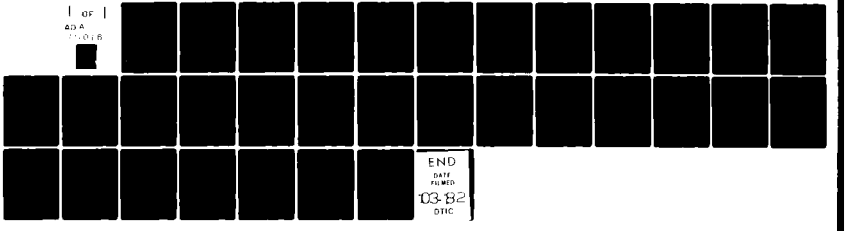
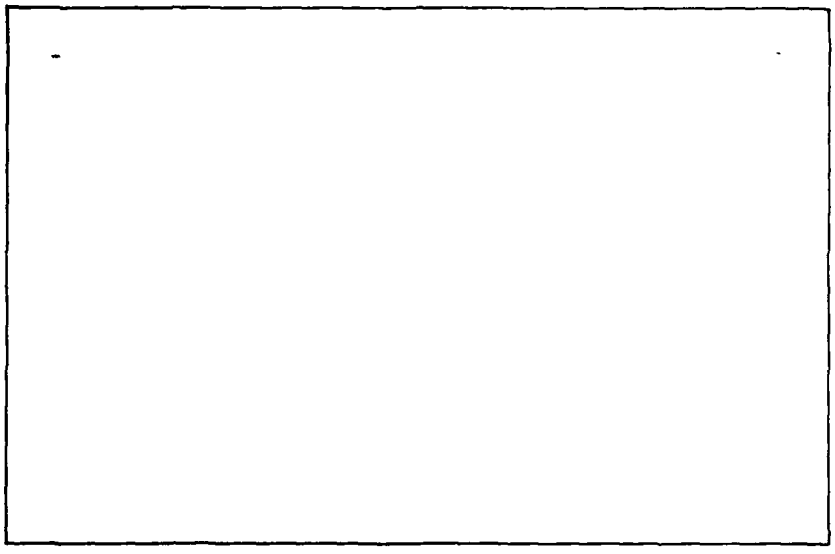
AD-A111 016 CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF STATISTICS F/G 12/1
RECENT ADVANCES IN THEORY AND METHODS FOR THE ANALYSIS OF CATEG--ETC(U)
DEC 81 S E FIENBERG N00014-80-C-0637
UNCLASSIFIED TR-229 NL



LEVEL II (1)

AD A111016

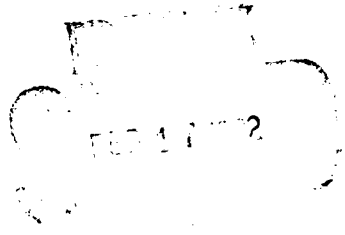


(12) 33

DEPARTMENT

OF

STATISTICS



DTIC FILE COPY

391140

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

82 02 17 061

JJ

RECENT ADVANCES IN THEORY AND METHODS FOR THE ANALYSIS OF CATEGORICAL DATA: MAKING THE LINK TO STATISTICAL PRACTICE

by

Stephen E. Fienberg

Departments of Statistics and Social Science
Carnegie-Mellon University
Pittsburgh, PA 15213, USA

Technical Report No. 229

December 1981

This paper is the text of the R.A. Fisher Lecture, presented at the 43rd Session of the International Statistical Institute in Buenos Aires, Argentina, November 30 - December 11, 1981, in a session named after recent Honorary Presidents of the ISI. It's preparation was supported in part by Office of Naval Research Contract N00014-80-C-0637 to Carnegie-Mellon University.

RECENT ADVANCES IN THEORY AND METHODS FOR THE ANALYSIS OF CATEGORICAL DATA: MAKING THE LINK TO STATISTICAL PRACTICE

STEPHEN E. FIENBERG

Departments of Statistics and Social Science
Carnegie-Mellon University
Pittsburgh, PA 15213, USA

Tell me whereon the *likelihood* depends.

Wm. Shakespeare
As You Like It
Act I, Scene 3, 56.

Life is the art of drawing *sufficient* conclusions from insufficient premises.

Samuel Butler
Notebooks

1. INTRODUCTION

It is a great honor to present a lecture named after Sir R.A. Fisher, especially at a session of the International Statistical Institute, an organization on whose behalf he expended so much energy. Fisher was one of the most productive and original statisticians of this century, and much of modern statistical theory and methods has its origins in his work. This is especially true of the current methods for the analysis of categorical data via loglinear models, the topic of my lecture.

The work I shall describe has as its foundation Fisher's notions of "likelihood" and "sufficiency," and the general theory for loglinear models is intimately linked to results for exponential families, that are implicit in some of Fisher's most profound theoretical papers. Amongst Fisher's contributions to statistical methodology are several papers on contingency table analysis and the distribution of chi-square statistics (see Fienberg, 1980a for a discussion of this work). These, along with Fisher's observations in other papers and suggestions by Fisher to his colleagues, serve as the precursors to the more general results that have been the focus of attention in recent years.

Fisher was not simply a great statistician, he was also a great scientist. And he worked hard at translating his theoretical statistical results into practical methods, of use to biologists and agricultural scientists with whom he worked. For example, it was for them that Fisher wrote Statistical Methods for Research Workers, a book that has served as a statistical bible for statisticians and non-statisticians alike, since it was first published in 1925. Thus, in the spirit of Fisher's own work, I shall discuss not only the basic statistical theory for the analysis of categorical data using loglinear models, but also the implications of this theory for general statistical practice in the reporting of tabular materials, and some of the exciting new substantive areas where the theory is currently being put to practice.

Sir R.A. Fisher was elected a member of the International Statistical Institute in 1931. Beginning at the end of World War II, he worked with Stuart A. Rice to revitalize and reorganize the Institute, which had been dominated up to that time by Europeans and by government statisticians. Over the next 11 years, Fisher struggled to open up the ISI membership to research statisticians and to integrate their activities with those statisticians of other persuasions. In her biography of Fisher, his daughter (Box, 1978) chronicles these activities, and quotes from a letter he wrote in 1956, as follows:

> We really have a terrifically long way to go in making the Institute as useful as it could be, since I think the great majority of our foreign membership quite take it for granted that it is primarily an assembly of officials concerned with national statistics, vital and economic, and of their more academic economic advisers. These people cannot deny the importance of mathematical statistics . . . and if we put in undeniably good mathematicians who insist on talking of the natural sciences and in terms of scientific research and holding sessions relevant to the applications of mathematical statistics to scientific research, we have done a pretty good generation's work.
>
> Box (1978, p.433).

Fisher was not completely successful in these attempts, but he continued to work on ISI activities, and participate in its meetings. In recognition of his many contributions, the

Institute elected Fisher as an honorary member in 1950, and along with P.C. Mahalanobis as Honorary President in 1957 (only two others had been previously so honored). Even in his "retirement," Fisher travelled to Japan to attend the 1960 ISI meetings, and to Paris to attend the 1961 meetings, the last ones held before his death in 1962.

The next section outlines the statistical theory for loglinear models in the analysis of categorical data, and links it to the more general theory of exponential families. We focus there on maximum likelihood estimation, its use of minimal sufficient statistics, and methods for assessing the goodness of fit of a model. Section 3 briefly describes the application of loglinear model methods for the analysis of multi-dimensional contingency tables, and then takes the form of an aside. In it we discuss the implications of loglinear model theory for the reporting of results from large-scale government sample surveys, especially in the form of tables of cross-classified counts. In Section 4, we turn to the applications of the results of Section 2 to "non-contingency table" problems in (a) the Bradley-Terry paired comparisons model, (b) the analysis of social networks, and (c) the use of the Rasch model in intelligence testing and its potential for innovative survey analysis. In each case, the non-contingency table problem is transformed and is re-represented as a problem in contingency-table form, whose solution has been studied previously.

Much of modern statistical practice relies heavily on the computational implementation of methodology. In Section 5 of this paper, we briefly summarize the state of the art of computation for loglinear model methods, and mention some topics of current research activity that may allow these methods to be of greater practical use in the future.

