Partitions of Pearson's Chi-square statistic for frequency tables

Citation for published version:
Loisel, S & Takane, Y 2015, 'Partitions of Pearson's Chi-square statistic for frequency tables: a comprehensive account', *Computational Statistics*. https://doi.org/10.1007/s00180-015-0619-1

Digital Object Identifier (DOI):
10.1007/s00180-015-0619-1

Published In:
Computational Statistics

Computational Statistics manuscript No.
(will be inserted by the editor)

Partitions of Pearson's chi-square statistic for frequency tables: A comprehensive account

Sébastien Loisel · Yoshio Takane

Received: date / Accepted: date

Abstract Pearson's chi-square statistic for frequency tables depends on what is hypothesized as the expected frequencies. Its partitions also depend on the hypothesis. Lancaster (1951) proposed ANOVA-like partitions of Pearson's statistic under several representative hypotheses about the expected frequencies. His expositions were, however, not entirely clear. In this paper, we clarify his method of derivations, and extend it to more general situations. A comparison is made with analogous decompositions of the log likelihood ratio (LR) statistic associated with log-linear analysis of contingency tables.

Keywords One-way tables · Two-way tables · Three-way tables · Orthogonal transformations · Metric matrix · Helmert-like contrasts · ANOVA-like partitions · Likelihood ratio (LR) statistic

Work reported in this has been supported by NSERC discovery grant to the second author.

Sébastien Loisel
Department of Mathematics
Heriot-Watt University
Edinburgh, EH14 4AS UK
Tel.: +44 181 451 3234
Fax: +44 131 451 3249
E-mail: sloisel@gmail.com

Yoshio Takane
Department of Psychology
University of Victoria
5173 Del Monte Ave., Victoria,
BC, V8Y 1X3, Canada
Tel.: 250 744 0076
Fax: 250 744 0076
email: yoshio.takane@mcgill.ca

1 Introduction

Pearson's chi-square statistic (Pearson 1900) measures an overall discrepancy between a set of observed frequencies and expected frequencies. Specifically, it is defined as

$$CS = \sum_{a=1}^{A} \frac{(f_a - \mathrm{E}(f_a))^2}{\mathrm{E}(f_a)}, \tag{1}$$

where f_a indicates the observed frequency of event (cell) a ($a = 1, \cdots, A$), and $\mathrm{E}(f_a)$ is the expected frequency of the corresponding event (cell) under some hypothesis. This statistic thus depends on the hypothesis from which the expected frequencies are derived. Lancaster (1951) proposed ANOVA-like partitions of Pearson's statistic under several representative hypotheses about the expected frequencies. When the CS above indicates a significant departure of observed data from the hypotheses, these partitions are useful in identifying what aspects are responsible for the departure. This is similar to multiple comparisons in ANOVA. Lancaster's (1951) expositions were, however, not entirely transparent. In this paper, we clarify his expositions, and extend his derivations to more general situations. We also compare the proposed partitions with those of the likelihood ratio (LR) statistic in analogous situations.

We first deal with one-way tables, through which we lay out the basic methodological tools for partitioning Pearson's chi-square statistic. Specifically, we introduce transformations of the constituent terms in the definition of the statistic. The transformations also serve as a basis for partitioning the statistic. We first discuss general properties of the transformations (Section 2.1), and then give special cases in which partitions are empirically better motivated (Section 2.2). In Section 3, we apply the same strategy to two-way tables. This amounts to reducing two-way tables to one by vectorising the original tables using vec operations and Kronecker products. In Section 4, we further extend the methodology to three-way and higher order designs. A variety of hypotheses are possible about the expected frequencies in higher order tables, which affect the partitioning of Pearson's statistic. We discuss two of the most commonly tested hypotheses in three-way contingency tables. One is complete independence in which all three factors are independent (Section 4.1), and the other is partial independence in which some but not all factors are independent, e.g., one factor is independent from the other two (Section 4.2). We then give a numerical example to demonstrate the partitions derived under these hypotheses (Section 4.3). In Section 5, we compare the proposed partitions with those of the likelihood ratio (LR) statistic in analogous situations.

2 One-way tables

In this section, we present a basic strategy for partitioning Pearson's chi-square statistic for one-way tables, which will be used as a building block for higher order tables. This strategy involves transformations of the terms in the definition of Pearson's statistic. We first discuss general properties of the transformations, and then their special cases, such as Irwin's (1949) construction of Helmert-type contrasts, for more empirically motivated partitions.

2.1 A general strategy for partitioning

Suppose there are A mutually exclusive events. Let p_a denote the probability of event a ($a = 1, \cdots, A$) in each trial. Suppose further that N such trials are independently replicated, where N is fixed. Let f_a denote the observed frequency of event a. A collection of f_a's ($a = 1, \cdots, A$) constitute a one-way frequency table. The generic form of Pearson's chi-square statistic given in (1) can be rewritten as

$$CS = \sum_{a=1}^{A} \frac{(f_a - Np_a)^2}{Np_a} = N \sum_{a=1}^{A} \frac{(\hat{p}_a - p_a)^2}{p_a}, \tag{2}$$

where $\hat{p}_a = f_a/N$. This statistic asymptotically follows the chi-square distribution with $A-1$ degrees of freedom (df) under the hypothesis that the p_a's are known completely. This is written as $CS \rightsquigarrow \chi^2_{A-1}$, where "$\rightsquigarrow$" denotes "asymptotically follows." The df are $A-1$ because N is fixed, so once $A-1$ frequencies are observed, the remaining one is known. The most commonly used hypothesis about the p_a's is that they are homogenous across all cells, that is, $p_a = 1/A$ for $a = 1, \cdots, A$, although other prescribed values are also permissible.

Let $\mathbf{x}$ represent the vector whose a-th element is given by

$$x_a = \sqrt{N}\frac{(\hat{p}_a - p_a)}{\sqrt{p_a}}. \tag{3}$$

Then,

$$\mathbf{x} = \sqrt{N}\mathbf{P}_A^{-1/2}(\hat{\mathbf{p}} - \mathbf{P}_A\mathbf{1}_A), \tag{4}$$

where $\hat{\mathbf{p}}$ is the vector of $\hat{p}_a$'s, $\mathbf{P}_A$ is the diagonal matrix with p_a as the a-th diagonal element, and $\mathbf{1}_A$ is the A-element vector of ones. We may rewrite (2) as

$$CS = \mathbf{x}'\mathbf{x} = N(\hat{\mathbf{p}}'\mathbf{P}_A^{-1}\hat{\mathbf{p}} - 1). \tag{5}$$

We show that $CS \rightsquigarrow \chi^2_{A-1}$. Note first that

$$\hat{\mathbf{p}} \rightsquigarrow \mathcal{N}(\mathbf{P}_A\mathbf{1}_A, (\mathbf{P}_A - \mathbf{P}_A\mathbf{1}_A\mathbf{1}_A'\mathbf{P}_A)/N), \tag{6}$$

where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal distribution with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. Since $N\mathbf{P}_A^{-1}$ is a g-inverse of $(\mathbf{P}_A - \mathbf{P}_A\mathbf{1}_A\mathbf{1}_A'\mathbf{P}_A)/N$, it follows that

$$\mathbf{x} \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{Q}_{P_A^{1/2}\mathbf{1}_A}), \tag{7}$$

where

$$\mathbf{Q}_{P_A^{1/2}\mathbf{1}_A} = \mathbf{I}_A - \mathbf{P}_A^{1/2}\mathbf{1}_A\mathbf{1}_A'\mathbf{P}_A^{1/2} \tag{8}$$

is the orthogonal projector onto the space orthogonal to $\mathrm{Sp}(\mathbf{P}_A^{1/2}\mathbf{1}_A)$ (where $\mathrm{Sp}(\mathbf{Z})$ indicates the space spanned by the column vectors of $\mathbf{Z}$). We further transform $\mathbf{x}$ into a vector whose elements are asymptotically independent standard normal variables. Let $\mathbf{T}_A$ denote an A by $A-1$ semi-orthogonal matrix, i.e.,

$$\mathbf{T}_A'\mathbf{T}_A = \mathbf{I}_{A-1}, \tag{9}$$

and such that it is also orthogonal to $\mathbf{P}_A^{1/2}\mathbf{1}_A$, i.e.,

$$\mathbf{T}_A'\mathbf{P}_A^{1/2}\mathbf{1}_A = \mathbf{0}_{A-1}. \tag{10}$$

Then,

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_{A-1} \end{pmatrix} \equiv \mathbf{T}_A'\mathbf{x} \rightsquigarrow \mathcal{N}(0, \mathbf{I}_{A-1}), \tag{11}$$

so that for $a = 1, \cdots, A-1$,

$$y_a^2 = (\mathbf{x}'\mathbf{t}_a)^2 \rightsquigarrow \chi_1^2 \tag{12}$$

independently from other a's, where $\mathbf{t}_a$ is the a-th column of $\mathbf{T}_A$, and

$$CS = \mathbf{y}'\mathbf{y} = \sum_{a=1}^{A-1} y_a^2 = \mathbf{x}'\mathbf{T}_A\mathbf{T}_A'\mathbf{x} \rightsquigarrow \chi_{A-1}^2. \tag{13}$$

