Austral. & New Zealand J. Statist. **40**(4), 1998, 465–477

PARTITIONING PEARSON'S CHI-SQUARED STATISTIC FOR A COMPLETELY ORDERED THREE-WAY CONTINGENCY TABLE

ERIC J. BEH[1*] AND PAMELA J. DAVY[1]

University of Wollongong

Summary

The paper presents a partition of the Pearson chi-squared statistic for triply ordered three-way contingency tables. The partition invokes orthogonal polynomials and identifies three-way association terms as well as each combination of two-way associations. This partition provides information about the structure of each variable by identifying important bivariate and trivariate associations in terms of location (linear), dispersion (quadratic) and higher order components. The significance of each term in the partition, and each association within each term can also be determined.

The paper compares the chi-squared partition with the log-linear models of Agresti (1994) for multi-way contingency tables with ordinal categories, by generalizing the model proposed by Haberman (1974).

Key words: location; dispersion and higher order components; orthogonal polynomials; three-way contingency table; ordinal log-linear models.

1. Introduction

A number of authors have successfully partitioned the classical Pearson chi-squared statistic for two-way contingency tables with ordered categories; see e.g. Lancaster (1953), Nair (1986) and Hirotsu (1978, 1982, 1983, 1986). Kendall & Stuart (1979, p. 607) and Lancaster (1951, 1980) discuss the partitioning of the chi-squared statistic for multi-way contingency tables.

In this paper we present the partition of the Pearson chi-squared statistic using orthogonal polynomials for a three-way contingency table where all three variables contain completely ordered categories. The partition isolates the location (or linear), dispersion (quadratic) and higher order components for each variable and determines the three-way association and each combination of two-way associations for the three variables. We show that such a partition can also be easily extended for the analysis of any multi-way contingency table which consists of ordered variables. The case when only one or two variables are ordered is considered in Beh & Davy (1999).

Section 2 of the present paper discusses the partition of the chi-squared statistic for three-way contingency tables with three ordered variables. Section 3 defines some models of association that can be used for goodness-of-fit, while Section 4 generalizes the partition given in Section 2 for any multi-way contingency table with all variables ordered. Section 5 compares the analysis using the partition with that of log-linear models of ordered variables. The log-linear models found in this section are extensions of the models proposed by Haberman

Received October 1997; revised April 1998; accepted April 1998.

[*] Author to whom correspondence should be addressed.
[1] School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia.
email: eric@uow.edu.au

Acknowledgments. The authors thank John C. W. Rayner for his very helpful comments and suggestions during the preparation of this paper, and anonymous reviewers for their comments.

(1974) who considered a two-way contingency table consisting of ordered rows and ordered columns. It is shown that the simplicity of modelling via the chi-squared partition approximates well the parameters from the more complicated log-linear analysis approach.

2. Chi-squared Partition for Three-way Tables

2.1. The Partition

Consider a three-way contingency table, N, whose grand total is n, with I ordered rows, J ordered columns, and, using the term of Kroonenberg (1989), K ordered tubes (Kendall & Stuart (1979) use the term 'layer'). The (i, j, k)th element of N is n_{ijk} for $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$. Define the (i, j, k)th cell probability as $p_{ijk} = n_{ijk}/n$ so that $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} p_{ijk} = 1$. Let $p_{i\cdot\cdot}$ be the ith row marginal probability so that $\sum_{i=1}^{I} p_{i\cdot\cdot} = 1$. Similarly let $p_{\cdot j\cdot}$ and $p_{\cdot\cdot k}$ be the jth column and kth tube marginal probabilities respectively so that $\sum_{j=1}^{J} p_{\cdot j\cdot} = \sum_{k=1}^{K} p_{\cdot\cdot k} = 1$.

Best & Rayner (1996) consider the chi-squared partition of a doubly-ordered two-way contingency table into bivariate components. This partition is generalized in this paper for multi-way contingency tables.

Consider our three-way contingency table, N. Then the Pearson chi-squared statistic can be partitioned, under the hypothesis of complete independence, so that

$$X^2 = \sum_{u=1}^{I-1}\sum_{v=1}^{J-1}\sum_{w=1}^{K-1} Z_{uvw}^2 + \sum_{u=1}^{I-1}\sum_{v=1}^{J-1} Z_{uv0}^2 + \sum_{u=1}^{I-1}\sum_{w=1}^{K-1} Z_{u0w}^2 + \sum_{v=1}^{J-1}\sum_{w=1}^{K-1} Z_{0vw}^2 \qquad (2.1)$$

where

$$Z_{uvw} = \sqrt{n}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} a_u(i)b_v(j)c_w(k)p_{ijk}\,. \qquad (2.2)$$

The sets $\{a_u(i)\colon u = 0, \ldots, (I-1)\}$, $\{b_v(j)\colon v = 0, \ldots, (J-1)\}$ and $\{c_w(k)\colon w = 0, \ldots, (K-1)\}$ are the orthogonal polynomials associated with the rows, columns and tubes respectively. These polynomials require the input of a set of scores which reflect the ordered structures of the categories. Beh (1998) defines these polynomials, and how they are affected by different scores.

