operator, consider the question of whether smoking causes cancer. It is not enough, of course, to compare $P(\text{cancer})$ to $P(\text{cancer}|\text{smoking})$, particularly because the event "smoking" here means the observation that someone smokes. As Pearl points out, the tobacco industry proposed a perfectly plausible probabilistic model based on a genetic predisposition to both nicotine cravings and cancer, which was consistent with all of the observational data. What we really want to know about is $P(\text{cancer}|\text{do}(\text{smoking}))$, as though we were setting up a controlled experiment. The reduction of expressions such as this to expressions that can be quantified by observational data is the key point of Pearl's methodology.

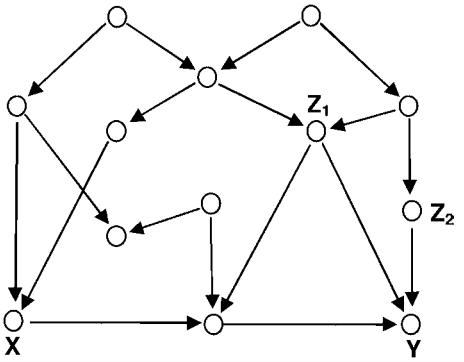

Figure 1. Adjustment Problem Causal Graph.

One of Pearl's examples offers the best introduction to this approach. Here the problem is the so-called "adjustment problem," where the issue is the analysis of observational data and where we must decide which covariates we should properly adjust for in analyzing the effect of variable X on variable Y. This of course is at the heart of Simpson's paradox as it arises in practice. There are many factors connected with variables X and Y and with each other, some of which may be unobservable, as suggested in the causal graph shown in Figure 1 (adapted from p. 357).

Two of the factors, Z_1 and Z_2 in the graph, happen to be among the observable ones, and we want to know whether these two would be a sufficient and appropriate set of covariates to adjust for. Pearl's system consists of a sequence of reductions of this graph to an endpoint from which we can read off our final conclusion. Can you figure it out, using a straightforward interpretation of the arrows? If you can and if you like this approach, then read the book to see it extensively developed. If you cannot but are curious to see how it works out, then read the book too!

Let me provide a brief sketch of the corresponding formal structures that underlie Pearl's approach. A causal model is defined as a triple $M = \langle U, V, F \rangle$, where the U form a set of background or exogenous variables determined by factors outside the model, the V form a set of model or endogenous variables, and the F are a set of functions. There is one element of F for each element of V, expressing this particular element of V as a function of the values of the U's and the values of each of the elements of V in a subset of V called the Markovian parents of the element under discussion. Such a model can incorporate probabilities by assigning probability distributions to U. Within this framework, one can define submodels, the do(x) operator, counterfactual statements, and other fundamental concepts. One can also develop theorems concerning the reduction and analysis of arguments or hypotheses. This allows for the investigation of many subtle variations that invariably arise in questions about causality, such as necessary versus sufficient causes, actual versus general causes, direct versus indirect causes, unobserved but likely causes, possible causes, intransitivity of causal dependencies, and criteria for exogeneity. Structural equations fit this framework and are a constant theme.

Reading this book is serious work and not for the faint-hearted. The prerequisites are modest, but the level is professional. The second half of the epilogue is probably the best introduction to the key ideas and would be good to read first. Although the main part of the text is expository, it is not written in the kind of pedagogical style that motivates ideas before delivering the details. There are no exercises to reinforce understanding or to try for practice, but a number of concrete examples and application areas are woven into the discussion. (The author has posted some exercises on the webpage referred to earlier.) I think that it might make a particularly good text for an ongoing seminar, where the participants would generally be able to supplement the material with insights and interpretations from their own areas of specialization.

In conclusion, make no mistake about it: This is an important book. Even if almost all of the content has appeared previously in diverse venues, it has been brought together here for all of us to think about. The field has no shortage of lively controversy and divergent opinion, but be that as it may, this is certainly one of the contributions that will bring this material further out of the closet and into the face of the broader statistical community, a move that we should welcome both as consumers and as testers of its utility.

