American Journal of Epidemiology
Vol. 146, No. 4
Printed in U.S.A.

BOOK REVIEWS

Bayesian Data Analysis

By A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin

ISBN 0-412-03991-5, Chapman and Hall, New York, New York (Telephone: 800-842-3636), 1995, 526 pp., $60.50 hardback

Bayes and Empirical Bayes Methods for Data Analysis

By Bradley P. Carlin and Thomas A. Louis

ISBN 0-412-05611-9, Chapman and Hall, New York, New York (Telephone: 800-842-3636), 1996, 399 pp., $59.05 hardback

Bayesian methods are designed for rational incorporation of prior information (information external to the data) into the process of statistical analysis. In doing so, they offer solutions to a number of vexing problems, such as how to analyze multiple exposures. As an example, suppose we have obtained diet histories from cases and controls in a study of breast cancer and we wish to search for effects of foods on breast-cancer risk. One conventional approach would be to enter food intakes in a logistic model for disease status, and to present the maximum-likelihood (ML) estimates of the coefficients along with their confidence limits. This approach takes no account of the known differences in nutrient contents of the foods or of the multiplicity of the comparisons (which, despite claims to the contrary, *is* an issue in such "fishing expeditions" (1–4)). One way to address these issues is to incorporate data on the nutrient contents of foods in a second level or stage of the analysis: After regressing disease on foods, we may regress the estimated food effects on the nutritional composition of the foods (5). The resulting food-effect estimates would, on average, be closer to the true effects than the conventional ML estimates (6).

The process just described is an example of two-stage hierarchical (multilevel) regression, also known as empirical-Bayes (EB) regression. The results could be further improved by constraining the nutrient effects in the food-nutrient example to fall within reasonable ranges; such an approach would be an example of a hierarchical Bayes or Bayes- empirical-Bayes regression. Aside from occasional application to disease mapping, these methods remain uncommon in epidemiology, despite the fact that several authors have argued and illustrated their advantages for epidemiologic studies (1–7).

I believe there are four major reasons for the underutilization of hierarchical methods. First, the methods are neglected in basic statistics education, and, until recently, most statisticians instructing and collaborating with epidemiologists had little experience with hierarchical models outside of the analysis-of-variance context. Second, the methods are more difficult to apply than conventional methods: They demand more details of model specification than, say, ordinary logistic or Poisson regression, and those details have to be considered in the interpretation of results. Third, until recently, packaged software suitable for hierarchical epidemiologic analysis was not widely available. With the arrival of procedures such as SAS proc GLIMMIX (SAS Institute, Cary, North Carolina), that has changed, but instruction in the use of such software is still needed and the outputs are not tailored to epidemiologic uses. Last but not least, there have been many negative perceptions of methods to which the name of Bayes is attached, encouraged by some noteworthy misinterpretations of the underlying philosophy.

Several authors have discussed the latter problem (8–12). A major misperception about Bayesian methods is that they are inherently more subjective than conventional methods. David Freedman, no Bayesian by any means, underscored the defect in such criticisms (13, pp. 23–24):

> Objectivists sometimes argue that they have the advantage, because science is objective. This is not serious: "objectivist" statistical analysis must often rely on judgement and experience; subjective elements come in.

For epidemiology, I would replace Freedman's "often" with "always." Epidemiologic studies *always* hinge on background knowledge and experience, from design questions ("What will be a good exposure measure?" "What will be important confounders?") to analysis decisions ("What model is appropriate for the analysis of these data?"). Unfortunately, the latter decisions are often left to persons (programmers and statisticians) with little or no understanding of the background science; those persons may in turn often leave decisions to mechanical algorithms such as stepwise regression. Hierarchical methods are more difficult to apply precisely because they do not (yet) offer such thoughtless substitutes for thoughtful model specification.

In recent times, the geometric growth of computing power and a gradual change in statistics philosophy have brought hierarchical methods to the forefront of statistical research. Many leading statisticians now consider such methods to be essential analytic tools. The books under review here, *Bayesian Data Analysis* by Gelman et al. and *Bayes and Empirical Bayes Methods for Data Analysis* by Carlin and Louis, are the latest fruits of these developments. Unfortunately, it seems unlikely that either book will find wide appreciation among epidemiologists. Both books are intended for graduate statistics courses, and neither provides examples that involve case-control studies or comparisons of nonrandomized cohorts.