Each of the Z terms defined by (2.2) is asymptotically standard normal and independent. Refer to the Appendix for the proof of (2.1) and (2.2). For the sake of simplicity, (2.1) is alternatively expressed as

$$X^2 = X_{IJK}^2 + X_{IJ}^2 + X_{IK}^2 + X_{JK}^2\,. \qquad (2.3)$$

2.2. Chi-squared Terms and Associated Values

The Pearson chi-squared statistic of (2.1) or (2.3) is partitioned into four terms. The first term, X_{IJK}^2, describes the trivariate association between rows, columns and tubes. Testing the significance of X_{IJK}^2 with the chi-squared distribution with $(I-1)(J-1)(K-1)$ degrees of freedom determines the level of association between the rows, columns and tubes. However, even if this term is not significant, associations described by its component Z values may be. The term Z_{uvw} is the deviation of the rows, columns and tubes up to the (u, v, w)th trivariate moment in the data from what might be expected under complete independence. For example,

Z_{111} is the linear-by-linear-by-linear association and assesses the trivariate location of the three variables. Z_{121} is the linear-by-quadratic-by-linear association of the row–column–tube interaction.

The remaining terms of (2.3) are the chi-squared statistics for each bivariate combination of the rows, columns and tubes. The second term, X^2_{IJ}, is the chi-squared statistic for the two-way doubly ordered contingency table created when the ordered tubes are collapsed. It is analogous to the statistic discussed in Rayner & Best (1996) and Beh (1997).

To show this, by definition

$$X^2_{IJ} = \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} Z^2_{uv0}\,, \tag{2.4}$$

where

$$Z_{uv0} = \sqrt{n} \sum_{i=1}^{I} \sum_{j=1}^{J} a_u(i) b_v(j) p_{ij\cdot}\,. \tag{2.5}$$

As $p_{ij\cdot} = \sum_{k=1}^{K} p_{ijk}$, X^2_{IJ} is just the chi-squared statistic applied to the contingency table formed by summing over the tubes.

The interpretation of X^2_{IJ} of (2.4) is then similar to (3.2.1) of Beh (1997) when summing over the tubes for each i and j. When compared with the chi-squared distribution, with $(I-1)(J-1)$ degrees of freedom, it is a measure of the departure from independence between the rows and columns of N assuming that the tubes are independent of the rows and columns. Similarly, X^2_{IK} and X^2_{JK} are the bivariate chi-squared statistics and are measures of the departure from marginal independence between the rows and tubes or columns and tubes respectively.

The interpretation of the Z values for the two-way terms is similar to that of Z_{uvw} in X^2_{IJK}. Consider the second term, X^2_{IJ}. The value Z_{uv0} defined by (2.5) is the value of the (u, v)th bivariate moment between the rows and columns when summing over the tubes. For example Z_{110} is the linear-by-linear association between the rows and columns independent of the tubes. Similarly for the term X^2_{IK}, Z_{u0w} is the (u, w)th marginal bivariate moment between the rows and tubes, while for the fourth term, X^2_{JK}, Z_{0vw} is the (v, w)th marginal bivariate moment between the columns and tubes. Even if it is found that a particular bivariate chi-squared statistic is not significant, it is possible to identify significant bivariate associations. Rayner & Best (1996) give additional interpretations of the bivariate Z values.

2.3. Component Values

The effect of the row location component on the three-way association is $\sum_{v=1}^{J-1} \sum_{w=1}^{K-1} Z^2_{1vw}$, while, in general, the uth row component for the association is $\sum_{v=1}^{J-1} \sum_{w=1}^{K-1} Z^2_{uvw}$. Testing for such components allows for an examination of the trend of the row categories, the trend being dictated by the uth order orthogonal polynomial. For example, the row location component describes how the row means (see μ_I in (3.6) later) affect the trivariate association. In a similar manner, the vth column component for the three-way association is $\sum_{u=1}^{I-1} \sum_{w=1}^{K-1} Z^2_{uvw}$, while the wth tube component is $\sum_{u=1}^{I-1} \sum_{v=1}^{J-1} Z^2_{uvw}$.

The component values for each two-way association are easily calculated. For the bivariate association between the rows and columns, the uth row component is $\sum_{v=1}^{J-1} Z^2_{uv0}$, while the vth column component is $\sum_{u=1}^{I-1} Z^2_{uv0}$. For example, the location component for the row categories is $\sum_{v=1}^{J-1} Z^2_{1v0}$; it describes the effect the row means have on the bivariate association between the rows and columns. Rayner & Best (1996) described the component values for the row–column bivariate moment.

Row, column and tube components for the remaining bivariate associations can be similarly calculated.

3. Models of Association

Models of association for a two-way contingency table have been extensively reviewed; see e.g. Goodman (1979, 1981, 1985, 1986, 1996), Gilula, Krieger & Ritov (1988) and Rom & Sarkar (1992). Multi-way contingency tables with ordinal categories have been analysed by Clogg (1982), Danaher (1991), Gilula & Haberman (1988) and Goodman (1970). This section proposes models of association for multi-way tables, based on the partition of (2.1)–(2.2).