2. LOGLINEAR MODELS AND EXPONENTIAL FAMILY THEORY

The analysis of categorical data, focuses on the fitting of models to collections of counts, often fashioned into the format of cross-classifications or contingency tables. For purposes of describing the loglinear model approach to such analyses we will consider a vector of observed counts falling into t cells.

$$\mathbf{x} = (x_1, x_2, \ldots, x_t). \tag{2.1}$$

These counts are realizations of a set of random variables

$$\mathbf{X} = (X_1, X_2, \ldots, X_t). \tag{2.2}$$

whose expectations and log-expectations are

$$\mathbf{m} = (m_1, m_2, \ldots, m_t) \tag{2.3}$$

and

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_t). \tag{2.4}$$

where

$$m_i = E(X_i) \quad i = 1,2,\ldots,t. \tag{2.5}$$

and

$$\lambda_i = \log m_i \quad i = 1,2,\ldots,t. \tag{2.6}$$

For the 2x2 contingency table $t=4$, and the observed counts are often displayed as

$$\begin{matrix} x_1 & x_2 \\ x_3 & x_4 \end{matrix}$$

Two basic sampling models, for probability distributions for the random variables X, have been the focus of attention in the literature on the analysis of contingency tables.

A. The Poisson Model. If the $\{x_i\}$ are observations from independent Poisson distributions the probability density or likelihood function is given by

$$\prod_{i=1}^{t} \frac{m_i^{x_i} e^{-m_i}}{x_i!} \tag{2.7}$$

This model can be thought of as appropriate when the counts represent the simultaneous record of t Poisson processes, observed for a fixed period of time.

B. Product-multinomial Model. Now suppose we partition the set of t cells into r sets, J_k, where the kth set contains t_k cells and

$$t = \Sigma_{k=1}^{r} t_k . \tag{2.8}$$

Then if the counts in these sets are observations from r independent multinomial distributions, the sums

$$n_k = \Sigma_{i \in J_k} X_i \qquad k = 1,2,\ldots,r. \tag{2.9}$$

are fixed by design. The probability density or likelihood function for this general situation is

$$\Pi_{k=1}^{r} \binom{n_k}{x_i} \Pi_{i \in J_k} \left(\frac{m_i}{n_k}\right)^{x_i} . \tag{2.10}$$

subject to the constraints

$$\Sigma_{i \in J_k} m_i = n_k \qquad \text{for } k = 1,2,\ldots,r. \tag{2.11}$$

Each of the constraints in (2.11) can be characterized by a vector whose components are 1 if $i \in J_k$ and 0 otherwise.

When r = 1, we have observations from a single multinomial. When r = 2 and t = 4, we have observations from two binomials. Thus, the product-multinomial includes two of the most widely used sampling models for the 2x2 table, i.e. the two-binomial model, and the single four-cell multinomial model.

Both the Poisson and product-multinomial sampling models, are special cases of the exponential family of distributions, introduced first by Fisher in his 1934 invited address to the Royal Statistical Society (Fisher, 1935), and elaborated upon by Darmois, Koopman, and Pitman. The general form of the exponential family density (e.g. see Andersen, 1980 or Barndorff-Nielsen, 1978) is

$$f(t_1,t_2,\ldots,t_p \mid \theta_1,\theta_2,\ldots,\theta_p) = [c(\theta_1,\theta_2,\ldots,\theta_p)]^{-1} \exp\{\Sigma_{i=1}^{p} \theta_i t_i\}\ h(t_1,t_2,\ldots,t_p) . \tag{2.12}$$

Both (2.7) and (2.10) can be written in this form, with $t_i = x_i$ and $\theta_i = \lambda_i$, although (2.10) is subject to the constraints (2.11) leading to the use of adjusted θ_i's based on the differences of λ_i's (for details, see Andersen, 1980, pp. 20-27). Exponential family theory suggests that the log-expectations λ should be the key parameters of interest. By reexpressing the λ_i's as linear functions of a reduced number of parameters, we arrive at the notion of loglinear models for the two basic sampling models.

A well-known result in basic probability, exploited by Fisher in much of his work on categorical data problems, links the Poisson and product-multinomial models:

RESULT 1. Suppose that $\mathbf{X}$ follows the Poisson sampling model. Then the conditional distribution of $\mathbf{X}$, given the restrictions (2.9), is that of the product multinomial in (2.10).

To specify a class of loglinear models, for the vector of expectations, $\mathbf{m}$, we need to specify a linear subspace of the t-dimensional space in which the vector of logexpectations, λ, lies. Call this subspace M (for model!). Thus we can represent the components of λ as linear combinations $\lambda = \lambda(\theta)$ of newly defined parameters θ, and we preserve the exponential family structure of (2.12). We now turn to the problem of maximum likelihood estimation of the loglinear parameters θ, and of $\lambda = \lambda(\theta)$ itself.

The following general results on maximum likelihood estimation for θ were originally developed by Birch (1963), and later extended by Bishop (1969), Haberman (1974), and others. They turn out to be special cases of more general results for exponential families as has been noted by Dempster (1971) and others.

RESULT 2. Corresponding to each parameter in θ there is a minimal sufficient statistic that is expressible as a linear combination of the $\{x_i\}$. (More formally, if M is used to denote the loglinear model specified by $\mathbf{m} = \mathbf{m}(\theta)$, then the MSS's are given by the projection of $\mathbf{x}$ onto M, i.e. $P_M\mathbf{x}$.)

RESULT 3. The maximum likelihood estimate under the Poisson model, $\hat{\mathbf{m}}$ of $\mathbf{m}$ = exp $\lambda(\theta)$, if it exists, is unique and satisfies the likelihood equations:

$$P_M\hat{\mathbf{m}} = P_M\mathbf{x} \ . \tag{2.13}$$

i.e. the MLE is found by setting the minimal sufficient statistics equal to their expectations.

We note that the MLE $\hat{\theta}$ of θ is defined implicitly via the MLE $\hat{\mathbf{m}}$ of $\mathbf{m}$ = exp $\lambda(\theta)$ in expression (2.13). In the statement of Result 3, we assume that $\hat{\mathbf{m}}$ exists. Necessary and sufficient conditions for the existence of MLE's are relatively complex, and we refer the interested reader to Haberman (1974) for details.

For product-multinomial sampling situations, the basic multinomial constraints (i.e., that the counts must add up to the multinomial sample sizes) must be taken into account. Thus we need to ensure that the constraints (2.11) are in fact satisfied. To do so, we let M_* be a loglinear model for m under product-multinomial sampling which

corresponds to a loglinear model M under Poisson sampling, such that the multinomial constraints, (2.11) "fix" a subset of the parameters, θ, used to specify M. Then

RESULT 4. The MLE of $\mathbf{m}$ under product-multinomial sampling for the model M_* is the same as the MLE of $\mathbf{m}$ under Poisson sampling for the model M.

Result 4 follows directly from Results 1, 2, and 3, and forms the basis of the unified approach to loglinear model problems, with and without multinomial constraints, as described in Bishop, Fienberg, and Holland (1975). Woolson and Brier (1981) show that a similar result holds for estimates of $\mathbf{m}$ (and thus θ) derived using the weighted least squares approach of Grizzle, Starmer, and Koch (1969). The key to the result in both cases is the loglinear structure of the parametric model, and the exponential family representation of the sampling model.

It is interesting to note that Fisher implicitly exploited Result 4 in his discussion of the degrees of freedom of the Pearson chi-square statistic for 2xN contingency tables (Fisher, 1922b). The generalization of Fisher's formulation of the chi-square problem has led to the following well-known theorem.

RESULT 5. If $\hat{\mathbf{m}}$ is the MLE of $\mathbf{m}$ under a loglinear model, and if the model is correct, then the statistics

$$X^2 = \Sigma_{i=1}^{t} (x_i - \hat{m}_i)^2 / \hat{m}_i \tag{2.14}$$

and

$$G^2 = 2 \Sigma_{i=1}^{t} x_i \log (x_i / \hat{m}_i) \tag{2.15}$$

have asymptotic χ^2 distributions with $t-s$ degrees of freedom, where s is the total number of independent constraints implied by the loglinear model and the multinomial sampling constraints, (2.11) (if any). If the model is not correct then X^2 and G^2, in (2.14) and (2.15), are stochastically larger than χ^2_{t-s}.