Nonetheless, the book by Gelman et al. has much to recommend it. The writing is exceptionally lucid, with ample text and numerical examples to illuminate the concepts behind the equations. Although the book does not offer detailed case studies, many of the examples are taken from biomedical data, including the first example, a simplified genetics illustration (p. 10). Although Gelman et al. do expect the reader to have had calculus, probability, and statistics (both applied and mathematical), they take pains to explain all their notation and terminology.

Perhaps the most interesting aspect of the book by Gelman et al. is its emphasis on Monte Carlo (simulation) analysis methods. Whether such an emphasis is best for practice remains a matter of debate (14), but Gelman et al. make a good case for its didactic value by providing ample graphs of simulation output. They also pay special attention to issues of causal inference from randomized and nonrandomized studies, although these issues are not addressed until the last sections of Chapter 7 (about halfway through the book).

Carlin and Louis are also devotees of Monte Carlo methods, but their book is written for students fluent in advanced mathematics and statistical theory. Only three numerical examples (two with real data) appear in the first four of the eight chapters, and conceptual issues are more often illustrated through equations and derivations. Although a few more examples are given in Chapters 5–7, a high level of abstraction is maintained until the all-too-brief final chapter (Chapter 8), which contains three case studies. Even here, it seems that a data-distant approach is maintained. In particular, the descriptions of the studies and their data are too sketchy to allow one to assess the relation of the results to the data, even though detailed statistical outputs are given. Most problematic, in my view, is that Carlin and Louis never explicate the issues involved in causal inferences from their methods, and in particular they ignore the difficulties that arise in matching statistical parameters with causal effects in nonexperimental studies.

Admittedly, Carlin and Louis provide an impressive compendium of the mathematical techniques underlying Bayes and empirical-Bayes methods. The book should thus be of great value for training statistics doctoral students. Nonetheless, unless terms such as *Cholesky factorization*, L_2 *convergence*, and *transition kernel* are part of your working vocabulary, you will not likely find the book useful.

Although both books begin with discussion of Bayesian philosophy, they cover just enough to motivate the ensuing developments. Again, Gelman et al. are clear and down-to-earth in their presentation, while Carlin and Louis focus on mathematical arguments. Readers who desire more detailed consideration of Bayesian philosophy should read the books by Good (9) and by Howson and Urbach (12), which involve only modest amounts of mathematics. Readers with more advanced training would also do well to read DeFinetti's classic *Theory of Probability* (15).

Both Gelman et al. and Carlin and Louis appear pragmatic in their philosophies. Both books emphasize that Bayesian methods make sense from a frequentist perspective, in that the methods can outperform classical frequentist approaches when evaluated in frequentist terms. They also emphasize that frequency evaluations are essential to ensure that a given Bayesian approach is trustworthy for a given application. Such pragmatic Bayesianism, or "Bayes/non-Bayes compromise," has long been advocated by some leading statisticians (9, 16), and is an immediate consequence of the general hierarchical viewpoint (17).

As emphasized by Good (9), there are differences among moderate Bayesian positions, and a comparison of the reviewed books will reveal several. Gelman et al. express considerable reservation about the use of so-called "noninformative" prior distributions (e.g., p. 56), and conclude that they are no more than convenient tools for avoiding the labor of producing or using a credible informative prior. In contrast, Carlin and

Louis seem to promote noninformative priors as a source of "objective" and "robust" methods.

I find Carlin and Louis's arguments for noninformative priors to be most unconvincing. One reason is that Carlin and Louis seem to believe that inferences can be based "solely on the data" (p. 33). In making such remarks, they forget that all analyses require specification of a functional form for the likelihood, which is not part of the data, and that noninformative priors can yield improper posterior distributions (18). Another flaw in their argument is that, in any epidemiologic problem (and I suspect, in any scientific problem), enough will be known about the parameters of interest to bound them in at least one direction, away from very large or very small values. As an example, consider the current controversy about whether induced abortion increases risk of breast cancer. The controversy concerns only rate ratios (RR) in the range from one to ten, yet the standard (log-uniform) noninformative prior for this problem would consider the rate ratio to be just as likely to fall between 10^{1000} and 10^{1001} as between 1 and 10; that is, it would imply that

$$\Pr(1 < \mathrm{RR} < 10) = \Pr(10^{1000} < \mathrm{RR} < 10^{1001}) \textit{ a priori.}$$

Such an equality would reflect no scientific opinion, for RR $= 10^{1000}$ would imply that almost every woman who had an induced abortion would get breast cancer soon after the abortion.