We can test for complete independence of the variables by considering the saturated trivariate model of association for a completely ordered three-way contingency table,

$$p_{ijk} = p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}\sum_{u=0}^{I-1}\sum_{v=0}^{J-1}\sum_{w=0}^{K-1}\left(\frac{Z_{uvw}}{\sqrt{n}}\right)a_u(i)b_v(j)c_w(k). \tag{3.1}$$

The model defined by (3.1) enables us to reconstitute the cell probability p_{ijk} for the $(i,\, j,\, k)$th cell, for all cells. The unsaturated model which can be used to approximate the $(i,\, j,\, k)$th cell probability under the hypothesis of complete independence is

$$p_{ijk} = p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}\sum_{u=0}^{M_1}\sum_{v=0}^{M_2}\sum_{w=0}^{M_3}\left(\frac{Z_{uvw}}{\sqrt{n}}\right)a_u(i)b_v(j)c_w(k), \tag{3.2}$$

where the first M_1 row components, M_2 column components, or M_3 tube components, are selected for $M_1 < I-1,\ M_2 < J-1$ and $M_3 < K-1$. Usually $M_1,\ M_2$ and M_3 are chosen to be equal to 2, so that N is analysed in terms of the location and dispersion components only. Unlike other modelling procedures, especially those where association values have an ordering constraint imposed such as for the singular value decomposition for two-way tables, the most important component values can be chosen so they improve the model. They may not include the first M_1 row components, M_2 column components or M_3 tube components. Instead they may include the most important M_1 row components, M_2 column components and M_3 tube components.

Expanding (3.2) gives the alternative unsaturated model,

$$\begin{aligned} p_{ijk} = p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}\Bigg[1 &+ \sum_{u=1}^{M_1}\sum_{v=1}^{M_2}\left(\frac{Z_{uv0}}{\sqrt{n}}\right)a_u(i)b_v(j) + \sum_{u=1}^{M_1}\sum_{w=1}^{M_3}\left(\frac{Z_{u0w}}{\sqrt{n}}\right)a_u(i)c_w(k) \\ &+ \sum_{v=1}^{M_2}\sum_{w=1}^{M_3}\left(\frac{Z_{0vw}}{\sqrt{n}}\right)b_v(j)c_w(k) + \sum_{u=1}^{M_1}\sum_{v=1}^{M_2}\sum_{w=1}^{M_3}\left(\frac{Z_{uvw}}{\sqrt{n}}\right)a_u(i)b_v(j)c_w(k)\Bigg]. \end{aligned} \tag{3.3}$$

In model (3.3) we consider each combination of two-way associations and three-way association terms and find that when the Z values are zero, the rows, columns and tubes are completely independent. If there is not complete independence, then model (3.3) can be used to identify which association(s) is significant. Model (3.3) is an extension of the model of

Rayner & Best (1996) who analysed two-way contingency tables. Note that when summing across the tubes, model (3.3) simplifies to

$$p_{ij\cdot} = p_{i\cdot\cdot}p_{\cdot j\cdot}\left[1 + \sum_{u=1}^{M_2}\sum_{v=1}^{M_1} a_u(i)\left(\frac{Z_{uv0}}{\sqrt{n}}\right)b_v(j)\right], \tag{3.4}$$

which is just the model of association in Beh (1997) and Rayner & Best (1996) for a two-way contingency table, and is also the model of association for the rows and columns of N when the tube categories are collapsed.

Similar models can be obtained for the association between the rows and tubes, and the columns and tubes.

Consider $Z_{uv0}/\sqrt{n}$ in model (3.3); it is defined by (2.5). Suppose that equally spaced integer value scores are used for the calculation of the orthogonal polynomials. Then, when $u = v = 1$ (that is when only the row and column location components are considered), (2.5) becomes

$$\frac{Z_{110}}{\sqrt{n}} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(i-\mu_I)}{\sigma_I}\frac{(j-\mu_J)}{\sigma_J} p_{ij\cdot}\,, \tag{3.5}$$

where

$$\mu_I = \sum_{i=1}^{I} i p_{i\cdot\cdot}\,, \quad \mu_J = \sum_{j=1}^{J} j p_{\cdot j\cdot}\,, \quad \sigma_I^2 = \sum_{i=1}^{I} i^2 p_{i\cdot\cdot} - \mu_1^2\,, \quad \sigma_J^2 = \sum_{j=1}^{J} j^2 p_{\cdot j\cdot} - \mu_J^2\,. \tag{3.6}$$

Equation (3.5) gives the correlation value for the I rows and J columns (see Danaher, 1991). Rayner & Best (1996) calculated the Pearson product moment correlation for two-way contingency tables using orthogonal polynomials. For three-way contingency tables, (3.5) offers a way of calculating the Pearson product moment correlation between the rows and columns, and can easily be generalized for any multi-way contingency table. Pearson product moment correlation values can be similarly calculated for the row–tube and column–tube interactions for our contingency table N. When midrank scores are used, (3.5) is also an extension of Spearman's rank correlation for the rows and columns of a three-way contingency table. While Best & Rayner (1996) determined Spearman's rank correlation for two-way contingency tables, (3.5) can be generalized for any multi-way contingency table.

Equation (3.5) is also the correlation between the rows and columns in Goodman's RC model (Goodman, 1985). When the scores used are the first non-trivial row and column singular vector from a simple correspondence analysis of the rows and columns when summing over the tubes, then (3.5) is the first singular value of the normalized cell probabilities (see Beh, 1998).

