SOME MULTIPLICATIVE MODELS FOR THE ANALYSIS OF CROSS CLASSIFIED DATA

LEO A. GOODMAN
UNIVERSITY OF CHICAGO

1. Introduction and summary

In the present article, we shall present some multiplicative models for the analysis of $R \times C$ contingency tables (that is, contingency tables with R rows and C columns), and shall apply these models to cross classified data in ways that will lead to a more complete analysis of these data than has heretofore been possible.

For the $R \times C$ contingency table, the usual model of "independence" in the table (that is, independence between the row classification and column classification of the table) is a simple example of a multiplicative model. For short, I shall call this model (that is, the model of independence between the row classification and column classification of the table) the I model. The model of "quasi-independence" in the $R \times C$ table, which was introduced and developed in my earlier work, and which I shall comment upon again later, is another example of a multiplicative model (see, for example, Goodman [9], [10], [12], [13], Caussinus [4], Bishop and Fienberg [2]). For short, I shall call this model the Q model. The various multiplicative models which I shall present here can be viewed as modifications or generalizations of the I model and/or the Q model.

To illustrate the application of these models, we shall analyze a 5×5 contingency table (Table I) in which there is a one to one correspondence between the five classes of the row classification and the five classes of the column classification, and in which the classes of the row (and column) classification can be ordered (from high to low). Although some of the particular models, which we shall consider herein (see Section 2), are particularly well suited to square contingency tables of this kind (in which there is this one to one correspondence and in which the classes of the row (and column) classification can be ordered), we wish to draw the reader's attention to the fact that the general class of multiplicative models presented in this article (see Sections 3 and 4) also includes a variety of models that can be applied more generally to rectangular contingency tables (as well as to square tables), where there may or may not be some kind of correspondence between the classes of the row and

This research was supported in part by Research Contract No. NSF GS 1905 from the Division of the Social Sciences of the National Science Foundation.

TABLE I

CROSS CLASSIFICATION OF BRITISH MALE SAMPLE ACCORDING TO EACH SUBJECT'S STATUS CATEGORY AND HIS FATHER'S STATUS CATEGORY, USING FIVE STATUS CATEGORIES

		Subject's Status				
		1	2	3	4	5
Father's Status	1	50	45	8	18	3
	2	28	174	84	154	55
	3	11	78	110	223	96
	4	14	150	185	714	447
	5	0	42	72	320	411

column classifications and where the classes of the row (and/or column) classifications may or may not be ordered.

Although the models presented here are described in different terms from the models in Haberman's fundamental work [20], it can be shown that the general theory and methods developed by Haberman are applicable to the various models considered herein. In most respects, the class of models considered by Haberman is broader than the class considered here; but by adopting a somewhat different perspective here and by confining our attention to a more limited class of models, we shall obtain some new results.

We shall formulate the models for the case where a random sample of n observations is drawn from the population cross classification table (the case where the sample size n is fixed), but the methods that we shall present here can also be applied to the case where a sample of $n_{i.}$ observations is drawn from the ith row, $i = 1, 2, \cdots, R$, of the population table (the case where the row marginals are fixed), or where a sample of $n_{.j}$ observations is drawn from the jth column, $j = 1, 2, \cdots, C$, of the population table (the case where the column marginals are fixed). The analysis in the case where the row marginals or the column marginals are fixed is similar, in most respects, to the analysis in the case where only the sample size n is fixed; but there are differences in the way some of the parameters of interest are defined in these cases, and there also are differences in the way these parameters are estimated. These differences will be discussed later (see Section 5).

The table that we shall use for illustrative purposes (Table I) presents data on intergenerational social mobility in Britain, which were collected by Glass and his coworkers [6]. The data in this table were obtained by a kind of stratified random sampling, but for our present exposition we view this table as a contingency table; that is, as if simple random sampling had been used. Table I was used earlier by Svalastoga [25], Levine [22], Mosteller [24], and Goodman [14] for purposes of comparison with a comparable 5×5 table (Table II) describing social mobility in Denmark, and it is a condensation of a 7×7 British table (Table III) that used a more detailed set of classes (status categories). Table I was formed from Table III by combining status category 2 with 3 and status category 6 with 7 in Table III, in order to make the status

TABLE II

CROSS CLASSIFICATION OF DANISH MALE SAMPLE ACCORDING TO EACH SUBJECT'S STATUS CATEGORY AND HIS FATHER'S STATUS CATEGORY, USING FIVE STATUS CATEGORIES

		Subject's Status				
		1	2	3	4	5
Father's Status	1	18	17	16	4	2
	2	24	105	109	59	21
	3	23	84	289	217	95
	4	8	49	175	348	198
	5	6	8	69	201	246

TABLE III

CROSS CLASSIFICATION OF BRITISH MALE SAMPLE ACCORDING TO EACH SUBJECT'S STATUS CATEGORY AND HIS FATHER'S STATUS CATEGORY, USING SEVEN STATUS CATEGORIES

		Subject's Status						
		1	2	3	4	5	6	7
Father's Status	1	50	19	26	8	18	6	2
	2	16	40	34	18	31	8	3
	3	12	35	65	66	123	23	21
	4	11	20	58	110	223	64	32
	5	14	36	114	185	714	258	189
	6	0	6	19	40	179	143	71
	7	0	3	14	32	141	91	106

categories more comparable to the corresponding categories in the Danish 5×5 table. For expository purposes, when illustrating the application of the models and methods presented here, we shall focus our attention on the analysis of Table I; but for the sake of completeness, we shall present corresponding results for Tables II and III as well, and shall also comment briefly upon these results (see Sections 6 and 7); which will further enrich our understanding of the data.

The main part of this paper will be concerned with the development of models and methods for the analysis of a given $R \times C$ table (or a given set of R or C multinomial populations). We shall also comment briefly in the final section (Section 8) on the extension and application of the models and methods presented to the analysis and comparison of two (or more) $R \times C$ cross classification tables, and to the analysis of multidimensional cross classification tables.

2. Some examples of multiplicative models

For expository purposes, we shall begin by considering first the usual model of "independence" between the row classification and the column classification in an $R \times C$ cross classification table (the I model), then the model of "quasi-independence" in the $R \times C$ table (the Q model), and then various modifications or generalizations of these models.

2.1. *The model of independence (the I model).* We shall now define the usual model of "independence" between the row classification and the column classification in an $R \times C$ population cross classification table. This model can be defined in various ways, and for our present development of the subject we shall proceed as follows. Let $\pi_{i,j}$ denote the probability that an individual in the $R \times C$ population table will fall in cell (i, j) of the table (that is, in the ith row and jth column), for $i = 1, 2, \cdots, R$, and $j = 1, 2, \cdots, C$. Then the row and column classifications are defined as "independent" if the probability $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \alpha_i \beta_j \qquad \text{for} \quad i = 1, 2, \cdots, R; \qquad j = 1, 2, \cdots, C, \tag{2.1.1}$$

for a set of positive constants α_i and β_j, for $i = 1, 2, \cdots, R$, and $j = 1, 2, \cdots, C$. When (2.1.1) is satisfied, α_i can be interpreted as the probability that an individual will fall in the ith row of the population table (when the α_i have been scaled so that $\Sigma_i \alpha_i = 1$), and a similar interpretation can be given to the β_j. If some of the rows or columns are empty, we consider the table consisting of the nonempty rows and columns.

For a sample of n individuals, let $f_{i,j}$ denote the number of individuals that fall in cell (i, j) of the table, and let $\hat{f}_{i,j}$ denote the maximum likelihood estimate of the expected number that would fall in cell (i, j) under a given model. Under the usual model of independence between the row classification and the column classification of the table, the $\hat{f}_{i,j}$ can be written as

$$\hat{f}_{i,j} = \frac{f_i^\alpha f_j^\beta}{n}, \tag{2.1.2}$$

where f_i^α and f_j^β denotes the ith row marginal and jth column marginal, respectively, in the table of the $f_{i,j}$. From (2.1.2) we see that the $\hat{f}_{i,j}$ can be written in the form

$$\hat{f}_{i,j} = a_i b_j, \tag{2.1.3}$$

where the a_i and b_j are such that

$$\sum_j \hat{f}_{i,j} = f_i^\alpha, \qquad \sum_i \hat{f}_{i,j} = f_j^\beta. \tag{2.1.4}$$

Conditions (2.1.4) can be rewritten as

$$\hat{f}_i^\alpha = f_i^\alpha, \qquad \hat{f}_j^\beta = f_j^\beta, \tag{2.1.5}$$

where the $\hat{f}_i^\alpha$ and $\hat{f}_j^\beta$ denote the ith row marginal and jth column marginal, respectively, in the table of the $\hat{f}_{i,j}$. Thus, the row and column marginals of the $\hat{f}$ fit the corresponding observed quantities.

Although the number of the a_i and the b_j in (2.1.3) is $R + C$, we can ignore one of these quantities (say, a_1) since $\hat{f}_{i,j}$ is unaffected by scaling the a_i and the b_j so that $a_1 = 1$; that is, by replacing a_i and b_j by $\tilde{a}_i = a_i/a_1$ and $\tilde{b}_j = a_1 b_j$, respectively. Similarly, although the number of restrictions described by (2.1.5) is $R + C$, we can ignore one of these restrictions (say, the first restriction), since

if $\hat{f}_j^\beta = f_j^\beta$ for $j = 1, 2, \cdots, C$, then $\Sigma_{i,j}\hat{f}_{i,j} = n$; and thus, if also $\hat{f}_i^\alpha = f_i^\alpha$ for $i = 2, 3, \cdots, R$, then $\hat{f}_1^\alpha = f_1^\alpha$. Thus, to calculate the degrees of freedom for testing the model of independence, we subtract $R + C - 1$ from $R \times C$, obtaining thereby the usual quantity, namely, $(R - 1)(C - 1)$.

2.2. *The model of quasi-independence (the Q model).* The model (2.1.1) in the preceding section applies to all $R \times C$ cells (i, j) for $i = 1, 2, \cdots, R$, $j = 1, 2, \cdots, C$. Now we shall consider a subset S of these cells; for example, the subset consisting of the cells that are not on the main diagonal of the table (that is, the cells (i, j) with $i \neq j$). For a given subset S, the row and column classification is defined as "quasi-independent" (with respect to S) if the probability $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \alpha_i\beta_j \qquad \text{for cells } (i, j) \text{ in } S, \tag{2.2.1}$$

for a set of positive constants α_i and β_j.

Since we are not concerned here with the cells that are not in S, we can assign zero probability to those cells. Thus, (2.2.1) can be rewritten as

$$\pi_{i,j} = \delta_{i,j}^S\alpha_i\beta_j \qquad \text{for} \quad i = 1, 2, \cdots, R; \qquad j = 1, 2, \cdots, C, \tag{2.2.2}$$

where

$$\delta_{i,j}^S = \begin{cases} 1 & \text{for cells } (i, j) \text{ in } S, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.3}$$

Letting $f_{i,j}$ denote the observed number of individuals that fall in cell (i, j), the maximum likelihood estimate $\hat{f}_{i,j}$ of the corresponding expected number (under model (2.2.1)) can be written as

$$\hat{f}_{i,j} = \delta_{i,j}^S a_i b_j, \tag{2.2.4}$$

where the a_i and b_j are such that

$$\hat{f}_i^\alpha = f_i^\alpha, \qquad \hat{f}_j^\beta = f_j^\beta, \tag{2.2.5}$$

and where now f_i^α and f_j^β are defined as

$$f_i^\alpha = \sum_j \delta_{i,j}^S f_{i,j}, \qquad f_j^\beta = \sum_i \delta_{i,j}^S f_{i,j}, \tag{2.2.6}$$

and the $\hat{f}_i^\alpha$ and $\hat{f}_j^\beta$ are defined similarly (with the $f_{i,j}$ in (2.2.6) replaced by the corresponding $\hat{f}_{i,j}$). Methods for calculating the $\hat{f}_{i,j}$ were discussed in, for example, Goodman [10], [13], and we shall return to them later herein.

To facilitate our understanding of matters that will be discussed later, we now introduce some new terminology. For each cell (i, j), let $\Lambda_{i,j}$ denote the set of parameters that appear in the formula for $\pi_{i,j}$ under a given model (see, for example, (2.2.1)). Thus, for the Q model, $\Lambda_{i,j}$ contains α_i and β_j if cell (i, j) is in S, and $\Lambda_{i,j}$ is empty if (i, j) is not in S. With this terminology, we can rewrite (2.2.6) as

$$f_i^\alpha = \sum_{g,h} \delta_{g,h}^{\alpha_i} f_{g,h}, \qquad f_j^\beta = \sum_{g,h} \delta_{g,h}^{\beta_j} f_{g,h}, \tag{2.2.7}$$

where

$$\delta_{g,h}^{\alpha_i} = \begin{cases} 1 & \text{if } \alpha_i \in \Lambda_{g,h}, \\ 0 & \text{otherwise,} \end{cases}$$

(2.2.8)

$$\delta_{g,h}^{\beta_j} = \begin{cases} 1 & \text{if } \beta_j \in \Lambda_{g,h}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that (2.2.7) states that f_i^α is the sum of the $f_{g,h}$ for cells (g, h) for which $\alpha_i \in \Lambda_{g,h}$; and that f_j^β is the corresponding sum for cells (g, h) for which $\beta_j \in \Lambda_{g,h}$. A similar statement applies to the $\hat{f}_i^\alpha$ and $\hat{f}_j^\beta$ (with the $f_{g,h}$ in (2.2.7) replaced by the corresponding $\hat{f}_{g,h}$).

To calculate the degrees of freedom for testing the model in this section, we consider first the case where the set S includes at least one cell from each row and column of the table, and where S is "inseparable" in the sense that it cannot be partitioned into two mutually exclusive (and exhaustive) subsets S_1 and S_2 that have no rows and no columns in common (see, for example, Goodman [13], Caussinus [4]). Note that S would be separable if the subsets S_1 and S_2 had no parameters in common; in other words, if the set of parameters that are contained in S_1 (that is, in $\Lambda_{i,j}$ for one or more of the cells (i, j) in S_1) and the set of parameters that are contained in S_2 (that is, in $\Lambda_{i,j}$ for one or more of the cells (i, j) in S_2) were mutually exclusive. For the case where S is inseparable, the remarks in the final paragraph of Section 2.1 can be directly applied; and in the present case, the degrees of freedom are obtained by subtracting $R + C - 1$ from $R \times C - V$, where V is the number of cells in the $R \times C$ table that are not included in S. Thus, as in the earlier literature on "quasi-independence," we find that there are $(R - 1)(C - 1) - V$ degrees of freedom.

In cases where entire rows and/or entire columns are not included in S, the above formula for the degrees of freedom can still be applied, except that "R" and "C" in that formula should be taken as the number of rows and columns, respectively, that contain at least one cell from S, and similarly the quantity V should be calculated for this (modified) "R" × "C" table. For cases in which S is separable, the results presented above can be applied separately to each subset that is itself not separable. A separable set S can always be partitioned into such subsets.

2.3. *The QO model, the QP model, the QN model, and the QPN model.* The results in the preceding section pertain to the case where S is any given subset of the cells in the $R \times C$ table. Consider now the case where the cross classification table is square (that is, $R = C$), where there is a one to one correspondence between the ith class of the row classification and the ith class of the column classification for $i = 1, 2, \cdots, R$, and where S is the set of cells that are not on the main diagonal (that is, the cells (i, j) with $i \neq j$). Since S consists of the off diagonal cells, we shall call the model of quasi-independence (with respect to this set S) the QO model. The remarks in the preceding section can be applied directly to the QO model. Note, for example, that the number of degrees of freedom for testing this model will be $(R - 1)(R - 1) - R = R^2 - 3R + 1$.

Consider now the case where the classes of the row (and column) classification in the $R \times R$ table can also be ordered from 1 to R, and where S is the set of cells (i, j) with $i > j$. For each cell in S, the difference $i - j$ is positive, and so we shall call this particular model of quasi-independence the QP model. From the remarks in the preceding section, we see that the number of degrees of freedom for testing the QP model will be $(R - 2)(R - 2) - [(R - 1)(R - 2)/2] = (R - 2)(R - 3)/2$. Note that the degrees of freedom are zero for $R = 3$; we shall be concerned here mainly with models for cases where $R > 3$.

Consider now the case where S is the set of cells (i, j) with $i < j$. For each cell in S, the difference $i - j$ is negative, and so we shall call this particular model of quasi-independence the QN model. As in the preceding paragraph, the number of degrees of freedom for testing the QN model will be $(R - 2)(R - 3)/2$.

Consider now the model in which the probability $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \begin{cases} \alpha_i \beta_j & \text{for } i > j, \\ \alpha'_i \beta'_j & \text{for } i < j. \end{cases} \tag{2.3.1}$$

This model states that both the QP model and the QN model are true, and so we shall call it the QPN model. Note that this model is not the same as the QO model, although both models pertain to the cells (i, j) with $i \neq j$. The QO model is a special case of the QPN model in which

$$\begin{aligned} \alpha'_i &= \Delta \alpha_i \qquad \text{for } i = 2, 3, \cdots, R - 1, \\ \beta'_j &= \Delta^* \beta_j \qquad \text{for } j = 2, 3, \cdots, R - 1, \end{aligned} \tag{2.3.2}$$

and in which

$$\Delta^* = \frac{1}{\Delta}. \tag{2.3.3}$$

The QPN model is not, strictly speaking, an example of a model of quasi-independence as this term was defined in the preceding section. Nevertheless, the methods developed earlier for the quasi-independence model (see, for example, Goodman [13]) can be applied to the QPN model, by analyzing separately the subsets S_1 and S_2, where S_1 is the set of cells (i, j) with $i > j$, and S_2 is the set of cells (i, j) with $i < j$; and applying the corresponding models (the QP and QN models) to those sets. Note that, although the sets S_1 and S_2 do have rows and columns in common, these sets are separable for the QPN model, since the set of parameters in S_1 (the α_i and β_j) and the set of parameters in S_2 (the α'_i and β'_j) are mutually exclusive.

