AN OVERVIEW OF DEPENDENCY MODELS FOR CROSS-CLASSIFIED CATEGORICAL DATA INVOLVING ORDINAL VARIABLES

BY RUTH DOUGLAS[1] AND STEPHEN E. FIENBERG[1]

Carnegie Mellon University

In the late 1970's the popularity of loglinear and logistic model techniques for cross-classified categorical data led to a resurgence of interest in models and methods which directly incorporate information about the ordinal structure of the categories corresponding to the classification variables. In this paper we present an overview of some of the models for dependence that have been the focus of interest in this recent literature. In particular, we consider a class of association models extensively developed by Goodman and we examine order restrictions on parameters corresponding to the ordinal structure of the underlying variables. We attempt to summarize what is known about how these order restrictions for association and other models characterize monotonicity constraints on the underlying cross-classification probabilities or marginal totals. The principle context for our discussion is the dependency structure for two-dimensional ordinal contingency tables, but extensions to multi-dimensional tables that build on loglinear model ideas are relatively direct.

1. Introduction. The study of dependence among continuous random variables has a long history in statistics. The corresponding issue regarding dependency for categorical random variables also has a long history going back to the work of Yule (1900) and Pearson (1900); it has only been since the 1960's that a coherent and elaborate literature has developed. Much of the emphasis before this period was on the development of measures of association (e.g., see Goodman and

[1]Supported in part by NSF Grant SES–8701606 to Carnegie Mellon University.

AMS 1980 subject classifications. Primary 62H20; secondary 62H17.

Key words and phrases. Association models; goodness-of-fit test statistics; likelihood estimation; monotonicity constraints; odds ratios.

A version of this paper was presented at the *Symposium on Dependence in Statistics and Probability* August 1-5, 1987. Zvi Gilula, Michael Meyer, and Allan Sampson provided ideas and materials that helped to shape our perspective on this topic.

Kruskal, 1954, 1979; Kruskal, 1958). In the mid-1960's attention of researchers interested in categorical data shifted to the study of loglinear models and most of the resulting literature was focused on the dependency or interaction structures for nominal categorical variables, i.e., with unordered categories. That literature does not exclude ordinal variables, and as was noted in Fienberg (1982), the ordinal nature of some categorical variables is often crucial to the structural organization of categorical data subjected to loglinear analysis (e.g., triangular arrays, social mobility tables, and tables representing age-period-cohort structures). The key feature of these models involving ordinal structures is that they are not permutation invariant, i.e., the categories of the variables cannot be permuted in an arbitrary way without affecting the parameters describing the dependency structures.

The standard loglinear model approach to two-dimensional $I \times J$ contingency tables represents the probability, P_{ij}, of an observation falling into the ith row and jth column as

$$\log P_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}, \tag{1}$$

where

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

and

$$\sum_i \sum_j P_{ij} = 1. \tag{3}$$

The model may be rewritten in multiplicative form as

$$P_{ij} = \alpha_i \beta_j \exp\{u_{12(ij)}\}. \tag{4}$$

The traditional approach for two-dimensional tables has been to treat the interaction terms, $u_{12(ij)}$, as being unrestricted (the so-called *saturated model*) or to set them equal to zero, thereby assuming that row classification is *independent* of column classification. Neither approach reflects any ordinal structure that may be present in the row or column categories.

In the 1970's two separate approaches to the study of dependency structures for ordinal variables emerged. The first of these was linked to the correspondence analysis approach developed by Benzécri and his associates (e.g. see Benzécri, et al. 1973; Greenacre, 1984) and focused on correlational-like ideas. This work was later picked up by Goodman (1981, 1985) and Gilula (1982). The second approach, proposed independently in the 1960's by Rasch (see Christiansen, 1966) and by Fienberg (1968), was developed extensively by Haberman (1974a, 1974b) and Goodman (1979, 1981, 1984, 1985, 1986) and linked in a formal way to ordinal variables by Agresti (1984), Agresti and Chuang (1985), and Fienberg (1982).

The association-model approach focuses attention on the interaction parameters, $u_{12(ij)}$, in model (1) and models them in terms of a reduced number of parameters that may be chosen to reflect ordinal structure. In this review, we concentrate our attention on these models and describe aspects of them that explicitly incorporate montonicity constraints corresponding to ordinal structures.

In Section 2 we describe the class of association models developed by Goodman, and in Section 3 we briefly outline the related correlation models. Then in Section 4 we turn to order restrictions imposed upon the parameters of association-models from Section 2.

In the final section of this paper, we briefly consider extensions to multidimensional tables, some of which are relatively direct.

2. Association Models. As we mentioned in the introduction, the class of association models was proposed by Rasch (see Christiansen, 1966) and Fienberg (1968) as a categorical analogue to the Tukey's one-degree-of-freedom model for nonadditivity and its generalizations. The first careful development of these models and their formal linkage to ideas on cross-product or odds ratios were given by Goodman (1979). Later elaborations by Goodman (1981, 1985) led to the general model described below. We use Goodman's notation wherever possible. Equivalent models and special cases have been formulated by several other authors (see for example the stereotype model by Anderson (1984), and the general base-comparison logit model by Cox and Chuang (1984)).

2.1. Model Formulation. Goodman (1985) puts forth the following reparametrization of the standard loglinear model which he refers to as the *saturated RC association model*:

$$P_{ij} = \alpha_i \beta_j \exp\{\sum_{m=1}^{M} \phi_m \mu_{im} \nu_{jm}\}, \tag{5}$$

where

- $\{\mu_{im}\}$ and $\{\nu_{jm}\}$ are standardized row scores and standardized column scores for row category i and column category j, respectively (these are parameters to be estimated from the data);
- the $\{\phi_m\}$ are measures of the "intrinsic association" (if $\phi_m = 0$ for $m = 1, 2, \ldots, M$ then the table exhibits independence of rows and columns);

(6)• $$M = \min(I - 1, J - 1).$$

Furthermore for m and $m^* = 1, 2, \ldots, M$, with $m \neq m^*$ we have the following identifying restrictions:

$$\sum_{i=1}^{I} \mu_{im} P_{i+} = \sum_{j=1}^{J} \nu_{jm} P_{+j} = 0, \tag{7}$$

(8) $$\sum_{i=1}^{I} \mu_{im}^2 P_{i+} = \sum_{j=1}^{J} \nu_{jm}^2 P_{+j} = 1,$$

(9) $$\sum_{i=1}^{I} \mu_{im}\mu_{im^*} P_{i+} = \sum_{j=1}^{J} \nu_{jm}\nu_{jm^*} P_{+j} = 0.$$

These are the marginal-weight versions of Becker and Clogg's (1989) generalized identification restrictions

(10) $$\sum_{i=1}^{I} h_i\mu_{im} = \sum_{j=1}^{J} g_j\nu_{jm} = 0,$$

(11) $$\sum_{i=1}^{I} h_i\mu_{im}^2 = \sum_{j=1}^{J} g_j\nu_{jm}^2 = 1,$$

where the $\{h_i\}$ and $\{g_j\}$ are row and column category weights, respectively, and some restrictions are applied to the cross dimension correlations

(12) $$\rho_{m,n} = \sum_{i=1}^{I} h_i\mu_{im}\mu_{in}$$

and

(13) $$\sigma_{m,n} = \sum_{j=1}^{J} g_j\nu_{jm}\nu_{jn}.$$

Becker and Clogg (1989) point out the importance of weighting systems both in measuring association and in comparing sets of contingency tables, and suggest some other possible choices for the weights.