The association values $Z_{101}/\sqrt{n}$ and $Z_{011}/\sqrt{n}$ have a similar interpretation as $Z_{110}/\sqrt{n}$, but relate to the rows and tubes, and columns and tubes respectively.

When only the location components are considered for each variable, (3.3) becomes

$$\begin{aligned} p_{ijk} = p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}\Bigg[1 &+ \left(\frac{Z_{110}}{\sqrt{n}}\right)\frac{(i-\mu_I)}{\sigma_I}\frac{(j-\mu_J)}{\sigma_J} + \left(\frac{Z_{101}}{\sqrt{n}}\right)\frac{(i-\mu_I)}{\sigma_I}\frac{(k-\mu_K)}{\sigma_K} \\ &+ \left(\frac{Z_{011}}{\sqrt{n}}\right)\frac{(j-\mu_J)}{\sigma_J}\frac{(k-\mu_K)}{\sigma_K} + \left(\frac{Z_{111}}{\sqrt{n}}\right)\frac{(i-\mu_I)}{\sigma_I}\frac{(j-\mu_J)}{\sigma_J}\frac{(k-\mu_K)}{\sigma_K}\Bigg]. \end{aligned} \tag{3.7}$$

Model (3.7) is similar to equation (5) in Danaher (1991). However, Danaher's model did not include the trivariate Z term that is in (3.7).

An alternative modelling procedure which can be used as an approximation to (3.7), even when the model is saturated, involves using the property $e^x \approx 1 + x$ when the Z values are small. Using this property, the (i, j, k)th cell probability can be approximated by

$$\begin{aligned} p_{ijk} \approx p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k} \exp\bigg[\bigg(\frac{Z_{110}}{\sqrt{n}}\bigg)\frac{(i-\mu_I)}{\sigma_I}\frac{(j-\mu_J)}{\sigma_J} + \bigg(\frac{Z_{101}}{\sqrt{n}}\bigg)\frac{(i-\mu_I)}{\sigma_I}\frac{(k-\mu_K)}{\sigma_K} \\ + \bigg(\frac{Z_{011}}{\sqrt{n}}\bigg)\frac{(j-\mu_J)}{\sigma_J}\frac{(k-\mu_K)}{\sigma_K} + \bigg(\frac{Z_{111}}{\sqrt{n}}\bigg)\frac{(i-\mu_I)}{\sigma_I}\frac{(j-\mu_J)}{\sigma_J}\frac{(k-\mu_K)}{\sigma_K}\bigg]. \end{aligned} \tag{3.8}$$

Model (3.8) is akin to Goodman's association model of Goodman (1986) for a three-way contingency table and can be used because the scores have been standardized. The advantage of this model is that when approximating a cell probability, negative values cannot arise. Equation (3.8) can also be generalized to consider quadratic and higher order components, as (3.1) does. However, the sum of all the reconstituted cell entries is not always unity.

4. Generalization for m-way Contingency Tables

The partition of (2.1) can be generalized for any m-way contingency table with m ordered variables. Suppose that the tth variable, v_t contains c_t categories (so that $1 \le t \le m$), then for a m-way contingency table, the chi-squared statistic can be partitioned so that

$$X^2 = \sum_{u_1=0}^{c_1-1}\sum_{u_2=0}^{c_2-1}\cdots\sum_{u_m=0}^{c_m-1} Z^2_{u_1u_2\ldots u_m} - n \tag{4.1}$$

where

$$Z^2_{u_1u_2\ldots u_m} = \sum_{v_1=1}^{c_1}\sum_{v_2=1}^{c_2}\cdots\sum_{v_m=1}^{c_m} a_{u_1}(v_1)b_{u_2}(v_2)\ldots c_{u_m}(v_m)p_{v_1v_2\ldots v_m}. \tag{4.2}$$

The Z terms are asymptotically standard normal and independent. (4.1)–(4.2) can be proved by generalizing the proof of (2.1)–(2.2) in the Appendix.

Equation (4.1) can easily be expanded so that it has the same form as (2.1), and includes $2^m - m - 1$ terms. When in this form, there is one term for the m-way association, m terms for the $(m-1)$-way association and so on down to $\frac{1}{2}m(m-1)$ two-way associations.

5. Comparison with Log-Linear Analysis

Haberman (1974) discussed log-linear models of two-way contingency tables in terms of linear, quadratic and higher order components. This section extends the analysis to deal with multi-way contingency tables, by considering the model for a three-way contingency table.

Fienberg (1977) and Agresti (1994) discuss the log-linear models for ordered categorical data. Agresti offers the model (5.1) for our three-way contingency table, N,

$$\begin{aligned} \log n_{ijk} = & u + u_{1(i)} + u_{2(j)} + u_{3(k)} + \beta_{IJ}(r_i - \bar{r})(c_j - \bar{c}) + \beta_{IK}(r_i - \bar{r})(t_k - \bar{t}\,) \\ & + \beta_{JK}(c_j - \bar{c})(t_k - \bar{t}\,) + \beta_{IJK}(r_i - \bar{r})(c_j - \bar{c})(t_k - \bar{t}\,), \end{aligned} \tag{5.1}$$