The number of degrees of freedom for testing the QPN model is the sum of the degrees of freedom for testing the QP model in S_1 and the QN model in S_2. Thus, there are $(R - 2)(R - 3)$ degrees of freedom for testing the QPN model. Note that the difference between the degrees of freedom for the QO model and the QPN model is $2R - 5$, which corresponds to the sum of the degrees of freedom associated with testing condition (2.3.2) and condition (2.3.3) (namely, $2(R - 3)$ degrees of freedom for (2.3.2) and one degree of freedom for (2.3.3)).

We now extend the concept of "quasi-independence" to include models of the kind described by (2.3.1). Let $S_1, S_2, \cdots, S_K$ denote mutually exclusive subsets of the cells (i, j) in an $R \times C$ cross classification table. The model of "quasi-independence" (with respect to the subsets S_k, for $k = 1, 2, \cdots, K$) is defined by the condition that the probability $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \alpha_i^{(k)}\beta_j^{(k)} \qquad \text{for cells } (i, j) \text{ in } S_k, \tag{2.3.4}$$

for $k = 1, 2, \cdots, K$. The sets $S_1, S_2, \cdots, S_K$ are separable for the model defined by (2.3.4), since the set of parameters in S_k (namely, $\alpha_i^{(k)}$ and $\beta_j^{(k)}$) and the set of parameters in $S_{k'}$ (namely, $\alpha_i^{(k')}$ and $\beta_j^{(k')}$) are mutually exclusive for $k \neq k'$. To analyze this more general model of "quasi-independence," we can apply the methods developed earlier for the more usual quasi-independence model, by analyzing separately the sets $S_1, S_2, \cdots, S_K$, applying the corresponding model of quasi-independence to each set.

2.4. *The triangles parameter model (the T model).* We return again to the square contingency table ($R = C$) in which there is a one to one correspondence between the classes of the row and column classification, and in which the classes of the row (and column) classification are ordered from 1 to R. Consider now the special case of the QPN model in which condition (2.3.2) is satisfied (but condition (2.3.3) may or may not be satisfied). This special case of the QPN model is equivalent to the model in which the probability $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \alpha_i\beta_j\tau_k \qquad \text{for cells } (i, j) \text{ in } S_k, \tag{2.4.1}$$

for $k = 1, 2$, where S_1 is the set of cells (i, j) with $i > j$, and S_2 is the set of cells (i, j) with $i < j$. For cell (i, j) in S_k, under model (2.4.1) the set $\Lambda_{i,j}$ contains α_i, β_j, and τ_k; and $\Lambda_{i,j}$ is empty if (i, j) is not in S_1 or S_2. The model (2.4.1) differs from the QO model in that it introduces an additional set of parameters τ_k that pertains differentially to the triangular subsets S_k, for $k = 1$ and 2, and so we call this model the triangles parameter model (the T model).

Note that the sets S_1 and S_2 are not separable for the T model defined by (2.4.1) since the set of parameters in S_1 (namely, α_i, β_j, and τ_1) and the set of parameters in S_2 (namely, α_i, β_j, and τ_2) are not mutually exclusive—the parameters α_i for $i = 2, 3, \cdots, R - 1$ and β_j for $j = 2, 3, \cdots, R - 1$ are included in both sets. Although the T model is not an example of a quasi-independence model as defined in Section 2.2 (nor of the more general model of "quasi-independence" defined by (2.3.4)), the remarks in Section 2.2 can be directly extended to the T model. By direct extension of (2.2.4) to (2.2.8), we find that the estimate $\hat{f}_{i,j}$ under the T model can be written as

$$\hat{f}_{i,j} = a_i b_j t_k \qquad \text{for cells } (i, j) \text{ in } S_k, \tag{2.4.2}$$

for $k = 1$ and 2, where the a_i, b_j, and t_k are such that

$$\hat{f}_i^{\alpha} = f_i^{\alpha}, \qquad \hat{f}_j^{\beta} = f_j^{\beta}, \qquad \hat{f}_k^{\tau} = f_k^{\tau}, \tag{2.4.3}$$

where the f_i^{α} and f_j^{β} are defined by (2.2.7), and f_k^{τ} is defined by

$$f_k^\tau = \sum_{g,h} \delta_{g,h}^{\tau_k} f_{g,h}, \tag{2.4.4}$$

where

$$\delta_{g,h}^{\tau_k} = \begin{cases} 1 & \text{if } \tau_k \in \Lambda_{g,h}, \\ 0 & \text{otherwise;} \end{cases} \tag{2.4.5}$$

and with the $\hat{f}_i^\alpha$, $\hat{f}_j^\beta$, $\hat{f}_k^\tau$ defined similarly (with the $f_{g,h}$ in (2.4.4) replaced by the corresponding $\hat{f}_{g,h}$).

Although there are two t_k in (2.4.2) (namely, t_1 and t_2), we can ignore one of them (say, t_1) since $\hat{f}_{i,j}$ is unaffected by scaling the t_k (and the a_i) so that $t_1 = 1$; that is, by replacing t_k and a_i by $\tilde{t}_k = t_k/t_1$ and $\tilde{a}_i = a_i t_1$, respectively. Similarly, although there are two restrictions described by the third condition of (2.4.3) for $k = 1$ and 2, we can ignore one of these restrictions (say, the first restriction) since if $\hat{f}_i^\alpha = f_i^\alpha$ for $i = 1, 2, \cdots, R$, then $\Sigma \hat{f}_{i,j} = n$ (where the $\hat{f}$ are summed over the off diagonal cells and n is similarly calculated for the f); and thus, if $\hat{f}_2^\tau = f_2^\tau$, then $\hat{f}_1^\tau = f_1^\tau$. Since there is one more parameter in the T model than in the QO model, the number of degrees of freedom for testing the T model will be $R^2 - 3R = R(R-3)$, for $R \geqq 3$.

2.5. *The diagonals parameter model (the D model) and related models.* Consider now the model in which the probability $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \alpha_i \beta_j \delta_k \qquad \text{for cells } (i,j) \text{ in } S_k', \tag{2.5.1}$$

where S_k' is the set of cells (i,j) with $i - j = k$, for $k = \pm 1, \pm 2, \cdots, \pm(R-1)$. The model (2.5.1) differs from the QO model in that it introduces an additional set of parameters (the δ_k) that pertains differentially to the minor diagonals S_k for $k = \pm 1, \pm 2, \cdots, \pm(R-1)$, and so we call this model the diagonals parameter model (the D model).

The T model (see (2.4.1)) is a special case of the D model in which the following condition is satisfied:

$$\delta_k = \begin{cases} \delta^* & \text{for } k = 1, 2, \cdots, R-1, \\ \delta^{**} & \text{for } k = -1, -2, \cdots, -(R-1). \end{cases} \tag{2.5.2}$$

The remarks about the analysis of the T model in the preceding section can be extended directly to the D model. For example, under the D model, the $\hat{f}_{i,j}$ can be written as

$$\hat{f}_{i,j} = a_i b_j d_k \qquad \text{for cells } (i,j) \text{ in } S_k', \tag{2.5.3}$$

for $k = \pm 1, \pm 2, \cdots, \pm(R-1)$, where the a_i, b_j and d_k are such that

$$\hat{f}_i^\alpha = f_i^\alpha, \qquad \hat{f}_j^\beta = f_j^\beta, \qquad \hat{f}_k^\delta = f_k^\delta, \tag{2.5.4}$$

where the f_i^α and f_j^β are defined by (2.2.7) and f_k^δ is defined similarly to (2.4.4) and (2.4.5), with δ replacing τ in those formulae. (The corresponding $\hat{f}$ are defined similarly.)

Although there are $2(R-1)$ statistics d_k in (2.5.3) (namely, d_k for $k = \pm 1, \pm 2, \cdots, \pm(R-1)$), we can ignore two of them (say, d_k for $k = \pm 1$) since $\hat{f}_{i,j}$ is unaffected by transforming the d_k (and the a_i and the b_j) so that $d_k = 1$ for $k = \pm 1$; that is, by replacing d_k by

$$\tilde{d}_k = \frac{d_k d_*^{k-1}}{d_1} \qquad \text{for} \quad k = \pm 1, \pm 2, \cdots, \pm(R-1), \tag{2.5.5}$$

with

$$d_* = \left(\frac{d_{-1}}{d_1}\right)^{1/2}, \tag{2.5.6}$$

and replacing a_i and b_j by

$$\tilde{a}_i = \frac{a_i d_1}{d_*^{i-1}}, \qquad \tilde{b}_j = b_j d_*^j, \tag{2.5.7}$$

respectively. Thus since there are $2(R-2)$ more parameters in the D model than in the QO model, the number of degrees of freedom for testing the D model will be $R^2 - 5R + 5$, for $R \geqq 4$. There will be zero degrees of freedom for testing the D model in the case where $R = 3$.

Consider now the special case of the D model in which the following condition is satisfied:

$$\delta_k = \delta_{k^*} \qquad \text{for} \quad k^* = -k, \text{ with } k = 1, 2, \cdots, R-1. \tag{2.5.8}$$

In this case, the parameter δ_k pertains to the pair of minor diagonals S'_k and S'_{k^*} with $k^* = -k$; that is, to the cells (i, j) for which the absolute value $|i - j|$ is equal to k. We shall call this case the DA model. The earlier remarks about the analysis of the D model can be directly extended to the DA model, where now δ_k pertains to the paired minor diagonals, for $k = 1, 2, \cdots, R-1$. The DA model has $R-2$ more parameters than the QO model for $R \geqq 4$, and so the number of degrees of freedom for testing the DA model will be $R^2 - 4R + 3 = (R-1)(R-3)$, for $R \geqq 4$. For $R = 3$, the DA model is equivalent to the QO model, so that in this special case there will be one degree of freedom for testing the model. Note that the difference between the degrees of freedom for the DA model and the D model is $R-2$ for $R \geqq 4$, which corresponds to the degrees of freedom associated with testing condition (2.5.8), for $k = 2, 3, \cdots, R-1$ (with $\delta_k = 1$, for $k = \pm 1$).

Consider now the special case of the D model in which the following condition is satisfied:

$$\delta_k = 1, \qquad \text{for} \quad k = -1, -2, \cdots, -(R-1). \tag{2.5.9}$$

In this case, the parameter δ_k pertains only to the minor diagonals S'_k for which k is positive; that is, to the cells (i, j) for which $i - j = k$ is positive. We shall call this case the DP model. The condition (2.5.9) is actually equivalent to the following condition, which might appear (at first sight) to be more general;

namely,

$$\delta_k = \delta^k \quad \text{for} \quad k = -1, -2, \cdots, -(R-1), \tag{2.5.10}$$

for some positive constant δ. Conditions (2.5.9) and (2.5.10) are equivalent because the $\pi_{i,j}$ of (2.5.1) is unaffected by replacing δ_k, α_i, and β_j by

$$\tilde{\delta}_k = \frac{\delta_k}{\delta^k}, \qquad \tilde{\alpha}_i = \alpha_i \delta^{i-1}, \qquad \tilde{\beta}_j = \frac{\beta_j}{\delta^{j-1}}, \tag{2.5.11}$$

respectively. When the parameters are transformed so that $\delta_k = 1$ for $k = \pm 1$, the condition (2.5.10) is replaced by the following condition:

$$\delta_k = (\delta')^{k+1} \quad \text{for} \quad k = -1, -2, \cdots, -(R-1), \tag{2.5.12}$$

for a positive constant δ'.

The earlier remarks about the analysis of the D model can also be directly extended to the DP model, where now δ_k pertains to the minor diagonals S'_k for which k is positive for $k = 1, 2, \cdots, R-1$. The DP model has $R-1$ more parameters than the QO model (namely, the δ_k for $k = 2, 3, \cdots, R-1$; and δ' from (2.5.12)), and so the number of degrees of freedom for testing the DP model will be $R^2 - 4R + 2$, for $R \geqq 4$. There will be zero degrees of freedom for testing the DP model in the case where $R = 3$.

Consider now the special case of the D model in which the following condition is satisfied:

$$\delta_k = 1 \quad \text{for} \quad k = 1, 2, \cdots, R-1. \tag{2.5.13}$$

We shall call this case the DN model. By modifying in an obvious manner the remarks (pertaining to the DP model) in the preceding two paragraphs, they can be applied to the DN model.

2.6. *The crossings parameter model (the C model).* We return again to the QPN model, and consider the special case where the following condition is satisfied:

$$\alpha_i \beta_i = \alpha'_i \beta'_i \quad \text{for} \quad i = 2, 3, \cdots, R-1. \tag{2.6.1}$$

This special case of the QPN model is equivalent to the model in which the probability $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \alpha_i \beta_j \gamma'_{i,j} \quad \text{for} \quad i \neq j, \tag{2.6.2}$$

where

$$\gamma'_{i,j} = \begin{cases} \prod_{u=j}^{i-1} \gamma_u & \text{for} \quad i > j, \\ \prod_{u=i}^{j-1} \gamma_u & \text{for} \quad i < j. \end{cases} \tag{2.6.3}$$

The model (2.6.2)–(2.6.3) differs from the QO model in that it introduces an additional set of parameters, namely, the γ_u, for $u = 1, 2, \cdots, R-1$, that

pertains differentially (and multiplicatively) to each crossing between adjacent classes (from class u to $u + 1$ or from class $u + 1$ to u). In this model, the parameter γ_u pertaining to the crossing from class u to $u + 1$ is equal to the parameter pertaining to the crossing from $u + 1$ to u. We shall call this model the crossings parameter model (the C model). Note that the symbol C will be used to refer to this model and also, as earlier, to the number of columns in a rectangular contingency table; the meaning of the symbol will be clear in either case.

The remarks in the preceding section about the analysis of the models considered there can be directly extended to the C model. Under the C model, the $\hat{f}_{i,j}$ can be written as

$$\hat{f}_{i,j} = a_i b_j c'_{i,j}, \tag{2.6.4}$$

where

$$c'_{i,j} = \begin{cases} \prod_{u=j}^{i-1} c_u & \text{for } i > j, \\ \prod_{u=i}^{j-1} c_u & \text{for } i < j, \end{cases} \tag{2.6.5}$$

and where the a_i, b_j, and c_u are such that

$$\hat{f}_i^{\alpha} = f_i^{\alpha}, \qquad \hat{f}_j^{\beta} = f_j^{\beta}, \qquad \hat{f}_u^{\gamma} = f_u^{\gamma}, \tag{2.6.6}$$

where f_i^{α} and f_j^{β} are defined by (2.2.7), and f_u^{γ} is defined similarly to (2.4.4) and (2.4.5) with τ replaced by γ in those formulae. The corresponding $\hat{f}$ are defined similarly.

Although there are $(R - 1)$ statistics c_u in (2.6.4) and (2.6.5) (namely, c_u, for $u = 1, 2, \cdots, R - 1$), we can ignore two of them (c_1 and c_{R-1}) since $\hat{f}_{i,j}$ is unaffected by setting $c_1 = c_{R-1} = 1$; that is, by replacing these two c_u by $\tilde{c}_1 = \tilde{c}_{R-1} = 1$ and by replacing a_1, a_R, b_1, b_R by $\tilde{a}_1 = a_1 c_1$, $\tilde{a}_R = a_R c_{R-1}$, $\tilde{b}_1 = b_1 c_1$, $\tilde{b}_R = b_R c_{R-1}$. Since there are $R - 3$ more parameters in the C model than in the QO model, the number of degrees of freedom for testing the C model is $R^2 - 4R + 4 = (R - 2)^2$. Note that the difference between the degrees of freedom for the C model and the QPN model is $R - 2$, which corresponds to the degrees of freedom associated with testing condition (2.6.1).

From (2.6.2) and (2.6.3) we see that the factor $\gamma'_{i,j}$ associated with a change from the ith class (with respect to the row classification) to the jth class (with respect to the column classification) was a product of the γ_u parameters pertaining to successive one step changes (crossings) from class i to j. Thus, the crossings parameter model (the C model) could also have been called the "one step Markov" model. (This terminology is appropriate, in a certain sense, since the parameter γ_u pertaining to an individual's one step change from class u to $u + 1$ depends only upon the class u and not upon the earlier history of changes that may have led to the individual's presence in class u. On the other hand, the terminology is not quite appropriate since the direction of change (from

class u to $u + 1$ or from class u to $u - 1$) does depend upon this earlier history.) Although the models in Haberman [20] are described there in different terms from the models presented here, it can be shown that the C model (see (2.6.4) and (2.6.5)) is equivalent, in most respect, to the "variable distance" model applied by Haberman. With the present formulation of the C model, some of the parameters and their estimates will be defined and calculated differently from the corresponding definitions and calculations that were applied to the "variable distance" model (see, for example, the related comments at the end of Section 7).

