values that are imputed, no matter by what mechanism the imputations were produced.

Comparative studies of various approaches in a variety of population structures will be very valuable. It would be very nice to be able to ask for theoretical analysis of realistic multivariate situations, but it is clear from these volumes that even univariate versions of the problem (which account for most of the calculations to date) require very intricate, if fairly elementary, manipulations. It seems therefore that much future work will be based on simulation studies and comparative exercises with specific, real surveys. Given the computer power now available, it seems likely that the model-based, bootstrap-like, multiple-imputation methods will gain support, as in recent work by Rubin and Schenker (1985) on the development of interval estimation procedures.

In spite of the impossibility of the problem, practical necessity will maintain the pressure for further research, and the material in these volumes represents an ideal springboard.

D. M. TITTERINGTON
University of Glasgow

REFERENCES

Greenlees, J. S., Reece, W. S., and Zieschang, K. D. (1982), "Imputation of Missing Values When the Probability of Response Depends on the Variable Being Imputed," *Journal of the American Statistical Association*, 77, 251–261.

Rubin, D. B. (1977), "Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys," *Journal of the American Statistical Association*, 72, 538–543.

Rubin, D. B., and Schenker, N. (1985), "Multiple Imputation for Interval Estimation From Simple Random Samples With Ignorable Nonresponse," unpublished manuscript.

Statistical Abstract of the United States: 1984 (104th ed.).

U. S. Bureau of the Census. Washington, DC: U.S. Department of Commerce, 1984. xxviii + 1015 pp. (paperback).

In 1,015 pages, the 1984 *Statistical Abstract of the United States* presents 1,542 statistical tables, supplemented by six appendixes and a front section of tables and charts on recent trends, and accessed by a 39-page subject index. The cover bears endorsements from Daniel Moynihan, James Michener, and Ben Bagdikian, who express, for all users of the *Statistical Abstract*, their admiration of its prudence, breadth, and objectivity, and its ability to change with the world it measures. This 104th edition includes 97 new tables and drops 110 tables that were in the prior issue. Appendix VI presents selected 1980 census data and the 1982 House election data for congressional districts.

The *Statistical Abstract* functions as (a) a reference to statistics on an enormous range of topics; (b) a guide to sources for the data that are shown, and more detailed data; and (c) a lexicon, offering guidance to the concepts that govern collection and presentation of the data. The *Statistical Abstract* claims the first and second of these functions and performs them very well. It performs the third function, that of a lexicon, with uneven success.

An important feature of the *Statistical Abstract* is its relationship to other publications in the *Statistical Abstract* family of publications and, beyond those, to the surveys and statistical programs that supply the numbers presented here. This point is apparent in the *Statistical Abstract*'s appendixes. Appendix I tells the reader where in the *Statistical Abstract* to look for the latest figure in series that are published in *Historical Statistics of the United States*. Appendix V tells where to look in the *State and Metropolitan Area Data Book* for state and metropolitan area statistics corresponding to national or state data shown in the *Statistical Abstract*.

Sources of data are indicated in the introductory text to each chapter, in footnotes to the tables, and in Appendix IV. Appendix IV lists for each subject the principal institutions and publications from which the data were drawn. Appendix III, entitled "Statistical Methodology and Reliability," presents short descriptions of the principal federal statistical programs that furnish much of the data shown here. It furnishes an excellent compact overview of 41 major data systems, such as the National Health Interview Survey. Thumbnail characterizations note the universe that is covered, survey design, data-collection procedures, imputations for missing data, examples of sampling error, and brief notations on the completeness of enumeration, response rates, and other sources of bias.

The *Statistical Abstract* does not work so well as a lexicon as it does as a reference and guide to sources. The degree of the shortfall may be seen if we examine a topic on which the *Statistical Abstract* distinguishes itself—namely, metropolitan population. Appendix II presents the concept and statistical definition of the metropolitan statistical area (MSA); a listing of each MSA and its component counties, together with their 1980 population figures; statistics describing the effects on MSA population of changing MSA definitions; and statistics on metropolitan growth attributable to changing boundaries and to growth of population within constant boundaries. (The 1982–1983 issue of the *Statistical Abstract* showed a map of metropolitan areas, too.) In words and in numbers, abstractly and concretely, by description and by enumeration, Appendix II conveys the metropolitan-area concept.