where $\sum_{i=1}^{I} u_{1(i)} = \sum_{j=1}^{J} u_{2(j)} = \sum_{k=1}^{K} u_{3(k)} = 0$. The value of r_i is the score associated with the ith row category, while c_j is the value of the score associated with the jth column category and t_k is the score associated with the kth tube category. These are the scores used in calculating the orthogonal polynomials. The values β_{IJ}, β_{IK}, β_{JK} describe the bivariate association term between the three variables and are calculated via maximum likelihood estimation (MLE). They correspond to the linear-by-linear associations between each pair of variables while β_{IJK} is the trivariate association term and corresponds to the linear-by-linear-by-linear association. To see this, consider model (3.8). Taking the natural log of both sides after multiplying by n yields

$$\log n_{ijk} = \log n + \log p_{i\cdot\cdot} + \log p_{\cdot j\cdot} + \log p_{\cdot\cdot k} + \left(\frac{Z_{110}}{\sqrt{n}}\right)\frac{(i-\mu_I)}{\sigma_I}\frac{(j-\mu_J)}{\sigma_J} + \left(\frac{Z_{101}}{\sqrt{n}}\right)\frac{(i-\mu_I)}{\sigma_I}\frac{(k-\mu_K)}{\sigma_K} + \left(\frac{Z_{011}}{\sqrt{n}}\right)\frac{(j-\mu_J)}{\sigma_J}\frac{(k-\mu_K)}{\sigma_K} + \left(\frac{Z_{111}}{\sqrt{n}}\right)\frac{(i-\mu_I)}{\sigma_I}\frac{(j-\mu_J)}{\sigma_J}\frac{(k-\mu_K)}{\sigma_K}. \tag{5.2}$$

Comparing models (5.1) and (5.2) for integer valued scores, the parameters calculated from the log-linear model of (5.1) can be approximated by

$$\hat{\beta}_{IJ} = \frac{Z_{110}}{\sigma_I\sigma_J\sqrt{n}}, \tag{5.3}$$

$$\hat{\beta}_{IK} = \frac{Z_{101}}{\sigma_I\sigma_K\sqrt{n}}, \tag{5.4}$$

$$\hat{\beta}_{JK} = \frac{Z_{011}}{\sigma_J\sigma_K\sqrt{n}}, \tag{5.5}$$

$$\hat{\beta}_{IJK} = \frac{Z_{111}}{\sigma_I\sigma_J\sigma_K\sqrt{n}}. \tag{5.6}$$

The models of Agresti (1994) and Fienberg (1977) consider the centring (about the mean) of the scores; for our analysis we standardize them.

Fienberg (1977, p. 47) points out that, while several authors have proposed techniques for selecting the optimum log-linear model, 'unfortunately, there is no all-purpose, best method of model selection'. However, for a three-way contingency table, the parameter approximations of (5.3)–(5.6) and the test for important bivariate and trivariate associations from the chi-squared partition of (2.1) allow the user to select the log-linear model which describes the various relationships between the variables.

The advantage of model (5.2) is that it can be generalized to consider not only the linear component, but also the quadratic and higher order component values. The model $\log n_{ijk}$ equals

$$\log n + \log p_{i\cdot\cdot} + \log p_{\cdot j\cdot} + \log p_{\cdot\cdot k} + \sum_{u=1}^{M_1}\sum_{v=1}^{M_2}\left(\frac{Z_{uv0}}{\sqrt{n}}\right)a_u(i)b_v(j) + \sum_{u=1}^{M_1}\sum_{w=1}^{M_3}\left(\frac{Z_{u0w}}{\sqrt{n}}\right)a_u(i)c_w(k) + \sum_{v=1}^{M_2}\sum_{w=1}^{M_3}\left(\frac{Z_{0vw}}{\sqrt{n}}\right)b_v(j)c_w(k) + \sum_{u=1}^{M_1}\sum_{v=1}^{M_2}\sum_{w=1}^{M_3}\left(\frac{Z_{uvw}}{\sqrt{n}}\right)a_u(i)b_v(j)c_w(k), \tag{5.7}$$

TABLE 1

Cross-classification of 1517 people according to happiness, schooling and number of siblings

Years of schooling	Number of siblings 0–1	2–3	4–5	6–7	8+
			Not too happy		
< 12	15	34	36	22	61
12	31	60	46	25	26
13–16	35	45	30	13	8
17+	18	14	3	3	4
			Pretty happy		
< 12	17	53	70	67	79
12	60	96	45	40	31
13–16	63	74	39	24	7
17+	15	15	9	2	1
			Very happy		
< 12	7	20	23	16	36
12	5	12	11	12	7
13–16	5	10	4	4	3
17+	2	1	2	0	1

where M_1, M_2 and M_3 are chosen in a manner similar to the corresponding parameters in model (3.2). Alternatively, either the M_1 most important row components, M_2 column components and/or M_3 tube components can be chosen to improve the model.

The advantage of this type of analysis of a multi-way contingency table is that the differences between categories of a variable, and the association and component values, can be represented graphically using correspondence analysis. Beh & Davy (1997) discuss such a method. With the link between models (5.1) and (5.2), and the generalization of (5.2) to higher order components (see model (5.7)), this correspondence analysis approach can also be viewed as a graphical method of log-linear analysis. Beh & Davy (1997) extends the correspondence analysis method for two-way contingency tables discussed in Beh (1997). Many authors have described the relationship between correspondence analysis of non-ordinal contingency tables and log-linear models; see e.g. van der Heijden, de Falguerolles & de Leeuw (1989), van der Heijden & de Leeuw (1985), van der Heijden & Worsley (1988) and Goodman (1986). Everitt (1992) discusses another method of graphically representing log-linear models proposed by Darroch, Lauritzen & Speed (1980).

