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) $\mathbf{u}_2$