Charles R. HADLOCK
Bentley College

REFERENCE

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Francisco: Morgan Kaufmann.

Bayesian Artificial Intelligence.

Keven B. KORB and Ann E. NICHOLSON. Boca Raton, FL: Chapman & Hall/CRC, 2004. ISBN 1-58488-387-1. 364 pp. $79.95.

The field of Bayesian networks, better known to the statistical community as probabilistic graphical models, is an area where the collaboration between researchers in statistics and artificial intelligence has been fruitful. Graphical models have been responsible for ground-breaking advances in statistics, such as the development of an amenable Gibbs sampler (Thomas, Spiegelhalter, and Gilks 1992), and for remarkable applications in artificial intelligence. The success of Bayesian networks has also been of critical cultural importance in artificial intelligence and statistics. In artificial intelligence, the enthusiasm for Bayesian networks has led to the lifting of a "30-year ban" against probability and statistics. In statistics, graphical models, by themselves or through their contribution to Gibbs sampling, have dramatically contributed to a better understanding and deeper appreciation of Bayesian methods. In both fields, graphical models have empowered statisticians and computer scientists to handle models of great complexity and have contributed to the switch in focus of many statistical endeavors from the purely formal, or the merely descriptive, to the important modeling aspects of the problems under study. Such modeling, together with the partially automated nature of Bayesian networks, has helped make them attractive to researchers and students in medicine and computer science and to those who do not necessarily have formal training in probability and statistics.

This book provides an introduction to Bayesian networks for this type of reader. The book starts ab imis fundamentis with an introduction to probability theory that assumes nothing from the reader except the notion of a function. The subsequent two chapters are devoted to the formal representation of Bayesian networks, their mathematical properties, and the leading algorithms in probabilistic reasoning, that is, computation of the posterior probability of a set of variables in the network given the values of other variables. The fourth chapter describes the decision-theoretic version of Bayesian networks commonly known as influence diagrams and describes how to augment the syntax of Bayesian networks with nodes representing decisions and utilities so that they can be directly used to represent decision problems. A detailed chapter describing applications of Bayesian networks closes the first part of the book.

The second part, comprising approximately one-quarter of the book, is entirely devoted to statistical methods to infer Bayesian networks from data, one of the most recent and successful uses of Bayesian networks. This part stays true to the approach of the book by including summary explanations of the most basic statistical concepts, making it accessible to readers with little or no statistical background. The third part is probably the most interesting and original, because it focuses on the use of Bayesian networks as modeling tools. It includes a chapter clearly describing validation and verification methods and fully worked out example applications that introduce the reader, in a problem-based fashion, to the techniques and heuristics of Bayesian networks modeling. For each chapter, the book also includes an exercise set consistent with its introductory nature. The book is augmented by the website *http://www.csse.monash.edu.au/bai*, which contains source code for some of the examples of the book and a repository of Bayesian networks ready to be studied.

The market position of this book as a graduate course textbook is unique. The landmark volume by Pearl (1990) has been followed by several comprehensive books on the subject. The overall outline of this book is similar to that of the introductory book by Castillo, Gutierrez, and Hadi (1997), but this book explores the statistical aspects of Bayesian networks and the methods to fit them from data in greater detail. Although a graduate course in statistics would probably be better served by the more sophisticated textbook of Cowell, Dawid, Lauritzen, and Spiegelhalter (1999), Lauritzen (1996), or the classic Whittaker (1990), this book's self-contained nature makes it quite appropriate for students in other disciplines. The emphasis on modeling practice, in particular, could make it attractive to students in computer science and engineering.

One concluding caveat: The title *Bayesian Artificial Intelligence* is somewhat misleading. Although the book focuses exclusively on Bayesian networks, Bayesian methods are today pervasive of artificial intelligence research, and