6. Example

Consider the three-way contingency table cited by Clogg (1982) from Davis (1977), which classifies 1517 people according to their reported happiness, number of completed years of schooling and the number of siblings. The data are reproduced in Table 1.

In this example we consider the ordering of the happiness, schooling and siblings variables to identify the important bivariate and trivariate moments and identify important location, dispersion and higher order components. Alternatively, we can regard happiness as a response variable and schooling and siblings as the explanatory variables. Beh & Davy (1997) consider such a problem when discussing partitioning the chi-squared statistic for partially ordered three-way contingency table.

TABLE 2

Partition of chi-squared statistic into component values

Term	Component	Value	df	P value
X^2_{IJ}	Column components			
	Location	222.2234	3	0
	Dispersion	7.7034	3	0.0528
	Error	5.3720	6	0.5035
	Row components			
	Location	209.9878	4	0
	Dispersion	24.2943	4	0.0001
	Error	1.0168	4	0.9156
		235.2988	12	0
X^2_{IK}	Tube components			
	Location	28.6582	3	0
	Dispersion	12.4805	3	0.0061
	Row component			
	Location	31.6954	2	0
	Dispersion	6.2973	2	0.0466
	Error	3.1460	2	0.2127
		41.1387	6	0
X^2_{JK}	Tube components			
	Location	8.7582	4	0.0770
	Dispersion	17.0634	4	0.0009
	Column component			
	Location	18.0973	2	0.0001
	Dispersion	0.9724	2	0.6172
	Error	6.7517	4	0.1492
		25.8215	8	0.0008
X^2_{IJK}	Tube components			
	Location	10.9158	12	0.5507
	Dispersion	15.3925	12	0.2250
	Column components			
	Location	4.0458	6	0.6683
	Dispersion	12.4921	6	0.0574
	Error	9.7705	12	0.6389
	Row components			
	Location	4.2737	8	0.8360
	Dispersion	18.9646	8	0.2000
	Error	3.0701	8	0.9245
		26.3084	24	0.3426
X^2		328.5674	50	0

The chi-squared statistic for Table 1 is 328.57 with integer row scores 1 to 4, integer column scores 1 to 5 and integer tube scores 1, 2 and 3. At 50 degrees of freedom the chi-squared value is highly significant, suggesting that there is an association between the happiness, number of years of schooling and number of siblings for a person.

Table 2 lists the values of each of the terms of the partition, their component values, degrees of freedom and permutation test Monte Carlo P values based on 10,000 simulations. If we take into consideration the ordered structure of the rows, columns and tubes, then by using the partition of (2.1), $X^2_{IJ} = 235.3$ and $X^2_{IK} = 41.14$ and both have zero Monte Carlo P values, while the P value of $X^2_{JK} = 25.82$ is 0.0008. Therefore all three bivariate

TABLE 3
Log-linear MLE and approximated parameter values of Table 1

	β_{IJ}	β_{IK}	β_{JK}
LLM Estimates	−0.3468	−0.2188	0.0726
Approximation	−0.3298	−0.2140	0.0725

terms are highly significant. The Monte Carlo P value of X^2_{IJK} is 0.3426 which is not significant. Thus the three-way association does not contribute all the variation present in the table. The interaction between the rows (number of years of completed schooling) and columns (number of siblings), X^2_{IJ}, is the most significant accounting for 72% of the total variation in Table 1. Clogg (1982) also concluded this association was the most significant, although for his analysis it contributed up to 77% of the total variation. Clogg's three factor interaction term was 24.30.

Table 2 shows the important row, column and tube components for each term of (2.1). When the tubes are summed across the rows and columns, the column location component value of 222.2234 has a Monte Carlo P value of zero. Therefore, the difference in Number of siblings is due to the difference at each level across Years of completed schooling when the happiness of the people is not of interest. In fact this column component accounts for 94.44% of the total row variation in X^2_{IJ}. The most important row component for the term X^2_{IJ}, location, with a value of 209.9878, also has a Monte Carlo P value of zero. Therefore, the difference in the levels of Years of completed schooling is due to the difference in the levels' mean values and accounts for 89.24% of the total variation in the rows for the term X^2_{IJ}.

When the columns are summed across, the location component is dominant for the tubes, and contributes to 69.99% of the total tube variation in the term X^2_{IK}. Similarly, for this term, the variation in Years of completed schooling levels can be explained by the difference in the mean schooling levels, as the row location component of 31.6954 is more dominant than other components with a zero Monte Carlo P value.

When the rows are summed across, the variation in the levels of Happiness can be best explained by the dispersion of the levels across the Number of siblings levels, as the tube dispersion component of 17.0634 has the only significant Monte Carlo P value of 0.0009 and accounts for 71.11% of the total tube variation. When the rows are of no interest, the variation in the Number of siblings levels seems to be caused by the difference in their mean values, as the tube location component has Monte Carlo P value of 0.0001 — highly significant.