In this context, it is pointless to demand that techniques be robust under possibilities as absurd as RR $= 10^{1000}$. Recognition that certain possibilities are absurd requires familiarity with the context, however. The true strength of the Bayesian approach is that it can encourage the user to adapt the calculations to the context, rather than to adopt an automated universal approach (16). Carlin and Louis mention such ideas early in their book, and even give a case study that uses priors elicited from clinicians. Nonetheless, they seem all too ready to excuse the use of absurd noninformative priors on the grounds that these priors yield robust results. In contrast, Gelman et al. (p. 57) underscore that such priors should not be used if they lead to conclusions different from those obtained using scientifically sensible priors.

Similar divergences arise in the distinction between Bayes and empirical-Bayes methods. Carlin and Louis emphasize the technical distinction between the approaches: Empirical-Bayes methods estimate the parameters in the final stage of a hierarchical model, whereas the corresponding Bayes methods add another level to the hierarchy and fix the parameters at this new level (17). For a frequentist, empirical Bayes is a legitimate approach for generating estimates with good frequency properties (even Jerzy Neyman (19), an orthodox frequentist, endorsed the empirical Bayes approach), and Carlin and Louis focus on these properties. In contrast, Gelman et al. do not focus on empirical-Bayes methods (and even eschew the term "empirical Bayes") because they consider such methods to be only approximations to fully Bayesian methods (p. 123). Here again, such an approximation should not be used if it leads to conclusions different from a sensible Bayesian analysis.

Carlin and Louis's emphasis may stem from their belief, stated as a fact, that "typically our knowledge at levels above the second prior stage is sufficiently vague that additional levels are of little benefit" (p. 24). In reality, higher-stage knowledge is often available in the background literature. For example, in the food-nutrient analysis by Witte et al. (5), it would have been possible to incorporate biochemical information about nutrients in a third level of the analysis (e.g., by regressing estimated nutrient effects on a measure of antioxidant activity).

A common tragedy of data analysis is that the persons familiar enough with the science to formulate reasonable priors and hierarchical models rarely understand Bayesian statistics well enough to do so. Of the two books reviewed here, I believe only Gelman et al. will contribute to closing this gap between scientists and statisticians. Neither book delves too deeply into issues of model selection, however. To paraphrase Freedman (13), "What justifies the use of any model in nonexperimental data?" Persons interested in such issues should examine the text by Leamer (20), which offers pointed and detailed Bayesian criticism of frequentist methods, and raises serious questions about Bayesian methods as well. These questions continue to arise (13, 21, 22) and should be considered in the course of every epidemiologic analysis, for even a simple analysis of a single 2-by-2 table invokes a probability model whose meaning is questionable in nonrandomized studies (22).

Both conventional and hierarchical methods share crucial and often questionable model assumptions. Nonetheless, I believe that hierarchical modeling is vastly superior to the current convention of single-stage modeling, especially in its ability to provide realistic estimates and measures of uncertainty in multivariate analyses (1–7, 9, 23). Epidemiologic illustrations of the approach (1–7) may help readers decide whether it is worthwhile pursuing the more thorough and technical treatments by Gelman et al. or Carlin and Louis.