This saturated RC model is essentially just an explicit rewriting of the original loglinear model of expressions (1) to (4) where the interaction terms, $\{u_{12(ij)}\}$ and the ANOVA-like constraints have been replaced by the sums

$$\sum_{m=1}^{M} \phi_m\mu_{im}\nu_{jm}$$

and somewhat different constraints.

If $\phi_m = 0$ for $m = M^* + 1, \ldots, M$, then we get the *unsaturated RC model* of order M^*. When $M^* = 1$, expression (5) reduces to

(14) $$P_{ij} = \alpha_i\beta_j \exp\{\phi\mu_i\nu_j\},$$

which is model II of Goodman (1979) expressed in terms of identifiable parameters. Expression (14) is what Goodman refers to as the general multiplicative row-and-column-effects (RC) model when the $\{\mu_i\}$ and $\{\nu_j\}$ are unspecified. Two special cases are (i) the row-effects (R) model when the $\{\mu_i\}$ are parameters and the $\{\nu_j\}$ are known, or they are ordered and the spacing between them is specified; (ii) the column-effects (C) model when the $\{\nu_j\}$ are parameters and the $\{\mu_i\}$ are known, or they are ordered and the spacing between them is specified. Gilula, Krieger, and Ritov (1988) provide an interpretation of ϕ as a measure of *stochastic order entropy.*

When the *orders* of the row and column categories are specified, and the rows and columns are appropriately ordered, the U association model may be considered. In the case in which the rows are equally spaced and the columns are equally spaced, we have (in Goodman's (1985) notation):

$$\mu_i - \mu_{i+1} = \Delta', \qquad i = 1, 2, \ldots, I-1, \tag{15}$$

$$\nu_j - \nu_{j+1} = \Delta'', \qquad j = 1, 2, \ldots, J-1. \tag{16}$$

The *local log odds ratios*, for the cells in the 2×2 subtables formed from the cells in adjacent rows and columns, are

$$\begin{aligned} \log\theta_{ij} &= \log\frac{P_{ij}P_{i+1,j+1}}{P_{i,j+1}P_{i+1,j}}, \\ &= \phi(\mu_i - \mu_{i+1})(\nu_j - \nu_{j+1}), \; i = 1, 2, \ldots, I-1, \; j = 1, 2, \ldots, J-1. \end{aligned} \tag{17}$$

For the equal spacing model of (15) and (16), we can rewrite expression (17) as

$$\begin{aligned} \log\theta_{ij} &= \log\frac{P_{ij}P_{i+1,j+1}}{P_{i,j+1}P_{i+1,j}} \\ &= \phi\Delta'\Delta'' \\ &= \text{constant.} \end{aligned} \tag{18}$$

The term *linear-by-linear association* is used to denote the generalized form of this model, for which

$$\mu_i - \mu_{i+1} = \Delta_i' \; i = 1, 2, \ldots, I-1, \tag{19}$$

$$\nu_j - \nu_{j+1} = \Delta_j'' \; j = 1, 2, \ldots, J-1, \tag{20}$$

but where the spacing values, $\{\Delta_i'\}$ and $\{\Delta_j''\}$ are known. For these association models there is an explicit (known) monotonic structure introduced by the fixed spacings.

The various models with their degrees of freedom are as follows:

Model	d.f.
Independence	$(I-1)(J-1)$
U	$(I-1)(J-1)-1$
R	$(I-1)(J-2)$
C	$(I-2)(J-1)$
RC	$(I-2)(J-2)$

2.2. Estimation of Parameters. Let the observed value in the ijth cell be x_{ij}. Then the maximum likelihood estimate $\hat{m}_{ij}$ of the expected frequency $m_{ij} = NP_{ij}$ for the RC association model will satisfy

$$\hat{m}_{i+} = x_{i+},\ i = 1, 2, \ldots, I, \tag{21}$$
$$\hat{m}_{+j} = x_{+j},\ j = 1, 2, \ldots, J, \tag{22}$$
$$\sum_j \hat{\nu}_j \hat{m}_{ij} = \sum_j \hat{\nu}_j x_{ij},\ i = 1, 2, \ldots, I, \tag{23}$$
$$\sum_i \hat{\mu}_i \hat{m}_{ij} = \sum_i \hat{\mu}_i x_{ij},\ j = 1, 2, \ldots, J, \tag{24}$$

when the x_{ij} follow any of the standard contingency table sampling models, i.e., Poisson, multinomial, product-multinomial (see, for example, Bishop, Fienberg, and Holland, 1975).

Goodman (1979) suggests the following iterative procedure for solving the likelihood equations. Let

$$\rho_i = \mu_i - \bar{\mu}, \tag{25}$$
$$\sigma_j = \nu_j - \bar{\nu}, \tag{26}$$
$$\bar{\mu} = \sum_i \mu_i / I, \tag{27}$$
$$\bar{\nu} = \sum_j \nu_j / J. \tag{28}$$

We denote the values of the estimates at any given stage in the iterative procedure by α_i^*, β_j^*, μ_i^*,and ν_j^*. Then we update the estimates one at a time, each update being followed by a recalculation of the values of m_{ij}. Denote the current expected frequencies by m_{ij}^*, then we replace $(m_{ij}^*, \alpha_i^*, \beta_j^*, \mu_i^*, \nu_j^*)$ by $(m_{ij}^{**}, \alpha_i^{**}, \beta_j^{**}, \mu_i^{**}, \nu_j^{**})$ where

$$\alpha_i^{**} = \alpha_i^* x_{i+} / m_{i+}^*, \tag{29}$$
$$\beta_j^{**} = \beta_j^* x_{+j} / m_{+j}^*, \tag{30}$$
$$\mu_i^{**} = \mu_i^* (\sum_j \sigma_j^* (x_{ij} - m_{ij}^*)) / (\sum_j \sigma_j^{*2} m_{ij}^*), \tag{31}$$

$$\nu_j^{**} \quad = \quad \nu_j^*(\sum_i \rho_i^*(x_{ij} - m_{ij}^*))/(\sum_i \rho_i^{*2} m_{ij}^*). \tag{32}$$

Here ρ_i^*, σ_j^*, ρ_i^{**}, σ_j^{**}, have the obvious meanings. The algorithm appears to work in practice but to our knowledge there is no formal proof for its convergence. Becker and Clogg (1989) extend this algorithm to deal with the case in which K two way tables are considered and compared.