Let us suppose now that we wished to generalize the C model by replacing (2.6.3) by

$$\gamma'_{i,j} = \begin{cases} \prod_{u=j}^{i-1} \gamma_u^* & \text{for } i > j, \\ \prod_{u=i}^{j-1} \gamma_u^{**} & \text{for } i < j. \end{cases} \tag{2.6.7}$$

This model distinguishes (at first sight) between the parameter γ_u^* pertaining to the crossing from the uth row class to the $(u + 1)$th column class and the parameter γ_u^{**} pertaining to the crossing from the $(u + 1)$th row class to the uth column class. Actually, this model is equivalent to the C model (with $\gamma_u^* = \gamma_u^{**}$) since the $\pi_{i,j}$ defined by (2.6.2) and (2.6.7) are unaffected by replacing the γ_u^* and γ_u^{**} by $\tilde{\gamma}_u = (\gamma_u^* \gamma_u^{**})^{1/2}$, and by replacing α_i and β_j by $\tilde{\alpha}_i = \alpha_i \gamma'_{i-1}$ and $\tilde{\beta}_j = \beta_j / \gamma'_{j-1}$, with

$$\gamma'_i = \begin{cases} 1 & \text{for } i = 1, \\ (\gamma_i^* / \gamma_i^{**})^{1/2} & \text{for } i = 2, 3, \cdots, R - 1. \end{cases} \tag{2.6.8}$$

Note, in particular, that the special case of the model defined by (2.6.2) and (2.6.7), in which $\gamma_u^{**} = 1$ for $u = 1, 2, \cdots, R - 1$, is equivalent to the C model as defined by (2.6.2) and (2.6.3). A similar remark applies for the special case of the model defined by (2.6.2) and (2.6.7) in which $\gamma_u^* = 1$ for $u = 1, 2, \cdots, R - 1$.

2.7. *The diagonals crossings parameter model (the DC model) and other combined models.* Consider now the model in which the probability $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \alpha_i \beta_j \gamma'_{i,j} \delta_k \qquad \text{for cells } (i, j) \text{ in } S'_k, \tag{2.7.1}$$

where S'_k is defined as in (2.5.1), and $\gamma'_{i,j}$ is defined as in (2.6.3). The remarks in Sections 2.5 and 2.6 can be directly extended to apply to this model (the DC model).

The DC model differs from the D model in that it includes an additional set of parameters (namely, the γ_u, for $u = 1, 2, \cdots, R - 1$). We noted earlier that γ_1 and γ_{R-1} could be ignored in the C model, and now for the DC model we also find this to be the case and that, in addition, this model is unaffected by a change in scale for the γ_u (with corresponding changes made in the other parameters). Thus, there are $R - 4$ more parameters in the DC model than in the D model, and so the number of degrees of freedom for testing the DC model will be $R^2 - 6R + 9 = (R - 3)^2$.

The C model can be combined in a similar way with the other models in Section 2.5, obtaining thereby the DAC model, the DPC model, and the DNC model. In these cases too the number of degrees of freedom for the models are obtained by subtracting $R - 4$ from the number for the corresponding model in Section 2.5. Thus, for the DAC model, we obtain $R^2 - 5R + 7$ degrees of freedom; and for the DPC and DNC models, we obtain $R^2 - 5R + 6 = (R - 2)(R - 3)$ degrees of freedom. For R = 3, the DAC model is equivalent to the DA model and the QO model. In addition, the T model of Section 2.3 can be combined with the DA model, the C model, and the DAC model. The number of degrees of freedom for these combined models are obtained by subtracting one from the number for the corresponding (uncombined) model. Thus, the degrees of freedom for the DAT model, the CT model, and the DACT model are $R^2 - 4R + 2$, $R^2 - 4R + 3 = (R - 1)(R - 3)$, and $R^2 - 5R + 6 = (R - 2)(R - 3)$, respectively. The latter two formulas apply for $R \geqq 3$, while the first formula applies for $R \geqq 4$. For $R = 3$, the DAT model actually has zero degrees of freedom, as do the other models (discussed in the present section and in Section 2.4) that include the triangles parameter τ_k.

All of the models considered in the present section and in Sections 2.4 to 2.6 are special cases of the DC model. For models that do not include the crossings parameters, we have

$$\gamma_u = 1 \qquad \text{for} \quad u = 2, 3, \cdots, R - 2, \tag{2.7.2}$$

in (2.7.1) (see (2.6.3)). For models that do not include the diagonals parameters, we have

$$\delta_k = 1 \qquad \text{for} \quad k = \pm 1, \pm 2, \cdots, \pm(R - 1), \tag{2.7.3}$$

in (2.7.1). For models that include the triangles parameter (but not the parameter pertaining to the paired minor diagonals), the δ_k in (2.7.1) will satisfy condition (2.5.2). Similarly, the δ_k in (2.7.1) will satisfy condition (2.5.8) for models that include parameters pertaining to the paired minor diagonals (but not the triangles parameter), and they will satisfy the following condition for models that include both the parameters pertaining to the paired minor diagonals and the triangles parameters:

$$\delta_k = \tau' \delta_{k^*} \qquad \text{for} \quad k^* = -k, \text{ with } k = 1, 2, \cdots, R - 1. \tag{2.7.4}$$

Furthermore, the δ_k in (2.7.1) will satisfy condition (2.5.9) for models that do not include parameters pertaining to the "negative" diagonals, and they will satisfy (2.5.13) for models that do not include parameters pertaining to the "positive" diagonals.

2.8. *The DC Model for the full table (the DCF model) and related models.* The models in Sections 2.3 to 2.7 were concerned with the analysis of the off diagonal cells, and were extensions of the QO model. Now we shall present models for the analysis of the full table—models that are extensions of the *I* model.

Consider the model in which the probability $\pi_{i,j}$ is given by (2.7.1), except that now we also apply (2.7.1) to the set S'_0 (that is, the cells (i,j) with $i - j = 0$), as well as to the sets S'_k as defined in (2.5.1) for $k = \pm 1, \pm 2, \cdots, \pm(R-1)$; and we set $\gamma'_{i,j} = 1$ for $i = j$. This model (the DCF model) differs from the DC model in that the R cells on the main diagonal are now included in the analysis, an additional diagonals parameter δ_0 is included, and two additional crossings parameters γ_1 and γ_{R-1} are also included. Although the probability $\pi_{i,j}$ for the DC model is unaffected by ignoring the parameters γ_1 and γ_{R-1} (that is, by setting $\gamma_1 = \gamma_{R-1} = 1$ and making corresponding changes in $\alpha_1, \alpha_R, \beta_1, \beta_R$), this is not so for the DCF model; the $\pi_{i,j}$ for the DCF model would be affected by ignoring γ_1 and γ_{R-1}. Therefore, the number of degrees of freedom for the DCF model can be obtained by adding $R - 3$ to the number for the DC model, thus obtaining $R^2 - 5R + 6 = (R-2)(R-3)$.

The relationship between the DC model and the DCF model, which we noted in the preceding paragraph, can be extended to other models. To each of the models introduced in Sections 2.4 to 2.7 for the analysis of the off diagonal cells, there is a corresponding model defined for the full table. The number of degrees of freedom for the model for the full table can be obtained by adding $R - 3$ to the number for the corresponding model (for the analysis of the off diagonal cells) if that model includes both diagonals and crossings parameters; by adding $R - 1$ to the number for the corresponding model if that model includes diagonals parameters but not crossings parameters, by adding $R - 2$ to the number for the corresponding model if that model includes crossings parameters but not diagonals parameters. (In the special case where $R = 3$, the above calculation is modified slightly if the corresponding model is one of those models for which the particular formula given earlier herein for the degrees of freedom does not apply when $R = 3$.) A similar calculation can be made when the model also includes the triangles parameter. Alternatively, the degrees of freedom for the models for the full table could be calculated directly using the same methods that were applied in Sections 2.4 to 2.7 (but without first calculating the results for the models for the off diagonal cells and then adding the appropriate quantities). The number of degrees of freedom for testing each of the models considered herein is given in Table IV.

3. The general case

We return now to the general $R \times C$ table. Let $\pi_{i,j}$ denote the probability that an observation will fall in cell (i,j), let $f_{i,j}$ denote the observed frequency in cell (i,j) for a sample of n observations, and let $\hat{f}_{i,j}$ denote the maximum likelihood estimate of the expected frequency under a given model. We shall now denote the parameters in the model as $\lambda_1, \lambda_2, \cdots, \lambda_W$ (with $\lambda_w > 0$ for $w = 1, 2, \cdots, W$), and we let Λ denote the full set of parameters; that is, $\Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_W\}$. For each cell (i,j) in that table, let $\Lambda_{i,j}$ denote a given subset of Λ. Consider now the model in which the probability $\pi_{i,j}$ can be

TABLE IV

THE DEGREES OF FREEDOM FOR TESTING VARIOUS MODELS APPLIED TO THE RECTANGULAR $R \times C$ TABLE AND TO THE SQUARE $R \times R$ TABLE (FOR $R \geqq 3$)

*The asterisk indicates that the formula does not apply for $R = 3$. For this special case, see comments in article.

**For the case of quasi-independence, see further details in the article.

Model	Degrees of Freedom
Independence	$(R - 1)(C - 1)$
Quasi-Independence	$[(R - 1)(C - 1) - V]$**
QO	$R^2 - 3R + 1$
QP (or QN)	$(R - 2)(R - 3)/2$
QPN	$(R - 2)(R - 3)$
T	$R(R - 3)$
D	$(R^2 - 5R + 5)$*
DA	$(R - 1)(R - 3)$*
DP (or DN)	$(R^2 - 4R + 2)$*
C	$(R - 2)^2$
DC	$(R - 3)^2$
DAC	$R^2 - 5R + 7$
DPC (or DNC)	$(R - 2)(R - 3)$
DAT	$(R^2 - 4R + 2)$*
CT	$(R - 1)(R - 3)$
DACT	$(R - 2)(R - 3)$
TF	$R^2 - 2R - 1$
DF	$(R - 2)^2$
DAF	$(R - 1)(R - 2)$
DPF (or DNF)	$R^2 - 3R + 1$
CF	$(R - 1)(R - 2)$
DCF	$(R - 2)(R - 3)$
DACF	$(R - 2)^2$
DPCF (or DNCF)	$(R - 1)(R - 3)$
DATF	$R^2 - 3R + 1$
CTF	$R(R - 3)$
DACTF	$(R - 1)(R - 3)$

written as

$$\pi_{i,j} = \prod_{\Lambda_{i,j}} \lambda_w, \tag{3.1}$$

where $\Pi_{\Lambda_{i,j}}$ denotes a product over the indices w for which $\lambda_w \in \Lambda_{i,j}$ with the product defined as zero when $\Lambda_{i,j}$ is empty.

Let f_w^λ and $\hat{f}_w^\lambda$ be defined by

$$f_w^\lambda = \sum_{g,h} \delta_{g,h}^{\lambda,w} f_{g,h}, \qquad \hat{f}_w^\lambda = \sum_{g,h} \delta_{g,h}^{\lambda,w} \hat{f}_{g,h}, \tag{3.2}$$

where

$$\delta_{g,h}^{\lambda,w} = \begin{cases} 1 & \text{if } \lambda_w \in \Lambda_{g,h}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

Then for the model (3.1), it is easy to show that $\hat{f}_{i,j}$ can be written as

$$\hat{f}_{i,j} = \prod_{\Lambda_{i,j}} \ell_w, \tag{3.4}$$

where the ℓ_w are such that

$$\hat{f}_w^{\lambda} = f_w^{\lambda} \quad \text{for} \quad w = 1, 2, \cdots, W. \tag{3.5}$$

To calculate the $\hat{f}_{i,j}$ defined by (3.4) we can proceed by the following iterative scaling method. At the initial step we define

$$\hat{f}_{i,j}(0) = \begin{cases} 0 & \text{if } \Lambda_{i,j} \text{ is empty,} \\ 1 & \text{otherwise.} \end{cases} \tag{3.6}$$

Then we set $v = w$ in the following formula, and we use (3.7) to calculate $\hat{f}_{i,j}(v)$ for $w = 1, 2, \cdots, W$:

$$\hat{f}_{i,j}(v) = \begin{cases} \hat{f}_{i,j}(v-1) f_w^{\lambda} / [\hat{f}(v-1)]_w^{\lambda} & \text{if } \lambda_w \in \Lambda_{i,j}, \\ \hat{f}_{i,j}(v-1) & \text{otherwise,} \end{cases} \tag{3.7}$$

where

$$[\hat{f}(v-1)]_w^{\lambda} = \sum_{g,h} \delta_{g,h}^{\lambda,w} \hat{f}_{g,h}(v-1). \tag{3.8}$$

This is the first cycle of iterations. For the second cycle, we set $v = W + w$ in (3.7), and we use (3.7) to calculate $\hat{f}_{i,j}(v)$ considering again $w = 1, 2, \cdots, W$. For the third cycle, we set $v = 2W + w$ in (3.7), and we proceed as in the preceding cycle of iterations. The cycles of iterations are continued until the $\hat{f}_{i,j}(v)$ satisfy condition (3.5).

The above method is a generalization of a corresponding procedure that was used earlier to calculate the maximum likelihood estimate $\hat{f}_{i,j}$ under the quasi-independence model and under other related models (see, for example, Caussinus [4], Bishop and Fienberg [2], Goodman [13], [17]). This method (as described above) does not provide estimates of the parameters λ_w. To calculate the ℓ_w in (3.4) we can proceed by any of the following three methods.

(1) After calculating the $\hat{f}_{i,j}$ by the iterative procedure given above, equation (3.4) can be solved for the ℓ_w. Explicit expressions for the ℓ_w as functions of the $\hat{f}_{i,j}$ can be obtained for the models of Sections 2.2 to 2.8. See the Appendix where such expressions are given.

(2) Instead of calculating the $\hat{f}_{i,j}$ by the iterative method (3.6) to (3.8), the ℓ_w in (3.4) can be calculated by a direct extension of the iterative procedure that was used earlier by Goodman [10], [13] to estimate the parameters in the quasi-independence model. To do this, we first note from (3.4) and (3.5) that

$$\hat{f}_w^{\lambda} = \ell_w \sum_{g,h} \delta_{g,h}^{\lambda,w} \prod_{\Lambda_{g,h,w}} \ell_u, \tag{3.9}$$

where $\Lambda_{g,h,w}$ is the set consisting of all those λ in $\Lambda_{g,h}$ except λ_w, and where $\Pi_{\Lambda_{g,h,w}}$

denotes a product over the indices u for which $\lambda_u \in \Lambda_{g,h,w}$. We can rewrite (3.5) as

$$\ell_w = \frac{f_w^\lambda}{\sum_{g,h} \delta_{g,h}^{\lambda,w} \prod_{\Lambda g,h,w} \ell_u}. \tag{3.10}$$

To start the iterative procedure for calculating ℓ_w, we define

$$\ell_w(0) = 1 \qquad \text{for} \quad w = 1, 2, \cdots, W. \tag{3.11}$$

For the first cycle of iterations, we use the following formula to calculate $\ell_w(v)$ for $v = 1, 2, \cdots, W$:

$$\ell_w(v) = \begin{cases} f_w^\lambda / \sum_{g,h} \delta_{g,h}^{\lambda,w} \prod_{\Lambda_{g,h,w}} \ell_u(v-1) & \text{if} \quad v = w. \\ \ell_w(v-1) & \text{otherwise.} \end{cases} \tag{3.12}$$

For the second cycle of iterations, we replace the condition that $w = v$, which appears on the right side of (3.12), by the condition that $w = v - W$, and then apply (3.12) for $v = W + 1, W + 2, \cdots, 2W$. For the third cycle, we replace the condition that $w = v$ by the condition that $w = v - 2W$, and then apply (3.12) for $v = 2W + 1, 2W + 2, \cdots, 3W$, and so on.

(3) Instead of calculating the ℓ_w in (3.4) by the methods described under (1) or (2) above, they can be determined by the following formula, making use of the terms $f_w^\lambda / [\hat{f}(v-1)]_w^\lambda$ calculated in the iterative scaling method described by (3.6) to (3.8):

$$\ell'_w(T) = \prod_{t=0}^{T-1} \left\{ \frac{f_w^\lambda}{[\hat{f}(tW + w - 1)]_w^\lambda} \right\} \tag{3.13}$$

where $\Pi_{t=0}^{T}$ denotes the product of the term in braces for $t = 0, 1, \cdots, T$; with $\Pi_{t=0}^{0}$ denoting the term in braces for $t = 0$. If the iterative scaling method is completed when T cycles of iterations have been carried out, then $\ell'_w(T)$ can be used as the ℓ_w in (3.4).

The following formula describes the relationship between the $\ell'_w(T)$ defined by (3.13) and the $\ell_w(v)$ defined by (3.12):

$$\ell'_w(T) = \ell_w(v) \qquad \text{for} \quad v = (T-1)W + w. \tag{3.14}$$

Formula (3.14) can be proved by mathematical induction on T. (For some related (but different) results on the relationship between the iterative scaling method described by (3.6) to (3.8) and the iterative method described by (3.11) and (3.12), see Goodman's article [13] on the model of quasi-independence, and Haberman's work [20] on more general models.)

