The Annals of Statistics
1976, Vol. 4, No. 4, 817–820

BOOK REVIEW

Y. M. M. BISHOP, S. E. FIENBERG AND P. W. HOLLAND, *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, 1975, x+557 pp. \$30.00.

Review by SHELBY J. HABERMAN
University of Chicago

Within the last fifteen years, the development of log-linear models has led to major advances in the statistical analysis of frequency data. Literature related to log-linear models has proliferated in journals, especially since 1968. Recently, books on the subject have begun to appear. One such book is *Discrete Multivariate Analysis: Theory and Practice*.

Discrete Multivariate Analysis is an ambitious attempt to present log-linear models to a broad audience. Exposition is quite discursive, and the mathematical level, except in Chapters 12 and 14, is very elementary. To illustrate possible applications, some 60 different sets of data have been gathered together from diverse fields. To aid the reader, an index of these examples has been provided.

Bishop, Fienberg and Holland provide a thorough discussion of a number of important topics in contingency table analysis rather than a comprehensive survey of log-linear models. Three important restrictions have been imposed. All log-linear models considered are hierarchical, iterative computations always use the Deming–Stephan (1940) iterative proportional fitting algorithm, and matrix inversions are never used to find asymptotic variances. These restrictions are closely related.

To indicate the nature of log-linear models that are hierarchical, consider an $r \times c$ contingency table

$$\mathbf{n} = \{n_{ij} : 1 \leqq i \leqq r, 1 \leqq j \leqq c\}$$

such that $\mathbf{n}$ is a multinomial vector with sample size N and positive cell probabilities $\mathbf{p} = \{p_{ij}\}$. Let $\mathbf{m} = N\mathbf{p}$ be the vector of expected cell frequencies. Then the vector $\boldsymbol{\mu} = \{\log m_{ij}\}$ satisfies the equation

$$\mu_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}, \qquad 1 \leqq i \leqq r, 1 \leqq j \leqq c,$$

for some u, $\mathbf{u}_1 = \{u_{1(i)} : 1 \leqq i \leqq r\}$, $\mathbf{u}_2 = \{u_{2(j)} : 1 \leqq j \leqq c\}$, and $\mathbf{u}_{12} = \{u_{12(ij)} : 1 \leqq i \leqq r, 1 \leqq j \leqq c\}$ such that

$$\textstyle\sum_i u_{1(i)} = \sum_j u_{2(j)} = \sum_i u_{12(ij)} = \sum_j u_{12(ij)} = 0 .$$

In a log-linear model, $\boldsymbol{\mu}$ is assumed to be in a known linear subspace Ω. Given the assumption that $\mathbf{n}$ has a multinomial distribution, the requirement is imposed that the unit vector $\mathbf{e}$ is in $\boldsymbol{\Omega}$, where e_{ij} is 1 for $1 \leqq i \leqq r$, $1 \leqq j \leqq c$.

Five basic hierarchical log-linear models exist: (1) No restrictions are imposed on $\mathbf{u}_1$, $\mathbf{u}_2$ or $\mathbf{u}_{12}$; (2) $\mathbf{u}_{12}$ is assumed $\mathbf{0}$; (3) $\mathbf{u}_1$ and $\mathbf{u}_{12}$ are assumed $\mathbf{0}$; (4) $\mathbf{u}_2$

and $\mathbf{u}_{12}$ are assumed $\mathbf{0}$; or (5) $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_{12}$ are all assumed $\mathbf{0}$. The term "hierarchical" arises since the higher-order term $\mathbf{u}_{12}$ is set to $\mathbf{0}$ whenever the lower-order term $\mathbf{u}_1$ is set to $\mathbf{0}$ or the lower-order term $\mathbf{u}_2$ is set to $\mathbf{0}$. The authors also note that additional hierarchical models can be obtained by ignoring some cells in the table or by using alternate systems of indexing. Although valuable, these hierarchical models provide no means to exploit such features of contingency tables as ordered categories. More general log-linear models are available which are specifically intended for ordered tables. For example, consider the logit model described by Finney (1971) in the context of biological assay. In this model, each column j has a corresponding score v_j. To simplify discussion, let $c \geqq 3$, let each v_j be distinct, and let

$$\sum v_j = 0 .$$

The logit

$$\lambda_j = \log (p_{1j}/p_{2j}) = \mu_{1j} - \mu_{2j} = \alpha + \beta v_j$$

for some unknown α and β. Equivalently, one may assume that

$$u_{12(ij)} = \tfrac{1}{2}\beta v_j .$$

Thus a restriction is placed on $\mathbf{u}_{12}$; however, $\mathbf{u}_{12}$ is not assumed to be zero.

One motivation for this restriction to hierarchical models is computational. Given the limitation to hierarchical log-linear models, it is possible to use iterative proportional fitting to compute all maximum likelihood estimates. Were more general log-linear models considered, more complex procedures such as the Newton–Raphson algorithm would be necessary.

In typical analyses involving log-linear models, estimated asymptotic variances and covariances of maximum likelihood estimates are obtained by matrix operations including inversions or solution of simultaneous linear equations. Such calculations are a by-product of computations performed with the Newton–Raphson algorithm; however, iterative proportional fitting provides no assistance in computation of estimated asymptotic variance and covariances. As a consequence, Bishop, Fienberg and Holland give somewhat less attention to estimation of parameters, estimation of asymptotic variances, and construction of approximate confidence intervals than to fitting models to data. This choice of emphasis may lead some readers astray. Typically, it is not enough to fit the data; in real problems, values of parameters and some assessment of accuracy of estimates are desired.