Alternative approaches to maximum likelihood estimation involve the use of general nonlinear maximization routines and other variants of standard Newton Raphson algorithms (e.g., see Chuang, 1980). See also Gilula (1982) who discusses the use of the singular value decomposition of the local odds ratio matrix to estimate the model parameters. It should, however, be noted that thus far in the literature, necessary and sufficient conditions for the existence of maximum likelihood estimates for the RC model have not been formulated.

2.3. Ordering Properties Implied by the Association Models. At this stage, we defer discussion of estimation under order restrictions. It is worth noting that tables fit by these models exhibit ordering properties which relate directly to traditional notions of dependence. We summarize some of these properties as they appear in the categorical data literature, and re-express them in the language of the dependence literature.

Let A and B denote the variables corresponding to rows and columns in the contingency table. Under the RC model, the association is isotropic, that is, rows and columns can be ordered in such a way that the *local odds ratios* have the property:

$$\theta_{ij} = \frac{P_{ij}P_{i+1,j+1}}{P_{i+1,j}P_{i,j+1}} \geq 1 \quad i = 1, \ldots, I-1, \;\; j = 1, \ldots, J-1. \tag{33}$$

Tables possessing the property (33) are *totally positive of order 2,* denoted TP_2 (see Schriever, 1986, or Gill and Schriever, 1987).

Agresti (1984) defines the local-global odds ratio

$$\theta_{ij}^r = \frac{(\sum_{k\leq j} P_{ik})(\sum_{k>j} P_{i+1\,k})}{(\sum_{k>j} P_{ik})(\sum_{k\leq j} P_{i+1\,k})}, \quad j = 1, 2, \ldots, J-1. \tag{34}$$

These odds ratios are local in the row variable, but "global" in the column variable, and may be defined for any two rows, a and b, rather than adjacent rows i and $i+1$. Analogous global-local odds ratios, global in the row variable and local in the column variable may also be defined for columns j and $j+1$ i.e.,

$$\theta_{ij}^c = \frac{(\sum_{k\leq i} P_{kj})(\sum_{k>i} P_{k\,j+1})}{(\sum_{k>i} P_{kj})(\sum_{k\leq i} P_{k\,j+1})}, \quad i = 1, 2, \ldots, I-1. \tag{35}$$

Then $\theta_{ij}^r > 1$ for each j implies that the conditional distribution in row $i+1$ is stochastically larger than the conditional distribution in row i, while $\theta_{ij}^c > 1$ for each i implies that the conditional distribution in row $j+1$ is stochastically larger than the conditional distribution in row j.

In fact, when the rows and columns are ordered appropriately such that (33) holds, we have the following stochastic ordering relationships (Goodman, 1981):

(i) The conditional distribution for the ith row of the table, P_{ij}/P_{i+} is stochastically smaller than the conditional distribution for the i'th row of the table if $i' > i$, i.e.,

$$1 \le i < i' \le I \Longrightarrow P(B \le j | A = i) \ \ge \ P(B \le j | A = i') \qquad j = 1, 2, \ldots, J. \tag{36}$$

Barlow and Proschan (1981) denote this property by *SI(B|A)*.

(ii) The conditional distribution for the jth column of the table, P_{ij}/P_{+j} is stochastically smaller than the conditional distribution for the j'th row of the table if $j' > j$, i.e.,

$$1 \le j < j' \le J \Longrightarrow P(A \le i | B = j) \ \ge \ P(A \le i | B = j') \qquad i = 1, 2, \ldots, I. \tag{37}$$

Barlow and Proschan denote this property by *SI(A|B)*.

Schriever (1983, 1986) refers to the variables A and B as being *double regression dependent of order 1 (DR_1)* where the order relationships (34) and (35) hold. Thus tables satisfying the RC association models are double regression dependent up to a permutation of rows and columns. This property Schriever refers to as *order dependence of order 1.* Schriever (1983, 1986) carries these arguments further noting that, for the RC association model with rows and columns reordered so that μ_i and ν_j are increasing in their indices, the table of probabilities is totally positive of order M where $M = min(I-1, J-1)$, i.e., the table is TP_M.

While the RC association model can always lead to a reordering of rows and columns such that TP_2 implies conditions (34) and (35), the reverse is not true. Thus there exist tables exhibiting order dependence for which the RC association model does not hold. This is a special case of results due to Schriever (1986) i.e.:

$$TP_k \Longrightarrow DR_{k-1}, \tag{38}$$

Thus for k=2, we get the implication that TP_2 implies DR_1 but the reverse is not necessarily true (except for $I \times 2$ and $J \times 2$ tables) since any rank 2 probability table can have its rows and columns permuted such that it is DR_1 (Schriever, 1986).

At this point, we note that there is clearly a hierarchical relationship between the dependence notions that have emerged thus far. In fact, we have identified the upper levels of Barlow and Proschan's (1975) tree of notions of bivariate dependence. These, and other dependence concepts forming nodes of the tree may be re-expressed in terms of conditions on the odds ratios of the contingency table or various subtables thereof. See Douglas et al. (1990) for a more detailed treatment of these linkages in tree-like form. It is relatively straightforward to elicit the implication structure once we view the dependence concepts in the contingency table framework.

Agresti, Chuang, and Kezouh (1987) show that for the row (columns) association model, since the estimated conditional distributions in the rows (columns) are stochastically ordered according to the values of the $\{\hat{\mu}_i\}$ and the $\{\hat{\nu}_j\}$, then the $\{\hat{\mu}_i\}$ and the $\{\hat{\nu}_j\}$ have the same ordering as the sample row and column means. This result follows from a rewriting of the likelihood equations from expressions (21), (22), (23), and (24).

3. Correlation Models. For completeness, we describe briefly the correlation models corresponding to the association models in Section 2 above. Goodman (1985, 1986) presents a much more careful treatment of these models and their relationships to the association models. He is quick to point out that although the association and correlation approaches are related not only may they yield different results, but also one approach may be preferable over the other in specific settings. The models are identical under independence of the row and column variables, and turn out to be reasonable approximations of each other when the intrinsic association (ϕ_m above) and correlation parameters (λ_m below) are close to zero in value.