It should be noted that, for a particular model, the ℓ_w in (3.4), which we can calculate by any of the methods described under (1), (2), or (3) above, may still need to be scaled (transformed) in order to obtain maximum likelihood estimates of the corresponding scaled (transformed) parameters. (The λ_w in (3.1) and the ℓ_w in (3.4) may not be uniquely defined until they have been scaled (transformed) in a suitable manner.) Thus, for a particular model, if the $\hat{f}_{i,j}$ are unaffected by

setting, say, $\ell_1, \ell_2, \cdots, \ell_U$ equal to one (where $U < W$) and by transforming $\ell_{U+1}, \ell_{U+2}, \cdots, \ell_W$ accordingly so that the transformed ℓ are uniquely defined, then these transformed quantities are the maximum likelihood estimates of the corresponding transformed parameters. (Actually, since (3.1) pertains to the probabilities $\pi_{i,j}$ whereas (3.4) pertains to the expected frequencies $\hat{f}_{i,j}$, the particular transformations that are used will determine whether a given transformed ℓ is an estimate of the corresponding transformed λ, or whether it is an estimate of the corresponding transformed λ multiplied by the sample size n.) For further comments on the transformed estimates, see Sections 4 and 7, and the Appendix.

When $f_{i,j} > 0$ for all cells (i,j) for which $\Lambda_{i,j}$ is not empty, then the results of Haberman [20] can be applied to show that the iterative procedures described here will converge to quantities that can be used to obtain the maximum likelihood estimates (that is, the $\hat{f}_{i,j}$ and the transformed ℓ). For cases where $f_{i,j} = 0$ for one or more cells (i,j) for which $\Lambda_{i,j}$ is not empty, the iterative procedures can still be used to obtain the maximum likelihood estimates, so long as the iterative procedures converge to a solution with $\hat{f}_{i,j} \neq 0$ when $\Lambda_{i,j}$ is not empty. The modified Newton-Raphson method, which was applied by Haberman [20] in his numerical example, has a more rapid convergence rate than do the iterative procedures described here (see Haberman [20]); but for data analysis in which a number of different models (of the kind presented here) are applied, the procedures described here have the advantage of being easier to program for a computer.

Before closing this section, we shall give an example of a multiplicative model that is not included within the class of models defined by (3.1), and that cannot be analyzed using the methods presented above. Consider again the QPN model in the special case where

$$\begin{aligned} \alpha'_i &= \Delta\alpha_i\phi^i \qquad \text{for} \quad i = 2, 3, \cdots, R-1, \\ \beta'_j &= \frac{\Delta^*\beta_j}{\phi^j} \qquad \text{for} \quad j = 2, 3, \cdots, R-1. \end{aligned} \tag{3.15}$$

This special case of the QPN model is equivalent to the special case of the DAT model in which the additional condition

$$\delta_k = \delta^k, \tag{3.16}$$

where $\delta = 1/\sqrt{\phi}$, is imposed. Although the DAT model, as defined earlier, can be analyzed by the methods presented in the present section, different methods are required for the special case in which condition (3.16) is imposed. The modified Newton-Raphson method presented in Haberman [20] can be applied to this special case. (The "fixed distance" model, which was applied in Haberman [20], is equivalent, in most respects, to the special case of the DA model in which condition (3.16) is imposed. In the present paragraph (see (3.15)) we have been considering the more general DAT model, rather than the DA model, in the special case where condition (3.16) is imposed.)

4. Unrestricted and restricted multiplicative models

The model defined by (3.1) was "intrinsically unrestricted" in the sense that no restrictions (or conditions) were imposed upon the parameters λ_w for $w = 1, 2, \cdots, W$, except for the fact that these parameters were such that

$$\sum_{i,j} \pi_{i,j} = 1. \tag{4.1}$$

(Of course, since the parameters λ_w were used in (3.1) to form the probabilities $\pi_{i,j}$ for the cells (i,j) where $\pi_{i,j} > 0$, we also assumed that $\lambda_w > 0$ for $w = 1, 2, \cdots, W$.) Formula (3.1) can also be written as

$$\pi_{i,j} = \Big[\prod_{\Lambda_{i,j}} \lambda_w\Big] (\Sigma)^{-1} \tag{4.2}$$

where Σ denotes the summation of the $\Pi_{\Lambda_{i,j}} \lambda_w$ over all cells (i,j) in the $R \times C$ table. Expressing the model in the form (4.2), we see that the λ_w in (4.2) are even unrestricted in the sense that the condition (4.1) does not impose any restrictions upon them.

For the various models introduced in Sections 2.1 to 2.8, we noted that the $\pi_{i,j}$ were unaffected by scaling (transforming) some of the λ parameters in certain ways. The parameters as described in formulas of the form (3.1) or (4.2) (see, for example, (2.1.1), (2.2.1), (2.3.1), and so on) were not uniquely defined until certain kinds of restrictions were imposed upon them (for example, the restriction that $\alpha_1 = 1$ in formula (2.1.1)). For each of the models in Sections 2.1 to 2.8, in order to calculate the degrees of freedom for testing the model, we described restrictions that could be imposed upon the λ parameters that (a) would uniquely define these parameters, and (b) would not affect the $\pi_{i,j}$. In some of the earlier sections, these restrictions were imposed upon the corresponding estimates of the λ parameters, but they could as well have been imposed upon the λ parameters. Since these particular kinds of restrictions did not affect the $\pi_{i,j}$, the models obtained when the restrictions were imposed were equivalent to the unrestricted models. This was the case for each of the models considered in Sections 2.1 to 2.8.

In addition to introducing restrictions upon the λ parameters that did not affect the $\pi_{i,j}$ (in order to calculate the degrees of freedom for testing the model or to uniquely define the parameters), we also introduced certain kinds of restrictions upon the parameters that did affect the $\pi_{i,j}$. For example, conditions (2.5.8), (2.5.9), and (2.5.10) were restrictions imposed upon the parameters that changed the D model into the DA model, the DP model, and the DN model, respectively. Despite the fact that these particular restrictions did affect the $\pi_{i,j}$, the models obtained when these restrictions were imposed could also be expressed in the general form (3.1) or (4.2), and so these models were also equivalent to unrestricted models. On the other hand, we noted at the end of Section 3, that if condition (3.16) were imposed upon the parameters of the DAT model (or the DA model, the DAC model, or the DACT model), the model obtained thereby would not be within the class of models defined by (3.1).

We shall call a model "intrinsically unrestricted" if it can be expressed in the general form (3.1) or (4.2). (Thus, the DA model, the DP model and the DN model are intrinsically unrestricted in this sense, despite conditions (2.5.8), (2.5.9), (2.5.10); whereas, the modification of the DAT model, which is obtained when condition (3.16) is imposed, is not unrestricted.) We shall now describe some of the kinds of restrictions that can be imposed upon a model of the general form (3.1) which are such that the modified models obtained thereby would also be intrinsically unrestricted.

Let H denote a given model of the form (3.1). The model H can be described by the set of parameters $\Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_W\}$, and by the subsets $\Lambda_{i,j}$ that are defined for each cell (i, j) in the table (see Section 3). Let Λ' be a given subset of the set Λ, and let H' denote the modification of model H that is obtained by imposing the condition

$$\lambda_u = 1 \qquad \text{for} \quad \lambda_u \in \Lambda'. \tag{4.3}$$

Despite the fact that this modification (that is, the model H') is a special case of model H that satisfied condition (4.3), it is also intrinsically unrestricted. Model H' can be expressed in the form (3.1) simply by deleting from Λ and from each set $\Lambda_{i,j}$ any λ_u that are included in Λ'. (For some examples of the modification H', note that the QO model and each of the models in Sections 2.4 to 2.7, except the models that include the paired minor diagonals parameters and/or the triangles parameter (for example, the DA model, the DAT model, and so forth), can be formed from the DC model by this type of modification.)

Now let $\Lambda_1^*, \Lambda_2^*, \cdots, \Lambda_K^*$ denote mutually exclusive subsets of the parameters in Λ; and for each parameter λ_w in Λ, let S_w^λ denote the set of cells (i, j) for which $\lambda_w \in \Lambda_{i,j}$. Consider the modification of model H that is obtained by imposing the conditions

$$\lambda_u = \lambda_{u'} \qquad \text{for all } \lambda_u \text{ and } \lambda_{u'} \in \Lambda_k^*, \tag{4.4}$$

for $k = 1, 2, \cdots, K$. This modification will be called a H'' type of modification if each Λ_k^* is such that the sets S_u and $S_{u'}$ are mutually exclusive for all λ_u and $\lambda_{u'}$ in Λ_k^*, for $k = 1, 2, \cdots, K$ (that is, if all λ_u and $\lambda_{u'}$ in Λ_k^* are "separable"). For some examples of the H'' type of modification, note that the QO model can be formed from the QPN model by this type of modification, and the models that include the paired minor diagonals parameters or the triangles parameter can also be formed from the corresponding models that include the diagonals parameters, by this type of modification. The model obtained by the H'' type of modification is also intrinsically unrestricted. It can be expressed in the form (3.1) by removing from Λ and from each $\Lambda_{i,j}$ any λ that are included in Λ_k^*, and by replacing each of these λ by a single parameter, say λ_k^* for $k = 1, 2, \cdots, K$.

The modification described by (2.7.4) might appear (at first sight) to differ from a H'' type of modification, but the change from the DC model to the DACT (or from the D model to the DAT model), which is described by the condition (2.7.4), can also be expressed in the following equivalent way. Without affecting the probabilities $\pi_{i,j}$ in the DC model (or the D model), the triangles

parameters could be included in the expression (2.7.1) (or the expression (2.5.1)) for the $\pi_{i,j}$ (as a multiplicative factor of the kind appearing in (2.4.1)); and with the inclusion of these parameters in the model, the change from the DC model to the DACT model (or from the D model to the DAT model) can be expressed by condition (2.5.8) (rather than (2.7.4)), which is obviously a condition of the H'' type. A similar kind of remark can be made about the change from the QPN model to the T model, the change from the QPN to the C model, and the change from the QPN model to the CT model.

For a given model H of the form (3.1), we noted in Section 3 that the estimate $\hat{f}_{i,j}$ of the expected frequency satisfied condition (3.5). For the modification H' of H obtained by imposing condition (4.3), the estimate $\hat{f}_{i,j}$ of the expected frequency under model H' will satisfy the following modification of (3.5):

$$\hat{f}_w^\lambda = f_w^\lambda \quad \text{if} \quad \lambda_w \notin \Lambda'. \tag{4.5}$$

For the modification H'' of H obtained by imposing condition (4.4), the estimate $\hat{f}_{i,j}$ of the expected frequency under model H'' will satisfy the following modification of (3.5):

$$\begin{aligned} &\hat{f}_w^\lambda = f_w^\lambda \quad \text{if} \quad \lambda_w \notin \Lambda_k^* \quad \text{for} \quad k = 1, 2, \cdots, K, \\ &\sum_{\Lambda_k^*} \hat{f}_u^\lambda = \sum_{\Lambda_k^*} f_u^\lambda \quad \text{for} \quad k = 1, 2, \cdots, K, \end{aligned} \tag{4.6}$$

where $\Sigma_{\Lambda_k^*}$ denotes summation over the indices u for which $\lambda_u \in \Lambda_k^*$.

We noted earlier that all of the models in Section 2.4 to 2.7 can be obtained from the DC model by using modification H' and/or H''. We shall now show how the models of Section 2.8 can be obtained from the DC model by using these kinds of modifications.

The DC model (as defined by (2.7.1)) was concerned with the analysis of the off diagonal cells, and so for simplicity we could set $\pi_{i,i} = 0$ in (2.7.1). On the other hand, we need not have imposed this restriction on the $\pi_{i,i}$, and could instead have defined the DC model by writing the probability $\pi_{i,j}$ as

$$\pi_{i,j} = \begin{cases} \alpha_i \beta_j \gamma'_{i,j} \delta_k & \text{for cells } (i,j) \text{ in } S'_k (k = \pm 1, \pm 2, \cdots, \pm(R-1)), \\ \pi_{i,i} & \text{for cells } (i,i) \text{ in } S'_0. \end{cases} \tag{4.7}$$

The $\pi_{i,i}$ on the right side of (4.7) can be viewed as "intrinsically unrestricted" λ parameters, as can the parameters α_i, β_j, γ_u, and δ_k. If the condition

$$\pi_{i,i} = \alpha_i \beta_i \delta_0 \tag{4.8}$$

is imposed upon the $\pi_{i,i}$ in (4.7), the DC model will be changed to the DCF model. To express condition (4.8) in the form (4.4) (that is, as a modification of the H'' type), we first note that the $\pi_{i,i}$ in (4.6) can be written as $\alpha'_i \beta'_i \delta_0$ (where the α'_i, β'_i, δ_0 are intrinsically unrestricted λ parameters) without affecting the probabilities in the DC model (see 4.7); and we then impose the condition

$$\alpha_i = \alpha'_i, \qquad \beta_j = \beta'_j \qquad \text{for} \quad i = 1, 2, \cdots, R; j = 1, 2, \cdots, R, \tag{4.9}$$

to change the DC model to the DCF model.

All of the models in Section 2.8 can be obtained from the DCF model by using modifications H' and/or H'' in the same ways that these modifications would be used to obtain the corresponding models of Sections 2.4 to 2.7 from the DC model. Since we noted in the preceding paragraph that the DCF model can be obtained from the DC model using a modification of the H'' type, we now see that all of the models in Section 2.8 can also be obtained from the DC model by using the modifications H' and/or H''. Furthermore, each of the models in Section 2.8 can be obtained from the corresponding model in Sections 2.4 to 2.7 by using a modification of the H''type in the same way that it was used above to obtain the DCF model from the DC model.

As we noted earlier, the present article is limited to multiplicative models of the general form (3.1), that is, models that are intrinsically unrestricted. These include all the models of Sections 2.1 to 2.8 and many others as well, but they do not include "restricted" models of the kind referred to at the end of Section 3.

5. The ratio index and the relative difference index for the cells on the main diagonal

Consider again the DC model for the off diagonal cells, and the corresponding model for the full table (that is, the DCF model). For the DCF model (as defined in Section 2.8), the probability $\pi_{i,j}$ can be written as

$$\mu_{i,j} = \alpha_i \beta_j \gamma'_{i,j} \delta_k \qquad \text{for cells } (i,j) \text{ in } S'_k, \tag{5.1}$$

for $k = 0, \pm 1, \pm 2, \cdots, \pm(R-1)$, with $\gamma'_{i,j} = 1$ for $i = j$. Thus, for both the DCF model and the DC model, the probability $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \alpha_i \beta_j \gamma'_{i,j} \delta_k \mu_{i,j} \qquad \text{for cells } (i,j) \text{ in } S'_k, \tag{5.2}$$

for $k = 0, \pm 1, \pm 2, \cdots, \pm(R-1)$, where the $\mu_{i,j}$ satisfy condition

$$\mu_{i,j} = 1 \qquad \text{for all cells } (i,j), \tag{5.3}$$

for the DCF model, and condition

$$\mu_{i,j} = \begin{cases} 1 & \text{for} \quad i \neq j, \\ \pi_{i,i}/(\alpha_i \beta_i \delta_0) & \text{for} \quad i = j \end{cases} \tag{5.4}$$

for the DC model. Note that the $\pi_{i,j}$ defined by (5.2) and (5.4) are equivalent to the $\pi_{i,j}$ defined by (4.7).

In view of (5.4), the quantity

$$\mu_i = \frac{\pi_{i,i}}{\alpha_i \beta_i \delta_0} \tag{5.5}$$

is of interest. We shall call μ_i the ratio index for the ith cell on the main diagonal.

(This index was defined earlier in Goodman [14] for the model of quasi-independence, and it was called there a "new index of immobility.") For models that do not include diagonals parameters (or the triangles parameter), we set $\delta_0 = 1$; for the other models of Section 2.4 to 2.7, the parameter δ_0 cannot be estimated from the data without introducing some additional assumptions (which we shall discuss in Section 7); for the models of Section 2.8, we set $\mu_i = 1$ for $i = 1, 2, \cdots, R$, and the parameter δ_0 (and the other δ_k) in (5.1) can be estimated from the data.

In models that include crossings parameters for the analysis of the off diagonal cells (for example, the C model or DC model), we noted earlier (see Sections 2.6 and 2.7) that the probability $\pi_{i,j}$ is unaffected by setting $\gamma_1 = \gamma_{R-1} = 1$, and by making corresponding changes in α_1, α_R, β_1, β_R. On the other hand, instead of setting $\gamma_1 = \gamma_{R-1} = 1$, the $\pi_{i,j}$ would also be unaffected by making some other assumptions about γ_1 and γ_{R-1} (with corresponding changes made in α_1, α_R, β_1, β_R); and with these assumptions (which we shall discuss in Section 7), the data could be used to estimate γ_1 and γ_{R-1}. These assumptions about γ_1 and γ_{R-1} will affect $\alpha_1\beta_1$ and $\alpha_R\beta_R$; and they thereby will affect μ_1 and μ_R.

For the DC model for analyzing the off diagonal cells, the remarks in the preceding two paragraphs indicate that there is an element of arbitrariness about μ_1, μ_R and δ_0 (that is, the additional assumptions, to which we referred in those paragraphs, will affect these quantities). This was not the case for the DCF model (see Section 2.8). Note that the difference between the degrees of freedom for the DCF model and the DC model is $R - 3$, which corresponds to the degrees of freedom associated with testing the hypothesis

$$\mu_i = \mu \qquad \text{for} \quad i = 2, 3, \cdots, R - 1, \tag{5.6}$$

where μ is unspecified. A similar kind of remark can be applied to the difference between the degrees of freedom for each of the models in Section 2.8 and the corresponding model from Sections 2.4 to 2.7. For the models in Section 2.8 that do not include crossings parameters, replace (5.6) by the same condition applied for $i = 1, 2, \cdots, R$ (rather than for $i = 2, 3, \cdots, R - 1$); and for models that do not include diagonals parameters (or the triangles parameter), replace the unspecified quantity μ in (5.6) by one.