In Result 5, X^2 is the usual Pearson χ^2 statistic for testing goodness of fit, and G^2 is minus twice the loglikelihood ratio comparing the restricted model $\mathbf{m} = \exp \lambda(\theta)$ to the unrestricted model. Fisher (1922a) had noted the asymptotic equivalence of X^2 and G^2 in certain situations, and suggested that the Pearson statistic X^2 achieved its validity because it is an approximation to the loglikelihood ratio statistic.

3. LOGLINEAR MODELS, MARGINAL TOTALS, AND THE REPORTING OF SURVEY DATA

The loglinear model theory described in the preceding section was developed primarily to deal with the analysis of multidimensional cross-classified tables of counts. In this section, we review how the results of Section 2 can be applied to such tables, and in the course of doing so we draw conclusions about the reporting of large scale national probability samples of the type carried by government agencies and others around the world.

We begin with a simple biomedical example. An experiment was designed to study the effects of two analgesic drugs on post-partum pain of women who had experienced normal deliveries. A total of 718 women were studied and they were assigned to one of four treatment groups:

A_1B_1 - 0 dosage of drug A and drug B, i.e. placebo
A_1B_2 - 100 mg. of drug B
A_2B_1 - 200 mg. of drug A
A_2B_2 - 200 mg. of drug A and 100 mg. of drug B.

The outcome variable for the study was reduction of pain (or change):

C_1 - no reduction
C_2 - reduction.

The resulting data form the 2x2x2 cross-classification given in Table 3-1, part (a).

TABLE 3-1
The Results of an Experiment Involving Two Analgesic Drugs Intended to Reduce Post-Partum Pain

(a) observed counts: $\{x_{ijk}\}$

Level of Drug B	Level of Drug A	Pain Change C_1	Pain Change C_2	Totals
	A_1	55	115	170
B_1	A_2	44	132	176
	A_1	33	154	187
B_2	A_2	25	160	185
			Grand Total	718

(b) estimated expected counts, $\{\hat{m}_{ijk}\}$ under model (3.2) and (3.3) subject to constraints (3.1).

Level of Drug B	Level of Drug A	C_1	C_2	Totals
	A_1	54.7	115.3	170
B_1	A_2	44.3	131.7	176
	A_1	33.3	153.7	187
B_2	A_2	24.7	160.3	185

For the data in Table 3-1, the totals for AxB are fixed by design (the totals differ somewhat from one another due to the manner in which the study was conducted). We are interested in the effects of drugs A and B on the response variable C. Let

x_{ijk} = no. women in group A$_i$B$_j$ who respond C$_k$.

Then the two-way totals, adding over k, are fixed, i.e.

$$m_{ij+} = x_{ij+}, \qquad i,j = 1,2. \tag{3.1}$$

where a "+" implies summation over the corresponding subscript. Expression (3.1) corresponds to the product-multinomial constraints (2.11).

One possible model for the data of Table 3-1 is

$$\log \frac{m_{ij2}}{m_{ij1}} = w + w_{1(i)} + w_{2(j)}, \tag{3.2}$$

where

$$\Sigma_i w_{1(i)} = \Sigma_j w_{2(j)} = 0. \tag{3.3}$$

Model (3.2) is referred to as a *logit* model and it postulates the additive effects of drugs A and B on the logarithm of the odds of pain change $m_{ij2} : m_{ij1}$. Using Result 4 of Section 2, we can also represent the logit model of (3.2) equivalently as a loglinear model for m_{ijk}, i.e.

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}, \tag{3.4}$$

with the usual ANOVA constraints that whenever a u-term is summed over a subscript the sum equals zero, e.g.

$$\Sigma_i u_{1(i)} = \Sigma_{i=j} u_{12(ij)} = 0. \tag{3.5}$$

Since (3.5) is subject to the constraints of equation (3.1), u, $\{u_{1(i)}\}$, $\{u_{2(j)}\}$, and $\{u_{12(ij)}\}$ are in effect fixed by design, while

$$w = 2\,u_{3(1)}, \qquad w_{1(i)} = 2\,u_{13(i1)}, \text{ and } w_{2(j)} = 2\,u_{23(j1)}. \tag{3.6}$$

The minimal sufficient statistics for model (3.2) (or (3.4) subject to (3.1)) are the three sets of two-way marginal totals:

$$\{x_{ij+}\}, \quad \{x_{i+k}\}, \quad \{x_{+jk}\}. \tag{3.7}$$

and, using Result 3, the likelihood equations are:

$$\hat{m}_{ij\cdot} = x_{ij\cdot} \qquad i,j = 1,2,$$
$$\hat{m}_{i\cdot k} = x_{i\cdot k} \qquad i,k = 1,2,$$
$$\hat{m}_{\cdot jk} = x_{\cdot jk} \qquad j,k = 1,2. \tag{3.8}$$

The solution to the likelihood equations does not have a closed-form expression and some form of numerical technique is required, such as iterative proportional fitting (e.g. see Andersen, 1980; Bishop, Fienberg, and Holland, 1975; or Haberman, 1974, 1978).

Part (b) of Table 3-1 displays the MLE's, $\{\hat{m}_{ijk}\}$, for our example. The goodness-of-fit statistics, (2.14) and (2.15), take values

$$X^2 = 0.014, \qquad G^2 = 0.014,$$

with 1 d.f. Comparing these values with various tail values of the χ^2_1 distribution, we see that model (3.2) fits the data *extremely* well. Thus the summary of the 2x2x2 array $\{x_{ijk}\}$ in terms of the minimal sufficient statistics (3.7) is a meaningful one. By reporting only the two-way marginal totals, we provide others with "sufficient information" to estimate the parameters of interest. In fact, reduced models also fit the data in Table 3-1 extremely well, and thus we can express the "sufficient information" even more compactly.

The ideas just described in the context of the 2x2x2 table generalize in a straightforward fashion to loglinear models for tables of more than 3 dimensions. Suppose we are interested in reporting the results of a national simple random sample of adults, age 25 or older, conducted to provide information on the interrelationship between educational achievement (variable 1 measured in terms of 4 categories), and occupational satisfaction (variable 2 with 3 categories), and how it varies with sex (variable 3 with 2 categories) and ethnic origin (variable 4 with, say, 8 categories). We have a single multinomial sample, but the models of interest are ones that condition on the "background variables," sex and ethnic origin. Thus, in analyzing the resulting 4x3x2x8 cross-classification, we would focus on models conditional on

$$m_{\cdot\cdot ks} = x_{\cdot\cdot ks} \qquad \begin{matrix} k = 1,2, \\ s = 1,2,\ldots,8. \end{matrix} \tag{3.9}$$

An example of a loglinear model for the array of expected cell counts $\{m_{ijks}\}$ is

$$\log m_{ijkl} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)}$$
$$+ u_{13(ik)} + u_{14(il)} + u_{23(jk)} + u_{24(jl)} + u_{34(kl)}. \qquad (3.10)$$

This model postulates simultaneous interrelationships between each of the two "response" variables (variables 1 and 2) and each of the two "explanatory" variables (variables 3 and 4), as well as between the two explanatory variables themselves. This model does not include any of the four terms that are interpretable as second-order interactions involving 3 variables, nor does it include the 4-variable, third-order interaction. Models containing such terms might be of interest to us, however, as they share with (3.10) several desirable features from the viewpoint of reporting of survey results.

For loglinear models of the sort being considered here, the minimal sufficient statistics always take the form of sets of marginal totals. In our particular example, they are the five two-dimensional marginal tables corresponding to the five two-factor terms in the model: the marginal tables for educational achievement by sex, $\{x_{+k+l}\}$, corresponding to $\{u_{34(kl)}\}$; educational achievement by ethnic group, $\{x_{++jl}\}$, corresponding to $\{u_{24(jl)}\}$; occupational satisfaction by sex, $\{x_{+i+l}\}$, corresponding to $\{u_{14(il)}\}$; occupational satisfaction by ethnic group, $\{x_{+ijl}\}$, corresponding to $\{u_{23(jk)}\}$; and sex by ethnic group, $\{x_{++k+}\}$, corresponding to $\{u_{13(ik)}\}$. If we were to report only these five two-way tables (*along with a description of our model*) then it would be possible for a reader with appropriate statistical training to construct a four-dimensional table sufficiently close to the observed table that he would suffer essentially zero information loss (in the Fisherian sense), provided that the model fits the data.