Let

$$\tilde{\mathbf{T}}_A = [\mathbf{P}_A^{1/2}\mathbf{1}_A, \mathbf{T}_A] = [\mathbf{t}_0, \mathbf{T}_A]. \tag{14}$$

This matrix is fully orthogonal (that is, $\tilde{\mathbf{T}}_A'\tilde{\mathbf{T}}_A = \tilde{\mathbf{T}}_A\tilde{\mathbf{T}}_A' = \mathbf{I}_A$). Since $\mathbf{1}_A'\mathbf{P}_A^{1/2}\mathbf{x}$ $= \mathbf{t}_0'\mathbf{x} = 0$ (which implies $\mathbf{P}_A^{1/2}\mathbf{1}_A$ is orthogonal to $\mathbf{x}$), we have $\tilde{\mathbf{T}}_A'\mathbf{x} = \begin{pmatrix} 0 \\ \mathbf{y} \end{pmatrix} \equiv$ $\tilde{\mathbf{y}}$, so that

$$CS = \mathbf{y}'\mathbf{y} = \tilde{\mathbf{y}}'\tilde{\mathbf{y}} = \mathbf{x}'\tilde{\mathbf{T}}_A\tilde{\mathbf{T}}_A'\mathbf{x} = \mathbf{x}'\mathbf{x} \rightsquigarrow \chi_{A-1}^2, \tag{15}$$

showing that $\mathbf{x}'\mathbf{x}$ indeed asymptotically follows the chi-square distribution with $A-1$ df. It should be observed that (12) and (13) also indicate a possible partition of χ_{A-1}^2.

Note that (5) indicates that CS is equal to the squared length of the vector $\mathbf{x}$ (i.e, $CS = \mathbf{x}'\mathbf{x} = ||\mathbf{x}||^2$). The elements of $\mathbf{x}$ are, however, not asymptotically independent or standard normal. An orthogonal transformation $\tilde{\mathbf{T}}_A$ given in (14) transforms the elements of $\mathbf{x}$ into an identically zero variable (this occurs because $\mathbf{t}_0 = \mathbf{P}_A^{1/2}\mathbf{1}_A$ is orthogonal to $\mathbf{x}$, the effect of $\mathbf{t}_0$ being *a priori* eliminated from $\mathbf{x}$) plus $A-1$ asymptotically standard normal variables y_a ($a = 1, \cdots, A-1$), while preserving the total length of $\mathbf{x}$ (i.e., $CS = ||\mathbf{x}||^2 = ||\tilde{\mathbf{y}}||^2$. Since the squares of standard normal variables follows the chi-square distribution with 1 df, and the sum of independent chi-square variables follows the chi-square distribution with df equal to the sum of df in constituent chi-square variables, CS asymptotically follows the chi-square distribution with $A-1$ df. Essentially the same strategy will be used repeatedly for higher order tables.

The matrix $\mathbf{T}_A$ above can be any A by $A-1$ matrix satisfying (9) and (10). Such a $\mathbf{T}_A$ can be easily obtained by the singular value decomposition (SVD) (or the eigen-decomposition) of $\mathbf{Q}_{P_A^{1/2}\mathbf{1}_A}$, namely

$$\mathbf{Q}_{P_A^{1/2}\mathbf{1}_A} = \mathbf{T}_A\mathbf{T}'_A. \tag{16}$$

The matrix $\mathbf{Q}_{P_A^{1/2}\mathbf{1}_A}$ has $A-1$ unit singular values and one zero singular value, and $\mathbf{T}_A$ represents the matrix of singular vectors corresponding to the unit singular values. For $A > 2$, this matrix is not unique. The vector $\mathbf{P}_A^{1/2}\mathbf{1}_A$, on the other hand, is the singular vector corresponding to the unique zero singular value. This vector appended to $\mathbf{T}_A$ (i.e., $\tilde{\mathbf{T}}_A$) comprises a complete set of orthogonal basis vectors spanning the Euclidean space of dimensionality A.

Note that (12) and (13) also imply possible partitioning of CS, depending on a choice of $\mathbf{T}_A$. As has been suggested above, there are infinitely many $\mathbf{T}_A$'s that satisfy (9) and (10) for $A > 2$. If $\mathbf{T}_A$ is derived by the procedure described above, each element of $\mathbf{y}$ is not likely to be empirically meaningful. While this may be satisfactory under some circumstances (e.g., if the sole purpose of the transformation is to obtain any asymptotically independent standard normal variables), more empirically meaningful components are desirable in other situations (e.g., when we are interested in the empirical significance of each component), particularly in one-way tables. (Unless the terms in the partition are meaningful, there is no point in partitioning chi-square in one-way tables.) The columns of $\mathbf{T}_A$ may be chosen to reflect such empirical interests, to which we now turn.

2.2 Helmert-like contrasts and other empirically motivated contrasts

We begin with Irwin's construction of Helmert-like contrasts. (We call them "Helmert-like" contrasts because they are Helmert contrasts orthogonal under unequal cell sizes.) While they may not exactly represent the contrasts of interest, they may provide many hints for constructing orthogonal contrasts that better reflect one' empirical interest. Each Helmert contrast has a specific meaning and orthogonality among the contrasts is assured. They are given explicitly in terms of hypothesized cell probabilities, so that they can be used no matter what these probabilities are by just plugging in specific values of cell probabilities. The matrices of Helmert contrasts depend on the number of cells in the table. When we need to explicitly indicate the number of cells, we put a parenthesized superscript on $\mathbf{T}_A$, as in $\mathbf{T}_A^{(j)}$, where j (≥ 2) indicates the number of cells. The subscript A on $\mathbf{T}$, on the other hand, is used as a generic name for a factor (to be explained) which could differ from the number of cells. In what follows, we give examples of matrices of Helmert-like contrasts for the first few values of j.

When there are only two cells ($j = 2$) in the table, we have

$$\tilde{\mathbf{T}}_A^{(2)} = [\mathbf{P}_A^{1/2}\mathbf{1}_2, \mathbf{T}_A^{(2)}] = \begin{bmatrix} \sqrt{p_1} & \sqrt{p_2} \\ \sqrt{p_2} & -\sqrt{p_1} \end{bmatrix}, \tag{17}$$

where the $\mathbf{p}_a$'s ($a = 1, 2$) are the diagonal elements of $\mathbf{P}_A$. The second column of $\tilde{\mathbf{T}}_A^{(2)}$ (or the 2 by 1 vector $\mathbf{T}_A^{(2)}$) takes the difference between the two cells in the table. For $j = 2$, this matrix is unique up to reflections of its columns.

When there are three cells ($j = 3$) in the table, we define

$$\tilde{\mathbf{T}}_A^{(3)} = [\mathbf{P}_A^{1/2}\mathbf{1}_3, \mathbf{T}_A^{(3)}] = \begin{bmatrix} \sqrt{p_1} & \sqrt{p_2/u_{12}} & \sqrt{p_1p_3/u_{12}} \\ \sqrt{p_2} & -\sqrt{p_1/u_{12}} & \sqrt{p_2p_3/u_{12}} \\ \sqrt{p_3} & 0 & -\sqrt{u_{12}} \end{bmatrix}, \tag{18}$$

where the p_a's ($a = 1, 2, 3$) are the diagonal elements of $\mathbf{P}_A$, and $u_{12} = p_1 + p_2$. The second column of $\tilde{\mathbf{T}}_A^{(3)}$ in (18) contrasts the first two cells of the table, while the third column contrasts a weighted average of the first two cells against the third. The $\tilde{\mathbf{T}}_A^{(3)}$ defined above reduces to $\tilde{\mathbf{T}}_A^{(2)}$ when $j = 2$ because $u_{12} = 1$, and there is no p_3.