Let us now denote $\alpha_i\beta_i\delta_0$ by $\pi^*_{i,i}$, and let $\pi^*_{i,j} = \pi_{i,j}$ for $i \neq j$, with the $\pi_{i,j}$ satisfying condition (4.7) for the DC model. With this notation, the index μ_i defined by (5.5) can be written as

$$\mu_i = \frac{\pi_{i,i}}{\pi^*_{i,i}}. \tag{5.7}$$

Instead of this ratio index, consider now the quantity

$$\mu^*_i = \frac{(\pi_{i,i} - \pi^*_{i,i})}{\pi_{i.}}, \tag{5.8}$$

where $\pi_{i.} = \Sigma_{j=1}^{R} \pi_{i,j}$, which we shall call the relative difference index for the ith cell on the main diagonal. This index was defined earlier in Goodman [16]

for the model of quasi-independence, and it was called there the "index of persistence."

Let $\tilde{\pi}_{i,j}$ and $\tilde{\pi}^*_{i,j}$ denote $\pi_{i,j}/\pi_{i\cdot}$ and $\pi^*_{i,j}/\pi^*_{i\cdot}$, respectively, where $\pi^*_{i\cdot} = \sum_{j=1}^{R} \pi^*_{i,j}$. From (4.7) and (5.8) we find that $\tilde{\pi}_{i,j}$ can be written as

$$\tilde{\pi}_{i,j} = \begin{cases} (1 - \mu^*_i)\tilde{\pi}^*_{i,j} & \text{for } i \neq j, \\ \mu^*_i + [(1 - \mu^*_i)\tilde{\pi}^*_{i,j}] & \text{for } i = j. \end{cases} \tag{5.9}$$

Because of (5.9), the index μ^*_i can be interpreted as the proportion of "stayers" among those individuals who are in the ith row in the population. This interpretation would apply when $\mu^*_i > 0$.

For the case where $\mu^*_i < 0$, we can rewrite (5.9) as

$$\tilde{\pi}_{i,j} = \begin{cases} \tilde{\pi}^*_{i,j}(1 + \nu_i) & \text{for } i \neq j, \\ \tilde{\pi}^*_{i,i} - \nu_i(1 - \tilde{\pi}^*_{i,i}) & \text{for } i = j, \end{cases} \tag{5.10}$$

where $\nu_i = -\mu^*_i$. Because of (5.10), the index ν_i (that is, the index $-\mu^*_i$) can be interpreted (when $\mu^*_i < 0$) as the proportion of individuals who have a "second chance" to move out of the ith row class among those individuals who are in the ith row in the population. This interpretation can be applied when $\tilde{\pi}^*_{i,i} > \nu_i$. The interpretations presented in the present paragraph, and in the preceding one, are extensions of the interpretations introduced in Goodman [16] for the model of quasi-independence.

From (5.7) and (5.8), we see that μ_i and μ^*_i are related as follows:

$$\begin{aligned} \mu_i &= \frac{\tilde{\pi}_{i,i}}{\tilde{\pi}_{i,i} - \mu^*_i} \\ \mu^*_i &= \frac{\tilde{\pi}_{i,i}(\mu_i - 1)}{\mu_i}. \end{aligned} \tag{5.11}$$

In addition, the index μ^*_i can be rewritten as

$$\mu^*_i = \frac{\tilde{\pi}_{i,i} - \tilde{\pi}^*_{i,i}}{1 - \tilde{\pi}^*_{i,i}}. \tag{5.12}$$

Thus, the index μ^*_i measures the difference $(\tilde{\pi}_{i,i} - \tilde{\pi}^*_{i,i})$ relative to the difference $(1 - \tilde{\pi}^*_{i,i})$.

From (5.12) we see that the index μ^*_i can be expressed as a function of $\tilde{\pi}_{i,i}$ and $\tilde{\pi}^*_{i,i}$ (rather than $\pi_{i,i}$ and $\pi^*_{i,i}$), whereas the corresponding ratio index μ_i was a function of $\pi_{i,i}$ and $\pi^*_{i,i}$. (Compare (5.12) with (5.7).) Note that $\tilde{\pi}_{i,j}$ is the conditional probability that an individual will fall in the jth column class of the population table, given that he is in the ith row class; and $\tilde{\pi}^*_{i,j}$ is the corresponding conditional probability obtained by replacing $\pi_{i,i}$ by $\alpha_i\beta_i\delta_0$. The conditional probabilities $\tilde{\pi}_{i,j}$ and $\tilde{\pi}^*_{i,j}$ are particularly relevant when the data to be analyzed have been obtained from a sample of $n_{i\cdot}$ individuals drawn from the ith row class for $i = 1, 2, \cdots, R$ (that is, when the row marginals are fixed), rather than from a sample of n individuals drawn from the population cross classification

table. It is in this context that the index μ_i^* (rather than the index μ_i) is particularly relevant.

When a sample of n individuals is drawn from the population cross classification table, the index μ_i may be viewed as one of the parameters of the model, since the $\pi_{i,j}$ can be expressed by (5.2) with the parameter $\mu_{i,j}$ in (5.2) set equal to μ_i for $i = j$ (see (5.4) and (5.5)). This remark can be applied to any of the models of Sections 2.4 to 2.7; and for the models of Section 2.8 the parameter μ_i is set equal to one. Similarly, when a sample of $n_{i\cdot}$ individuals is drawn from the ith row class for $i = 1, 2, \cdots, R$, the index μ_i^* may be viewed as one of the parameters of the model, since the $\tilde{\pi}_{i,j}$ can be expressed by (5.9).

Note that the index μ_i is "symmetric" in the sense that it is invariant when the row classes are interchanged with the column classes. Of course, this will not be the case for the index μ_i^*. In addition to the index μ_i^* defined herein for the case where the row marginals are fixed, there is the relative difference index that would be defined in a directly analogous way for the case where the column marginals are fixed.

Applying the general methods of estimation described in Section 3 to the DC model, we can calculate the estimate

$$\hat{f}_{i,j} = \begin{cases} a_i b_j c'_{i,j} d_k & \text{for } i \neq j, \\ f_{i,i} & \text{for } i = j, \end{cases} \tag{5.13}$$

of the expected frequency under the model. (Compare (5.13) with (4.7).) Note that d_0 does not appear in (5.13), and the corresponding δ_0 did not appear in (4.7) or (2.7.1). For models that do not include diagonals parameters (or the triangles parameter), we set $d_0 = 1$; for models that do include these parameters, the value of d_0 will be estimated from the data after we introduce some additional assumptions (which we shall discuss in Section 7). Denoting $a_i b_i d_0$ by $f^*_{i,i}$, and letting $f^*_{i,j} = \hat{f}_{i,j}$ for $i \neq j$ (with $\hat{f}_{i,j}$ defined by (5.13)), we can estimate the index μ_i, which was defined by (5.7).

$$m_i = \frac{f_{i,i}}{f^*_{i,i}}. \tag{5.14}$$

Similarly, the index μ_i^*, which was defined by (5.8), can be estimated as

$$m_i^* = \frac{f_{i,i} - f^*_{i,i}}{f_{i\cdot}} \tag{5.15}$$

where $f_{i\cdot} = \Sigma_{j=1}^R f_{i,j}$. Formulae that are directly analogous to (5.11) and (5.12) can be obtained for m_i and m_i^*, by replacing μ_i, μ_i^*, $\tilde{\pi}_{i,i}$, and $\tilde{\pi}^*_{i,i}$ in (5.11) and (5.12) by m_i, m_i^*, $f_{i,i}/f_{i\cdot}$, and $f^*_{i,i}/f^*_{i\cdot}$, respectively, where $f^*_{i\cdot} = \Sigma_{j=1}^R f^*_{i,j}$. The estimates m_i and m_i^* can be calculated for the QO model and for each of the models in Sections 2.4 to 2.7.

6. The comparison of observed frequencies and expected frequencies under the various models

The usual chi square goodness of fit statistic for comparing the observed frequency $f_{i,j}$ with the corresponding estimate $\hat{f}_{i,j}$ of the expected frequency under a given model, can be written as

$$\sum \frac{(f_{i,j} - \hat{f}_{i,j})^2}{\hat{f}_{i,j}}, \tag{6.1}$$

and the corresponding chi square statistic based upon the likelihood ratio criterion can be written as

$$2\sum f_{i,j} \log\left(\frac{f_{i,j}}{\hat{f}_{i,j}}\right), \tag{6.2}$$

where the summation in (6.1) and (6.2) is taken over the off diagonal cells for the QO model and for the models of Sections 2.4 to 2.7, over all the cells in the table for the model of "independence" and for the models of Section 2.8, and over the set (or sets) of cells that are to be analyzed for a given model of quasi-independence of the kind described in Section 2.2 and 2.3. Both the statistic (6.1) and the statistic (6.2) have an asymptotic chi square distribution under the given model, with the degrees of freedom equal to the number of parameters in the model (calculating this number after the parameters have been uniquely defined). In certain contexts, the statistic (6.2) has some advantages over (6.1) (see, for example, Bahadur [1], Good [7], Goodman [13], [15], [17], [19], and Hoeffding [21]). We shall give in Table V the numerical values of both (6.1) and (6.2), for various models applied to the British 5×5 table (Table I), the Danish 5×5 table (Table II), and the British 7×7 table (Table III); but, for the sake of simplicity, when we discuss Table V later in the present section, we shall confine our attention to the numerical values of the statistic (6.2).

Since the estimate $\hat{f}_{i,j}$ of the expected frequency under each of the models considered here can be expressed in the general form (3.4), we see that the statistic (6.2) can also be written as

$$2[\sum f_{i,j} \log f_{i,j} - \sum_{w=1}^{W} f_w^{\lambda} \log \ell_w], \tag{6.3}$$

where the first summation sign $\sum$ in (6.3) has the same meaning that it did in (6.1) and (6.2); namely, the summation over all cells (i, j) for which $\Lambda_{i,j}$ is not empty. Note that formula (6.3) provides a method for calculating the statistic (6.2) using the W terms ℓ_w for $w = 1, 2, \cdots, W$ without calculating the $\hat{f}_{i,j}$ terms.

Now let H denote a given model of the form (3.1), and let H^+ denote the modified model that is obtained by applying to H a given modification of the H' and/or H'' type (see Section 4). As we noted in Section 4, the model H^+ is also "intrinsically unrestricted." Let $[H^+ | H]$ denote the hypothesis that H^+

TABLE V

COMPARISON OF THE OBSERVED FREQUENCIES AND THE EXPECTED FREQUENCIES UNDER VARIOUS MODELS APPLIED TO THE BRITISH AND DANISH 5 × 5 TABLES, AND THE BRITISH 7 × 7 TABLE

British and Danish 5 × 5 Tables					
		British Sample		Danish Sample	
Model	Degrees of Freedom	Goodness of Fit Chi Square	Likelihood Ratio Chi Square	Goodness of Fit Chi Square	Likelihood Ratio Chi Square
Ind	16	1199.4	811.0	754.1	654.2
QO	11	328.7	249.4	270.3	248.7
QP	3	8.5	12.6	6.9	7.4
QN	3	1.3	1.4	2.4	2.5
QPN	6	9.9	14.0	9.4	9.9
T	10	313.1	242.3	269.3	248.5
C	9	11.9	15.4	12.2	12.8
DA	8	15.9	19.1	6.7	6.9
CT	8	10.2	14.1	12.1	12.7
DAT	7	14.4	17.8	6.6	6.8
DP	7	10.3	10.6	6.8	7.0
DN	7	18.6	23.8	10.3	10.9
DAC	7	8.6	11.1	6.4	6.6
DACT	6	7.2	10.0	6.4	6.5
DPC	6	2.2	2.2	6.6	6.9
DNC	6	9.5	13.4	10.0	10.5
D	5	9.0	9.5	4.7	4.8
DC	4	1.5	1.6	4.4	4.5
DCF	6	6.7	6.9	6.3	6.3
DPCF	8	7.6	7.7	8.2	8.3
DACF	9	13.8	16.7	8.1	8.4
DF	9	52.1	50.4	10.1	10.2
DAF	12	59.5	60.6	12.2	12.4

is true assuming that H is true. Since H^+ is a modification of H in which some given conditions of the type (4.3) and/or (4.4) are imposed, the hypothesis $[H^+|H]$ states that the given conditions are true, assuming the H is true. Let $\chi^2(H)$ and $\chi^2(H^+)$ denote the statistic (6.2) with $\hat{f}_{i,j}$ calculated under H and H^+, respectively. The statistics $\chi^2(H)$ and $\chi^2(H^+)$ can be used to test the models H and H^+, respectively, and the following statistic can be used to test the hypothesis $[H^+|H]$:

$$(6.4)\qquad \chi^2(H^+|H) = \chi^2(H^+) - X^2(H) = 2\sum f \log\left[\frac{\hat{f}}{\hat{f}^+}\right] = 2\sum \hat{f} \log\left[\frac{\hat{f}}{\hat{f}^+}\right],$$

where $\hat{f}$ and $\hat{f}^+$ are the estimated expected frequencies under H and H^+, respectively. The final equality in (6.4) holds because the $\hat{f}$ satisfy (3.5) and

TABLE V (Continued)

Model	Degrees of Freedom	Goodness of Fit Chi Square	Likelihood Ratio Chi Square
	British 7 × 7 Table		
Ind	36	1361.7	897.5
QO	29	523.0	408.4
QP	10	9.4	13.4
QN	10	7.4	7.5
QPN	20	16.7	20.9
T	28	517.8	404.1
C	25	20.6	24.6
DA	24	20.1	22.1
CT	24	19.9	24.1
DAT	23	19.5	21.6
DP	23	21.3	22.3
DN	23	19.3	23.6
DAC	21	15.1	17.1
DACT	20	14.6	16.6
DPC	20	14.6	15.8
DNC	20	13.8	18.0
D	19	13.6	14.6
DC	16	8.4	9.4
DCF	20	25.0	26.3
DPCF	24	38.8	39.7
DACF	25	31.7	33.8
DF	25	54.2	54.8
DAF	30	59.8	61.9

the $\hat{f}^+$ satisfy conditions of the form (4.5) and/or (4.6). The statistic (6.4) is the chi square statistic based upon the likelihood ratio criterion for testing the hypothesis $[H^+|H]$. This statistic has an asymptotic chi square distribution under the hypothesis $[H^+|H]$, with the degrees of freedom equal to the difference between the corresponding number of degrees of freedom for testing H^+ and H, respectively.

Let us now examine the numerical values of $\chi^2(H)$, which are given in Table V for each of the models in Sections 2.3 to 2.7, and for some of the models in Section 2.8, applied to Table I. For the DC model, we see that $\chi^2(\mathrm{DC}) = 1.6$ with 4 degrees of freedom, which indicates that this model fits the data very well. Comparing the D model with the DC model using the statistic (6.4), we obtain $\chi^2(\mathrm{D}|\mathrm{DC}) = 7.9$ with one degree of freedom, which indicates that the crossings parameter makes a statistically significant contribution. In other words, assuming that the DC model is true, a test of the null hypothesis (2.7.3) would lead to rejection of the hypothesis. Comparing the C model with the DC model, we obtain $\chi^2(\mathrm{C}|\mathrm{DC}) = 13.8$ with 5 degrees of freedom, which indicates that the diagonals parameters make a statistically significant contribution. In other words, assuming that the DC model is true, a test of the null hypothesis (2.7.2) would lead to the rejection of the hypothesis.

Comparing the CT model, the DAC model, and the DNC model with the DC model, we obtain $\chi^2(\mathrm{CT}|\mathrm{DC}) = 12.5$, $\chi^2(\mathrm{DAC}|\mathrm{DC}) = 9.5$, and $\chi^2(\mathrm{DNC}|\mathrm{DC}) = 11.8$, with 4, 3, and 2 degrees of freedom, respectively. Thus, assuming that the DC model is true, a test of each of the null hypotheses (2.5.2), (2.5.8), and (2.5.13) would lead to their rejection. A similar result is obtained comparing the DACT model with the DC model. Comparing the DCF model and the DPC model with the DC model, we obtain $\chi^2(\mathrm{DCF}|\mathrm{DC}) = 5.3$ and $\chi^2(\mathrm{DPC}|\mathrm{DC}) = 0.6$ each with 2 degrees of freedom. Thus, assuming that the DC model is true, a test of the null hypothesis (5.6) would lead to rejection of the hypothesis at the 10 per cent level of significance, and a test of the null hypothesis (2.5.9) would lead to acceptance of that hypothesis. Indeed, the DPC model fits the data very well.

Having noted that the DPC model fits the data well, we now compare various models with the DPC model. Comparing the DP and the C models with the DPC model, we obtain $\chi^2(\mathrm{DP}|\mathrm{DPC}) = 8.4$ and $\chi^2(\mathrm{C}|\mathrm{DPC}) = 13.2$, with 1 and 3 degrees of freedom, respectively, which indicates that both the crossings parameters and the parameters pertaining to the "positive" diagonals in the DPC model make a statistically significant contribution. Comparing the DPCF model with the DPC model, we obtain $\chi^2(\mathrm{DPCF}|\mathrm{DPC}) = 5.5$ with 2 degrees of freedom. Thus, assuming that the DPC model is true, a test of the null hypothesis (5.6) would lead to rejection of the hypothesis at the 10 per cent level of significance.

The preceding comments pertained to the results given in Table V for the analysis of Table I. The corresponding results, which are given in Table V for the analysis of Tables II and III, do not lead to conclusions that are as clear cut as those obtained for Table I. We shall now comment briefly on the results for Tables II and III without presenting a full analysis of them.