The implications of the use of loglinear models for the analysis and reporting of multidimensional cross-classified survey data are thus relatively clear:

(1) By the use of model building we are often led to particular forms of summary appropriate for our data.

(2) In the case of cross-classified data and loglinear models this summary takes the form of certain sets of marginal totals, specified by the model.

(3) If we report all of the marginal totals appropriate for a loglinear model that fits the data well, then another investigator can, in effect, reconstruct the data with little or no loss in information.

Few government or other survey organizations adopt such a model-based approach to analysis and reporting, and we are usually left to ponder the relevance of tables that are reported.

The approach to reporting just described for survey-based cross-classified data assumed that we are dealing with either a simple random sample, or perhaps with a stratified random sample, where the variables underlying the strata (if they have any intrinsic interest) are included amongst the explanatory variables in the loglinear models. The analysis and reporting of categorical data from sample designs involving clustering or unequal probabilities of selection is more complex (see e.g., Brier, 1980; Fellegi, 1980; and Rao and Scott, 1981), but the principles behind the reporting remain the same. We should not report summaries of a survey involving categorical variables only in a form which prevents others from reconstructing what is essentially an equivalent version of the original data or some subset thereof (i.e. summaries that do not include an appropriate set of minimal sufficient statistics). This is the type of practical advice that I believe Fisher might have given had he been more extensively involved in the analysis of survey data!

4. THE USE OF LOGLINEAR MODELS FOR SOME "NON-CONTINGENCY" TABLE PROBLEMS

The application of the loglinear model results from Section 2 to multidimensional contingency tables focussed on models where each set of the parameters in the logarithmic scale is associated with one or more dimensions of the table. One of the values of general theoretical results is that they are often applicable to specific settings beyond those which led to the formulation of the general structure. This is certainly true for results on the analysis of categorical data problems. Fortunately many of the "non-contingency table" applications of the loglinear model results have contingency table-like representations so that we can *interpret* the results of our analyses using whatever intuition we have gleaned from the analysis of contingency table data using loglinear models.

4.1 THE BRADLEY-TERRY PAIRED COMPARISONS MODEL

To illustrate this approach let us consider the Bradley-Terry model for binary paired comparisons, a statistical topic which has been studied extensively for almost three decades (for an excellent review of this literature see Bradley, 1976). Suppose t items (e.g., different types of chocolate pudding) or treatments, labeled T_1, T_2,.....T_t, are compared in pairs by sets of judges. (Or suppose that t football teams compete in pairs in a series of matches.) The Bradley-Terry model postulates that the probability of T_i being preferred to T_j is

$$\Pr(T_i \text{ preferred to } T_j) = \frac{\pi_i}{\pi_i + \pi_j}, \qquad i,j = 1,2,\ldots,t, \quad i \neq j. \tag{4.1}$$

where each $\pi_i \geq 0$ and we add the constraint that $\sum_{i=1}^{t} \pi_i = 1$. The model assumes independence of the same pair by different judges and different pairs by the same judge. In the example of the football matches we assume the independence of outcomes of the matches.

TABLE 4-1
Layout for Data in Paired-Comparisons Study with $t = 4$

		Against			
		T_1	T_2	T_3	T_4
For	T_1	--	x_{12}	x_{13}	x_{14}
	T_2	x_{21}	--	x_{23}	x_{24}
	T_3	x_{31}	x_{32}	--	x_{34}
	T_4	x_{41}	x_{42}	x_{43}	--

In the typical paired comparison experiment, T_i is compared with T_j $n_{ij} \geq 0$ times, and we let x_{ij} be the observed number of times T_i is preferred to T_j in these n_{ij} comparisons. Table 4-1 shows the typical layout for the observed data when $t = 4$, with preference (for, against) defining rows and columns. Clearly the binomial constraint,

$$x_{ij} + x_{ji} = n_{ij}, \tag{4.2}$$

is of the form (2.9), and we can apply Result 4 of Section 2 to convert (4.1) into a model for expected values for a Poisson sampling setting, i.e.

$$\log m_{ij} = \alpha_i + \beta_j + \gamma_{ij} \tag{4.3}$$

where

$$\gamma_{ij} = \gamma_{ji}, \tag{4.4}$$

with suitable side constraints. But this, as was noted in Fienberg and Larntz (1976), is

simply the model of quasi-symmetry in a square contingency table (see Bishop, Fienberg, and Holland, 1975, Chapter 8). The minimal sufficient statistics are (from Result 1)

$$\{x_{i+}\}, \quad \{x_{+j}\}, \quad \{x_{ij} + x_{ji}\} \tag{4.5}$$

(actually either the row or column totals are redundant), and we can use a trick, suggested in Bishop, Fienberg, and Holland, to transform the problem to one for a three-way table of expected counts. We generate duplicate tables and set

$$m_{ijk} = \begin{cases} m_{ij} & k = 1, \\ m_{ji} & k = 2, \end{cases} \tag{4.6}$$

and, for the observed counts,

$$x_{ijk} = \begin{cases} x_{ij} & k = 1, \\ x_{ji} & k = 2. \end{cases} \tag{4.7}$$

Then the loglinear version of the Bradley-Terry model given by (4.3) and (4.4) becomes the model of no-second-order interaction in the new 3-dimensional table, whose minimal sufficient statistics are $(\{x_{ij+}\}, \{x_{i+k}\}, \{x_{+jk}\})$. Thus we can analyze the fit of the model and variations on it in a familiar contingency table setting of the sort described in Section 3.

These results on the loglinear representation for the Bradley-Terry model are by now reasonably well-known, and they can be extended to more complex settings involving ties, multiple comparisons, and rankings. Recent results by Meyer (1981) are of special use in given contingency table representations to some of these generalizations. For the remainder of this section we describe two other classes of categorical data problems where loglinear models are proving to be useful, and for which standard contingency table representations are especially helpful for both theoretical and computational reasons.

4.2. MODELS FOR SOCIAL NETWORKS

A directed graph consists of a set of g nodes, and a collection of directed arcs connecting pairs of nodes. Such graphs have been used to depict social networks describing relationships between pairs of individual actors. Figure 4-1 contains an example of such a graph for the relationship "social *friendship*," for 12 5th grade boys.

Each boy was asked to name the two boys with whom he was the friendliest outside the classroom. Table 4-2 summarizes the information from the directed graph of Figure 4-1 in the form of a 12x12 *sociomatrix* or *adjacency matrix*, $\mathbf{x}$, with elements

$$x_{ij} = \begin{cases} 1 & \text{if } i \text{ chooses } j \text{ as his friend} \\ 0 & \text{otherwise.} \end{cases} \tag{4.8}$$

where by convention, the diagonal terms $x_{ii} = 0$.

Holland and Leinhardt (1981) note that for any pair or *dyad* in a network, with adjacency matrix $\mathbf{x}$,

$$x_{ij}x_{ji} + x_{ij}(1-x_{ji}) + (1-x_{ij})x_{ji} + (1-x_{ij})(1-x_{ji}) = 1 , \tag{4.9}$$

for $i \neq j$, and that exactly one of the terms on the left hand side of (4.9) is 1 and the remaining three are 0. They then suggest the following model to describe these outcomes (using $\mathbf{X}$ as the matrix of random variables of which the adjacency matrix $\mathbf{x}$ is a realization):

$$\begin{aligned} \log \Pr[(1-X_{ij})(1-X_{ji}) = 1] &= \lambda_{ij} \\ \log \Pr[(1-X_{ij})X_{ji} = 1] &= \lambda_{ij} + \alpha_i + \beta_j + \theta \\ \log \Pr[X_{ij}(1-X_{ji}) = 1] &= \lambda_{ij} + \alpha_j + \beta_i + \theta \\ \log \Pr[X_{ij}X_{ji} = 1] &= \lambda_{ij} + \alpha_i + \alpha_j + \beta_i + \beta_j + 2\theta + \rho_{ij} , \end{aligned} \tag{4.10}$$

where the $\{\lambda_{ij}\}$ are "dyadic" effects included here (but only implicitly in Holland and Leinhardt) to assure that the multinomial constraint (4.9) is satisfied, and where

$$\Sigma_i \alpha_i = \Sigma_j \beta_j = 0 . \tag{4.11}$$

There are too many parameters in this model for complete identification, and so Holland and Leinhardt set

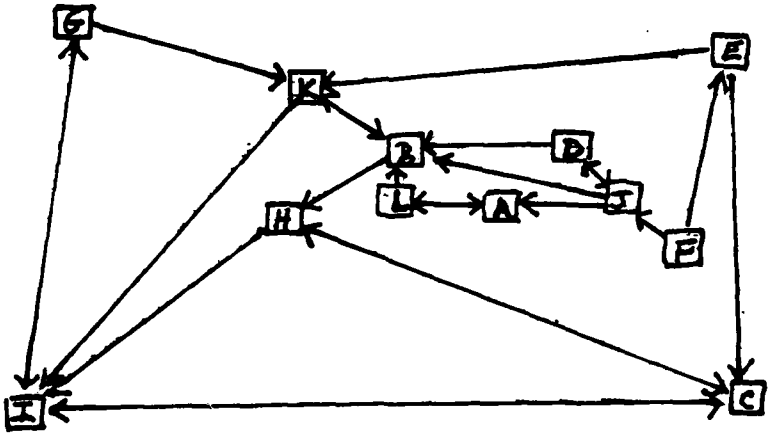
$$\rho_{ij} = \rho . \tag{4.12}$$

They refer to the resulting model as p_1.