For $j = 4$, we define

$$\begin{aligned} \tilde{\mathbf{T}}_A^{(4)} &= [\mathbf{P}_A^{1/2}\mathbf{1}_4, \mathbf{T}_A^{(4)}] \\ &= \begin{bmatrix} \sqrt{p_1} & \sqrt{p_2/u_{12}} & \sqrt{p_1p_3/u_{12}u_{123}} & \sqrt{p_1p_4/u_{123}} \\ \sqrt{p_2} & -\sqrt{p_1/u_{12}} & \sqrt{p_2p_3/u_{12}u_{123}} & \sqrt{p_2p_4/u_{123}} \\ \sqrt{p_3} & 0 & -\sqrt{u_{12}/u_{123}} & \sqrt{p_3p_4/u_{123}} \\ \sqrt{p_4} & 0 & 0 & -\sqrt{u_{123}} \end{bmatrix}, \end{aligned} \tag{19}$$

where the p_a's ($a = 1, \cdots, 4$) are the diagonal entries of $\mathbf{P}_A$, $u_{12} = p_1 + p_2$ as before, and $u_{123} = p_1 + p_2 + p_3$. The second column of $\tilde{\mathbf{T}}_A^{(4)}$ contrasts the first two cells, and the third column contrasts an average of the first two cells versus the third cell, while the fourth column contrasts an average of the first three cells against the fourth cell. The matrix $\tilde{\mathbf{T}}_A^{(4)}$ given above reduces to $\tilde{\mathbf{T}}_A^{(3)}$ when there are only three cells, since $u_{123} = 1$, and there is no p_4. The matrix of Helmert-like contrasts for $j > 4$ can be derived, following the instructions given by Irwin (1949).

Note that the first column of the $\tilde{\mathbf{T}}_A$'s given above pertains to the effect of mean probability (frequency) regardless of the number of cells. This corresponds to $\mathbf{t}_0$ in (14), and as has been noted, its effect is always *a priori* eliminated. This is analogous to the ANOVA situation in which only differences in group means are of interest, but not the grand mean. It has no significant role in the analysis of one-way tables. This column, however, plays an important role in defining non-highest order interaction effects when we deal with higher order tables.

When $j > 2$, the Helmert contrasts are not the only possible candidates for $\mathbf{T}_A$, whose columns are also meaningful. For example, we may choose

$$\tilde{\mathbf{T}}_A^{(4)} = \begin{bmatrix} \sqrt{p_1} & \sqrt{p_2/u_{12}} & 0 & \sqrt{p_1 u_{34}/u_{12}} \\ \sqrt{p_2} & -\sqrt{p_1/u_{12}} & 0 & \sqrt{p_2 u_{34}/u_{12}} \\ \sqrt{p_3} & 0 & \sqrt{p_4/u_{34}} & -\sqrt{p_3 u_{12}/u_{34}} \\ \sqrt{p_4} & 0 & -\sqrt{p_3/u_{34}} & -\sqrt{p_4 u_{12}/u_{34}} \end{bmatrix}, \tag{20}$$

where $u_{34} = p_3 + p_4$. The first two columns of this matrix remain the same as in (19). The third column contrasts between the last two cells, while the fourth column contrasts between a weighted average of the first two cells and that of the last two cells. This leads to a different partition of CS from the one that results from (19). In general, specifications of contrast vectors are quite flexible, and other empirically meaningful contrasts may also be specified.

The contrasts we have encountered so far may be classified into a small number of groups. One type, like the second column in (19), takes a difference between two cells ignoring all others. Another type, like the third column in (19), takes a difference between an average of the two cells whose difference has been taken already (by the second column in (19)) and a third cell. A third type of contrast, like the fourth column in (20), takes a difference between two averages each taken over two cells whose differences are taken already (by the second and third columns in (20)). By combining these three types of contrasts and with appropriate permutations of cells, we are able to construct most, if not all, of the common type of orthogonal contrasts.

If, however, one wants to find contrasts which do not fit to the above profiles, one may resort to a numerical mean, given specific values of cell probabilities. While this is theoretically less satisfactory, it has no real drawback in practice. The method to be presented below assumes that we have a set of orthogonal contrasts for equal cell probabilities. These contrasts are turned into orthogonal contrasts under unequal cell probabilities, which are usually the case in contingency table analysis. As an example, let us derive $\tilde{\mathbf{T}}_A^{(4)}$ in (19) above. Let $\mathbf{t}_0 = \mathbf{P}_A^{1/2}\mathbf{1}_A$, which will be the first column of $\tilde{\mathbf{T}}_A^{(4)}$, and let

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -2 & 1 \\ 0 & 0 & -3 \end{bmatrix} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]. \tag{21}$$

This is a matrix of Helmert contrasts for equal cell probabilities. The resultant matrix is denoted as $[\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3]$. We first derive $\mathbf{t}_1$ by orthogonalizing $\mathbf{v}_1$ to $\mathbf{t}_0$, while preserving zeros in $\mathbf{v}_1$. Let $\mathbf{w}_1$ denote the nonzero parts of $\mathbf{v}_1$, and let $\mathbf{Y}_1$ denote the vector of the corresponding elements of $\mathbf{t}_0$, that is,

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \text{ and } \mathbf{Y}_1 = \begin{pmatrix} \sqrt{p_1} \\ \sqrt{p_2} \end{pmatrix}.$$

Then, calculate

$$\mathbf{z} = \mathbf{Q}_{Y_1}\mathbf{w}_1, \tag{22}$$

where $\mathbf{Q}_{Y_1}$ is the orthogoal projector onto $\mathrm{Ker}(\mathbf{Y}_1')$, i.e., $\mathbf{Q}_{Y_1} = \mathbf{I} - \mathbf{Y}_1(\mathbf{Y}_1'\mathbf{Y}_1)^{-}\mathbf{Y}_1'$. We then normalize $\mathbf{z}$ by

$$\mathbf{t}_1^* = \mathbf{z}/||\mathbf{z}||, \tag{23}$$

if necessary. The vector $\mathbf{t}_1$ is obtained by appending appropriate numbers of zeros at the bottom of $\mathbf{t}_1^*$, namely $\mathbf{t}_1 = \begin{pmatrix} \mathbf{t}_1^* \\ \mathbf{0}_2 \end{pmatrix}$.

We next obtain $\mathbf{t}_2$ by orthogonalizing $\mathbf{v}_2$ to $[\mathbf{t}_0, \mathbf{t}_1]$. We define $\mathbf{w}_2$ to be the nonzero parts of $\mathbf{v}_3$, and $\mathbf{Y}_2$ to be the matrix of the corresponding elements of $\mathbf{t}_0$ and $\mathbf{t}_1$, namely

$$\mathbf{w}_2 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \text{ and } \mathbf{Y}_2 = \begin{bmatrix} \mathbf{Y}_1 & \mathbf{t}_1^* \\ \sqrt{p_3} & 0 \end{bmatrix}.$$

We then apply (22) and (23) with $\mathbf{w}_1$ and $\mathbf{Y}_1$ replaced by $\mathbf{w}_2$ and $\mathbf{Y}_2$, respectively, to obtain $\mathbf{t}_2^*$, from which $\mathbf{t}_2$ is obtained by appending a zero at the bottom. We finally obtain $\mathbf{t}_3$ by orthogonalizing $\mathbf{v}_3$ to $[\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2]$ in a similar way as before. We define

$$\mathbf{w}_3 = \mathbf{v}_3, \text{ and } \mathbf{Y}_3 = \begin{bmatrix} \mathbf{Y}_2 & \mathbf{t}_2^* \\ [\sqrt{p_4}, 0] & 0 \end{bmatrix}.$$

The vector $\mathbf{t}_3 = \mathbf{t}_3^*$ is calculated by (22) and (23) with $\mathbf{w}_1$ and $\mathbf{Y}_1$ replaced by $\mathbf{w}_3$ and $\mathbf{Y}_3$, respectively. The $\tilde{\mathbf{T}}_A^{(4)}$ given in (20) can also be derived in a similar way. This method of deriving a set of orthogonal contrasts under unequal cell probabilities works quite generally, so far as we have a set of orthogonal contrasts under equal cell probabilities.

In the discussion above, we have assumed that we are dealing with one-way tables. However, the basic strategy of transforming constituent terms in chi-squares into asymptotically independent components works for higher order tables as well, where the elementary transformations given above are used as building blocks. In such cases, what is referred to as "cells" above should be replaced by "levels of a factor."