Using all three variables, we can determine the cause of bivariate associations. Table 2 shows that while the three-way term X^2_{IJK} is not significant, the relationship between the Years of Schooling and Happiness levels is affected by the dispersion in the Number of Siblings levels, as the Monte Carlo P value is 0.0574 — almost significant. No other component values significantly affect any other bivariate relationships.

The MLE parameter estimations of the log-linear model (5.1) compare very well with the approximations of (5.3)–(5.6). As the three-way chi-squared term is not significant, it is safe to say that the three-way term of (5.5) need not be calculated. Table 3 gives the bivariate MLE parameter values from the model (5.1) after 500 iterations, and their approximations of (5.3), (5.4) and (5.5).

Therefore, instead of selecting a log-linear model by trial and error as is the situation when fitting, testing and refitting a model, the approximations offered by equations (5.3)–(5.6) give a far quicker and relatively accurate method of parameter estimation. Also a far more informative model than that of Agresti (1994) or Fienberg (1977) can be obtained by selecting significant parameters which involve moments above the location, as model (5.7) does.

The linear-by-linear association for the row-by-column interaction (Z_{110}) is highly significant with a value of -14.42. Hence, those with many siblings tend to finish school earlier than those with few siblings. Clogg (1982) reached the same conclusion, showing that this particular association is 'negative overall'. The row-by-tube interaction is also dominated by the linear-by-linear association (Z_{101}) with a significant value of -4.93. Thus, those who are very happy tend to be those with fewer years of completed schooling. The linear-by-quadratic association for the column-by-tube interaction (Z_{012}) is also significant with a value of 3.47. Therefore, as the number of siblings increases, happiness tends to decrease, then increase. Thus those with a few siblings and those with a lot of siblings tend to be happier than those with a moderate number (4 to 5) of siblings.

Appendix

Proof of (2.1)–(2.2). To prove the partition of X^2 described by (2.1)–(2.2) above, consider the classical Pearson chi-squared statistic:

$$X^2 = n\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{(p_{ijk}-p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k})^2}{p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}} = n\left\{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\frac{p_{ijk}^2}{p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}}\right)-1\right\}. \quad \text{(A.1)}$$

The row, column and tube orthogonal polynomials have the following property:

$$\sum_{u=0}^{I-1}a_u(i)a_u(i') = \begin{cases}1/p_{i\cdot\cdot} & \text{for } i=i',\\ 0 & \text{otherwise;}\end{cases} \quad \text{(A.2a)}$$

$$\sum_{v=0}^{J-1}b_v(j)b_v(j') = \begin{cases}1/p_{\cdot j\cdot} & \text{for } j=j',\\ 0 & \text{otherwise;}\end{cases} \quad \text{(A.2b)}$$

$$\sum_{w=0}^{K-1}c_w(k)c_w(k') = \begin{cases}1/p_{\cdot\cdot k} & \text{for } k=k',\\ 0 & \text{otherwise.}\end{cases} \quad \text{(A.2c)}$$

Therefore,

$$\frac{p_{ijk}^2}{p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}} = \sum_{u=0}^{I-1}\sum_{v=0}^{J-1}\sum_{w=0}^{K-1}a_u(i)b_v(j)c_w(k)p_{ijk}^2$$
$$= \sum_{i'=1}^{I}\sum_{j'=1}^{J}\sum_{k'=1}^{K}\sum_{u=0}^{I-1}\sum_{v=0}^{J-1}\sum_{w=0}^{K-1}a_u(i)a_u(i')b_v(j)b_v(j')c_w(k)c_w(k')p_{ijk}p_{i'j'k'}\,.$$

Hence, $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\frac{p_{ijk}^2}{p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}}\right)$ equals $\sum_{i=1}^{I}\sum_{i'=1}^{I}\sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{u=0}^{I-1}\sum_{v=0}^{J-1}\sum_{w=0}^{K-1}a_u(i)a_u(i')\times$ $b_v(j)b_v(j')c_w(k)c_w(k')p_{ijk}p_{i'j'k'} = \frac{1}{n}\sum_{u=0}^{I-1}\sum_{v=0}^{J-1}\sum_{w=0}^{K-1}Z_{uvw}^2$, where

$$Z_{uvw} = \sqrt{n}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}a_u(i)b_v(j)c_w(k)p_{ijk}\,,$$

which is (2.2) above.

Therefore, Pearson's chi-squared statistic becomes:

$$X^2 = \sum_{u=0}^{I-1} \sum_{v=0}^{J-1} \sum_{w=0}^{K-1} Z_{uvw}^2 - n \,. \tag{A.3}$$

But

$$Z_{000} = \sqrt{n} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} p_{ijk} = \sqrt{n}$$

and

$$Z_{u00} = \sqrt{n} \sum_{i=1}^{I} a_u(i)\, p_{i\cdot\cdot} = 0 \,.$$

Similarly, $Z_{0v0} = Z_{00w} = 0$.