As was the case for Table I, the DC model fits the data very well for Table III; and it fits the data rather well for Table II, but not as well as for Tables I and III. Among the models that fit the data well (or rather well), we find the DA model for Table II and the DAC model for Table III. Comparing the models for the analysis of the full tables with the corresponding models for the analysis of the off diagonal cells, we find that a test of the null hypothesis (5.6) would lead to rejection of the hypothesis for Table III and acceptance of the hypothesis for Table II. Comparison of the QP and QN models indicates that the QN model fits the data better than the QP model for Tables II and III (and also for Table I).

For each chi square statistic, the corresponding number of degrees of freedom can be obtained from Tables IV and V. These tables give the degrees of freedom of the corresponding asymptotic distribution under the null hypothesis. This is the appropriate number of degrees of freedom to use in testing the hypothesis *if* the hypothesis were decided upon before the data were studied. On the other hand, if a set of hypotheses were tested simultaneously (or if the particular hypothesis that was tested was contained within a larger set of hypotheses that were studied), the degrees of freedom could be adjusted in a

similar way to the adjustment made in calculating simultaneous confidence intervals and simultaneous tests in the present context (see Goodman [11], [14], [19]). This adjustment will limit the risks of rejecting hypotheses that are true, even when the hypotheses are suggested by the data. Of course, the risks of accepting false hypotheses are also affected if the hypotheses are suggested by the data.

The various hypotheses and models that we have tested and compared here could also have been assessed by adapting to the present context some of the concepts that arise in stepwise regression (for example, some of the concepts of backward regression and/or forward regression). For a discussion of this kind of adaptation, and for an example of its application, the reader is referred to Goodman [18].

7. The estimated parameters and indices

In this section we shall comment briefly upon the estimates of some of the parameters, and of the ratio and relative difference indices, which are obtained when some of the models of Section 2 are applied to Tables I, II, and III. These estimates will be presented in Tables VI to X later in this section.

For the models that included the triangles parameters τ_k for $k = 1$ and 2 (see, for example, (2.4.1)), we noted earlier that these parameters were not uniquely defined until one restriction was imposed upon them. This restriction could be expressed in several different ways; for example, as (a) the condition that $\tau_1 = 1$ (as we did in Section 2.4), or as (b) the condition that $\tau_2 = 1$, or as (c) the condition

$$\tau_1\tau_2 = 1. \tag{7.1}$$

For our present purposes, it is convenient to impose condition (7.1), and to take τ_1 as the uniquely defined triangles parameter τ. From (7.1), we see that τ can also be expressed as

$$\tau = \left(\frac{\tau_1}{\tau_2}\right)^{1/2} \tag{7.2}$$

(In contexts where condition (7.1) is replaced by one of the other conditions given above, the quantity defined by (7.2) would also be replaced as the triangles parameter.)

Table VI gives the maximum likelihood estimate t of τ obtained under the four models that include the triangles parameter; namely, the T model, the CT model, the DAT model, and the DACT model. Note that the estimate t is less than one for each of the cases considered in Table VI, which indicates that, aside from the effects of the other parameters in the model, the estimated triangles parameter will diminish the estimated expected frequencies for the cells in set S_1 (where $i - j > 0$) relative to the estimated expected frequencies for the cells in set S_2 (where $i - j < 0$). For each of the three mobility tables

TABLE VI

THE ESTIMATE OF THE TRIANGLES PARAMETER UNDER VARIOUS MODELS APPLIED TO THE BRITISH AND DANISH SAMPLES

	Models			
	T	CT	DAT	DACT
British 5 × 5 Table	0.855	0.935	0.935	0.937
Danish 5 × 5 Table	0.970	0.981	0.980	0.983
British 7 × 7 Table	0.904	0.966	0.966	0.968

(Tables I, II, and III), this would indicate that the estimated τ parameter has the direct effect of introducing "downward mobility" (as expressed in the fact that $(\tau_2/\tau_1)^{1/2}$ is estimated as being larger than one), over and above the indirect effects (with respect to upward or downward mobility) that are due to the other parameters. The effect of t appears to be more pronounced for the T model (particularly for Table I), and it becomes less pronounced as more parameters (for example, the crossings parameters and/or the parameters pertaining to the paired minor diagonals) are included in the model. To test whether $\tau = 1$ in, say, the DACT model applied to Tables I, II, and III, we compare the DAC model with the DACT model using the results given in Table V, and we find that the τ parameter does not have a statistically significant effect when the other parameters are included in the model.

With respect to the diagonals parameters (Table VII), we note that d_k decreases as $|k|$ increases in all cases, except for d_4 for Table II under the DPC, D, and DC models. This would indicate that aside from the effects of the other parameters in the model, generally speaking the estimated diagonals parameters diminish the estimated expected frequencies in a progressively more pronounced way for cells that are on minor diagonals that are further away from the main diagonal. In other words, the estimates d_k have the direct effect of introducing "status inertia" in the mobility table. With respect to Table II, it is worth noting (see Table V) that the modifications of the DA model and the DAC model that are obtained by distinguishing diagonals parameters on the "positive" diagonals from those on the "negative" diagonals (for example, the D model, the DC model, and the DPC model) did not improve the fit markedly. Note also that the difference in Table VII between d_k and the corresponding d_{-k} is, generally speaking, smaller for Table II than for the other tables. When comparing the DA model with the DAC model (or the D model with the DC model), we find that the effect of d_k is somewhat more pronounced in the former model than in the latter one (which included the crossings parameters as well). To facilitate the comparison of models in Table VII, condition (2.5.12) was used, rather than the equivalent (2.5.9), in the DPC model; and condition (7.3) below was used in the DAC, DPC, and DC models.

With respect to the crossings parameters (Table VIII), for models that include both the γ_u and δ_k parameters, we noted earlier that the γ_u, for $u =$

TABLE VII

THE ESTIMATE OF THE DIAGONALS PARAMETERS UNDER VARIOUS MODELS APPLIED TO THE BRITISH AND DANISH SAMPLES

	Models							
	DA	DAC	DPC		D		DC	
British 5 × 5 Table	$d_k = d_{-k}$	$d_k = d_{-k}$	d_k	d_{-k}	d_k	d_{-k}	d_k	d_{-k}
$k = 1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$k = 2$	0.59	0.64	0.74	0.54	0.68	0.52	0.73	0.56
$k = 3$	0.26	0.32	0.39	0.29	0.32	0.22	0.39	0.28
$k = 4$	0.08	0.13	0.00	0.16	0.00	0.11	0.00	0.17
Danish 5 × 5 Table								
$k = 1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$k = 2$	0.48	0.51	0.52	0.45	0.49	0.47	0.52	0.50
$k = 3$	0.15	0.16	0.14	0.20	0.13	0.16	0.14	0.17
$k = 4$	0.11	0.12	0.17	0.09	0.15	0.06	0.17	0.06
British 7 × 7 Table								
$k = 1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$k = 2$	0.65	0.70	0.69	0.61	0.64	0.65	0.69	0.70
$k = 3$	0.34	0.40	0.51	0.38	0.41	0.28	0.49	0.34
$k = 4$	0.20	0.26	0.30	0.23	0.21	0.18	0.28	0.25
$k = 5$	0.08	0.11	0.08	0.14	0.05	0.09	0.08	0.13
$k = 6$	0.03	0.04	0.00	0.09	0.00	0.04	0.00	0.06

TABLE VIII

THE ESTIMATE OF THE CROSSINGS PARAMETERS UNDER VARIOUS MODELS APPLIED TO THE BRITISH AND DANISH SAMPLES

	Models			
	C	DAC	DPC	DC
British 5 × 5 Table				
c_2	0.40	0.46	0.45	0.46
c_3	0.60	0.64	0.63	0.64
Danish 5 × 5 Table				
c_2	0.46	0.51	0.48	0.51
c_3	0.43	0.47	0.46	0.48
British 7 × 7 Table				
c_2	0.54	0.60	0.55	0.60
c_3	0.52	0.58	0.53	0.58
c_4	0.60	0.67	0.61	0.66
c_5	0.64	0.70	0.65	0.70

$2, 3, \cdots, R - 2$, were not uniquely defined until one restriction was imposed upon them—a restriction that has the effect of fixing the scale of the γ_u. This restriction could be expressed in several different ways; for example, as (a) the condition that

$$\max_k \gamma_k = 1 \qquad \text{for} \quad 2 \leqq k \leqq R - 2, \tag{7.3}$$

which is equivalent to replacing the γ_u by $\bar{\gamma}_u = \gamma_u/\gamma^*$, where $\gamma^* = \max_k \gamma_k$ for $2 \leqq k \leqq R - 2$; or (b) the condition that

$$\delta_2 \delta_{-2} = 1, \tag{7.4}$$

which will also have the effect of fixing the scale of the γ. For models that included the δ_k parameters, we noted earlier that the restriction that $\delta_1 = \delta_{-1} = 1$ would uniquely define the δ_k: this restriction together with condition (7.4) would uniquely define the δ_k and γ_u for $u = 2, 3, \cdots, R - 2$, in models that included both sets of parameters. In order to facilitate comparison between the crossings parameters for the C model (for which the γ_u for $u = 2, 3, \cdots, R - 2$, are uniquely defined without imposing any restrictions upon them) and the corresponding parameters in models that also include the δ_k parameters, the restriction (7.4) was used in calculating the results presented in Table VIII. To make the c_u in Table VIII consistent with the corresponding d_k in Table VII for a model that includes both sets of parameters (for example, the DAC, the DPC, and the DC models), it is only necessary to divide each of the c_u in Table VIII by $\max_k c_k$ for $2 \leqq k \leqq R - 2$, for the given model (see condition (7.3)). The estimates c_u in Table VIII have the direct effect of introducing "status barriers" in the mobility tables. Comparing the results in Table VIII for the 5×5 tables (that is, Tables I and II), we note that the effect of c_2 is more pronounced than that of c_3 for Table I, and the reverse is true for Table II. For Table III, the effect of c_u becomes more pronounced as u decreases, except for $u = 2$. From Table VIII we see that the relative difference between the c is less for Table II than for Table I. From Table V we also take note of the fact that the modification of the various models (for example, the DA, the DP, the D models) that is obtained by including the crossing parameters (thus obtaining the DAC, the DPC, and the DC models) did not improve the fit markedly for Table II, but it did for Table I. Finally, we note that the corresponding c_u in Table VIII are very similar for the DAC model and the DC model.

With respect to the ratio index and the relative difference index (Tables IX and X), we first note that the numerical values obtained for the QO model (which did not fit the mobility tables well) are grossly misleading. For models that fit the data well, the numerical values obtained for these indices can differ greatly from the values obtained with models that do not fit the data. Note also that the corresponding values for the DA model and the D model are very similar, and so are the corresponding values of the DAC and the DC models. With the introduction of the additional parameters (for example, the δ_k and/or the γ_u) into the QO model, the effect of the m_i diminishes (see Table IX).

TABLE IX

THE ESTIMATE OF THE RATIO INDEX FOR THE CELLS ON THE MAIN DIAGONAL OF THE $R \times R$ CROSS CLASSIFICATION TABLE, UNDER VARIOUS MODELS APPLIED TO THE BRITISH AND DANISH SAMPLES

	Models					
	QO	DA	DAC	DPC	D	DC
British 5 × 5 Table						
m_1	34.5	3.8	4.6	4.5	3.8	4.5
m_2	4.0	0.9	0.6	0.6	0.9	0.6
m_3	1.7	1.0	1.0	1.0	1.0	1.0
m_4	1.0	0.7	0.8	0.7	0.8	0.8
m_5	2.9	0.9	1.1	1.1	0.9	1.1
Danish 5 × 5 Table						
m_1	13.8	1.1	1.2	1.2	1.1	1.2
m_2	4.8	0.9	1.0	1.0	0.9	1.0
m_3	1.8	0.8	0.8	0.8	0.8	0.8
m_4	1.2	0.7	0.7	0.7	0.7	0.7
m_5	3.4	0.7	0.8	0.7	0.7	0.8
British 7 × 7 Table						
m_1	35.0	2.4	2.6	2.4	2.4	2.6
m_2	8.6	1.5	1.3	1.2	1.5	1.3
m_3	2.2	0.7	0.6	0.5	0.7	0.6
m_4	1.7	1.0	1.0	1.0	1.0	1.0
m_5	1.2	0.8	0.8	0.7	0.8	0.8
m_6	2.3	1.3	1.4	1.3	1.3	1.4
m_7	2.9	0.7	0.8	0.7	0.7	0.8

TABLE X

THE ESTIMATE OF THE RELATIVE DIFFERENCE INDEX FOR THE CELLS ON THE MAIN DIAGONAL OF THE $R \times R$ CROSS CLASSIFICATION TABLE UNDER VARIOUS MODELS APPLIED TO THE BRITISH AND DANISH SAMPLES

	Models					
	QO	DA	DAC	DPC	D	DC
British 5 × 5 Table						
m_1^*	0.38	0.29	0.30	0.30	0.29	0.30
m_2^*	0.26	−0.05	−0.23	−0.24	−0.04	−0.22
m_3^*	0.09	−0.01	0.00	0.00	−0.01	0.00
m_4^*	−0.01	−0.17	−0.16	−0.16	−0.16	−0.14
m_5^*	0.32	−0.04	0.05	0.04	−0.04	0.05
Danish 5 × 5 Table						
m_1^*	0.29	0.02	0.06	0.04	0.02	0.06
m_2^*	0.26	−0.02	0.01	−0.01	−0.02	0.01
m_3^*	0.18	−0.09	−0.09	−0.13	−0.09	−0.09
m_4^*	0.08	−0.16	−0.16	−0.19	−0.16	−0.16
m_5^*	0.33	−0.22	−0.15	−0.20	−0.21	−0.14
British 7 × 7 Table						
m_1^*	0.38	0.23	0.24	0.23	0.23	0.24
m_2^*	0.24	0.09	0.06	0.05	0.09	0.06
m_3^*	0.10	−0.07	−0.11	−0.17	−0.07	−0.12
m_4^*	0.09	0.01	0.01	−0.01	0.00	0.01
m_5^*	0.08	−0.14	−0.12	−0.20	−0.14	−0.12
m_6^*	0.17	0.06	0.08	0.08	0.06	0.08
m_7^*	0.18	−0.14	−0.06	−0.09	−0.14	−0.06

To test hypotheses about the μ_i, we return to our earlier discussion of (5.6). As we noted there, for a given model that does not include crossings parameters (but does include diagonals parameters or the triangles parameter), the hypothesis that the relative differences among the μ_i are nil for $i = 1, 2, \cdots, R$, can be tested by comparing the corresponding model for the analysis of the full table with the model for the analysis of the off diagonal cells (for example, comparing the DF model with the D model, or the DAF model with the DA model). Also, for a given model that does include crossings parameters (and also diagonals parameters and/or the triangles parameter), the hypothesis that the relative differences among the μ_i are nil for $i = 2, 3, \cdots, R - 2$, excluding $i = 1$ and $i = R$, can be tested by a similar kind of comparison (for example, the comparison of the DCF model with the DC model, or the DPCF model with the DPC model, or the DACF model with the DAC model). Examination of the corresponding relative magnitudes in Table IX shed further light on the results obtained by comparing the corresponding chi squares given in Table V. Recall, for example, that the comparison of the DCF model with the DC model indicated that the relative differences among the μ_i were statistically significant for Table III; they were not as statistically significant for Table I (but they were significant at the 10 percent level); and they were not statistically significant for Table II.

We remarked in Section 5 that, in calculating the m_i and m_i^* for models that include diagonals parameters and/or the triangles parameter, we must first make some assumptions about δ_0. For the calculations in Tables IX and X, we have assumed that

$$\frac{\delta_0}{\delta_1''} = \frac{\delta_1''}{\delta_2''}, \tag{7.5}$$

where δ_k'' is defined as

$$\delta_k'' = (\delta_k \delta_{-k})^{1/2}. \tag{7.6}$$

Denoting δ_0 by δ_0'', condition (7.5) states that $\delta_k''/\delta_{k+1}''$ is constant, for $k = 0$ and 1. Although there is an element of arbitrariness in this way of defining δ_0, there are good reasons for using (7.5) here. (An alternative procedure would be to set $\delta_0 = 1$, which would yield the same result as obtained with (7.5) in cases where the model does not include diagonals parameters (and/or the triangles parameter), and also in cases where the model includes both the crossings parameters and diagonals parameters (and/or the triangles parameter) when the parameters are uniquely defined by condition (7.4) together with the condition that $\delta_1 = \delta_{-1} = 1$. In addition, for models that include both γ_u and δ_k (and/or τ_k), the m_i for $i = 2, 3, \cdots, R - 1$, would be the same when $\delta_0 = 1$ as the corresponding quantities obtained when the parameters are uniquely defined by condition (7.3) together with the condition that $\delta_1 = \delta_{-1} = 1$.) Since we set $\delta_1 = \delta_{-1} = 1$ in all models that include diagonals parameters, condition (7.5) can be simplified to

$$\delta_0 = \frac{1}{\delta_2''}. \tag{7.7}$$

In order to calculate m_1 and m_R (and also m_1^* and m_R^*) in Tables IX and X, we have set $\gamma_1 = \gamma_{R-1} = \gamma^*$, where $\gamma^* = \max_k \gamma_k$ for $2 \leqq k \leqq R - 2$. This procedure provides a way of calculating m_1 and m_R (and also m_1^* and m_R^*) that is conservative in the sense that, while it acknowledges the possible existence of the crossings parameters γ_1 and γ_{R-1}, it estimates their effects as being equal to the least pronounced estimated effects among the γ_u for $u = 2, 3, \cdots, R - 2$. (In his analysis of the "variable distance" model, Haberman [20] made the assumption that $\mu_1 = \mu_R = 1$ (that is, that $\log \mu_1 = \log \mu_R = 0$), and under this assumption he then estimated parameters corresponding to γ_1 and γ_{R-1} (namely, $\log \gamma_1$ and $\log \gamma_{R-1}$). In contrast to this, in the present article the μ_1 and μ_R (and also the μ_1^* and μ_R^*) are estimated from the data as indicated in the first sentence of this paragraph. Also, in contrast to the earlier analysis of the "variable distance" model, the parameters μ_i^* for $i = 1, 2, \cdots, R$, introduced here provide an additional index of interest (see Section 5).)