3 Two-way tables

We extend the method in the previous section to two-way tables. We begin by rewriting the statistic for two-way tables using matrices and vectors. Let there be two discrete variables (factors), Factor A with A levels and Factor B with B levels, by which events are cross-classified. This gives rise to an A by B contingency table $\mathbf{F}$. Define

$$\hat{\mathbf{P}} = \mathbf{F}/N, \tag{24}$$

where N indicates the total frequency. As before, N is assumed fixed. Let

$$\mathbf{X} = \sqrt{N}\mathbf{P}_A^{-1/2}(\hat{\mathbf{P}} - \mathbf{P}_A\mathbf{1}_A\mathbf{1}_B'\mathbf{P}_B)\mathbf{P}_B^{-1/2}, \tag{25}$$

where $\mathbf{P}_A$ and $\mathbf{P}_B$ are the diagonal matrices of hypothesized marginal probabilities of the levels of Factor A and Factor B, respectively. The above definition assumes that the two factors are independent, that is, $\mathrm{E}[\hat{p}_{ab}] = p_{ab} = p_a p_b$, where p_{ab} is the joint probability of the a-th level of Factor A and the b-th level of Factor B, and p_a and p_b are the marginal probabilities of the a-th level of Factor A and the b-th level of Factor B, respectively. As has been emphasized earlier, partitions of Pearson's statistic depend on what is hypothesized for $\mathbf{P}_A$ and $\mathbf{P}_B$. The two most commonly employed hypotheses are:

Scenario 1. Diagonal elements of $\mathbf{P}_A$ and $\mathbf{P}_B$ are prescribed. Any prescribed numbers may be used, as long as they are positive and add up to unity within each factor. The most representative case is the one in which we set $\mathbf{P}_A = \mathbf{I}_A/A$ and $\mathbf{P}_B = \mathbf{I}_B/B$, which is called marginal homogeneity (under independence).

Scenario 2. Diagonal elements of $\mathbf{P}_A$ and $\mathbf{P}_B$ are estimated from the data, and are equal to the row and column totals of $\hat{\mathbf{P}}$, respectively. In this case, marginal probabilities are perfectly fitted.

Note that Scenario 2 is the standard assumption made in correspondence analysis (Greenacre 1984; Nishisato 1980) of two-way contingency tables. Note also that the cases in which some marginals are fixed also fall into this category. This may sound somewhat counterintuitive because marginal probabilities are completely known in such cases. However, using the known marginal probabilities is in effect equivalent to estimating them from the data. Marginal probabilities are perfectly fitted in either case. Combinations of the two scenarios above are also possible. For example, we may assume Scenario 1 for $\mathbf{P}_A$ and Scenario 2 for $\mathbf{P}_B$ (that is, $\mathbf{P}_A = \mathbf{I}_A/A$, but the diagonal elements of $\mathbf{P}_B$ are estimated from the data).

By vectorizing $\mathbf{X}$ above using vec operators and Kronecker products, we obtain

$$\mathbf{x} \equiv \mathrm{vec}(\mathbf{X}) = \sqrt{N}(\mathbf{P}_B \otimes \mathbf{P}_A)^{-1/2}(\hat{\mathbf{p}} - (\mathbf{P}_B \otimes \mathbf{P}_A)\mathbf{1}_{AB}), \tag{26}$$

where $\hat{\mathbf{p}} = \mathrm{vec}(\hat{\mathbf{P}})$, and $\mathbf{1}_{AB} = \mathbf{1}_B \otimes \mathbf{1}_A$ is the AB-component vector of ones. Note that (26) is essentially of the same form as (4) with $\mathbf{P}_A$ and $\mathbf{1}_A$ in the latter replaced by $\mathbf{P}_B \otimes \mathbf{P}_A$ and $\mathbf{1}_{AB}$, respectively. Pearson's statistic for two-way tables can now be written as

$$CS = \mathbf{x}'\mathbf{x} \tag{27}$$

analogously to (5). The df for the above chi-square depend on the hypotheses about $\mathbf{P}_A$ and $\mathbf{P}_B$. As will be shown shortly, the df are equal to $AB - 1$ under Scenario 1, and to $(A-1)(B-1)$ under Scenario 2. In either case, they are equal to AB minus the number of estimated (or fixed) quantities in the definition of the statistic.

We now transform $\mathbf{X}$ separately for rows and columns. Let $\tilde{\mathbf{T}}_A$ and $\tilde{\mathbf{T}}_B$ denote the transformation matrices of appropriate sizes for Factor A and Factor

B, as introduced in the previous section. Let

$$\begin{aligned}\tilde{\mathbf{Y}} &= \tilde{\mathbf{T}}'_A \mathbf{X} \tilde{\mathbf{T}}_B \\ &= \sqrt{N}\tilde{\mathbf{T}}'_A \mathbf{P}_A^{-1/2}(\hat{\mathbf{P}} - \mathbf{P}_A \mathbf{1}_A \mathbf{1}'_B \mathbf{P}_B)\mathbf{P}_B^{-1/2}\tilde{\mathbf{T}}_B.\end{aligned} \tag{28}$$

By vectoring $\tilde{\mathbf{Y}}$, we obtain

$$\begin{aligned}\tilde{\mathbf{y}} &= \text{vec}(\tilde{\mathbf{Y}}) \\ &= \sqrt{N}(\tilde{\mathbf{T}}'_B \mathbf{P}_B^{-1/2} \otimes \tilde{\mathbf{T}}'_A \mathbf{P}_A^{-1/2})(\hat{\mathbf{p}} - (\mathbf{P}_B \otimes \mathbf{P}_A)\mathbf{1}_{AB}) \\ &= \sqrt{N}(\tilde{\mathbf{T}}_B \otimes \tilde{\mathbf{T}}_A)'(\mathbf{P}_B \otimes \mathbf{P}_A)^{-1/2}(\hat{\mathbf{p}} - (\mathbf{P}_B \otimes \mathbf{P}_A)\mathbf{1}_{AB}) \\ &= \sqrt{N}\tilde{\mathbf{T}}'_{AB}\mathbf{x},\end{aligned} \tag{29}$$

where

$$\tilde{\mathbf{T}}_{AB} \equiv \tilde{\mathbf{T}}_B \otimes \tilde{\mathbf{T}}_A. \tag{30}$$

The matrix $\sqrt{N}(\tilde{\mathbf{T}}_B \otimes \tilde{\mathbf{T}}_A)$ is fully orthogonal, i.e., $\tilde{\mathbf{T}}'_{AB}\tilde{\mathbf{T}}_{AB} = \tilde{\mathbf{T}}_{AB}\tilde{\mathbf{T}}'_{AB} = \mathbf{I}$ (because the Kronecker product of fully orthogonal matrices is also fully orthogonal), so that

$$\hat{\mathbf{y}}'\hat{\mathbf{y}} = \mathbf{x}'\tilde{\mathbf{T}}_{AB}\tilde{\mathbf{T}}'_{AB}\mathbf{x} = \mathbf{x}'\mathbf{x} = CS. \tag{31}$$

Let $\tilde{\mathbf{T}}_{AB}$ be partitioned as

$$\tilde{\mathbf{T}}_{AB} = [\mathbf{t}_0, \mathbf{T}_{(A)}, \mathbf{T}_{(B)}, \mathbf{T}_{(AB)}], \tag{32}$$

where

$$\mathbf{t}_0 = \mathbf{P}_B^{1/2}\mathbf{1}_B \otimes \mathbf{P}_A^{1/2}\mathbf{1}_A, \tag{33}$$

$$\mathbf{T}_{(A)} = \mathbf{P}_B^{1/2}\mathbf{1}_B \otimes \mathbf{T}_A, \tag{34}$$

$$\mathbf{T}_{(B)} = \mathbf{T}_B \otimes \mathbf{P}_A^{1/2}\mathbf{1}_A, \tag{35}$$

and

$$\mathbf{T}_{(AB)} = \mathbf{T}_B \otimes \mathbf{T}_A. \tag{36}$$

Then, $CS(\text{A}) \equiv \mathbf{x}'\mathbf{T}_{(A)}\mathbf{T}'_{(A)}\mathbf{x}$ pertains to the chi-square representing the main effects of Factor A, $CS(\text{B}) \equiv \mathbf{x}'\mathbf{T}_{(B)}\mathbf{T}'_{(B)}\mathbf{x}$ pertains to the chi-square representing the main effects of Factor B, and $CS(\text{AB}) \equiv \mathbf{x}'\mathbf{T}_{(AB)}\mathbf{T}'_{(AB)}\mathbf{x}$ pertains to the chi-square representing the interaction effects between A and B. These CS's are called part chi-squares. The $(\mathbf{x}'\mathbf{t}_0)^2$ related to the effect of mean frequency, on the other hand, is identically equal to zero (being eliminated *a priori*). Note that

$$\begin{aligned}CS &= \tilde{\mathbf{y}}'\tilde{\mathbf{y}} = \mathbf{x}'\mathbf{T}_{AB}\mathbf{T}'_{AB}\mathbf{x} \\ &= \mathbf{x}'\mathbf{T}_{(A)}\mathbf{T}'_{(A)}\mathbf{x} + \mathbf{x}'\mathbf{T}_{(B)}\mathbf{T}'_{(B)}\mathbf{x} + \mathbf{x}'\mathbf{T}_{(AB)}\mathbf{T}'_{(AB)}\mathbf{x}, \\ &= CS(\text{A}) + CS(\text{B}) + CS(\text{AB}).\end{aligned} \tag{37}$$

The terms on the righthand side of the last two equations above defines a family-wise partition of CS.