REFERENCES

1. Thomas DC. The problem of multiple inference in identifying point-source environmental hazards. Environ Health Perspect 1985;62:407–14.
2. Greenland S, Robins JM. Empirical Bayes adjustments for multiple comparisons are sometimes useful. Epidemiology 1991;2:244–51.
3. Greenland S, Poole C. Empirical-Bayes approaches to occupational and environmental-health surveillance. Arch Environ Health 1994;48:9–16.
4. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. (Invited paper). Environ Health Perspect 1994;102(suppl 8):33–9.
5. Witte JS, Greenland S, Haile RW, et al. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. Epidemiology 1994;5:612–21.
6. Efron B, Morris CN. Data analysis using Stein's estimator and its generalizations. J Am Stat Assoc 1975;70:311–19.
7. Yanagimoto T, Kashiwagi N. Empirical Bayes methods for smoothing data and for simultaneous estimation of many parameters. Environ Health Perspect 1990;87:109–14.
8. Good IJ. Alleged objectivity: a threat to the human spirit? Int Stat Rev 1978;46:65–6.
9. Good IJ. Good thinking. Minneapolis, MN: University of Minnesota Press, 1983.
10. Berger JO, Berry DA. Statistical analysis and the illusion of objectivity. Am Scientist 1988;76:159–65.
11. Greenland S. Probability versus Popper: an elaboration of the insufficiency of current Popperian approached in epidemiologic analysis. In: Rothman KJ, ed. Causal inference. Chestnut Hill, MA: Epidemiology Resources Inc, 1988: 95–104.
12. Howson C, Urbach P. Scientific reasoning: the Bayesian approach. 2nd ed. LaSalle, IL: Open Court, 1993.
13. Freedman DA. Some issues in the foundations of statistics (with discussion). Found Science 1995;1:19–83.
14. Leonard T. Monte Carlo and bust. (Letter). RSS News 1996; 23:8.
15. DeFinetti B. Theory of probability (in two volumes). New York: Wiley, 1974.
16. Box GEP. An apology for ecumenism in statistics. In: Box GEP, Leonard T, Wu CF, eds. Scientific inference, data analysis, and robustness. New York: Academic Press, 1983: 51–84.
17. Good IJ. Hierarchical Bayesian and empirical Bayesian methods. (Letter). Am Statistician 1987;41:92.
18. Hobert JP, Casella G. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. J Am Stat Assoc 1996;91:1461–73.
19. Neyman J. Two breakthroughs in the theory of statistical decision making. Rev Int Stat Inst 1962;30:11–27.
20. Leamer EE. Specification searches. New York: Wiley, 1978.
21. Robins JM, Greenland S. The role of model selection in causal inference from nonexperimental data. Am J Epidemiol 1986; 123:392–402.
22. Greenland S. Randomization, statistics, and causal inference. Epidemiology 1990;1:421–9.
23. Rothman KJ, Greenland S. Modern epidemiology. 2nd ed. Philadelphia: JB Lippincott Co, 1997 (in press).

Sander Greenland
UCLA School of Public Health
Department of Epidemiology
Box 951772
Los Angeles, CA 90095–1772

Understanding Medical Research: A Practitioner's Guide

By Jane L. Garb

ISBN 0–316–29169–2, Little, Brown, and Co., Inc., Boston, Massachusetts (Telephone: 617–859–5500), 1996, 276 pp., $24.95 paperback

Epidemiological Research Methods

By Don McNeil

ISBN 0–471–96196–5, John Wiley & Sons, Ltd., New York, New York (Telephone: 212–850–6336), 1996, 305 pp., $34.95 paperback

As a statistician who interacts with epidemiologists and other health researchers on a daily basis, I am always on the lookout for an introductory statistics text to recommend to my colleagues and to use as a reference text for teaching. Hitting the right balance between scientific rigor and interpretability is a challenge, however, and I have yet to find a text with which I have been totally satisfied.

The two books reviewed here are two of the latest entries in the field and present marked contrasts in style and content. In part, this reflects the authors orientations in writing the books. Jane Garb has written a book specifically for clinicians. As she notes in her Preface and Introduction, the book's purpose is to "de-mystify statistics" (p. xi) and "provide clinicians with the practical skills to evaluate studies in the medical literature or to conduct original studies for publication" (p. xiii). Don McNeil's book, by contrast, is designed for use by final year undergraduate- or Master's-level statistics students and for medical scientists. He rightly notes that the book ambitiously "attempts to cover in reasonable depth the concepts of statistical modeling of epidemiological data. . . with minimal statistical prerequisites" (p. x). The McNeil text, though meaty in places, pretty much delivers on its promise. Garb's text, by contrast, misses the mark.