TABLE 4-2
Sociomatrix for Social Friendship Among 12 5th Grade Boys

	A	B	C	D	E	F	G	H	I	J	K	L
A					1						1	
B									1	1		
C							1			1		
D									1		1	
F		1								1		
G				1					1			
I		1								1		
J		1					1					
K		1					1					
L	1			1								
M			1				1					
N	1	1										

FIGURE 4-1
Sociogram or Directed Graph Representing Data in Sociomatrix of Table 4-2



If we assume that the dyads are independent, then we have a product-multinomial sampling model with one observation per multinomial. (This model doesn't yet take into account the extra constraints in the data of Table 4-2 where the row sums of x are all restricted to equal 2). Holland and Leinhardt make direct use of the exponential family theory results on maximum likelihood estimation (c.f. Section 2) to estimate the parameters in p_1. Fienberg and Wasserman (1981a, 1981b) note, however, that there is a direct link between the p_1 model and a loglinear model for a multi-dimensional table representation of the probabilities in (4.10). In particular, they work with the four-dimensional array:

$$\begin{aligned} X_{ij11} &= X_{ij} X_{ji} \\ X_{ij10} &= X_{ij}(1-X_{ij}) \\ X_{ij01} &= (1-X_{ij})X_{ji} \\ X_{ij00} &= (1-X_{ij})(1-X_{ji}). \end{aligned}$$

(4.13)

Note that

$$X_{ijk\ell} = X_{ji\ell k} .$$

(4.14)

because the dyad (i,j) is the same as the dyad (j,i). Thus, if $\{x_{ijk\ell}\}$ is a realization of $\{X_{ijk\ell}\}$ we only need to consider one "triangle" of $\{x_{ijk\ell}\}$ in which $i > j$. But by retaining all $4g^2$ cells in the gxgx2x2 table we are able to express the minimal sufficient statistics for the parameters of p_1 as marginal totals of $\{x_{ijk\ell}\}$:

$$\begin{aligned} X_{++11} &= \sum_{i<j} X_{ij} X_{ji} \\ X_{i++1} &= X_{i+} \qquad i = 1,2,\ldots,g. \\ X_{+i+1} &= X_{+i} \qquad i = 1,2,\ldots,g. \\ X_{++10} &= X_{++} - X_{++11} . \end{aligned}$$

(4.15)

Finally, by coupling (4.15) with (4.9) and (4.14), and then reexpressing, we can get an alternative set of sufficient statistics:

$$\{x_{ij++}\}, \quad \{x_{i+k+}\}, \quad \{x_{+jk+}\}, \quad \{x_{i+++}\}, \quad \{x_{+ik+}\}, \quad \{x_{++k+}\} .$$

(4.16)

(allowing for redundances resulting from symmetries and duplications). But (4.16) and the set of six two-dimensional marginal totals of the four-dimensional array, and it can be shown (Meyer, 1981) that fitting p_1 to $x = \{x_{ij}\}$ is equivalent to fitting the no-second-order interaction model to the newly created redundant array $\{x_{ijk\ell}\}$.

This standard contingency table representation for Holland and Leinhardt's p_1 model leads to superior numerical solutions to the likelihood equations. It also leads naturally to a generalization of p_1 where

$$\rho_{ij} = \rho + \rho_i + \rho_j \qquad i > j. \tag{4.17}$$

Fitting this model to $\{x_{ij}\}$ is equivalent to fitting the standard loglinear model to $\{x_{ijk}\}$ with minimal sufficient statistics

$$\{x_{ij\cdot}\}, \quad \{x_{i\cdot k}\}, \quad \{x_{\cdot jk}\}. \tag{4.18}$$

We now return to the data in Table 4-2 on social friendships amongst 12 grade 5 boys, and recall that the row totals were fixed to equal 2, by design. This leads to a relatively complex hypergeometric sampling scheme, but we can approximate results for it by using the methods for p_1 just described and then focus only on the parameters $\{\beta_i\}$ and ρ. Our analysis of the data in Table 4-2 is relatively straightforward. Measuring the fit of Holland and Leinhardt's p_1 model using the likelihood ratio criterion of expression (2.15), we get $G^2_{p_1} = 104.15$ with 98 d.f. (The general formula for d.f. is $g(g-1)$ and $g = 12$, but we need to adjust here for the zero marginal total in the 6th column.) Next we fit the "differential reciprocity" model, (4.17), whose fitted is summarized by $G^2_{d\cdot} = 92.84$ with 87 d.f. (the d.f. calculation here is quite problematic, but the results do not depend on a precise calculation). Thus we can check on the fit of p_1 to the data in Table 4-2 by taking

$$\Delta G^2 = G^2_{p_1} - G^2_{d\cdot} = 11.31$$

with "approximately" 11 d.f. The p_1 model fits reasonably well. The boys who attract the most friendship (e.g. boys 2, 3, 9, 10, and 11) do not appear to *reciprocate* in a differential manner from those who attract little friendship, given that we adjust for their differing levels of attractiveness.

What is especially attractive about the multi-dimensional contingency table representation of the social network data problem as outlined here is that it carries over to networks involving multiple relationships. For details, see Fienberg, Meyer, and Wasserman (1981). Yet this type of representation is not a panacea. The sparseness of the array $\{x_{ijk}\}$ makes the application of the usual asymptotics, and in particular Result 5 of Section 2, problematic at best. The array $\{x_{ijk}\}$ is of size $4g^2$ but $x_{\cdot\cdot\cdot} = 2g(g-1)$, and the p_1 model has $2g$ parameters. For a more detailed discussion of the relevant asymptotics for this problem see Fienberg and Wasserman (1981a) and Haberman (1981).

4.3 THE RASCH MODEL

We now turn to yet another problem which begins with a representation as a two-way table of 0's and 1's, and ends up as a relatively standard multi-dimensional contingency table problem. The results of *ability tests* are often structured in the form of sequences of 1's for correct answers and 0's for incorrect answers. For a test with k problems or items administered to n individuals, we let

$$Y_{ij} = \begin{cases} 1 & \text{if individual } i \text{ answers item } j \text{ correctly} \\ 0 & \text{otherwise.} \end{cases}$$

(4.19)

Thus we have a two-way table of random variables $\{Y_{ij}\}$ with realizations $\{y_{ij}\}$. An alternative representation of the data is in the form of a $n \times 2^k$ table $\{W_{i j_1 j_2 \ldots j_k}\}$ where the subscript i still indexes individuals and now $j_1, j_2, \ldots, j_k$ refer to the correctness of the responses on items 1,2,....k, respectively, i.e.

$$W_{i j_1 j_2 \ldots j_k} = \begin{cases} 1 & \text{if } i \text{ responds } (j_1, j_2, \ldots, j_k) \\ 0 & \text{otherwise.} \end{cases}$$

(4.20)

The Rasch model (Rasch, 1960 as reprinted in 1980; Birnbaum, 1957) for the $\{Y_{ij}\}$ is

$$\log \frac{P(Y_{ij}=1)}{P(Y_{ij}=0)} = \gamma + \mu_i + \nu_j ,$$

(4.21)

where

$$\Sigma \mu_i = \Sigma \nu_j = 0 .$$

(4.22)

Differences of the form $\mu_i - \mu_r$ are typically described as measuring the relative abilities of individuals i and r, while those of the form $\nu_j - \nu_s$ are described as measuring the relative difficulties of items j and s. Expression (4.21) is a *logit* model in the usual contingency table sense for a 3-dimensional array whose first layer is $\{y_{ij}\}$ and whose marginal totals adding across layers is an nxk table of 1's. Because the Rasch model depends on the item parameters in a non-linear way, it is not at all clear whether we can collapse the array $\{w_{i j_1 j_2 \ldots j_k}\}$ by adding over subjects for estimation purposes. We return to this matter below.