The three terms in the partition above depend on the hypotheses regarding $\mathbf{P}_A$ and $\mathbf{P}_B$, as described earlier. Under Scenario 1 above, $CS(\mathrm{A}) \rightsquigarrow \chi^2_{A-1}$, $CS(\mathrm{B}) \rightsquigarrow \chi^2_{B-1}$, and $CS(\mathrm{AB}) \rightsquigarrow \chi^2_{(A-1)(B-1)}$ independently from each other. However, under Scenario 2, $CS(\mathrm{A}) = CS(\mathrm{B}) = 0$ (identically), while $CS(\mathrm{AB}) \rightsquigarrow \chi^2_{(A-1)(B-1)}$ as before. Note that while the asymptotic behavior of $CS(\mathrm{AB})$ remains the same under the two scenarios, its values calculated from a particular contingency table are likely to be different. This means that not only those that become zero, but also all other part chi-squares are affected by adopting different null hypotheses. It should be kept in mind that the tests of part chi-squares based on a single partition all assume the same null hypothesis under which the partition is derived.

Note that essentially the same applies to mixed scenario cases mentioned above. In the case in which the marginal homogeneity is assumed for $\mathbf{P}_A$ but $\mathbf{P}_B$ is estimated, for example, $CS(\mathrm{A}) \rightsquigarrow \chi^2_{A-1}$, but $CS(\mathrm{B}) = 0$, and $CS(\mathrm{AB}) \rightsquigarrow \chi^2_{(A-1)(B-1)}$. The $CS(\mathrm{A})$ and $CS(\mathrm{AB})$ have the same asymptotic properties as $CS(\mathrm{A})$ and $CS(\mathrm{AB})$ derived under Scenario 1. However, the values of these statistics calculated from a specific table are likely to be different from those obtained under Scenario 1.

Note also that, if desired, the part chi-squares in the above partitions, whenever they are associated with df larger than 1, can be further partitioned into asymptotically independent chi-square variables with fewer df. For example, let $\mathbf{t}_i$ be the i-th column vector of $\mathbf{T}_{(AB)}$. Then, $CS(\mathrm{AB}) = \sum_i (\mathbf{x}'\mathbf{t}_i)^2$, where $(\mathbf{x}'\mathbf{t}_i)^2 \rightsquigarrow \chi^2_1$ independently from each other.

4 Three-way and higher order tables

It is now rather straightforward to extend the method developed earlier to higher order tables, particularly when all factors are assumed independent (complete independence). For three-way tables, this means that $p_{abc} = p_a p_b p_c$ ($a = 1, \cdots, A$, $b = 1, \cdots, B$, and $c = 1, \cdots, C$), where p_{abc} is the joint probability of the a-th level of Factor A, the b-th level of Factor B, and the c-th level Factor C, and p_a, p_b, and p_c are their respective marginal probabilities. In three-way or higher order tables, however, it is not uncommon to hypothesize partial independence. In three-way tables, for example, we may hypothesize that one factor is independent from the other two (e.g., Factor A is independent from B an C, which in turn are not independent from each other), or that two factors are independent given the levels of a third factor. These two conditions (complete and partial independence) lead to quite distinct situations in partitioning Pearson's statistic, and we treat them separately in the following two subsections.

4.1 Complete independence

Let $\mathbf{P}_C$ represent the diagonal matrix of marginal probabilities of Factor C (i.e., p_c for $c = 1, \cdots, C$). This matrix is analogous to $\mathbf{P}_A$ and $\mathbf{P}_B$ introduced earlier. Then, under the three-way independence hypothesis, the diagonal matrix of hypothesized joint probabilities of cells in three-way tables can be expressed as

$$\mathbf{K} \equiv \mathbf{P}_C \otimes \mathbf{P}_B \otimes \mathbf{P}_A. \tag{38}$$

Let $\hat{\mathbf{p}}$ denote the vector of observed joint probabilities $\hat{p}_{abc}$ arranged in such a way that a is the fastest moving index, and c is the slowest moving index. In a 2 by 2 by 2 table, for example, this vector looks like $\hat{\mathbf{p}} = (\hat{p}_{111}, \hat{p}_{211}, \hat{p}_{121}, \hat{p}_{221}, \hat{p}_{112}, \hat{p}_{212}, \hat{p}_{122}, \hat{p}_{222})'$. Then, the total chi-square for the table is obtained by

$$CS = \mathbf{x}'\mathbf{x}, \tag{39}$$

where

$$\mathbf{x} = \sqrt{N}\mathbf{K}^{-1/2}(\hat{\mathbf{p}} - \mathbf{K}\mathbf{1}_{ABC}), \tag{40}$$

where $\mathbf{1}_{ABC} = \mathbf{1}_C \otimes \mathbf{1}_B \otimes \mathbf{1}_A$. These expressions are analogous to (5) and (27), and to (4) and (26), respectively.

As before, the df for the above chi-square depends on the hypotheses about $\mathbf{P}_A$, $\mathbf{P}_B$, and $\mathbf{P}_C$. The two most commonly used hypotheses are:

Scenario 1. Diagonal elements of $\mathbf{P}_A$, $\mathbf{P}_B$, and $\mathbf{P}_C$ are prescribed, the most representative case of which states $\mathbf{P}_A = \mathbf{I}_A/A$, $\mathbf{P}_B = \mathbf{I}_B/B$, and $\mathbf{P}_C = \mathbf{I}_C/C$ (Marginal homogeneity under independence).

Scenario 2. Diagonal elements of $\mathbf{P}_A$, $\mathbf{P}_B$, and $\mathbf{P}_C$ are estimated from the data. They are the observed marginal probabilities of the three factors.

Under Scenario 1, the df for the above chi-square is $ABC - 1$, while under Scenario 2, it is $ABC - A - B - C + 2$. As before, Scenario 2 includes the cases in which marginal probabilities are fixed for some but not all factors. There are a variety of combinations of the above two scenarios which we may call partial homogeneity cases, e.g., $\mathbf{P}_A = \mathbf{I}_A/A$ and $\mathbf{P}_B = \mathbf{I}_B/B$ as in Scenario 1, but $\mathbf{P}_C$ is estimated from the data as in Scenario 2, or $\mathbf{P}_A = \mathbf{I}_A/A$ as in Scenario 1, while $\mathbf{P}_B$ and $\mathbf{P}_C$ are estimated from the data as in Scenario 2, etc. See the bottom half of Figure 1 where all possible combinations are listed. The df in mixture cases vary between the above two extreme cases.

To partition the chi-square defined in (39) and (40), we transform $\mathbf{x}$ into asymptotically independent standard normal variables in a manner similar to before. That is,

$$\tilde{\mathbf{y}} = \tilde{\mathbf{T}}'_{ABC}\mathbf{x}, \tag{41}$$

where

$$\tilde{\mathbf{T}}_{ABC} = \tilde{\mathbf{T}}_C \otimes \tilde{\mathbf{T}}_B \otimes \tilde{\mathbf{T}}_A. \tag{42}$$

Here, the transformation matrices of appropriate sizes must be chosen for $\tilde{\mathbf{T}}_C$, $\tilde{\mathbf{T}}_B$, and $\tilde{\mathbf{T}}_A$. The matrix $\tilde{\mathbf{T}}_{ABC}$ is fully orthogonal, so that

$$\tilde{\mathbf{y}}'\tilde{\mathbf{y}} = \mathbf{x}\tilde{\mathbf{T}}_{ABC}\tilde{\mathbf{T}}'_{ABC}\mathbf{x} = \mathbf{x}'\mathbf{x} = CS. \tag{43}$$