Despite the restrictions the authors have imposed on themselves, they do consider a wide variety of topics, including in later chapters some subjects not directly related to log-linear models. After a brief introductory chapter, Chapters 2, 3 and 4 provide a detailed discussion of complete factorial tables. Chapter 2 provides a general introduction to log-linear models. Chapter 3 explores computation of maximum likelihood estimates. Closed-form expressions are provided when available, and iterative proportional fitting is introduced.

Chapter 4 examines formal and informal tests of fit. Chi-square tests, residual analysis, and standardized rates are considered. Chapter 5 examines incomplete factorial tables. The methods of Chapter 5 are then used in Chapter 6 to develop methods for using a multiple-recapture census to estimate the size of a closed population. Chapter 7 explores Markov chain models, and Chapter 8 considers models for square contingency tables which exploit symmetry. Chapter 9 considers some special problems involved in fitting models.

The last five chapters in the book are not primarily concerned with log-linear models. Chapter 10 is an examination of alternate approaches to those used in the rest of the book. Methods briefly described include information-theoretic approaches, minimum chi-square approaches, probit models, logit models, multivariate logit models and exact procedures. References are provided for all methods discussed. No mention is made of simultaneous confidence intervals, and latent-structure models are mentioned but not discussed. Chapter 11 discusses measures of association and agreement for two-way tables.

Compared to previous chapters, Chapter 12 is much more theoretically oriented and far more mathematically demanding. Chapter 12 reflects Fienberg and Holland's work toward the development of contingency table estimation procedures analogous to the James–Stein estimates used in multivariate analysis. Unlike other chapters, this chapter relies on a specific loss function used in the context of decision theory. Statisticians skeptical concerning use of loss functions are as likely to question the value of the procedures advocated in this chapter as they are likely to question other statistical procedures based on decision theory.

Chapter 13 provides a review of sampling models for discrete data, and Chapter 14 provides a detailed discussion of asymptotic theory. The discussion is well done but most of the material is more appropriate in a general work on mathematical statistics than in an otherwise elementary exposition of methods of contingency table analysis.

Two problems may cause difficulties for some readers. A number of inaccurate statements are present, and the authors have frequently failed to cite sources of results or sources of related material. The book's value would be increased if these problems were alleviated in future editions.

For most statisticians, the book's treatment of degrees of freedom for incomplete multiway tables presents the greatest practical difficulties. Some related problems also arise in Chapter 3 when the authors consider adjustments of degrees of freedom for tables with cell counts equal to zero. To illustrate the problem, let $\mathbf{n} = \{n_{ijk}: 1 \leqq i \leqq 2, 1 \leqq j \leqq 2, 1 \leqq k \leqq 2\}$. Assume that $\mathbf{n}$ is multinomial with expected value $\mathbf{m} = \{m_{ijk}\}$. Let

$$\begin{aligned} m_{ijk} &= 0\,, \quad i = j = k = 1\,, \\ &= 0\,, \quad i = j = k = 2\,, \\ &> 0\,, \quad \text{otherwise.} \end{aligned}$$

Assume that if $(i, j, k) \neq (1, 1, 1)$ and $(i, j, k) \neq (2, 2, 2)$, then

$$\mu_{ijk} = \log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} ,$$

where

$$\textstyle\sum_i u_{1(i)} = \sum_j u_{2(j)} = \cdots = \sum_j u_{23(jk)} = \sum_k u_{23(jk)} = 0 .$$

If a chi-square test of this log-linear model is made, then the correct degree of freedom is zero. On page 218, the authors provide rules for computation of degrees of freedom which indicate that there are -1 degrees of freedom.

The authors make some errors in Section 3.5 when dealing with computational procedures. Contrary to assertion, the Newton–Raphson algorithm is determined by the sufficient statistics of the log-linear model. In addition, incomplete proofs are provided on pages 85–87 and 93–94 for convergence of the iterative proportional fitting algorithm.

On page 2, the authors state, "Finally, of course, we expect the book to provide a reference source for the methods collected in it." This goal is not always achieved. Citation practices vary considerably in quality depending on the particular chapter involved. With a few exceptions, citations are carefully provided in Chapters 6–14; however, citations in Chapters 2–5 are often omitted. Consequently, the interested reader often has little access to further details or to alternate treatments of the subject.

Despite the reservations expressed concerning accuracy and citation practices, it should be emphasized that the book contains a wealth of material on important topics. Its numerous examples are especially valuable. Because of limitations imposed by emphasis on iterative proportional fitting, the book does not constitute a comprehensive survey of methods for contingency table analysis; however, its contents should prove useful for many researchers. To date, it is the most ambitious attempt to present log-linear models to a general audience.

REFERENCES

[1] DEMING, W. E. and STEPHAN, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **11** 427–444.

[2] FINNEY, D. J. (1971). *Probit Analysis*, 3rd ed. Cambridge Univ. Press.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
1118 EAST 58TH STREET
CHICAGO, ILLINOIS 60637