Duncan (1982) has proposed that we should view certain types of survey data in much the same way as we do ability test data. For example, he describes a 4-item scale included in a survey pertaining to beliefs about effects of marijuana. If we can consider these items in isolation from the rest of the survey questions (see the discussion of this in Section 3 on reporting), then we can display the relevant data as an nx4 array of the form (4.19), and we can explore the appropriateness of the Rasch model as a description of the observed data. In the context of Duncan's examples the individual parameters, $\{\mu_i\}$, can be thought of as values for a "latent trait" of the survey respondents in much the same way as psychometricians have intepreted these parameters as measuring the single latent trait, ability. Duncan discusses the matter, not considered here, of structuring the μ_i's according to multiple dimensions, and he links the notion of background variables and stratification to differing latent trait structures.

Maximum likelihood estimation for the parameters of the Rasch model (4.21) has been the focus of several authors including Rasch and Andersen. Unconditional maximum likelihood (UML) estimates can be derived but they have rather problematic asymptotic properties, e.g. the estimates are inconsistent as $n \to \infty$ and k remains moderate, although they are consistent when both n and $k \to \infty$ (Haberman, 1977).

Before turning to an alternative to the UML approach, we point out a recently-derived result for UML estimates for the Rasch model which links up in yet another way with loglinear structures for contingency tables. In order to derive necessary and sufficient conditons for the existence of UML estimates (a problem not really discussed for any of the data structures in this paper), Fischer (1981) embeds the matrix $\mathbf{y} = \{y_{ij}\}$ into a larger (n+k)x(n+k) matrix of the form:

$$\mathbf{A} = \{a_{ij}\} = \begin{bmatrix} 0 & \mathbf{e}-\mathbf{y}' \\ \mathbf{y} & 0 \end{bmatrix} \tag{4.23}$$

where e is an nxk matrix of 1's, so that, for all (i,j),

$$a_{ij} = a_{ji} = 1 \, . \tag{4.24}$$

Then he notes that the Rasch model of (4.21) is transformed into an incomplete version of the Bradley-Terry model of expression (4.1) discussed at the beginning of this section, i.e.

$$P(a_{ij}=1) = \frac{\pi_i}{\pi_i + r} \qquad \begin{array}{l} i = k+1,\ldots,k+n. \\ j = 1,2,\ldots,k. \end{array} \tag{4.25}$$

and similarly for the other non-zero block of entries in **A**, where

$$\log \frac{\pi_i}{\pi_r} = \mu_i - \mu_r \qquad \begin{array}{l} i,r = 1,2,\ldots,n. \\ i \neq r. \end{array} \tag{4.26}$$

and

$$\log \frac{\pi_j}{\pi_s} = \nu_j - \nu_s \qquad \begin{array}{l} j,s = 1,2,\ldots,k. \\ j \neq s. \end{array} \tag{4.27}$$

Thus, using a three-dimensional representation for **A** alluded to at the beginning of this section, we can show that estimation results for the UML approach to the Rasch model correspond to those of for the no-second-order interaction model applied to an incomplete three-dimensional contingency consisting of two zero blocks of dimension kxkx2 and nxnx2, and a duplicated version of the nxkx2 table with layers y_{ij} and $e - y_{ij}$.

Now, we turn to a conditional approach to likelihood estimation (CML) advocated initially by Rasch, who noted that the conditional distribution of **Y** given the individual marginal totals $\{y_{i\cdot} = y_{i\cdot}\}$ depends only on the item parameters, $\{\nu_j\}$. Then each of the row sums $\{y_{i\cdot}\}$ can take only k+1 distinct values corresponds to the number of correct responses. Next, we recall the alternate representation of the data in the form of an $nx2^k$ array, $\{W_{j_1 j_2 \ldots j_k}\}$, as given by expression (4.20). Adding across individuals we create a 2^k contingency table, **X**, with entries

$$\mathbf{X}_{j_1 j_2 \ldots j_k} = W_{j_1 j_2 \ldots j_k}. \tag{4.28}$$

Earlier, we asked the question of whether we could work with this collapsed array. The answer is yes, since all of the information we need to preserve is the response pattern, i.e. $\{j_1, j_2, \ldots, j_k\}$, and the number of "correct" responses that correspond to that pattern. Such information allows us to completely reconstruct the original matrix of responses, **Y**, except for the labelling of individuals, and thus we can use the 2^k array **x** to represent the conditional distribution of **X** given $\{Y_{i\cdot} = y_{i\cdot}\}$.

Duncan (1982) and Tjur (1981) independently noted that we can estimate the item parameters for the Rasch model of (4.21) using the 2^k array **x**, and the loglinear model

$$\log m_{j_1 j_2 \ldots j_k} = u + \sum_{i=1}^{k} \delta(j_i, 1) u_i + u_{j_1 j_2 \ldots j_k} \quad (4.29)$$

where the subscript $j_i = \sum_{i=1}^{k} j_i$, $\delta(j_i, 1) = 1$ if $j_i = 1$ and is 0 otherwise, and

$$\sum_{j_i=0}^{1} u_{j_1 j_2 \ldots j_k} = 0 \quad (4.30)$$

The *amazing* result, due to Tjur (1981), is that maximum likelihood estimation of the 2^k contingency table of expected values, $\mathbf{m} = \{m_{j_1 j_2 \ldots j_k}\}$ using a Poisson sampling scheme and the loglinear model (4.29), produces the conditional maximum likelihood estimates of $\{u_i\}$ for the original Rasch model. Tjur proves this equivalence by (1) assuming that the individual parameters are independent identically distributed random variables from some completely unknown distribution, π; (2) integrating the conditional distribution of $\mathbf{Y}$ given $\{Y_{i+}=y_{i+}\}$ over the mixing distribution, π; (3) embedding this "random effects" model in an "extended random model"; and (4) noting that the likelihood for the extended model is equivalent to that for (4.29) applied to $\mathbf{x}$ (using Result 4 of Section 2 above).

TABLE 4-3
Multiplicative Representation of Expected Values of Model (4.29) for the Case k = 3

		Item C: Yes		Item C: No	
		Item A		Item A	
		Yes	No	Yes	No
Item B	Yes	$abcS_3$	abS_2	bcS_2	bS_1
	No	acS_2	aS_1	cS_1	S_0

For k=3, the loglinear version of the Rasch model for the 2^3 table, i.e. (4.29), can be represented in multiplicative form for the expected values $\mathbf{m}$ as in Table 4-3. The minimal sufficient statistics are

$$\{x_{i+}\}, \quad \{x_{+j}\}, \quad \{x_{++k}\} \quad (4.31)$$

and

$$\{x_{111}, \quad x_{110} + x_{101} + x_{011}, \quad x_{100} + x_{010} + x_{001}, \quad x_{000}\} \quad (4.32)$$

But these are the minimal sufficient statistics of the model of quasi-symmetry preserving one-dimensional marginal totals which was first proposed by Bishop, Fienberg, and Holland (1975, Chapter 8). Indeed, that model is equivalent to (4.29).

Thus following the prescription of Bishop, Fienberg, and Holland (1975, p.305), we can re-represent the data in a 4-dimensional redundant form (as a 2x2x2x6 table) and estimate the Rasch model item parameters using a standard loglinear model fitted to a 4-way table (although not the 4-way table **w** of expression (4.20)). Additional simplifications ensue here because

$$\hat{m}_{111} = x_{111} ,$$
$$\hat{m}_{0jk} = x_{0jk} .$$

(4.33)

Plackett (1981), in a very brief section of the 2nd edition of his monograph on categorical data analysis, notes that the Q-statistic of Cochran (1950) can be viewed as a means of testing that the item parameters in the Rasch model are all equal and thus zero, i.e. $\lambda_j = 0$ for all j. This observation is intimately related to the results just described, and our original data representation in the form of an nxk (individual by item) array **y** is exactly the same representation used by Cochran. By carrying out a conditional test for the equality of marginal proportions given model (4.29), i.e. quasi-symmetry preserving one-dimensional marginals, we get a test that is essentially equivalent to Cochran's test. But this is also the test for $\{\lambda_j = 0\}$ within model (4.29).