Let $\tilde{\mathbf{T}}_{ABC}$ be partitioned as

$$\hat{\mathbf{T}}_{ABC} = [\mathbf{t}_0, \mathbf{T}_{(A)}, \mathbf{T}_{(B)}, \mathbf{T}_{(AB)}, \mathbf{T}_{(C)}, \mathbf{T}_{(AC)}, \mathbf{T}_{(BC)}, \mathbf{T}_{(ABC)}], \tag{44}$$

where $\mathbf{t}_0 = \mathbf{P}_C^{1/2}\mathbf{1}_C \otimes \mathbf{P}_B^{1/2}\mathbf{1}_B \otimes \mathbf{P}_A^{1/2}\mathbf{1}_A$, $\mathbf{T}_{(A)} = \mathbf{P}_C^{1/2}\mathbf{1}_C \otimes \mathbf{P}_B^{1/2}\mathbf{1}_B \otimes \mathbf{T}_A$, $\mathbf{T}_{(B)} = \mathbf{P}_C^{1/2}\mathbf{1}_C \otimes \mathbf{T}_B \otimes \mathbf{P}_A^{1/2}\mathbf{1}_A$, $\mathbf{T}_{(AB)} = \mathbf{P}_C^{1/2}\mathbf{1}_C \otimes \mathbf{T}_B \otimes \mathbf{T}_A$, $\mathbf{T}_{(C)} = \mathbf{T}_C \otimes \mathbf{P}_B^{1/2}\mathbf{1}_B \otimes \mathbf{P}_A^{1/2}\mathbf{1}_A$, $\mathbf{T}_{(AC)} = \mathbf{T}_C \otimes \mathbf{P}_B^{1/2}\mathbf{1}_B \otimes \mathbf{T}_A$, $\mathbf{T}_{(BC)} = \mathbf{T}_C \otimes \mathbf{T}_B \otimes \mathbf{P}_A^{1/2}\mathbf{1}_A$, and $\mathbf{T}_{(ABC)} = \mathbf{T}_C \otimes \mathbf{T}_B \otimes \mathbf{T}_A$. The vector $\mathbf{t}_0$ pertains to the effect of mean probability , $\mathbf{T}_{(A)}$, $\mathbf{T}_{(B)}$, and $\mathbf{T}_{(C)}$ to the A, B, and C main effects, $\mathbf{T}_{(AB)}$, $\mathbf{T}_{(AC)}$, and $\mathbf{T}_{(BC)}$ to the AB, AC, and BC two-way interaction effects, and $\mathbf{T}_{(ABC)}$ to the three-way interaction effects.

The part chi-square pertaining to the effect of mean frequency $(\mathbf{x}'\mathbf{t}_0)^2$ is identically equal to zero as before. (This effect is always *a priori* eliminated.) The other part chi-squares depend on the hypotheses about the marginal probabilities. Under Scenario 1,

$$\begin{aligned}
CS(\text{A}) &\equiv \mathbf{x}'\mathbf{T}_{(A)}\mathbf{T}'_{(A)}\mathbf{x} \rightsquigarrow \chi^2_{A-1}, \\
CS(\text{B}) &\equiv \mathbf{x}'\mathbf{T}_{(B)}\mathbf{T}'_{(B)}\mathbf{x} \rightsquigarrow \chi^2_{B-1}, \\
CS(\text{C}) &\equiv \mathbf{x}'\mathbf{T}_{(C)}\mathbf{T}'_{(C)}\mathbf{x} \rightsquigarrow \chi^2_{C-1}, \\
CS(\text{AB}) &\equiv \mathbf{x}'\mathbf{T}_{(AB)}\mathbf{T}'_{(AB)}\mathbf{x} \rightsquigarrow \chi^2_{(A-1)(B-1)}, \\
CS(\text{AC}) &\equiv \mathbf{x}'\mathbf{T}_{(AC)}\mathbf{T}'_{(AC)}\mathbf{x} \rightsquigarrow \chi^2_{(A-1)(C-1)}, \\
CS(\text{BC}) &\equiv \mathbf{x}'\mathbf{T}_{(BC)}\mathbf{T}'_{(BC)}\mathbf{x} \rightsquigarrow \chi^2_{(B-1)(C-1)},
\end{aligned}$$

and

$$CS(\text{ABC}) \equiv \mathbf{x}'\mathbf{T}_{(ABC)}\mathbf{T}'_{(ABC)}\mathbf{x} \rightsquigarrow \chi^2_{(A-1)(B-1)(C-1)}.$$

These statistics may be used to test the significance of the respective effects against the hypothesis assumed under Scenario 1. Under Scenario 2, on the other hand, the part chi-squares pertaining to the main effects are all eliminated and become zero, while the remaining effects have the same asymptotic distributional properties as their corresponding effects under Scenario 1. As in two-way tables, however, the values of these statistics calculated from an observed contingency table are likely to be different under the two scenarios. The factors for which marginal probabilities are estimated (or fixed) determines which main effects are eliminated and become zero. As before, any part chi-squares with df larger than 1 may be further partitioned into finer components.

Further extensions of the method to higher-order tables are now fairly apparent under the complete independence assumptions. We define $\mathbf{K} = \cdots \otimes \mathbf{P}_D \otimes \mathbf{P}_C \otimes \mathbf{P}_B \otimes \mathbf{P}_A$, and $\mathbf{T}_{ABCD\cdots} = \cdots \otimes \mathbf{T}_D \otimes \mathbf{T}_C \otimes \mathbf{T}_B \otimes \mathbf{T}_A$. We only need to be careful about arranging the elements of $\hat{\mathbf{p}}$ appropriately. Our convention has been to make the index of the newest factor the slowest moving index.

4.2 Partial independence

The situation becomes radically different when we assume only partial independence. There are two distinct cases subsumed under partial independence for three-way tables. One assumes that one factor is independent from the other two (one-factor independence), and the other assumes that two factors are independent given a third factor (conditional independence). See the top half of Figure 1. Both of these cases are weaker than complete independence discussed in the previous section. In contrast to the complete independence cases, part chi-squares in ANOVA-like partitions of Pearson's statistic are not mutually independent, and the order in which they are taken into account matters.

We begin with the former (one-factor independence). There are three possible subcases: 1. $p_{abc} = p_c p_{ab}$, 2. $p_{abc} = p_b p_{ac}$, and 3. $p_{abc} = p_a p_{bc}$, where p_{abc} is, as before, the joint probability of the a-th, b-th, and c-th levels of Factors A, B, and C, respectively, p's with two subscripts indicate the joint marginal probabilities indexed by the two subscripts, and p's with one subscript indicate the marginal probabilities indexed by the single subscript. We assume that these marginal probabilities are estimated from the data, although in rare cases they may also be prescribed. We only discuss the first case in some detail. (The other two cases are similar.) The total chi-square in this case is calculated by

$$CS = \sum_{a,b,c} \frac{(p_{abc} - \hat{p}_c \hat{p}_{ab})^2}{\hat{p}_c \hat{p}_{ab}}, \tag{45}$$

which asymptotically follows the chi-square distribution with $ABC - AB - C + 1$ df under the Case 1 hypothesis above.

We first explain why part chi-squares in the ANOVA-like partition of the total chi-square are usually not independent from each other under Case 1 above. In this case, we may factorially combine levels of Factors A and B and create a new factor which may be called Factor AB. We then have a two-way table, in which the columns represent levels of Factor C and the rows represent levels of Factor AB. Define $\mathbf{K} = \mathbf{P}_{AB} \otimes \mathbf{P}_C$, where $\mathbf{P}_{AB}$ is the diagonal matrix with marginal probabilities of the levels of Factor AB as diagonal elements. We may also define $\tilde{\mathbf{T}}_{\underline{AB}C} = \tilde{\mathbf{T}}_{AB} \otimes \tilde{\mathbf{T}}_C$, where the matrices on the righthand side are chosen to be orthogonal transformation matrices of appropriate sizes similar to the ones used in previous sections. The matrix $\tilde{\mathbf{T}}_{\underline{AB}C}$ is fully orthogonal, as is $\tilde{\mathbf{T}}_{ABC}$ defined in (42). The problem is that the matrix $\tilde{\mathbf{T}}_{AB}$, constructed according to the methods described in Section 2, does not reflect the factorial structure of the rows of the table, while the rows of the table were in fact constructed by factorially combining levels of Factors A and B. It may be possible to incorporate such structure into this matrix. For example, when Factor A and Factor B both have only two levels, we may define