8. Extensions and applications to the analysis of two (or more) cross classification tables, and to the analysis of multidimensional tables

The results presented in Sections 3 and 4 for the general case of the $R \times C$ table can be directly extended to the analysis of two $R \times C$ tables (that is, to the analysis of the $R \times C \times 2$ table), to the analysis of G such $R \times C$ tables (that is, to the analysis of the $R \times C \times G$ table), and more generally to the analysis of multidimensional contingency tables. Indeed, the results presented in those sections can be applied directly to the multidimensional table simply by replacing each reference to the cells (i, j) of the $R \times C$ table throughout those sections by a corresponding reference to the cells of the multidimensional table.

I shall give here only a few examples (although there are many) of the possible extensions of the particular models introduced earlier to the analysis of three way tables (or to the analysis of contingency tables of higher dimensions). Suppose we applied, say, the DC model to G different $R \times R$ tables and found that the model fit the data for each of these tables. For the gth table, $g = 1, 2, \cdots$, the parameters $\alpha_i^{(g)}$, $\beta_j^{(g)}$, $\delta_k^{(g)}$, $\gamma_u^{(g)}$ could be estimated by the methods described earlier, and we could consider various hypotheses of the following kind:

$$\delta_k^{(g)} = \delta_k \quad \text{for} \quad g = 1, 2, \cdots, G, \tag{8.1}$$

$$\gamma_u^{(g)} = \gamma_u \quad \text{for} \quad g = 1, 2, \cdots, G, \tag{8.2}$$

$$\delta_k^{(g)} = \delta_k \quad \text{and} \quad \gamma_u^{(g)} = \gamma_u \quad \text{for} \quad g = 1, 2, \cdots, G. \tag{8.3}$$

Each of these hypotheses can be expressed as an "intrinsically unrestricted" model for the $R \times C \times G$ table, since the restrictions (8.1), (8.2), and (8.3)

are of the H'' type (as defined in Section 4). Thus, the method presented earlier could also be used to analyze the model H'' obtained by modifying the DC model for each of the G tables by imposing conditions of the kind described above (that is, (8.1) to (8.3)).

Let H_g denote a given model (for example, the DC model) that will be applied to the gth $R \times R$ table for $g = 1, 2, \cdots, G$, and let H denote the hypothesis (model) that states that the H_g are true for $g = 1, 2, \cdots, G$. Let H'' denote the model obtained by modifying H by imposing a given condition (for example, a condition of the kind described under (8.1) to (8.3)) that still permits H'' to be expressed as an "intrinsically unrestricted" model for the $R \times C \times G$ table. Let $\chi^2(H_g)$, $\chi^2(H)$, and $\chi^2(H'')$ denote the chi square statistic based upon the likelihood ratio criterion for testing H_g, H, and H'', respectively. Each of these statistics can be calculated by the methods presented earlier applied to the G different $R \times R$ tables (for calculating $\chi^2(H_g)$) and to the $R \times R \times G$ table (for calculating $\chi^2(H)$ and $\chi^2(H'')$). Note that

$$\chi^2(H) = \sum_{g=1}^{G} \chi^2(H_g). \tag{8.4}$$

Letting $[H''|H]$ denote the hypothesis that H'' is true assuming that H is true, we see that $[H''|H]$ states that the given condition, which was used to modify H to form H'', is true assuming that H is true. The following statistic is the chi square statistic based upon the likelihood ratio criterion for testing $[H''|H]$:

$$\begin{aligned} \chi^2(H''|H) &= \chi^2(H'') - \chi^2(H) \\ &= 2\sum f \log\left[\frac{\hat{f}}{\hat{f}''}\right] = 2\sum \hat{f} \log\left[\frac{\hat{f}}{\hat{f}''}\right], \end{aligned} \tag{8.5}$$

where $\hat{f}$ and $\hat{f}''$ denote the estimated expected frequencies in the $R \times C \times G$ table under H and H'', respectively. Note that the $\hat{f}$ pertaining to the gth $R \times R$ table can be calculated separately for each of the G tables, but the $\hat{f}''$ need to be calculated from the $R \times R \times G$ table (if H'' is a hypothesis of the kind described by (8.1) to (8.3)).

The statistic (8.5) has an asymptotic chi square distribution under the hypothesis $[H''|H]$, with the degrees of freedom equal to the difference between the corresponding number of degrees of freedom for testing H'' and H, respectively. The degrees of freedom of $\chi^2(H''|H)$ can also be calculated from the number of restrictions on the parameters (calculating this number after the parameters have been uniquely defined). Thus, for example, for testing hypotheses (8.1), 8.2), and (8.3), the corresponding number of degrees of freedom will be $G - 1$ multiplied by $2(R - 2) - 1 = 2R - 5$, $(R - 3) - 1 = R - 4$, and $2(R - 2) + (R - 4) = 3R - 8$, for hypotheses (8.1), (8.2), and (8.3), respectively.

We shall now comment briefly on the relationship between hypotheses of the kind considered above for the three way table (or for tables of higher dimension) and the usual hypothesis H_0 of zero three factor interaction in the three way

table (or the corresponding kind of hypotheses for the table of higher dimension) (see, for example, Goodman [15], [17]). The usual hypothesis H_0 in the three way $R \times C \times G$ table states that the probability $\pi_{i,j,g}$ can be written as

$$\pi_{i,j,g} = \alpha_i^{(g)}\beta_j^{(g)}\theta_{i,j} \quad \text{for } i = 1, 2, \cdots, R, \quad j = 1, 2, \cdots, C, \quad g = 1, 2, \cdots, G. \tag{8.6}$$

For the $R \times R \times G$ table, consider now the hypothesis H that states that the DCF model holds true for the gth $R \times R$ table, $g = 1, 2, \cdots, G$, and let H'' denote the modification of H obtained by imposing condition (8.3) on model H. The model H'' can be expressed as

$$\pi_{i,j,g} = \alpha_i^{(g)}\beta_j^{(g)}\gamma'_{i,j}\delta_k \quad \text{for cells } (i, j, g) \text{ for which } (i, j) \text{ is in } S'_k, \tag{8.7}$$

for $k = 0, \pm 1, \pm 2, \cdots, \pm(R-1)$; and with $\gamma'_{i,j}$ defined by (2.6.3), with $\gamma'_{i,j} = 1$ for $i = j$. To test the hypotheses H_0, H, and H'' in the $R \times R \times G$ table, the corresponding number of degrees of freedom will be $(R-1)^2(G-1)$, $G(R-2)(R-3)$, and $G(R-1)^2 - (3R-5)$, respectively. Since there are $(R-2)(R-3)$ degrees of freedom for testing the DCF model (see Table IV), the corresponding number of degrees of freedom for the H model will be $G(R-2)(R-3)$. Note also that the degrees of freedom for the H'' model can be calculated by subtracting from GR^2 the number of estimated parameters under the model (namely, $G(2R-1) + (3R-5)$) or by adding $(G-1)(3R-5)$ to $G(R-2)(R-3)$, since condition (8.3) actually imposes $(G-1)(3R-5)$ restrictions upon the parameters of H. The hypothesis $[H''|H]$ states that condition (8.3) is true, assuming that H is true; and the hypothesis $[H''|H_0]$ states that the following condition is true, assuming that H_0 is true:

$$\theta_{i,j} = \gamma'_{i,j}\delta_k \quad \text{for } (i, j) \text{ in } S'_k, \tag{8.8}$$

for $k = 0, \pm 1, \pm 2, \cdots, \pm(R-1)$, where the parameters in (8.8) are defined by the expressions (8.6) and (8.7). To test the hypotheses $[H''|H]$ and $[H''|H_0]$, the corresponding number of degrees of freedom will be $(G-1)(3R-5)$ and

$$(R-1)^2 - (3R-5) = R^2 - 5R + 6 = (R-2)(R-3), \tag{8.9}$$

respectively.

Note that the same number of degrees of freedom are obtained for the hypothesis $[H''|H_0]$ as for the DCF model in the $R \times R$ table. The hypothesis $[H''|H_0]$ states that the parameters $\theta_{i,j}$ in (8.6) will satisfy condition (8.8); and the DCF model states that the probabilities $\pi_{i,j}$ can be written as

$$\pi_{i,j} = \alpha_i\beta_j\theta'_{i,j}, \tag{8.10}$$

where the $\theta'_{i,j}$ are of the form

$$\theta'_{i,j} = \gamma'_{i,j}\delta_k \quad \text{for cells } (i, j) \text{ in } S'_k, \tag{8.11}$$

for $k = 0, \pm 1, \pm 2, \cdots, \pm(R-1)$. Similarly, each of the hypotheses considered in Sections 2.1 to 2.8 states that the probabilities $\pi_{i,j}$ can be written as

(8.10) where the $\theta'_{i,j}$ are subject to certain specified conditions. For example, the DC model states that the $\theta'_{i,j}$ are of the form

$$\theta'_{i,j} = \begin{cases} \gamma'_{i,j}\delta_k & \text{for cells } (i,j) \text{ in } S'_k \text{ for } k = \pm 1, \pm 2, \cdots, \pm(R-1), \\ \mu'_i & \text{for cells } (i,i) \text{ in } S'_0, \end{cases} \tag{8.12}$$

where μ'_i is an "intrinsically unrestricted" parameter.

For the usual hypothesis H_0 of zero three factor interaction in the $R \times C \times G$ table (see (8.6)), the probabilities $\pi_{i,j,g}$ can be rewritten as

$$\pi_{i,j,g} = \alpha_{i,g}\beta_{j,g}\theta_{i,j}. \tag{8.13}$$

In the present section, we have considered hypotheses about the form of the $\theta_{i,j}$ in the $R \times R \times G$ table. It should also be noted that the general methods presented in Sections 3 and 4 can be directly applied to any given hypothesis about the form of the $\alpha_{i,g}$, $\beta_{j,g}$, and/or $\theta_{i,j}$ in the $R \times C \times G$ table, as long as the corresponding model is "intrinsically unrestricted" in the three way table.

The methods presented here can be directly applied not only to models for the three way table that are formed by imposing conditions on the parameters in (8.13), but also more generally to any hypothesis about the probabilities $\pi_{i,j,g}$ in the three way table (or in a table of higher dimension), as long as the corresponding model is intrinsically unrestricted. As an example of such models (that is, intrinsically unrestricted models) that are not formed by imposing conditions on the parameters in (8.13), see Goodman [13], [14] where the hypothesis of zero three factor interaction is extended to the case where a given subset of the cells in the three way table are deleted. These models for the three way table can be further extended in the same ways as we have here extended the model of quasi-independence in the two way table. For example, in the same way that the T model was formed by introducing the triangles parameters into the QO model, we could also introduce the three dimensional analogues of the triangles parameters (considering now parameters pertaining to certain triangles and tetrahedra in the three way table) into the models for the three way table considered in Goodman [13] (that is, into models of zero three factor interaction applied to a given subset of the cells of the table). In the same way that the triangles parameters are introduced in order to describe a particular kind of two factor interaction (between the row and column classifications) in the two way table, the three dimensional analogues of the triangles parameters can be introduced in order to describe a particular kind of three factor interaction in the three way table. For three way tables that do not conform to the usual hypothesis H_0 of zero three factor interaction, we can now provide a wide variety of multiplicative models (that is, the three dimensional analogues of the models in Sections 2.2 to 2.8) that can be used to analyze the data. In addition, for three way tables that do conform to the usual hypothesis H_0, we noted earlier in the present section that the general methods presented here can be used to test given hypotheses about the form of the parameters in the H_0 model (that is, hypotheses about the form of the two factor interactions in the three way table).

The preceding remarks can be directly extended to the multidimensional contingency table. We noted earlier that the usual hypothesis H_0 of zero three factor interaction in the three way table could be expressed as an intrinsically unrestricted multiplicative model, and similarly each of the hierarchical hypotheses described in Goodman [17] for the multidimensional table can also be expressed as an intrinsically unrestricted multiplicative model. The various extensions and modifications of H_0, which we described earlier in the present section for the analysis of the three way table, can also be directly extended in order to provide further extensions and modifications of the hierarchical hypotheses that were considered in the earlier literature on multidimensional contingency tables. Since these extensions and modifications are directly analogous to those already presented in the present article, we need not discuss this further here.

Before closing this section, we return for a moment to the models of Sections 2.3 to 2.8. Note that (2.3.4) can be viewed as a model for an $R \times C \times K$ table (R rows, C columns, K layers) that describes conditional quasi-independence between the row and column classifications, given the kth layer classification, $k = 1, 2, \cdots, K$, when certain cells (i, j, k) have been deleted, namely, the cells (i, j, k) for which the (i, j) is not in S_k (see, for example, Goodman [13], [17]). In particular, (2.3.1) is a model that describes conditional quasi-independence in the $R \times R \times 2$ table, where one layer pertains to positive $(i - j)$ and the other layer pertains to negative $(i - j)$. Similarly, (2.4.1) is a model for quasi-mutual independence in the $R \times R \times 2$ table (that is, a model of "complete independence" among the three variables in the three way table when certain cells have been deleted); and (2.5.1) is a model of quasi-mutual independence in an $R \times R \times [2(R - 1)]$ table. The other models in Sections 2.3 to 2.8 can also be viewed as models in three way or multi-way tables.

APPENDIX

Explicit Formulae for the Estimates of the Parameters in the Models

For any given model of the kind described in Sections 2.2 to 2.8, we shall now show how the maximum likelihood estimates of the parameters in the model can be expressed explicitly as functions of the estimates $\hat{f}_{i,j}$ of the expected frequencies under the model. Other ways to calculate the maximum likelihood estimates of the parameters were described in Section 3. The results, which we shall now present, will provide (a) further insight into the meaning of the parameters (expressed explicitly as functions of the probabilities $\pi_{i,j}$), and (b) a method for calculating the maximum likelihood estimates of the parameters (after the estimates $\hat{f}_{i,j}$ have been calculated as described in Section 3) which some readers may find easier to apply than the other methods that were described in Section 3 for estimating the parameters.

For simplicity, let us first consider the QO model (see Section 2.3). For this model, the maximum likelihood estimate $\hat{f}_{i,j}$ can be written as follows (see (2.2.4)):

$$\hat{f}_{i,j} = a_i b_j \quad \text{for} \quad i \neq j. \tag{A.1}$$

As we noted earlier (see Sections 2.1 to 2.3), the $\hat{f}_{i,j}$ in (A.1) will be unchanged if a_1 is set equal to one (and the other a_i and b_j in (A.1) are changed accordingly). With $a_1 = 1$, we see from (A.1) that the b_j can be written as

$$b_j = \begin{cases} \hat{f}_{1,j} & \text{for} \quad j = 2, 3, \cdots, R, \\ \hat{f}_{3,1}\hat{f}_{1,2}/\hat{f}_{3,2} & \text{for} \quad j = 1. \end{cases} \tag{A.2}$$

Similarly, having calculated b_1 from (A.2), we obtain the following formula for the a_i

$$a_i = \begin{cases} 1 & \text{for} \quad i = 1, \\ \hat{f}_{i,1}/b_1 & \text{for} \quad i = 2, 3, \cdots, R. \end{cases} \tag{A.3}$$

If we are interested in the a_i after they have been scaled so that $\Sigma_i a_i = 1$, then a_i of (A.3) can be replaced by $\tilde{a}_i = a_i/\Sigma_{h=1}^{R} a_h$. The $\tilde{a}_i$ are the maximum likelihood estimates of the scaled parameters $\tilde{\alpha}_i = \alpha_i/\Sigma_{h=1}^{R} \alpha_h$, which can be interpreted (for the QO model) as the hypothetical proportion of individuals in the ith row class in the hypothetical population in which none of cells on the main diagonal needs to be deleted, and there is independence between the row and column classifications in the table (see Goodman [13], [14]). A similar comment applies to b_j and to the corresponding $\tilde{b}_j$ and $\tilde{\beta}_j$. For the QO model, we find that

$$\tilde{\beta}_j = \tilde{\pi}^*_{i,j} \quad \text{for} \quad i = 1, 2, \cdots, R, \quad j = 1, 2, \cdots, R, \tag{A.4}$$

where $\tilde{\pi}^*_{i,j} = \pi^*_{i,j}/\pi_{i.}$ is the hypothetical conditional probability defined in Section 5. Thus, by applying the mover-stayer interpretation described by (5.9) to the QO model, we see from (A.4) that $\tilde{\beta}_j$ can be interpreted as the hypothetical probability that a "mover" will be in the jth column class (see, for example, Goodman [16]). By interchanging the row and column classifications, and then applying the mover-stayer interpretation described by (5.9), we see that the $\tilde{\alpha}_i$ can be interpreted in a similar way to the above interpretation of the $\tilde{\beta}_j$. (Analogous kinds of interpretations of the $\tilde{\beta}_j$ and $\tilde{\alpha}_i$ can be obtained when (5.10) is applicable, rather than (5.9).)