Duncan (1982) gives several examples of the application of the Rasch model to survey research problems, and he presents several extensions of the model, indicating how they can be represented in a multi-dimensional table format such as that of Table 4-3.

5. COMPUTATION FOR LOGLINEAR MODEL METHODS

As we noted in Section 3 on multi-dimensional contingency tables, we do not necessarily get closed-form estimates of the MLE's $\hat{\mathbf{m}}$ of the expected counts. Thus some form of iterative numerical procedure is often required. The most popular numerical procedure for calculating MLE's is the method of *iterative proportional fitting* (IPFP): which iteratively adjusts the entries of a contingency table to have marginal totals specified by the likelihood equations.

To illustrate the algorithm we consider a three-way table of independent Poisson counts, $\mathbf{x} = \{x_{ijk}\}$. Suppose we wish to fit the loglinear model of no-second-order interaction for the mean **m**, i.e. the model given by expression (3.4). The basic IPFP takes an initial table $\mathbf{m}^{(0)}$, such that log $(\mathbf{m}^{(0)})$ satisfies the model (typically we would use $m^{(0)}_{ijk} = 1$ for all i,j, and k) and sequentially scales the current fitted table to satisfy the three sets of the two-way margins of the observed table, **x**. The νth iteration consists of three steps which form:

$$m^{t,1}_{i+} = m^{t-1,3}_{i+} \cdot A_{i+} / m^{t-1,3}_{i+} \,,$$

$$m^{t,2}_{ij} = m^{t,1}_{ij} \cdot A_{ij} / m^{t,1}_{ij} \,,$$

$$m^{t,3}_{ijk} = m^{t,2}_{ijk} \cdot A_{ijk} / m^{t,2}_{ijk} \,,$$

(5.1)

(The first superscript refers to the iteration number, and the second to the step number within iterations). The algorithm continues until the observed and fitted margins are sufficiently close. For a detailed discussion of convergence and some of the other properties of the algorithm, see Bishop, Fienberg and Holland (1975) or Haberman (1974).

Common alternatives to the IPFP are versions of Newton's method or other algorithms which use information about the second derivatives of the likelihood function. While such methods have quadratic convergence properties compared to the linear properties of the IPFP, and are often quite efficient (see e.g. Haberman (1974), or Fienberg, Meyer and Stewart (1979)), they are of limited use for models of high dimensionality because of storage requirements. Newton's method also automatically produces an estimate of the variance-covariance matrix of the parameters, but this is what requires all of the storage space. *Currently, the most* widely-used computer program that employs a Newton-like algorithm is GLIM, which is distributed by the Numerical Algorithms Group of the United Kingdom (Baker and Nelder, 1978).

Recent research on numerical procedures for maximum likelihood estimation in loglinear models has focussed on alternative algorithms that will handle the types of large data arrays that arise in practical problems. For example, Fienberg, Meyer, and Wasserman (1981) describe an application of the social network methodology of Section 4.3 in which the basic data consist of *three correlated* 73x73 adjacency matrices. We briefly outline three different approaches that have been proposed to handle large data arrays.

One approach to increasing the storage capacity of current problems is found in work in progress by Fienberg, Meyer, and Stewart (1981), who have been developing programs for both loglinear and logit models using a variant of Newton's method. Their algorithms involve the construction of the upper half of a pxp weighted cross-product matrix where p is the dimension of the parameter vector θ, and take full advantage of the sparseness of the nxp design matrix without actually constructing it. The algorithms proceed via Newton's method with variable step length, using a Cholesky decomposition

with pivoting. A special feature of these algorithms is a subroutine that checks for the existence of MLE's, $\hat{m}$, by performing a pivoted Cholesky decomposition on a substantially reduced problem. It should be possible to use these algorithms, when they become available, as replacements for the Newton-like algorithms in programs such as GLIM.

McIntosh (1981) has proposed the use of yet another alternative to IPFP, the method of *conjugate gradients*. Unlike Newton's method which uses the full matrix of second derivatives of the likelihood function, the method of conjugate gradients works by carrying out an "optimal" sequence of one-dimensional maximizations. The method of conjugate gradients has storage requirements similar to that of IPFP, but has "superlinear" convergence properties. McIntosh (1981) provides numerical comparisons of different algorithms for several contingency table examples, but these fail to demonstrate the areas of superiority of the current versions of his conjugate gradient algorithms, which have been implemented within GLIM.

Finally, we note the recent work of Meyer (1981), who considers generalizations of IPFP due to both Haberman (1975) and Csiszar (1975). Meyer has developed a new method for *estimating* MLE's that is especially attractive for large problems and which combines the advantages of both Newton's method and IPFP. Basically, his approach is to break the large problem into manageable but overlapping subproblems. Then he iterates in an IPFP-like manner amongst the subproblems, for each of which he uses Newton's method.

All of the computational approaches just discussed are currently under active development. We expect that these and other efforts will ultimately expand the scope and size of categorical data problems that can be analyzed using loglinear model methods.

6. CONCLUDING REMARKS

In this lecture I have examined a variety of categorical data problems using models that are linear in the logarithms of the expected cell values. The methods and models are linked to a small core of theoretical statistical results depending on exponential family theory, and the concepts of minimal sufficient statistics and maximum likelihood estimation. All of these results have as their foundation research work of Sir R.A. Fisher.

The building of bridges from statistical theory to statistical practice is an activity which Fisher thought to be especially appropriate for ISI Meetings. I hope that many of you will have crossed such a bridge with me today, and in the process gained an appreciation for the richness of the theoretical results on loglinear models for categorical data analysis, and the many different practical areas to which they may be applied.

7. ACKNOWLEGEMENTS

The preparation of this paper was partially supported by Office of Naval Research Contract N00014-80-C-0637 at Carnegie-Mellon University. Reproduction in whole or part is permitted for any purpose of the United States Government. Michael Meyer provided helpful comments on an earlier draft, as well as computational assistance. Much of the ideas on transforming tables in Section 4 was stimulated by his work. I am also indebted to Dudley Duncan whose ideas on the use of the Rasch model are reflected in Section 4.3. Sam Leinhardt provided assistance in producing Figure 4-1, and Margie Krest prepared the manuscript. Finally I am indebted to Fred Mosteller who introduced me to the world of contingency tables 15 years ago, with the specific task of working on the reporting problem.

BIBLIOGRAPHY

Andersen, E.B. (1980). *Discrete Statistical Models with Social Science Applications.* Amsterdam: North Holland.

Baker, R.J. and Nelder, J.A. (1978). *The GLIM System, Release 3 Manual.* Oxford, England: Numerical Algorithms Group.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* New York: Wiley.

Birch, M.W. (1963). "Maximum likelihood in three-way contingency tables." *J. Roy. Statist. Soc. B*, **25**, 229-233.

Birnbaum, A. (1957). "On the estimation of mental ability." Report No. 15. Randolph Air Force Base, USAF School of Aviation Medicine.

Bishop, Y.M.M. (1969). "Full contingency tables, logits, and split contingency tables." *Biometrics*, **25**, 383-400.

Bishop, Yvonne, M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass.: MIT Press.

Box, J.F. (1978). *R.A. Fisher, The Life of a Scientist.* New York: Wiley.

Bradley, R.A. (1976). "Science, statistics, and paired comparisons." *Biometrics*, **32**, 213–232.

Brier, S.S. (1980). "Analysis of contingency tables under cluster sampling." *Biometrika*, **67**, 591–596.

Cochran, W.G. (1950). "The comparison of percentages in matched samples." *Biometrika* **37**, 256–66.

Csiszar, I. (1975). "I-divergence geometry of probability distributions and minimization problems." *The Annals of Probability*, **3**, 146–158.

Dempster, A.P. (1971). "An overview of multivariate data analysis." *J. Mult. Anal.*, **1**, 316–346.

Duncan, O.D. (1982). "Rasch measurement in survey research: further examples and discussion." To appear, in C.F. Turner and E. Martin (eds.) *Survey Measurement of Subjective Phenomena*, Vol. 2. (forthcoming).