The *saturated RC canonical correlation model* may be formulated as

$$P_{ij} = P_{i+}P_{+j}(1 + \sum_{m=1}^{M} \lambda_m x_{im} y_{jm}) \tag{39}$$

where

- x_{im} and y_{jm} are standardized row scores and standardized column scores respectively, to be estimated from the data.
- λ_m measures the correlation between x_{im} and y_{jm} and

$$\lambda_1 \geq \ldots \geq \lambda_M, \tag{40}$$

  and

$$M = \min(I-1, J-1). \tag{41}$$

- Furthermore for $m = 1, 2, \ldots, M$, we have the following identifying restrictions:

(42) $$\sum_{i=1}^{I} x_{im} P_{i+} = \sum_{j=1}^{J} y_{im} P_{+j} = 0,$$

(43) $$\sum_{i=1}^{I} x_{im}^2 P_{i+} = \sum_{j=1}^{J} y_{im}^2 P_{+j} = 1,$$

(44) $$\sum_{i=1}^{I} x_{im} x_{im^*} P_{i+} = \sum_{j=1}^{J} y_{jm} y_{jm^*} P_{+j} = 0.$$

Maximum likelihood estimation is direct for these saturated correlation models; not so for their unsaturated counterparts. Estimation for the unsaturated models is described in Goodman (1985).

The saturated RC canonical correlation model is simply a reparametrization of the *saturated RC correspondence analysis* model (see, e.g., Greenacre, 1984), i.e.

(45) $$P_{ij} = P_{i+} P_{+j} (1 + \sum_{m=1}^{M} x'_{im} y'_{jm} / \lambda_m),$$

where $x'_{im} = \lambda_m x_{im}$ and $y'_{jm} = \lambda_m y_{jm}$.

Unsaturated RC correlation models, with $\lambda_m = 0$ for $m = m^* + 1, \ldots, M$ and *U correlation models* may be formulated as in the association approach. Their order restriction properties have been studied in depth by Schriever (1983, 1984). For example, the RC correlation model with $m^* = 1$ also demonstrates order dependence of order 1 (i.e. expressions (34) and (35) hold after a suitable reordering of rows and columns). Moreover, Schriever (1983, 1986) proves the following result:

THEOREM. *Suppose that the row and column variables, A and B are double regression dependent. Then the first set of correspondence analysis row and column scores , that is , for m=1, can be chosen to satisfy*

(46) $$x_{11} \leq x_{21} \ldots \leq x_{I1}, \quad y_{11} \leq y_{21} \leq \ldots \leq y_{J1}.$$

Strict inequalities in (30) and (31) imply strict inequalities in (46).

Schriever also generalizes this theorem to higher order sets of scores (i.e. $m > 1$). Gilula, Krieger, and Ritov (1988) provide an interpretation of λ_1 for the RC model with $m^* = 1$, as a measure of *stochastic order extremity.* They also note the link to Kimeldorf and Sampson's (1978) coefficient of monotonic dependence.

Canonical correlation analysis has also been used to develop tests for independence of rows and columns. Haberman (1981) assigns scores ϕ_i and ψ_j to row category i and column category j respectively, and maximizes the correlation

(47)
$$R_1 = \sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij}\phi_i\psi_j$$

where $p_{ij} = n_{ij}/N$, the observed relative frequency in cell j. NR_1^2 is shown to be approximately asymptotically distributed according to the distribution of the maximum eigenvalue of a central Wishart matrix with $J-1$ degrees of freedom. Approximate critical values for this distribution are available from existing tables.

In the case in which we consider the $I \times J$ table to represent observations on a discrete bivariate random vector (X, Y) taking values (i, j), $i = 1, \ldots, I$, $j = 1, \ldots, J$, Sethuraman (1977) has shown the above distribution to be the limiting distribution of nR_n^*. Here R_n^* is the sample Renyi maximum correlation

(48)
$$R_n^* = R(x, y) = \max_{f(x), g(y)} \rho(f(x), g(y))$$

where

- the sample consists of n observations on the bivariate variable (X, Y),
- the maximum is over all functions f of X and g of Y such that Ef^2 and Eg^2 are finite,
- ρ denotes correlation, and
- the following conditions are satisfied:
    1. if $p_i = P(X = i)$, $I \leq J$, the $I \times I$ matrix with diagonal elements $p_i - p_i^2$, and off diagonal elements $-p_i p_j$ is of rank $(I-1)$, and
    2. $R(X, Y) = 0$, i.e., X and Y are independent.

4. Order-Restricted Association Models. In this section we consider association models of the form (5) in which one or both of the following constraints are assumed to hold:

(49)
$$\mu_1 \leq \mu_2 \leq \ldots \leq \mu_I,$$

(50)
$$\nu_1 \leq \nu_2 \leq \ldots \leq \nu_J.$$

Before detailing parameter estimation and goodness of fit procedures, we present three important illustrations of these models in order to better familiarize the reader with their structure.

4.1. The Stereotype Model. Anderson's (1984) stereotype model is a special case of the qualitative logistic regression model

(51)
$$\Pr(y = y_s \mid z) = \frac{\exp(\eta_{0s}^* + \eta_s^T z)}{\sum_{t=1}^{k} \exp(\eta_{0t}^* + \eta_t^T z)}, \quad s = 1, \ldots, k,$$

where

- y is an ordered categorical response with categories $y_1 \dots y_k$.
- $\eta_s^T = (\eta_{s1} \dots \eta_{sk})$ gives the regression coefficients for the odds of $y = y_s$ relative to $y = y_k$.

In the stereotype model, the $\{\eta_s\}$ are taken to be parallel, i.e.,

$$(52) \qquad \eta_s = -\xi_s \eta, \quad s = 1, \dots, k,$$

and the $\{\xi_s\}$ are taken to be monotone decreasing, i.e.,

$$(53) \qquad 1 = \xi_1 > \xi_2 > \dots > \xi_k = 0.$$

The resulting model is expressible as

$$(54) \qquad \Pr(y = y_s \mid z) = \frac{\exp(\eta_{0s}^* - \xi_s \eta^T z)}{\sum_{t=1}^k \exp(\eta_{0t}^* - \xi_t \eta^T z)} \quad s = 1, \dots, k.$$

This model is one dimensional in that only one linear function, viz. $\eta^T z$, is required to describe the relationship between y and z for all categories y_s. The importance of this model lies not only in the fact that it is ordered, but also that it can be extended to incorporate a multidimensional regression relationship. Anderson applies this model to the $2 \times k$ contingency table, showing that the stereotype model can be applied if and only if the cross product ratios are monotone increasing or decreasing.