$$\tilde{\mathbf{T}}_{AB} = \begin{bmatrix} \sqrt{p_1} & \sqrt{p_1 u_{24}/u_{13}} & \sqrt{p_1 u_{34}/u_{12}} & \sqrt{p_1 u_{23}/u_{14}} \\ \sqrt{p_2} & -\sqrt{p_2 u_{13}/u_{24}} & \sqrt{p_2 u_{34}/u_{12}} & -\sqrt{p_2 u_{14}/u_{23}} \\ \sqrt{p_3} & \sqrt{p_3 u_{24}/u_{13}} & -\sqrt{p_3 u_{12}/u_{34}} & -\sqrt{p_3 u_{14}/u_{23}} \\ \sqrt{p_4} & -\sqrt{p_4 u_{13}/u_{24}} & -\sqrt{p_4 u_{12}/u_{34}} & \sqrt{p_4 u_{23}/u_{14}} \end{bmatrix}, \qquad (46)$$

where the p_i's $(i = 1, \cdots 4)$ are the marginal probabilities of the levels of Factor AB, and $u_{jk} = p_j + p_k$ $(j, k = 1, \cdots, 4)$. The first column of this matrix pertains to the effect of mean probability, while the second column represents the A main effect, the third column the B main effect, and the fourth column the AB interaction effect. This matrix, together with $\tilde{\mathbf{T}}_C$, may be used to obtain part chi-squares in the ANOVA-like family-wise partition of the total chi-square for this case. The problem is that the last three columns of the matrix above are generally not orthogonal to each other except for the very special cases in which $\mathbf{P}_{AB}$ can be factored into $\mathbf{P}_B \otimes \mathbf{P}_A$ (which in fact implies complete independence). Consequently, part chi-squares generated by $\tilde{\mathbf{T}}_{\underline{AB}C}$ are not mutually independent. There are three nonzero part chi-squares, CS(AC), CS(BC), and CS(ABC). The order in which these effects are taken into account, however, makes a difference. There are six different ways to order three items. We may impose the restriction that no higher-order interactions are considered before any lower-order interactions, in which case this number is reduced to two. There are two families of effects of the same order, and there are two ways to order them: AC before BC, and BC before AC. In the former case, we take into account the AC interaction effects first, then the BC interaction effects, followed by the ABC interaction effects. The resultant effects are called AC ignoring BC, BC eliminating AC, and ABC eliminating AC and BC. These effects are written as AC, BC|AC, and ABC|AC,BC, respectively. The second case is analogous: BC ignoring AC (simply written as BC), AC eliminating BC (written as AC|BC), and ABC eliminating AC and BC (this is the same as before). There are thus (at least) two distinct family-wise partitions of the total chi-square. As before, any asymptotically chi-squared variables with df larger than 1 can ultimately be partitioned into the sum of asymptotically independent chi-squared variables each with a single df. However, non-orthogonality among the family-wise effects considerably complicates the situation. Contrasts may be constructed which are orthogonal within the same families, but non-orthogonal across different families.

As noted above, there are two other cases of one-factor independence. Non-orthogonality of part chi-squares under these hypotheses remain essentially the same as above. There are thus at least six distinct partitions under the one-factor independence hypotheses. A greater variety of hypotheses and a larger number of possible partitions of total chi-squares are possible in higher order tables.

As has been mentioned, there are other partial independence hypotheses in three-way tables. These are called conditional independence conditions in which two factors are independent given levels of a third factor. Again, there are three such cases in three-way tables: 1. $p_{abc} = p_{ab}p_{ac}/p_a$ (Factors B and C are independent given levels of Factor A), 2. $p_{abc} = p_{ac}p_{bc}/p_c$ (A and B are independent given C), and 3. $p_{abc} = p_{ab}p_{bc}/p_b$ (A and C are independent given B). These conditions are depicted at the second from the top level in Figure 1. The probabilities on the righthand side of these hypotheses are usually estimated from the data, although, as before, they could also be prescribed. We only discuss Case 1 above in some detail. (The other cases are similar.) This case subsumes both Cases 1 and 2 of the one-factor independence conditions. It reduces to $p_{abc} = p_c p_{ab}$ when $p_{ac}/p_a = p_c$, and to $p_{abc} = p_b p_{ac}$ when $p_{ab}/p_a = p_b$. The total chi-square is calculated by

$$CS = \sum_{a,b,c} \frac{(\hat{p}_{abc} - \hat{p}_{ac}\hat{p}_{bc}/\hat{p}_c)^2}{\hat{p}_{ac}\hat{p}_{bc}/\hat{p}_c}, \quad (47)$$

which asymptotically follows the chi-square distribution with $ABC - AB - AC + A$ df under the Case 1 hypothesis. Similar partitions of the total chi-square are possible as before, although many of the part chi-squares in the partitions are identically zero. There are only two nonzero part chi-squares, namely CS(BC) and CS(ABC|BC). As will be seen in the next section, the values of these part chi-squares calculated from an observed table are usually not equal to the corresponding part chi-squares calculated under different hypotheses.

4.3 A numerical example

We give numerical examples of the theoretical results presented in the previous sections. The example data set we use has been previously analyzed by many authors including Snedecor (1958), Cheng et al (2006), and Takane and Zhou (2013), and comparisons of our results to theirs are of interest. The observed frequencies and probabilities of cells in a 2 by 2 by 2 table are given in Table 1, where the cells are arranged in the order suggested earlier, that is, the index of Factor A moves fastest and that of Factor C slowest.

Table 1 The 2 by 2 by 2 table used in the numerical demonstration.

Factors			Observed Frequencies	Observed Probabilities	
C	B	A	**f**	$\hat{p}_{abc}$	$\hat{\mathbf{p}}$
C_1	B_1	A_1	79	$\hat{p}_{111}$	.0945
		A_2	177	$\hat{p}_{211}$	.2117
	B_2	A_1	62	$\hat{p}_{121}$	.0742
		A_2	121	$\hat{p}_{221}$	.1447
C_2	B_1	A_1	73	$\hat{p}_{112}$	.0873
		A_2	81	$\hat{p}_{212}$	.0969
	B_2	A_1	168	$\hat{p}_{122}$	.2010
		A_2	75	$\hat{p}_{222}$	.0897

Results of the computation are reported in Table 2. The first column of this table labels various part chi-squares. Recall that a single alphabetic symbol indicates the main effects (e.g., A indicates the main effects of A), while concatenated alphabetic symbols indicate interaction effects (e.g., AB indicates the interaction effects between A and B). Multiple groups of symbols separated by a comma indicate joint effects (e.g., A,AB indicates the joint effects of the A main and AB interaction effects). Multiple groups of symbols separated by a vertical bar indicate the effects placed on the left side of the bar eliminating the effects on the right (e.g., AB|AC indicates the effects of AB eliminating the effects of AC).

The second column of Table 2 gives the breakdown of the total chi-square under the complete marginal homogeneity (under independence) condition, that is, $\mathbf{P}_A = \mathbf{P}_B = \mathbf{P}_C = (1/2)\mathbf{I}_2$, and $\tilde{\mathbf{T}}_{ABC} = \tilde{\mathbf{T}}_C \otimes \tilde{\mathbf{T}}_B \otimes \tilde{\mathbf{T}}_A$, where

$$\tilde{\mathbf{T}}_C = \tilde{\mathbf{T}}_B = \tilde{\mathbf{T}}_A = \begin{bmatrix} \sqrt{.5} & \sqrt{.5} \\ \sqrt{.5} & -\sqrt{.5} \end{bmatrix}.$$

Under this hypothesis, part chi-squares are all asymptotically independent, so that the effects ignoring and the effects eliminating other effects are identical (the values of the latter are given in parentheses in the table), and the total chi-square is uniquely partitioned into the sum of part chi-squares pertaining to the three (families of) main effects (A, B, and C), the three (families of) two-way interaction effects (AB, AC, and BC), and the (single family) of three-way interaction effects (ABC).

The third column of Table 2, on the other hand, shows the breakdown of the total chi-square under Scenario 2 (Three-way independence without marginal homogeneity). This implies that

$$\mathbf{P}_A = \begin{bmatrix} .4569 & 0 \\ 0 & .5431 \end{bmatrix}, \ \mathbf{P}_B = \begin{bmatrix} .4904 & 0 \\ 0 & .5096 \end{bmatrix},$$

$$\text{and } \ \mathbf{P}_C = \begin{bmatrix} .5251 & 0 \\ 0 & .4749 \end{bmatrix},$$

and

$$\tilde{\mathbf{T}}_A = \begin{bmatrix} .6760 & .7369 \\ .7369 & -.6760 \end{bmatrix}, \ \tilde{\mathbf{T}}_B = \begin{bmatrix} .7003 & .7138 \\ .7138 & -.7003 \end{bmatrix},$$

$$\text{and } \ \tilde{\mathbf{T}}_C = \begin{bmatrix} .7247 & .6891 \\ .6891 & -.7247 \end{bmatrix}.$$

The main effects are all identical to zero, and the total chi-square in this case only reflects the two-way and three-way interaction effects. Note that nonzero part chi-squares in the third column are subtly different from the corresponding part chi-squares in the second column. For example, $CS(\text{AB})$ is 23.45 in the second column, while it is 24.10 in the third column. As remarked earlier, this is precisely what is meant by the dependence of partitions of Pearson's statistic on the hypothesis about the expected frequencies.