Perhaps it would be asking too much to expect all topics to receive such splendid treatment. In that case, the book has other models to offer, notably those instances in which it offers help with the concepts by the way it presents the data. For example, early in the chapter on income and wealth, we are given a table (No. 737) that in 18 lines shows the relationships among gross national product (GNP), national income, personal income, and related quantities, and another table (No. 735) showing the relationships among the components of GNP. This is a big help in grasping the concepts. Moreover, it enables us to establish the connections among other, more detailed, tables showing the objects of personal-consumption expenditures or the origins of national income. Tables on the objects and sources of health-care expenditures perform a similar function, showing the relations among components and furnishing a means of connecting the detailed tables in the chapter on health and nutrition. Table 7, on components of population change, shows how births, deaths, immigration, and emigration combine to increase population, and many later tables may be interpreted as various ways of disaggregating (by age, state, etc.) the quantities in that table. The point is that, at a number of places in this volume, well-designed integrating tables furnish a conceptual guide to the subject of the chapter and enable the reader to move with confidence among the tables of the chapter.

No table offers an equivalent integration of the concepts governing statistics on households and families. Those concepts are complex enough to require such a table, and the classification of individuals by family status offers a means of constructing it. Lacking an integrating tabulation, we get very interesting data without the setting that could make transparent their connection to other quantities in the chapter. We are, of course, given verbal explanations in the text. But text can present only one item at a time, and texts usually present concepts without all the corresponding quantities. So we need a "map" that relates the parts to one another while attaching numbers to the concepts. These are the functions of the integrating tables, and there should be more of them. The volume can improve its function as a lexicon by bringing all of the book up to the best practice it has already established.

The functioning of the *Statistical Abstract* as a reference merits some study by the editors. It is puzzling to find data on voting and registration, but not on the political affiliation of the population; such data are available from academic surveys going back to the 1940s. It is puzzling to find data that religious bodies report regarding their finances and numbers of communicants, but none of the excellent survey data on the religious affiliation of the population and the characteristics of religious groups. The absence of such data cannot reflect an aversion to nongovernmental data sources, since such sources are already employed in this book. Nor can this absence reflect the *Statistical Abstract*'s commitment to data on the social, economic, and political *organization* of the United States; few features are more central to our political and social structure. In recent years the Reagan administration has proposed to increase the National Science Foundation's social and economic science budget (which it once slashed), primarily on the grounds that the surveys it supports furnish a continuing data base that is indispensable to federal policymaking. The National Elections Study, the General Social Survey, the Income Dynamics Survey, and kindred programs deserve another look from the editors.

ROBERT PARKE
National Cancer Institute

E. T. Jaynes: Papers on Probability, Statistics, and Statistical Physics.

R. D. Rosenkrantz (ed.). Dordrecht, the Netherlands: D. Reidel, 1983. xxiv + 434 pp. $49.50.

This book contains 13 articles written by E. T. Jaynes over a span of 25 years (1957–1980). They are reprinted in chronological order, and Jaynes has added introductory comments to each of them.

Having the collection of Jaynes's papers and his comments in one volume is of great value for statisticians. It will help to disseminate Jaynes's ideas, which are not yet well known by most statisticians because Jaynes is (primarily) a physicist and his articles have usually appeared in journals not traditionally read by researchers in statistics.

The articles in the book provide a review of the author's necessarian approach (L. J. Savage's label) to the theory of probability. Jaynes, as the other necessarian statisticians (e.g., H. Jeffreys) do, proposes a subjective approach to inference, based on a nonfrequentist concept of probability, but an objective method for assigning prior distributions. The priors are derived from general principles by logical analysis of prior information. This is viewed by the personalists (e.g., the L. J. Savage school) as trying to get something from nothing. In this book, Jaynes very clearly shows with several examples what you give and what you get with the necessarian position.