We can apply the stereotype model to the $I \times J$ table if we consider the ordered categorical response y with categories $y_1, \dots, y_J$ and the single predictor, z, with values $z_1, \dots, z_I$. Denoting the cell probabilities by P_{ij}, we get

$$(55) \qquad \frac{P_{ij}}{P_{iJ}} = \frac{\Pr(y = y_j \mid z = z_i)}{\Pr(y = y_J \mid z = z_i)} = \exp(\eta_{0j}^* - \xi_j \eta z_i).$$

Thus

$$(56) \qquad \log(\frac{P_{ij}}{P_{iJ}}) = \eta_{0j}^* - \xi_j \eta z_i.$$

The RC association model of expression (5) gives

$$(57) \qquad \log(\frac{P_{ij}}{P_{iJ}}) = \log(\eta_j - \eta_J) + \xi \mu_i (\nu_j - \nu_J).$$

Thus we have that Anderson's model is really just the RC association model with the monotonicity constraint of expression (53) and the following correspondence of component parameters:

Stereotype Model	RC model
β_{0j}^*	$\beta_j - \beta_J$
ϕ_j	$\nu_j - \nu_J$
z_i	μ_i
β	ϕ

The stochastic ordering properties of expressions (34) and (35) described in Section 2.3 above are noted by Anderson for his model.

4.2. The Proportional Odds Model. Another regression-like model for the analysis of ordinal data is proposed by McCullagh (1980), who distinguishes clearly between explanatory and response variables. The response has k ordered categories, with probabilities $\pi_1(\underline{x}), \pi_2(\underline{x}), \ldots \pi_k(\underline{x})$, where $\underline{x}$ is the vector of covariates. The proportional odds model is expressed as

$$\text{(58)} \qquad \log \frac{\gamma_j(\underline{x})}{1-\gamma_j(\underline{x})} = \theta_j - \underline{\beta}^T \underline{x} \quad 1 \leq j < k$$

where

- $\gamma_j(\underline{x}) = \pi_1(\underline{x}) + \ldots \pi_j(\underline{x})$
- $\underline{\beta}$ is the vector of unknown parameters.

The model is thus a cumulative logit model and induces the stochastic ordering properties of expressions (34) and (35) as follows. The difference: $\text{logit}(\gamma_j(\underline{x}_1))-$ $\text{logit}(\gamma_j(\underline{x}_2))$ is equal to $\underline{\beta}^T(\underline{x}_2 - \underline{x}_1)$, thus constant over j. McCullagh denotes this difference by Δ. Since the logit function is monotonic, the sign of Δ determines whether $\gamma_j(\underline{x}_1) > \gamma_j(\underline{x}_2)$ or $\gamma_j(\underline{x}_1) < \gamma_j(\underline{x}_2)$ for all j.

4.3. The Monotone Scores Association Model. Chuang and Agresti (1986) give a detailed treatment of a model in which the row variable is nominal, and the column variable is regarded as an ordinal response, with score parameters for the columns constrained to be monotone increasing. The model is

$$\text{(59)} \qquad \log P_{ij} = \mu + a_i + b_j + \mu_i \nu_j,$$

where

$$\text{(60)} \qquad \sum_{i=1}^{I} a_i = \sum_{j=1}^{J} b_j = \sum_{i=1}^{I} \mu_i = 0,$$

and

$$\text{(61)} \qquad 1 = \nu_1 \leq \ldots \leq \nu_J = J,$$

Model (58) is simply the RC association model of Section 2, with the added ordering constraint (61). This constraint produces a stochastic ordering among the response distributions of the rows. Suppose that rows correspond to two drugs a and b, with $\mu_a > \mu_b$. Then, because of the ordering of the $\{\nu_j\}$, the log odds ratios for adjacent responses,

$$\text{(62)} \qquad \log \frac{P_{aj} P_{b,j+1}}{P_{a,j+1} P_{bj}} = (\mu_b - \mu_a)(\nu_{j+1} - \nu_j),$$

are nonnegative, and the response distribution for the bth drug, $\{P_{bj}/P_{b+}\}$ is stochastically greater than that for the ath drug, $\{P_{aj}/P_{a+}\}$. Chuang and Agresti (1986) use the model to analyze a data set in which the column variable is the response (on an ordinal scale) to treatment with different drugs (the row variables).

4.4. Parameter Estimation. In the monotone scores case we are essentially concerned with fitting the RC association model of expression (5) under a monotonicity constraint on the $\{\nu_j\}$. Here we consider estimation for the order restricted R, C, and RC models i.e. the R model under the restriction $\mu_1 \leq \ldots \leq \mu_I$, the C model under the restriction $\nu_1 \leq \ldots \leq \nu_J$, and the RC model under one or both of the preceding restrictions. Results will be stated for the most general case when available, i.e., in terms of the RC model; if not, then in terms of the R model (in which case it is implied that they hold analogously for the C model) or the C model (in which case it is implied that they hold analogously for the R model).

If the true parameters are strictly monotone, i.e. they do not lie on the boundary of the parameter space, then the estimates for the order-restricted RC model and the RC model are asymptotically equivalent.

For estimation under order restrictions, Goodman (1985) discusses an approximation yielding an ordered solution for the R or C models with $\phi > 0$ when the actual maximum likelihood estimates for the unrestricted model, are not in the correct order. He notes that for each violated restriction the likelihood is maximized on the boundary where these adjacent values are equal to each other. Then in his iterative procedure, he combines rows (or columns) corresponding to pairs of scores violating the order constraints, and refits the model. This, in effect, enforces an equality constraint on the score parameters for the rows (columns) concerned. Thus the order restricted solution is the same as the ordinary ML solution for the appropriately collapsed table.

For the R model, the following necessary and sufficient conditions that completely characterize this collapsing are given in Agresti, Chuang, and Kezouh (1987):

$$\hat{m}^*_{i+} = x_{i+} \quad i = 1, \ldots, I, \tag{63}$$

$$\hat{m}^*_{+j} = x_{+j} \quad j = 1, \ldots, J, \tag{64}$$

$$\sum_{i \leq b} (\sum_j \nu_j \hat{m}^*_{ij}) \leq \sum_{i \leq b} (\sum_j \nu_j x_{ij}) \quad b = 1, \ldots, I, \tag{65}$$

$$\sum_{i \leq r_k} (\sum_j \nu_j \hat{m}^*_{ij}) = \sum_{i \leq r_k} (\sum_j \nu_j x_{ij}) \quad k = 1, \ldots, a, \tag{66}$$

where $\{r_1, \ldots, r_a\}$ are such that

$$\hat{\mu}^*_1 = \ldots = \hat{\mu}^*_{r_1} < \hat{\mu}^*_{r_1+1} = \ldots = \hat{\mu}^*_{r_2} < \ldots < \hat{\mu}^*_{r_{a-1}+1} = \ldots = \hat{\mu}^*_{r_a}. \tag{67}$$

A heuristic argument for conditions (65) and (66) may be formulated as follows: In the unrestricted case the inequality (65) is an equality (23). When order constraints are applied, the restricted maximum is lower than the unrestricted maximum; this is manifested by inequalities in the likelihood equations with equalities only on the boundaries of the level sets. For the RC model, however, there are no sufficient conditions. This is due to the fact that, because the RC model is not loglinear, the log likelihood is not necessarily concave.