Consider now the QP and QN models (see Section 2.3). For each of these models, the parameters can be estimated by formulae similar to (A.2) and (A.3), based upon the fact that the b_j are proportional to the $\hat{f}_{R,j}$ for $j = 1, 2, \cdots, R - 1$, and the a_i are proportional to the $\hat{f}_{i,1}$ for $i = 2, 3, \cdots, R$, for the QP model, and upon the fact that the b_j are proportional to the $\hat{f}_{1,j}$ for $j = 2, 3, \cdots, R$ and the a_i are proportional to the $\hat{f}_{i,R}$ for $i = 1, 2, \cdots, R - 1$, for the QN model. More generally, for any model of quasi-independence for an inseparable set of cells (or for the inseparable subsets of a separable set of cells), the parameters can also be estimated by formulae similar to (A.2) and (A.3).

Now consider the DC model (see Section 2.7). For this model, the maximum likelihood estimate $\hat{f}_{i,j}$ can be written as follows (see (2.7.1)):

$$\hat{f}_{i,j} = a_i b_j c'_{i,j} d_k \qquad \text{for cells } (i,j) \text{ in } S'_k, \tag{A.5}$$

for $k = \pm 1, \pm 2, \cdots, \pm(R-1)$, where

$$c'_{i,j} = \begin{cases} \prod_{u=j}^{i-1} c_u & \text{for } i > j, \\ \prod_{u=i}^{j-1} c_u & \text{for } i < j. \end{cases} \tag{A.6}$$

We shall let v_i, w_i, y_i, and z_i denote the quantities

$$v_i = \frac{(\hat{f}_{i,i+2}\hat{f}_{i+1,i-1})}{(\hat{f}_{i,i-1}\hat{f}_{i+1,i+2})} \qquad \text{for } i = 2, 3, \cdots, R-2, \tag{A.7}$$

$$w_i = \frac{(\hat{f}_{i,1}\hat{f}_{2,i+1})}{(\hat{f}_{2,1}\hat{f}_{i,i+1})} \qquad \text{for } i = 3, 4, \cdots, R-1, \tag{A.8}$$

$$y_i = \left[\prod_{j=1}^{i}\left(\frac{\hat{f}_{j,j+1}}{\hat{f}_{j+1,j}}\right)\right]\left[\left(\frac{\hat{f}_{i+1,1}}{\hat{f}_{1,i+1}}\right)\right] \qquad \text{for } i = 2, 3, \cdots, R-1, \tag{A.9}$$

$$z_i = \begin{cases} \dfrac{(\hat{f}_{R,1}\hat{f}_{R-1,2})}{(\hat{f}_{R-1,1}\hat{f}_{R,2})} & \text{for } i = 1, \\ \dfrac{(\hat{f}_{1,R}\hat{f}_{2,R-1})}{(\hat{f}_{1,R-1}\hat{f}_{2,R})} & \text{for } i = -1. \end{cases} \tag{A.10}$$

(For simplicity, we shall first assume that $\hat{f}_{i,j} > 0$ for $i \neq j$.) From (A.5) to (A.10), we find that

$$v_i = \frac{c_i^2 d'_2}{d'_1} \qquad \text{for } i = 2, 3, \cdots, R-2, \tag{A.11}$$

$$w_i = \frac{d'_{i-1}\left(\prod_{j=2}^{i-1} c_j^2\right)}{d'_1} \qquad \text{for } i = 3, 4, \cdots, R-1, \tag{A.12}$$

$$y_i = \left(\frac{d_{-1}}{d_1}\right)^i \left(\frac{d_i}{d_{-i}}\right) \qquad \text{for } i = 2, 3, \cdots, R-1, \tag{A.13}$$

$$z_i = \begin{cases} \dfrac{d_{R-1}d_{R-3}}{d_{R-2}^2} & \text{for } i = 1, \\ \dfrac{d_{-(R-1)}d_{-(R-3)}}{d_{-(R-2)}^2} & \text{for } i = -1, \end{cases} \tag{A.14}$$

where d'_i is defined by

$$d'_i = d_i d_{-i}. \tag{A.15}$$

As we noted earlier (see Section 7), the c_u in (A.5) and (A.6) can be defined uniquely for $u = 2, 3, \cdots, R - 2$, either by introducing a restriction to be imposed directly upon them (for example, that the maximum c_u be set equal to one) or by introducing an additional restriction to be imposed upon the d_k (for example, that $d'_2 = 1$, in addition to the restriction that $d_1 = d_{-1} = 1$). The former kind of restriction would be appropriate when the DC model is viewed as an extension of the D model, and the latter kind of restriction would be appropriate when the DC model is viewed as an extension of the C model. If the maximum c_u is equal to one for $u = 2, 3, \cdots, R - 2$, then from (A.11) we obtain the following formula for the c_u,

$$c_u = \left(\frac{v_u}{v^*}\right)^{1/2} \qquad \text{for} \quad u = 2, 3, \cdots, R - 2, \tag{A.16}$$

where

$$v^* = \max_i v_i \qquad \text{for} \quad 2 \leqq i \leqq R - 2. \tag{A.17}$$

If instead of the above restriction on the maximum c_u, we set $d'_2 = 1$ (in addition to setting $d_1 = d_{-1} = 1$), we see from (A.11) that c_u would be calculated by a modified form of (A.16) in which v^* is replaced by one. Since the $\hat{f}_{i,j}$ in (A.5) are unaffected if d_1 and d_{-1} are set equal to one (and the a_i, b_j, and other d_k are changed accordingly), we see from (A.12) that the d'_i can be calculated by

$$d'_i = \frac{w_{i+1}}{\prod_{j=2}^{i} c_j^2} \qquad \text{for} \quad i = 2, 3, \cdots, R - 2, \tag{A.18}$$

where the c_j are calculated from (A.16). From (A.13) and (A.15), we see that the d_i can be calculated as

$$d_i = \begin{cases} (d'_i y_i)^{1/2} & \text{for} \quad i = 2, 3, \cdots, R - 2, \\ (d'_i / y_i)^{1/2} & \text{for} \quad i = -2, -3, \cdots, -(R - 2), \end{cases} \tag{A.19}$$

where d'_i is calculated from (A.18). From (A.14) we obtain the following formulae for d_{R-1} and $d_{-(R-1)}$:

$$d_i = \begin{cases} z_1 d^2_{R-2} / d_{R-3} & \text{for} \quad i = R - 1, \\ z_{-1} d^2_{-(R-2)} / d_{-(R-3)} & \text{for} \quad i = -(R - 1). \end{cases} \tag{A.20}$$

We next consider

$$\hat{f}'_{i,j} = \frac{\hat{f}_{i,j}}{d_k c'_{i,j}} \qquad \text{for cells } (i, j) \text{ in } S'_k, \tag{A.21}$$

for $k = \pm 1, \pm 2, \cdots, \pm(R - 1)$, where the d_k are calculated from (A.19) and (A.20) and the $c'_{i,j}$ are calculated from (A.6), with c_u calculated from (A.16). From (A.5) and (A.21), we see that

(A.22) $$\hat{f}'_{i,j} = a_i b_j \qquad \text{for} \quad i \neq j.$$

Since the $\hat{f}'_{i,j}$ are of the same form as described by (A.1), the a_i and b_j can be calculated from (A.2) and (A.3) by replacing $\hat{f}_{i,j}$ by $\hat{f}'_{i,j}$ in these formulae. Thus, by applying the methods described in the present and preceding paragraphs, all of the parameters in the DC model can be estimated.

In the preceding discussion of the DC model, we assumed that $\hat{f}_{i,j} > 0$ for $i \neq j$. In cases where this assumption is not true, the above methods require some modification. For example, if $f_{R,1} = 0$ (as is the case for Tables I and III), then $\hat{f}_{R,1} = 0$ and $d_{R-1} = 0$ when the DC model is applied to these tables, and $\hat{f}'_{R,1}$ will be undefined. In this case, the $\hat{f}'_{i,j}$ that are well defined will have the same form as (A.22), and the methods described earlier for estimating the parameters in the model of quasi-independence can be applied to this set of $\hat{f}'_{i,j}$.

The methods described in this section can be applied, either directly or indirectly, to estimate the parameters in any of the models of Sections 2.2 to 2.8. For example, with the DPC model, by a direct application of the above formulae, we would obtain, among other things, the estimate d_k of the δ_k that satisfy (2.5.12). By an indirect application, we would first set $d_k = 1$ for $k = -1, -2, \cdots, -(R-1)$, in accordance with condition (2.5.9), and then from (A.11), (A.12), and (A.15), we would obtain

(A.23) $$d_1 = \frac{v_2}{y_2 c_2^2},$$

with c_2 calculated from (A.16). (For the DPC model, if the d_k are set at one for $k = -1, -2, \cdots, -(R-1)$, then d_1 will not be set at one.) To calculate d_k for $k = 2, 3, \cdots, R-1$, for this model, we then obtain, from (A.12) and (A.15),

(A,24) $$d_i = d_1 d'_i, \qquad \text{for} \quad i = 2, 3, \cdots, R-2,$$

with d_1 and d'_i calculated from (A.23) and (A.18), respectively. The value of d_{R-1} can be calculated from (A.20), using the d_i, for $i = R-2$ and $R-3$, calculated from (A.24).

In cases where the given model includes the triangles parameter (for example, in the DACT model), the parameter τ can be expressed as

(A.25) $$\tau = \left(\frac{\delta_k}{\delta_{-k}}\right)^{1/2} \qquad \text{for} \quad k = 1, 2, \cdots, R-1,$$

where $\tau = \sqrt{\tau'}$, for τ' defined by (2.7.4). From (A.13) and (A.25) we would obtain the following formula for the maximum likelihood estimate t of τ:

(A.26) $$t = \frac{1}{\sqrt{y_2}}.$$

If the model includes both the triangles parameter and the parameters pertaining to the paired minor diagonals, then after introducing the τ parameter explicitly

into the model, the model will be unaffected by setting $\delta_k = \delta_{-k}$ for $k = 1, 2, \cdots, R-1$. With the parameter τ estimated by (A.26), we can estimate $\delta_k = \delta_{-k}$ by

$$d_k = \sqrt{d'_k} \quad \text{for} \quad k = 2, 3, \cdots, R-2, \tag{A.27}$$

where d'_k is calculated from (A.18). We can set $d_1 = d_{-1} = 1$ in this model, and d_{R-1} can be calculated from (A.20), using d_i for $i = R-2$ and $R-3$ calculated from (*A*.27).

For the QO model or for any of the models described in Sections 2.4 to 2.7, the ratio index μ_i and the relative difference index μ_i^*, which we defined in Section 5, can be estimated by the corresponding quantities m_i and m_i^* (see (5.14) and (5.15)), with the a_i and b_j calculated as described in the present section, and with d_0 as described in Section 7. For the models described in Section 2.8, we set $\mu_i = 1$, and therefore

$$\hat{f}_{i,i} = a_i b_i d_0 \quad \text{for} \quad i = 1, 2, \cdots, R. \tag{A.28}$$

With the $\hat{f}_{i,j}$ calculated by the iterative scaling method (for all cells (i, j) in the $R \times R$ table under the models of Section 2.8), we can calculate d_0 from

$$d_0 = \frac{\hat{f}_{i,i}}{a_i b_i}, \tag{A.29}$$

with the a_i and b_i calculated as described earlier in the present section.

For any given model of the kind described in Section 2.8 (for example, the DCF model), an alternative method for calculating d_0 can be based upon the fact that

$$d_0^2 = d'_1 c_i^2 u_i, \tag{A.30}$$

where

$$u_i = \frac{(\hat{f}_{i,i}\hat{f}_{i+1,i+1})}{(\hat{f}_{i,i+1}\hat{f}_{i+1,i})} \quad \text{for} \quad i = 1, 2, \cdots, R-1, \tag{A.31}$$

with the c_i calculated as earlier for $i = 2, 3, \cdots, R-2$. Formula (*A*.30) can be used to calculate d_0 (applying the formula for any given value of $i = 2, 3, \cdots, R-2$), and then c_1 and c_{R-1} can be calculated by rewriting (A.30)

$$c_i = \frac{d_0}{(d'_1 u_i)^{1/2}}. \tag{A.32}$$

Now for all cells (i, j) in the table, we can consider

$$\hat{f}'_{i,j} = \frac{\hat{f}_{i,j}}{d_k c'_{i,j}} \quad \text{for cells } (i, j) \text{ in } S'_k, \tag{A.33}$$

for $k = 0, \pm 1, \pm 2, \cdots, \pm(R-1)$. (Compare (A.33) with (A.21).) The $\hat{f}'_{i,j}$ will be of the form

$$\hat{f}'_{i,j} = a_i b_j \quad \text{for all cells } (i, j) \tag{A.34}$$

(see (A.22)), and so the a_i and b_j can be calculated here by the same methods used to estimate the parameters in the usual model of independence between the row and column classifications. For the models of Section 2.8, the method just described provides an alternative to the method described following (A.22) for calculating a_i and b_j. It should also be noted, as we did earlier, that in cases where the assumption that $\hat{f}_{i,j} > 0$ is not met for all cells in the table (under a given model of Section 2.8), the above methods require modifications of the kind which we described in the paragraph following (A.22).

The results of the present section can be applied to any of the models in Section 2.2 to 2.8 whenever the maximum likelihood estimate $\hat{f}_{i,j}$ exists. For comments concerning the existence of the $\hat{f}_{i,j}$, see Section 3 herein and Haberman [20].

For helpful comments, the author is indebted to R. Fay, S. Fienberg, S. Haberman, and T. Pullum.

REFERENCES

[1] R. R. Bahadur, "Rates of convergence of estimates and test statistics," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 303–324.

[2] Y. M. M. Bishop and S. E. Fienberg, "Incomplete two-dimensional contingency tables," *Biometrics*, Vol. 25 (1969), pp. 118–128.

[3] I. Blumen, M. Kogan, and P. J. McCarthy, *The Industrial Mobility of Labor as a Probability Process*, Cornell Studies in Industrial and Labor Relations, Vol. 4, Ithaca, Cornell University, 1955.

[4] H. Caussinus, "Contribution à l'analyse statistique des tableaux de corrélation," *Annales de la Faculté des Sciences de l'Université de Toulouse*, Vol. 29 (1965), pp. 77–182.

[5] S. E. Fienberg, "Quasi-independence and maximum-likelihood estimation in incomplete contingency tables," *J. Amer. Statist. Assoc.*, Vol. 65 (1970) pp. 1610–1616.

[6] D. V. Glass (editor), *Social Mobility in Britain*, Glencoe, Free Press, 1954.

[7] I. J. Good, "Saddle-point methods for the multinomial distribution," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 861–881.

[8] L. A. Goodman, "Statistical methods for the mover-stayer model," *J. Amer. Statist. Assoc.*, Vol. 56 (1961), pp. 841–868.

[9] ———, "Statistical methods for the preliminary analysis of transaction flows," *Econometrica*, Vol. 31 (1963), pp. 197–208.

[10] ———, "A short computer program for the analysis of transaction flows," *Behav. Sci.*, Vol. 9 (1964), pp. 176–186.

[11] ———, "Simultaneous confidence limits for cross-product ratios in contingency tables," *J. Roy. Statist. Soc., Ser. B*, Vol. 26 (1964), pp. 86–102.

[12] ———, "On the statistical analysis of mobility tables," *Am. J. Sociology*, Vol. 70 (1965), pp. 564–585.

[13] ———, "The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries," *J. Amer. Statist. Assoc.*, Vol. 63 (1968), pp. 1091–1131.

[14] ———, "How to ransack social mobility tables and other kinds of cross-classification tables," *Am. J. Sociology*, Vol. 75 (1969), pp. 1–40.

[15] ———, "On partitioning χ^2 and detecting partial association in three-way contingency tables," *J. Roy. Statist. Soc., Ser. B*, Vol. 31 (1969), pp. 486–498.
[16] ———, "On the measurement of social mobility: An index of status persistence," *American Sociological Review*, Vol. 34 (1969), pp. 832–850.
[17] ———, "The multivariate analysis of qualitative data: Interactions among multiple classifications," *J. Amer. Statist. Assoc.*, Vol. 65 (1970), pp. 226–256.
[18] ———, "The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications," *Technomet.*, Vol. 13 (1971), pp. 33–61.
[19] ———. "A simple simultaneous test procedure for quasi-independence in contingency tables," *J. Roy. Statist. Soc., Ser. C*, Vol. 20 (1971).
[20] S. HABERMAN, "The general log-linear model," Ph.D. thesis, University of Chicago, 1970.
[21] W. HOEFFDING, "Asymptotically optimal tests for multinomial distributions," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 369–408.
[22] J. LEVINE, "Measurement in the study of intergenerational status mobility," Ph.D. thesis, Harvard University, 1967.
[23] N. MANTEL, "Incomplete contingency tables," *Biometrics*, Vol. 26 (1970), pp. 291–304.
[24] F. MOSTELLER, "Association and estimation in contingency tables," *J. Amer. Statist. Assoc.*, Vol. 63 (1968), pp. 1–28.
[25] K. SVALASTOGA, *Prestige, Class and Mobility*, London, William Heinemann, 1959.