The best known and most controversial of the necessarian principles is undoubtedly the maximum entropy principle (MEP), of which Jaynes is a major advocator. In the discrete case the MEP states that the a priori distribution (P_1, P_2, . . . , P_n) that is maximally noncommital with the missing information maximizes the entropy, that is,

$$-\sum_{i=1}^{n} P_i \log P_i,$$

subject to the statistical conditions (e.g., a priori information) we wish to retain. If we have no a priori information on the P_i's other than $\sum P_i = 1$, then the ME distribution is the uniform $P_i = 1/n$. Hence the MEP generalizes the old principle of insufficient reason of Bernoulli and Laplace.

I think that one of the greatest successes of the nonfrequentist theory of probability has, paradoxically, been in physics. With respect to his application of the MEP to statistical mechanics, Jaynes writes:

> The conclusion, it seemed to me, was inescapable. We can have our justification for the rules of statistical mechanics, in a way that is incomparably simpler than anyone had thought possible, if we are willing to pay the price. The price is simply that we must loosen the connections between probability and frequency, by returning to the original viewpoint of Bernoulli and Laplace. (p. 236).

In fact this new freedom obtained from the nonfrequentist theory of probability permitted extending the range of validity of statistical mechanics to irreversible processes.

Two groups of papers in the book are clearly recognizable—that is, those whose subject is statistical mechanics and those dealing with statistics and probability theory. The former group is outside of the general background knowledge of the average statistician. However, all of the fundamental probabilistic concepts found in the statistical-mechanics papers are explained and extended in the latter group. On the other hand, it is a must for any serious worker in the foundations of statistics (Bayesian or not) to know for its historical and practical importance the fundamental concepts of statistical mechanics. For them, Jaynes's papers on the subject, reprinted in this book, will add to the work of Maxwell, Boltzman, and Gibbs.

One of Jaynes's major achievements has been to recognize the importance of Shannon's work on information theory in the foundations of statistical mechanics and probability. In his first paper, "Information Theory and Statistical Mechanics I," [ITSM(I)], Jaynes shows that it is possible to reobtain all of Gibbs's distributions from the dynamics of the thermodynamic system and the MEP only, without using additional ergodic assumptions. A great part of this scholar's life has been spent in clarifying and extending the revolutionary ideas presented in ITSM(I) and its sequel, ITSM(II), and this is the subject of the book.

Even though Jaynes starts by considering the MEP as fundamental, he then recognizes the "problem invariances" (i.e., symmetries) and, later, marginalization theory as more basic. In the introductory comment to his 1963 Brandeis lectures, Jaynes says:

> The Brandeis lectures have also the recognition that the continuous information measure as given by Shannon was not derived by him from any desideratum, but only written down by analogy with the discrete measure; and if we derive it by a limiting process from the discrete case there is an extra term $m(x)$. Of course, expressions of this type had been given three years earlier by Kullback, and sixty years earlier by Gibbs, but they were not given this motivation: Recognition of $m(x)$ restored the invariance of the theory under parameter changes, which had been a minor problem for Shannon but a major one for us. (p. 39)

The desideratum for consistency (mentioned before) for the continuous information measure has recently been found by Shore and Johnson (1980). Their work has rooted Jaynes's theory on even more solid ground.

Jaynes's paper on marginalization and prior probabilities deserves special recognition among statisticians. He clearly shows that the marginalization paradoxes of Dawid, Stone, and Zidek (1973) cannot be used to blame the improper priors considered there. Hence, even though there is still no "proper" mathematical theory of how to handle these "Jeffreys' beasts," we cannot use Dawid, Stone, and Zidek's work (as many still think) to get rid of them!

In summary, this book presents E. T. Jaynes's "objective Bayesian" theory, which is built on group invariance, marginalization theory, and the maximum entropy principle. Jaynes's writing style is clear and always to the point, and the mathematical level of his statistical papers should present no problem to readers familiar with calculus and elementary probability theory. The book also contains a supplementary bibliography with enough items to show that the maximum entropy formalism has been successful in attacking problems in a variety of subjects, including biology, chemistry, communication theory, decision making, econometrics and forecasting, engineering and operations research, geology and geophysics, spectral analysis, statistics, thermodynamics, and traffic networks. We can now add to the list nonparametric probability density estimation (see Rodriguez 1984).