Agresti and Chuang (1985) and Agresti, Chuang, and Kezouh (1987) note that the partition for the order restricted ML solution $R_1, \ldots, R_a$ above, where $R_k = \{r_{k-1}+1, \ldots, r_k\}$, is identical to the partition of level sets obtained in using the *Pool Adjacent Violators* algorithm *(PAV)* to obtain the regression of the sample row means (weighted by the row totals) in the class of functions isotonic with respect to the simple order in the rows (see Barlow, et al., 1972).

Gilula (personal communication) has pointed out the kinds of problems that come with the *PAV* approach. He considers the following 3 by 3 table:

1	1	1
1	2	1
1	1	1

and notes what happens when we fit the row-effects model of expression (14) with the following fixed values of $\{\nu_j\}$ under constraint (61):

$$\nu_1 = 0, \quad \nu_2 = 1, \quad \nu_1 = 1.$$

Then the unrestricted MLE's of $\{\mu_i\}$ are:

$$\hat{\mu}_1 = -0.2582, \quad \hat{\mu}_2 = 0.3873, \quad \hat{\mu}_3 = -0.2582$$

and

$$\hat{\phi} = 0.6424.$$

Now, if we impose a monotonicity constraint on the $\{\mu_i\}$ and apply the *PAV* algorithm we get

$$\hat{\mu} = (-0.4833, 0.2065, 0.2065)$$

and

$$\hat{\phi} = 0.3193$$

But the same maximum value of the likelihood is achieved by

$$\hat{\mu} = (0.4833, -0.2065, -0.2065)$$

and

$$\hat{\phi} = -0.3193.$$

Ritov and Gilula (1987) point out that the problem with direct application of *PAV* to the score parameters lies in the fact that these parameters are mutually dependent. Thus reordering a pair of μ_i's may result in a violation of the order constraint on the ν_j's. A solution is to apply the *PAV* algorithm to specific functions of the parameters, viz.

$$\tilde{E}_i(\hat{\nu}) = \sum_{j=1}^{J} \tilde{P}_{ij}\hat{\nu}_j / \tilde{P}_{i.} \tag{68}$$

and

$$\tilde{F}_j(\hat{\mu}) = \sum_{i=1}^{I} \tilde{P}_{ij}\hat{\mu}_i / \tilde{P}_{.j} \tag{69}$$

where $\tilde{P}_{ij}$ represents the empirical distribution in the observed contingency table. The quantities $(\tilde{E}_i - \tilde{E}_{i-1})$ and $(\tilde{F}_j - \tilde{F}_{j-1})$ are asymptotically uncorrelated and amalgamation can be done separately (independently) for rows and columns. The unrestricted maximum likelihood estimates for the collapsed table, in which the $\tilde{E}_i$ and $\tilde{F}_j$ follow the desired order, are asymptotically equivalent to the order restricted maximum likelihood estimates.

Dykstra and Lemke (1988) discuss this restricted maximization problem, and its dual I-projection problem. They note the applicability of a general algorithm from Dykstra (1985) to solve the I-projection problem, and thus the maximum likelihood problem here.

4.5. Goodness-of-Fit Statistics. When the true parameters for the order restricted model are in fact monotone, then the Pearson chisquare and Likelihood Ratio chisquare are an asymptotic χ^2. When there are equalities between adjacent parameters, the goodness of fit statistics are distributed as mixtures of chisquares. For example, it follows from Agresti, Chuang, and Kezouh (1987) that in the RC model, under monotonicity constraints on the $\{\mu_i\}$, when the $\{\mu_i\}$ are strictly monotone, except for one identical adjacent pair, the likelihood ratio statistic has an asymptotic distribution that is an equal mixture of the one for the RC model and one with an additional degree of freedom. This is because when one pair is equal then

- with limiting probability 0.5, asymptotically, the ordinary MLE's will follow the order restriction, and G^2 under the order restriction will be the same as the unrestricted G^2 with $\chi^2_{(I-2)(J-2)}$ distribution.
- with limiting probability 0.5, the ordinary MLE's will have the estimates for the pair of parameters concerned out of order. The order restricted solution will be the same as the ordinary solution for the table with the two rows concerned, collapsed.

From Agresti, Chuang, and Kezouh (1987), Theorem 4, the likelihood ratio chi-square statistic in this case can be decomposed as:

$$G^2(RC^*) = G^2(RC') + G^2(I), \tag{70}$$

where

- $G^2(RC^*)$ denotes the fit of the order restricted model to the original table,
- $G^2(RC')$ denotes the fit of the ordinary RC model to the collapsed table, and
- $G^2(I)$ denotes the fit of the model of independence to the $2 \times J$ table formed from the pair of rows with parameters out of order.

The first of the quantities on the right hand side has an asymptotic $\chi^2_{(I-3)(J-2)}$ distribution, and the second a $\chi^2_{(J-1)}$ distribution, and they are asymptotically independent. Thus their sum is asymptotically distributed as $\chi^2_{(I-2)(J-2)+1}$.

In general, however, the foregoing result does not tell us the asymptotic distribution of G^2 when the order constraints are included in the model because we do not know the values of the true parameters and hence the equalities/inequalities that may hold among them.

Ritov and Gilula (1987) show that the chisquare statistic for testing H_0: Order restricted RC model against H_1: Unrestricted RC model has the following asymptotic distribution:

$$P(X^2 > c) = \sum_{k=1}^{I'+J'-2} \beta_k P(\chi^2_{(k)} > c) \tag{71}$$

where I' and J' denote the maximum number of row and column parameters that can be equal and the mixture probabilities β_k depend on the marginals $P_{i.}$ and $P_{.j}$.

The general problem of inference under order restrictions has been treated in detail by Barlow et al. (1972), and, most recently by Robertson, Wright, and Dykstra (1988) which includes some of the material in Barlow et al. and also covers subsequent developments in the area of isotonic methods. Raubertas, Lee, and Nordheim (1986) consider hypothesis tests for linearly constrained normal means, and Robertson (1978) and Lee (1987) deal with tests for order restrictions on multinomial parameters.

Robertson considers the following three hypotheses:

- $H_0: \quad p = q,$ where $p = (p_1, \ldots, p_k)$ is a probability vector of unknown values and q is a known probability vector.
- $H_1: \quad p \neq q,$ but p satisfies some order restriction O on its components.
- $H_2: \quad p \epsilon R^k, \; \sum p_i = 1.$

The asymptotic distribution of the test statistic for testing H_0 versus H_1 is of the $\bar{\chi}^2$ form (see Barlow et al., 1972, ch. 3), i.e., a weighted sum of standard chi-squares, depending on the alternate hypothesis through the weights. The test of H_1 versus H_2 is not similar in that it depends on the particular p satisfying H_1; however, asymptotically, an upper bound on the significance level of the test of H_1 versus H_2 can be found. Raubertas, Lee, and Nordheim (1986) give similar results for the normal means case. In both cases, the maximum likelihood estimates that appear in the test statistics turn out to be projections onto polyhedral cones.