Fellegi, I.P. (1980). "Approximate tests of independence and goodness-of-fit based on stratified multistage samples." *J. Amer. Statist. Assoc.* **75**, 261–268.

Fienberg, S.E. (1980a). "Fisher's contributions to the analysis of categorical data." In S.E. Fienberg and D.V. Hinkley (eds.) *R.A. Fisher: An Appreciation*. Lecture Notes in Statistics, **1**. New York: Springer-Verlag, 75–84.

Fienberg, S.E. (1980b). *The Analysis of Cross-Classified Categorical Data* (2nd ed.). Cambridge, Mass.: The MIT Press.

Fienberg, S. E. and Larntz, K. (1976). "Loglinear representation for paired and multiple comparisons models." *Biometrika*, **63**, 245–254.

Fienberg, S.E., Meyer, M.M., and Stewart, G.W. (1979). "Alternative computational methods for estimation in multinomial logit response models." Technical Report 348, School of Statistics, University of Minnesota.

Fienberg, S.E., Meyer, M.M., and Stewart, G.W. (1979). "The numerical analysis of contingency table data." Unpublished manuscript.

Fienberg, S.E., Meyer, M.M., and Wasserman, S.S. (1981). "Analyzing data from multivariate directed graphs: an application to social networks." In Vic Barnett (ed.), *Looking at Multivariate Data*, Chichester, England: Wiley, 289–306.

Fienberg, S.E. and Wasserman, S.S. (1981a). "Comment on 'An exponential family of probability distributions for directed graphs'." *J. Amer. Statist. Assoc.* **76**.

Fienberg, S.E. and Wasserman, S.S. (1981b). "Categorical data analysis of single sociometric relations." In S. Leinhardt (ed.) *Sociological Methodology 1981*, San Francisco: Jossey-Bass, 156–192.

Fischer, G.H. (1981). "On the existence and uniqueness of maximum-likelihood estimates in the Rasch model." *Psychometrika*, **46**, 59-78.

Fisher, R.A. (1922a). "On the mathematical foundations of theoretical statistics." *Phil Trans A*, **222**, 309-368.

Fisher, R.A. (1922b). "On the interpretation of χ^2 from contingency tables, and the calculation of p." *J. Roy. Statist. Soc.*, **85**, 87-94.

Fisher, R.A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Fisher, R.A. (1935). "The logic of inductive inference." *J. Roy Statist. Soc.* **98**, 39-54.

Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). "Analysis of categorical data by linear models." *Biometrics*, **25**, 489-504.

Haberman, S.J. (1974). *The Analysis of Frequency Data*. Chicago: The University of Chicago Press.

Haberman, S.J. (1977). "Maximum likelihood estimates in exponential response models." *Ann. Statist.* **5**, 815-841.

Haberman, S.J. (1978). *Analysis of Qualitative Data, Volume 1 (Introductory Topics)*. New York: Academic Press.

Haberman, S.J. (1979). *Analysis of Qualitative Data, Volume 2 (New Developments)*. New York: Academic Press.

Haberman, S.J. (1981). "Comment on 'An exponential family of probability distributions for directed graphs.'" *J. Amer. Statist. Assoc.*, **76**, 60-61.

Holland, P.W. and Leinhardt, S. (1981). "An exponential family of probability distributions for directed graphs." *J. Amer. Statist. Assoc.*, **76**, 33-50.

McIntosh, A. (1981). *Fitting Linear Models: An Application of Conjugate Gradient Algorithms*. Ph.D. Thesis, Department of Statistics, University of Toronto.

Meyer, M.M. (1981). *Applications and Generalizations of the Iterative Proportional Fitting Procedure*. Ph.D. Thesis, School of Statistics, University of Minnesota.

Plackett, R.L. (1981). *The Analysis of Categorical Data*. 2nd edition. London: Griffin.

Rao, J.N.K., and Scott, A.J. (1981). "The analysis of categorical data from complex surveys: chi-squared tests for goodness of fit and independence in two-way tables." *J. Amer. Statist. Assoc.*, **76**, 221-230.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests.* (Reprinting of the 1960 book.) Chicago: Univ. of Chicago Press.

Tjur, T. (1981). "A connection between Rasch's item analysis model and a multiplicative Poisson model." *Scan. J. Statistics* (in press).

Woolson, R.F. and Brier, S.S. (1981). "Equivalence of certain chi-squared test statistics." *Amer. Statist.*, 35, 250–253.

SUMMARY

The past 20 years have seen an enormous growth in the statistical literature on the analysis of categorical data, much of it based on the use of loglinear models. This paper reviews some of the general results on maximum likelihood estimation for loglinear models and links them back to ideas that have their foundations in the work of Sir R.A. Fisher. These results have special relevance for the analysis of multidimensional contingency tables, and for the reporting of data from large-scale sample surveys. In addition, the results are applicable to other categorical data problems that are often representable in contingency table form. The paper concludes with a brief description of the state of the art of computation for loglinear model methods.

RÉSUMÉ

Les vingt années précédentes ont assisté à une croissance considérable de la littérature statisque traitant l'analyse des tables de contingence, souvent en utilisant des modèles log-linéaires. Cet article passe en revue quelques résultats généraux sur l'estimation maximum de vraisemblance pour les modèles log-linéaires, et les relie à des idées provenant de l'oeuvre de Sir R.A. Fisher. Ces résultats ont un rapport particulier à l'analyse des tables de contingences multidimensionelles, et au reportage des données d'enquêtes étendues. En plus, ces résultats peuvent servir à l'analyse d'autres données categoriques qui permittent une présentation tabulaire. L'article se conclut avec une courte description des méthodes numeriques utilisées à présent pour l'analyse des modèles log-linéaires.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical Report No. 229 — 2. GOVT ACCESSION NO.: AD-A111016	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): RECENT ADVANCES IN THEORY AND METHODS FOR THE ANALYSIS OF CATEGORICAL DATA: MAKING THE LINK TO STATISTICAL PRACTICE	5. TYPE OF REPORT & PERIOD COVERED: December, 1981
	6. PERFORMING ORG. REPORT NUMBER: TR #229
7. AUTHOR(s): Stephen E. Fienberg	8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0637
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA 15213	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Contracts Office, Carnegie-Mellon University, Pittsburgh, PA 15213	12. REPORT DATE
	13. NUMBER OF PAGES: 30
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

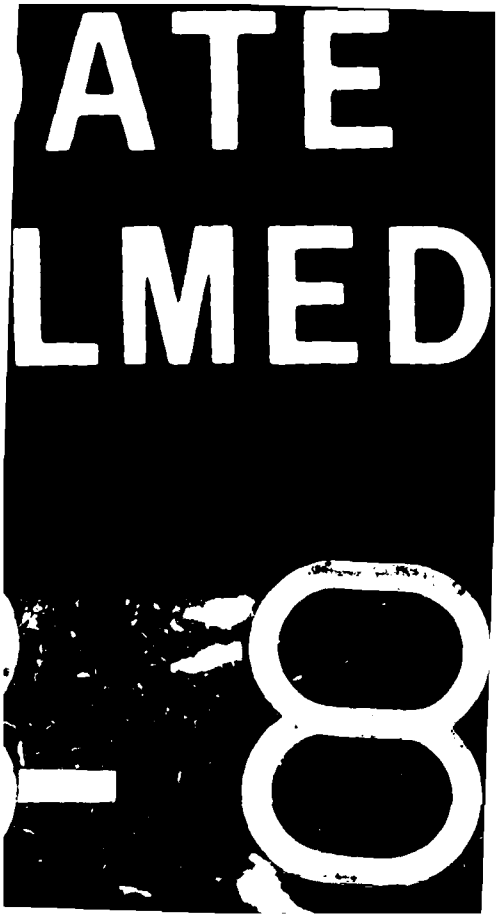
Loglinear models, sufficient statistics, maximum likelihood analysis, exponential family, Rasch model

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper reviews some of the general results on maximum likelihood estimation for loglinear models and links them back to ideas that have their foundations in the work of Sir R.A. Fisher. These results have special relevance for the analysis of multidimensional contingency tables, and for the reporting of data from large-scale sample surveys. In addition, the results are applicable to other categorical data problems that are often representable in contingency table form.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)


ATE
LMED
8