R. D. Rosenkrantz has beautifully summarized his view of Jaynes' ideas. He writes:

> The problems with which he (Jaynes) has grappled—the Bertrand paradoxes, the marginalization paradoxes of statistical theory, and the seemingly intractable problems of irreversible thermodynamics—have withstood the efforts of many powerful minds (and given many others an attack of vertigo!). These are not puzzles that beckon one with a promise of easy gold at the hand of fairy or elf. One must plunge into murky depths and risk lying suspended indefinitely in an agony of confusion. Yet, in every case, Jaynes has managed to lay hold of a constructive principle that can steer us towards the light. The fruits of such hard-won gains always go beyond the mere harvesting of new scientific findings or the forging of powerful new instruments of inquiry; rather, advances at this fundamental level advance our understanding of thinking itself. (pp. viii–ix)

I agree, and I heartily recommend this book to everyone.

CARLOS C. RODRIGUEZ
State University of New York at Albany

REFERENCES

Dawid, A. P., Stone, M., and Zidek, J. (1973), "Marginalization Paradoxes in Bayesian and Structural Inference," *Journal of the Royal Statistical Society*, Ser. B, 35, 189–233.

Rodriguez, C. (1984), "Maximum Entropy Histograms," unpublished Ph.D. dissertation, State University of New York at Stony Brook.

Shore, J. E., and Johnson, R. W. (1980), "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Transactions on Information Theory*, 26, 26–37.

Planning and Analysis of Observational Studies.

William G. Cochran. New York: John Wiley, 1983. xiii + 145 pp. $21.95.

The planning and analysis of observational studies is central to many areas of the social sciences and medicine, and yet compared to the design and analysis of both experiments and surveys, the statistical literature on the subject is comparatively slight. The reason for this is almost certainly that the central problem is intractable, has not yielded to systematic advances, and lacks a framework for statistical inference comparable to both experiments and surveys. Although those who analyze data from observational studies will always be left with the possibility of incorrect inference due to uncontrolled sources of influence on the variables of interest, this fact should strengthen the case for a thorough methodology and appropriate statistical practice, not diminish it.

Throughout his career W. G. Cochran made contributions to the design and analysis of observational studies, and this volume was edited by L. E. Moses and F. Mosteller from Cochran's almost-completed manuscript after his death. From the editors' introduction it is clear that much of the material is just as Cochran wrote it and the book is characterized by the clear, lucid style that we associate with Cochran's writing. There are, however, disadvantages in making as little change as possible. The two that are the most conspicuous are the incomplete Chapter 7 on simple study structures and the lack of references to the most recent developments. (There are very few references to work after 1973.) The lack of recent references presumably occurred because the manuscript was written over a number of years, and as Cochran's health failed there were few additions.

The contents are organized as follows:

Chapter 1, "Variation Control and Bias." An introduction is given to basic statistical ideas in drawing conclusions from experimental data, the principal sources of variation, and methods of control. Thus blocking, matching, and covariate adjustment are first introduced, as well as sampled and target populations, concepts that reappear throughout the book.

Chapter 2, "Statistical Introduction." An elementary account of the statistical methods used in all statistical studies is presented and then emphasis is placed on observational studies and the problem of bias due to lack of experimental control.

Chapter 3, "Preliminary Aspects of Planning." This is a brief account and yet full of insight and sound advice on the planning phase of observational studies.

Chapter 4, "Further Aspects of Planning." The technical part of this chapter is largely concerned with sample-size determination corresponding to tests and estimation. The ideas are extended to a components-of-variance model for two-stage sampling for clusters. All of this is as relevant to other types of studies as it is to observational studies. One brief section puts these results in the context of the bias that may come from observational studies. The final sections of the chapter are nontechnical and contain further insights and advice on nonresponse and other methodological issues.