5. Parameter Constraints in Terms of Table Margins. Yet another way to approach information on ordering is to incorporate it into the model in the form of order restrictions on the marginal totals of the cross classification. This has been explored by Eddy, Fienberg, and Meyer (1982). The motivation behind this approach is similar to the motivation behind the structural zero method for modelling contingency tables in which some cell frequencies are known to be zero (see Bishop, Fienberg and Holland, 1975). There appears to be an interesting link between this approach and ideas associated with loglinear models.

6. Association Models for Multi-Dimensional Tables. Natural extensions of the association models from Section 2 to three and higher-dimensional contingency tables are reasonably direct and have been explored previously by Chuang (1980), Clogg (1982), Fienberg (1982), Agresti (1983), and Goodman (1986). Rather than attempting an exhaustive treatment of the topic we present some illustrative examples and features of these extensions, noting the links to the association models for $I \times J$ tables and the kinds of stochastic ordering features described in Section 2.

For the $I \times J \times K$ cross-classification involving variables A, B, and C, let P_{ijk} be the probability of an observation falling into the ith row, jth column and kth layer. The saturated loglinear model for this 3-dimensional table is usually written in the form:

$$(72) \quad \log P_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$$

with identifying constraints requiring the sum of any subscripted u-term over each subscript to be equal to zero. Extensions of the association models to this situation typically involve

1. setting one or more subscripted u-term equal to zero;

2. representing some of the remaining u-terms with 2 or more subscripts as products of parameters.

For both (1) and (2) the choices must be in accord with the generalized hierarchy principle (e.g., see Fienberg, 1980, p. 43 and p. 100) associated with ANOVA-like models wherein restriction terms must be compatible with restrictions or structures for the lower-order terms which are marginal to them. Thus representing u_{12} in multiplicative form implies related multiplication forms for u_{123}. In particular setting $u_{12} = 0$ implies that variables A and B not be linked for any multiplicative component used to model related higher-order terms, e.g. u_{123}.

Chuang (1980), for example, described several choices for association models based on (72) where the 3-factor term is represented as

$$\text{(a) } u_{123(ijk)} = \lambda v_{1(i)}\, v_{2(j)}\, v_{3(k)}, \tag{73}$$

$$\text{(b) } u_{123(ijk)} = \lambda v_{1(i)}\, v_{23(jk)}. \tag{74}$$

In the multiplicative notation of Goodman (1985) these models are representable as

$$\text{(a) } P_{ijk} = \alpha_{ij}^{AB}\alpha_{ik}^{AC}\alpha_{jk}^{BC} \exp\,(\phi\mu_i^A\mu_j^B\mu_k^C), \tag{75}$$

$$\text{(b)} P_{ijk} = \lambda_{ij}^{AB}\lambda_{ik}^{AC}\lambda_{jk}^{BC} \exp\,(\phi\mu_i^A\mu_j^{BC}). \tag{76}$$

Some other examples of loglinear-association models are:

$$\begin{aligned} \text{(c) } u_{123(ijk)} &= 0, \\ u_{12(ij)} &= \lambda_1 v_{1(i)} v_{2(j)}, \\ u_{13(ik)} &= \lambda_2 v'_{1(i)} v'_{2(j)}, \\ u_{23(jk)} &= \lambda_3 v^*_{2(j)} v^*_{3(j)}, \end{aligned} \tag{77}$$

$$\begin{aligned} \text{(d) } u_{12(ij)} &= 0, \\ u_{123(ijk)} &= \lambda v_{1(i)} v_{23(jk)}, \end{aligned} \tag{78}$$

which are representable in Goodman's notation as

$$\text{(c) } P_{ijk} = \alpha_i^A\alpha_j^B\alpha_k^C \exp(\phi_1\mu_i^A\mu_j^B + \phi_2\nu_i^A\nu_k^C + \phi_3\alpha_j^B\alpha_k^C), \tag{79}$$

$$\text{(d) } P_{ijk} = \alpha_{ik}^{AC}\alpha_{jk}^{BC} \exp(\phi\mu_i^A\mu_{jk}^{BC}), \tag{80}$$

(see also Clogg, 1982).

Maximum likelihood estimates for these loglinear-association models require some form of iterative procedure. Chuang (1980) proposes a variant on the Newton-Raphson method and direct generalizations of Goodman's iterative method for two-way tables are available.

The various mixed loglinear-association models described above can all be reformulated as models for the key parameters for three-way arrays, i.e. ratios of odds-ratios of the form

$$\theta_{ijk} = \frac{P_{ijk}P_{i+1,jk}}{P_{i,j+1,k}P_{i+1,j+1,k}} \Big/ \frac{P_{ij,k+1}\ P_{i+1,j,k+1}}{P_{i,j+1,k}\ P_{i+1,j+1,k+1}}. \tag{81}$$

The (θ_{ijk}) do not readily lend themselves to any total postivity order restrictions and, to date, no-one has considered appropriate generalizations of the stochastic ordering properties discussed in Section 2.2.

No one in the categorical data analysis literature has yet presented methods for estimation of the higher dimensional loglinear-association models subject to order on the multiplicative parameters such as in expression (49) and (50). The estimation problems for the RC model alluded to in Section 4.4 come home in spades when we move to higher dimensions and more than two sets of order restructures. Again the log likelihood is not necessarily concave and approaches modelled after the Pooled Adjacent Violators algorithm will not necessarily work.

Schriever (1983, 1986) considers some multi-dimensional generalizations of the correlation models of Section 3, and related stochastic ordering properties. His description allows a two-dimensional matrix representation involving submatrices which are two-dimensional marginals. Thus his approach, while far more restrictive than the one suggested here for loglinear-association models, does allow for the direct examination of total-positivity and order-dependence properties.

REFERENCES

AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data.* Wiley, New York.

AGRESTI, A. (1983). A survey of strategies for modeling cross-classifications having ordinal variables. *J. Amer. Statist. Assoc.* **78** 184–198.

AGRESTI, A. and CHUANG, C. (1985). Bayesian and maximum likelihood approaches to order-restricted inference for models for ordinal categorical data. *Proceedings of the Symposium on Order Restricted Statistical Inference* held in Iowa City, Iowa, Sept. 11–13, 1985, 6–27.

AGRESTI, A., CHUANG, C., and KEZOUH, A. (1987). Order restricted score parameters in association models for contingency tables. *J. Amer. Statist. Assoc.* **82** 619–623.