So by expanding (A.3), the classical Pearson chi-squared statistic simplifies:

$$\begin{aligned} X^2 &= \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} Z_{uvw}^2 + \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} Z_{uv0}^2 + \sum_{u=1}^{I-1} \sum_{w=1}^{K-1} Z_{u0w}^2 + \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} Z_{0vw}^2 + (\sqrt{n})^2 - n \\ &= \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} Z_{uvw}^2 + \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} Z_{uv0}^2 + \sum_{u=1}^{I-1} \sum_{w=1}^{K-1} Z_{u0w}^2 + \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} Z_{0vw}^2 \,, \end{aligned}$$

thereby providing the partition of (2.1).

The chi-squared partitions of Rayner & Best (1996) and Best & Rayner (1996), which consider singly and doubly ordered two-way contingency tables respectively, can be proved in a manner similar to the partition of (2.1)–(2.2).

References

AGRESTI, A. (1994). *Analysis of Ordinal Categorical Data.* New York: Wiley.

BEH, E.J. (1997). Simple correspondence analysis of ordinal cross-classifications using orthogonal polynomials. *Biometrical J.* **39**, 1–25.

— (1998). A comparative study of scores for correspondence analysis with ordered categories. *Biometrical J.* (to appear).

— & DAVY, P.J. (1997). Multiple correspondence analysis of ordinal multi-way contingency tables using orthogonal polynomials. (In preparation.)

— & — (1999). Partitioning Pearson's chi-squared statistic for an ordered three-way contingency table: Part 2. *Austral. & New Zealand J. Statist.* (to appear).

BEST, E.J. & RAYNER, J.C.W. (1996). Nonparametric analysis for doubly ordered two-way contingency tables. *Biometrics* **52**, 1153–1156.

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

DANAHER, P.J. (1991). Canonical expansion model for multivariate media exposure distributions: a generalisation of the 'Duplication of Viewing Law'. *J. Marketing Res.* **28**, 361–367.

DARROCH, J.N., LAURITZEN, S.L. & SPEED, T.P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8**, 539–552.

DAVIS, J.A. (1977). *Codebook for the 1977 General Social Survey.* Chicago: National Opinion Research Centre.

EVERITT, B.S. (1992). *The Analysis of Contingency Tables.* London: Chapman & Hall.

FIENBERG, S.E. (1977). *The Analysis of Cross-Classified Categorical Data.* Cambridge: MIT Press.

GILULA, Z. & HABERMAN, S.J. (1988). The analysis of multivariate contingency tables by restricted canonical and restricted association models. *J. Amer. Statist. Assoc.* **83**, 760–771.

—, KRIEGER, A.M. & RITOV, Y. (1988). Ordinal association in contingency tables: some interpretive aspects. *J. Amer. Statist. Assoc.* **83**, 540–545.

GOODMAN, L.A. (1970). The multivariate analysis of quantitative data: interactions among multiple classifications. *J. Amer. Statist. Assoc.* **65**, 226–256.

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

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

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

— (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *Internat. Statist. Rev.* **54**, 243–309.

— (1996). A single general method for the analysis of cross-classified data: reconciliation and synthesis of some methods of Pearson, Yule and Fisher, and also some methods of correspondence analysis and association analysis. *J. Amer. Statist. Assoc.* **91**, 408–428.

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

HIROTSU, C. (1978). Ordered alternatives for interaction effects. *Biometrika* **65**, 561–570.

— (1982). Use of cumulative efficient scores for testing ordered alternatives in discrete models. *Biometrika* **69**, 567–577.

— (1983). Defining the pattern of association in two-way contingency tables. *Biometrika* **70**, 579–589.

— (1986). Cumulative chi-square statistic as a tool for testing goodness of fit. *Biometrika* **73**, 165–173.

KENDALL, M.G. & STUART, A. (1979). *The Advanced Theory of Statistics*, Vol.2. London: Charles Griffin.

KROONENBERG, P.M. (1989). Singular value decomposition of interactions in three-way contingency tables. In *Multiway Data Analysis*, eds R. Coppi & S. Bolasco, 169–184. Amsterdam: North Holland.

LANCASTER, H.O. (1951). Complex contingency tables treated by the partition of χ^2. *J. Roy. Statist. Soc. Ser. B* **13**, 242–249.

— (1953). A reconstitution of χ^2 considered from metrical and enumerative aspects. *Sankhyā* **13**, 1–10.

— (1980). Orthogonal models for contingency tables. In *Development in Statistics*, Vol.3, ed. P. R. Krishnaiah, 99–157. New York: Academic Press.

NAIR, V.N. (1986). Testing industrial experiments with ordered categorical data. *Technometrics* **28**, 283–311.

RAYNER, J.C.W. & BEST, D.J. (1996). Smooth extensions of Pearson's product moment correlation and Spearman's rho. *Statist. Probab. Lett.* **30**, 171–177.

ROM, D. & SARKAR, S.K. (1992). A generalised model for the analysis of association in ordinal contingency tables. *J. Statist. Plann. Inference* **33**, 205–212.

van der HEIJDEN, P.G.M. & de LEEUW, J. (1985). Correspondence analysis used complimentary to log-linear analysis. *Psychometrika* **50**, 429–447.

— & WORSLEY, K.J. (1988). Comment on 'Correspondence analysis used complimentary to log-linear analysis'. *Psychometrika* **53**, 287–291.

—, de FALGUEROLLES, A. & de LEEUW, J. (1989). A combined approach to contingency table analysis using correspondence analysis and log-linear analysis. *Appl. Statist.* **38**, 249–292.