ANDERSON, J.A. (1984). Regression and ordered categorical variables. *J. Roy. Statist. Soc. B* **46** 1–30.

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M., and BRUNK, H.D. (1972). *Statistical Inference Under Order Restrictions.* John Wiley and Sons, New York.

BARLOW, R.E. and PROSCHAN, F. (1981). *Statistical Theory of Reliability and Life Testing: Probability Models.* To Begin With, Silver Spring, Maryland.

BECKER, M.P. and CLOGG, C.C. (1989). Analysis of sets of two-way contingency tables using association models. *J. Amer. Statist. Assoc.* **84** 142–151.

BENZÉCRI, J.P., et al. (1973). *L'Analyse des Données, Vol. 1: La Taxinomie, Vol. 2: L'Analyse des Correspondances.* Dunod, Paris (Second Edition 1976).

BISHOP, Y., FIENBERG, S.E., and HOLLAND, P.W. (1975). *Discrete Multivariate Analysis.* M.I.T. Press, Cambridge.

CHRISTIANSEN, U. (1966). *The Lectures by Professor G. Rasch over the Theory of Statistics* (in Danish). University Foundation for the Printing of Textbooks, Copenhagen.

CHUANG, C. and AGRESTI, A. (1986). A new model for an ordinal pain data set from a pharmaceutical study. *Statist. in Medicine* **5** 15–20.

CHUANG, C. (1980). Analysis of categorical data with ordered categories. Ph.D. Dissertation, School of Statistics, University of Minnesota.

CLOGG, C.C. (1982). Some models for the analysis of association in multiway cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **77** 803–815.

COX, C. and CHUANG, C. (1984). A comparison of chi-square partitioning and two logit analyses of ordinal pain data from a pharmaceutical study. *Statist. in Medicine* **3** 273–285.

DOUGLAS, R.B., FIENBERG, S.E., LEE, M.-L.T., SAMPSON, A.R., and WHITAKER, L. (1990). Positive dependence concepts and contingency table parametrizations. In *Topics in Statistical Dependence* (H.W. Block, A.R. Sampson, T.H. Savits, eds.), *IMS Lecture Notes/Monograph Series*. Hayward, CA.

DYKSTRA, R.L. (1985). An iterative procedure for obtaining I-projections onto the intersection of convex sets. *Ann. Prob.* **13**, 975–984.

DYKSTRA, R.L. and LEMKE, J.H. (1988). Duality of I-projections and maximum likelihood estimation for log-linear models under cone constraints. *J. Amer. Statist. Assoc.* **83** 546–554.

EDDY, WM. F., FIENBERG, S.E., and MEYER, M.M. (1982). Contingency table estimation with order restrictions on the margins. Unpublished manuscript.

FIENBERG, S.E. (1968). The estimation of cell probabilities in two-way contingency tables. Ph.D. Dissertation, Department of Statistics, Harvard University.

FIENBERG, S.E. (1980). *The Analysis of Cross-classified Categorical Data.* (2nd ed.) M.I.T. Press, Cambridge.

FIENBERG, S.E. (1982). Using information on ordering for loglinear model analysis of multidimensional contingency tables. In *Proceedings of XIIth International Biometrics Conference.* Tonteruse, France, 133–139.

GILL, R.D. and SCHRIEVER, B.F. (1986). Comment on "Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables." *Int. Statist. Rev.* **54** 289–291.

GILULA, Z. (1982). A note on the analysis of association in cross-classifications having ordered categories. *Commun. Statist.-Theor. Meth.* **11**(11) 1233–1240.

GILULA, Z., KRIEGER, A.M., and RITOV, Y. (1988). Ordinal association in contingency tables: Some interpretive aspects. *J. Amer. Statist. Assoc.* **83** 540–545.

GOODMAN, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **74** 537–552.

GOODMAN, L.A. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **76** 320–334.

GOODMAN, L.A. (1984). *The Analysis of Cross-Classified Data Having Ordered Categories.* Harvard University Press, Cambridge.

GOODMAN, L.A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.* **13** 10–69.

GOODMAN, L.A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables (with discussion). *Int. Statist. Rev.* **54** 243–309.

GOODMAN, L.A. and KRUSKAL, W.H. (1954). Measures of association for cross classification. *J. Amer. Statist. Assoc.* **49** 732–764.

GOODMAN, L.A. and KRUSKAL, W.H. (1979). *Measures of Association for Cross Classifications.* Springer-Verlag, New York.

GREENACRE, M.J. (1984). *Correspondence Analysis.* Academic Press, New York.

HABERMAN, S.J. (1974a). Log-linear models for frequency tables with ordered classifications. *Biometrics* **30** 589–600.

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

HABERMAN, S.J. (1981). Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction. *Ann. Statist.* **9** 1178–1186.

KIMELDORF, G. and SAMPSON, A. (1978). Monotone dependence. *Ann. Statist.* **6** 895–903.

KRUSKAL, W.H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.* **53** 814–861.

LEE, C.C. (1987). Chi-squared tests for and against an order restriction in multinomial parameters. *J. Amer. Statist. Assoc.* **82** 611–618.

MCCULLAGH, P. (1980). Regression models for ordinal data. *J. Roy. Statist. Soc. B* 109–142 (with discussion).

PEARSON, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag.* **50** No. 5, 157–175.

RAUBERTAS, R.F., LEE, C.C., and NORDHEIM, E.V. (1986). Hypothesis tests for normal means constrained by linear inequalities. *Commun. Statist.-Theor. Meth.* **15**(9) 2809–2833.

RITOV, Y. and GILULA, Z. (1987). The order restricted RC model for ordered contingency tables: Estimation and testing for fit. Manuscript.

ROBERTSON, T. (1978). Testing for and against an order restriction on multinomial parameters. *J. Amer. Statist. Assoc.* **73** 197–202.

ROBERTSON, T., WRIGHT, F.T., and DYKSTRA, R.L. (1988) *Order Restricted Statistical Inference.* Wiley.

SCHRIEVER, B.F. (1983). Scaling of order dependent categorical variables with correspondence analysis. *Int. Statist. Rev.* **51** 225–238.

SCHRIEVER, B.F. (1986). *Order Dependence.* CWI-tract 20. Centre for Mathematics and Computer Science, Amsterdam.

SETHURAMAN, J. (1977). The limit distribution of the Renyi maximum correlation, with applications to contingency tables and correspondence analysis. In *Proceedings of the 41st Session of the International Statistical Institute*, Book 4, 701–703.

YULE, G.U. (1900). On the association of attributes in statistics: with illustration from the material of the childhood society, c. *Philos. Trans. Roy. Soc. Ser. A* **194** 257–319.

DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PA 15213