Table 2 Part chi-squares in partitions of Pearson's statistic under various hypotheses regarding the expected frequencies (probabilities) for the data set in Table 1.

	Hypothesis about p_{cba}							
Effects	1/8	$\hat{p}_a\hat{p}_b\hat{p}_c$	$\hat{p}_c\hat{p}_{ab}$		$\hat{p}_b\hat{p}_{ac}$		$\hat{p}_a\hat{p}_{bc}$	
A	6.20	0	0	0	0	0	0	0
B	0.31	0	0	0	0	0	0	0
C	2.11	0	0	0	0	0	0	0
AB	23.45	24.10	0*	0*	24.10		24.10	
AB|AC	(23.45)	(24.10)						11.27
AB|BC	(23.45)	(24.10)				11.81		
AB|AC,BC	(23.45)	(24.10)						
AC	70.05	68.66	68.66		0*	0*		68.66
AC|AB	(70.05)	(68.66)					55.83	
AC|BC	(70.05)	(68.66)		55.30				
AC|AB,BC	(70.05)	(68.66)						
BC	31.39	31.80		31.80		31.80	0*	0*
BC|AB	(31.39)	(31.80)			19.51			
BC|AC	(31.39)	(31.80)	18.44					
BC|AB,AC	(31.39)	(31.80)						
ABC	4.60	7.45						
ABC|AB,AC	(4.60)	(7.45)					7.06	7.06
ABC|AB,BC	(4.60)	(7.45)			6.35	6.35		
ABC|AC,BC	(4.60)	(7.45)	6.63	6.63				
ABC|AB,AC,BC	(4.60)	(7.45)						
AB,AC,ABC	(98.10)	(100.21)					86.99	86.99
AB,BC,ABC	(59.44)	(63.35)			49.96	49.96		
AC,BC,ABC	(101.04)	(107.91)	93.93	93.93				
AB,BC,AC,ABC	129.49	131.99						
+A,B,C (Total)	138.11							

It may be pointed out that Lancaster's (1951) method for calculating part chi-squares under these two scenarios works fine because of the orthogonality of all the terms in the partitions. This method involves reducing an original three-way table into two-way marginal tables to calculate part chi-squares for the main effects and two-way interaction effects, and subtracting the sum of these part chi-squares from the total chi-square to calculate the part chi-square for the three-way interaction effects. The values of part chi-squares reported in Table 2 of Snedecor (1958) were presumably obtained by Lancaster's (1951) method under the three-way independence hypothesis. The values of $CS(\text{AC}) = 68.30$ and $CS(\text{ABC}) = 7.80$ reported in his paper seem to be incorrect. Our values are 68.66 and 7.45, respectively. The total chi-square is correct in Snedecor (1958), however. Since the three-way interaction effect was presumably calculated by subtracting the sum of all two-way interaction effects from the total chi-square, the cause of the miscalculation of the three-way interaction effects seems to be due to the miscalculation of $CS(\text{AC})$.

The last six columns of Table 2 give partitions of Pearson's statistic under the one-factor independence hypotheses. As noted in the previous section, there are three such hypotheses, and within each hypothesis, there are two alternative partitions due to the non-orthogonality of the effects involved. The values of part chi-squares reported in the table are taken from Takane and Zhou (2013), who provided computational formula easier to use in calculating these quantities. (Note that there is a 0 with an asterisk in each of the last six columns of Table 2. Nonzero values were given for these part chi-squares in Takane and Zhou (2013). However, these effects are *a priori* eliminated under the respective hypotheses, and the value of zero is more appropriate, as given here.)

Table 3 shows part chi-squares under three conditional independence hypotheses. The part chi-quare due to the BC interaction effects in the second column of this table are analogous to BC|AB and BC|AC under the hypotheses that A is independent from BC, and B is independent from AC, respectively, in Table 2. As has been noted earlier, however, the former part chi-square is not identical to either of the latter two. Similar observations can be made for the AB and AC interactions in columns 3 and 4 of the table. Likewise, none of the three-way interaction effects in Table 3 are exactly the same as the values of the three-way interaction effects in Table 2, although the values are fairly close.

Table 3 Part chi-squares in partitions of Pearson's statistic under conditional independence conditions for the data set in Table 1

	Hypothesis about p_{abc}		
Effects	$p_{ab}p_{ac}/p_a$	$p_{ac}p_{bc}/p_c$	$p_{ab}p_{bc}/p_b$
A	0	0	0
B	0	0	0
C	0	0	0
AB	0	62.29	0
AC	0	0	57.45
BC	20.10	0	0
ABC|AB		6.82	
ABC|AC			6.07
ABC|BC	6.87		
AB,ABC|AB		69.11	
AC,ABC|AC			63.52
BC,ABC|BC	26.93		

5 Comparison with the LR statistic

Finally, we compare partitions of Pearson's statistic presented in previous sections with those of the (log) likelihood ratio (LR) statistic. To simplify our presentation, we mostly restrict our attention to three-way tables. In this case, the LR statistic is defined as

$$LR = 2N \sum_{a,b,c} \hat{p}_{abc} \log\left(\frac{\hat{p}_{abc}}{p_{abc}}\right), \tag{48}$$

where $\hat{p}_{abc}$ is the observed joint probability, and p_{abc} is the corresponding expected probability under some hypothesis. The LR statistic defined above asymptotically follows the chi-square distribution with df that depend on the hypothesis about p_{abc}. Part LR chi-squares are defined as the difference between LR's calculated under two nested hypotheses about p_{abc}.

Figure 1 summarizes the most representative hypotheses about p_{abc} in hierarchical form, a majority of which have already been discussed in the context of Pearson's statistic. There are seven layers of hypotheses in Figure 1. Specific hypotheses about p_{abc} are indicated in framed boxes. In general, hypotheses placed in lower layers represent more stringent hypotheses, while those in upper layers represent less stringent hypotheses. For example, complete homogeneity (under independence) at the bottom is the most stringent hypotheses, while the saturated model placed at the top does not impose any restrictions on p_{abc}. The figure is roughly divided into two parts: The layers above the three-way independence hypotheses represent partial independence hypotheses, while those below all assume complete independence plus some degrees of (partial or complete) marginal homogeneity. Some hypotheses in adjacent layers are connected by line segments. These indicate that the hypotheses in lower layers are special cases of the ones in upper layers. For example, $p_{abc} = 1/ABC$ is a special case of $p_{abc} = p_a/BC$, in which p_a in the latter is assumed equal to $1/A$. The label A attached to this line segment indicates that the difference between the two connected hypotheses represents the main effects of A. The difference in the (log) LR statistics associated with the two hypotheses indicates the part LR chi-square due to the main effects of A denoted by $LR(\text{A})$. As another example, $p_{abc} = p_a p_b p_c$ (the complete independence hypothesis) is a special case of $p_{abc} = p_c p_{ab}$ (C is independent from A and B). The difference between the two corresponds to the hypothesis that $p_{ab} = p_a p_b$, representing the AB interaction effects. Figure 1 indicates that the three-way (complete) independence hypothesis is also a special case of two other one-factor independence hypotheses.

In log-linear analysis to which the LR statistic is closely linked, the effects representing interactions of different orders are generally not independent from each other. For example, two-way interaction effects are not independent from main effects. We use the same convention that we used for Pearson's statistic, in which we always eliminate the effects of all lower order interactions when we consider higher order interactions. For example, when we consider two-way interaction effects we always eliminate the main effects (e.g., AB is always AB|A,B,C), and when we consider three-way interaction effects, we eliminate all the main effects and two-way interaction effects (e.g., ABC|AB,AC,BC is actually ABC|AB,AC,BC,A,B,C, but to avoid clutters of symbols, we use the former to denote the latter.) This convention effectively orthogonalizes all interaction effects of different orders. (Main effects can be regarded as order-one interaction effects.) Note that Cheng et al. (2006) also uses AB in the sense of AB|A,B,C, and ABC in the sense of ABC|AB,AC,BC,A,B,C. In log-linear analysis, the three (families of) main effects are mutually orthogonal without eliminating other main effects (e.g., A = A|B = A|C = A|B,C). The relation-

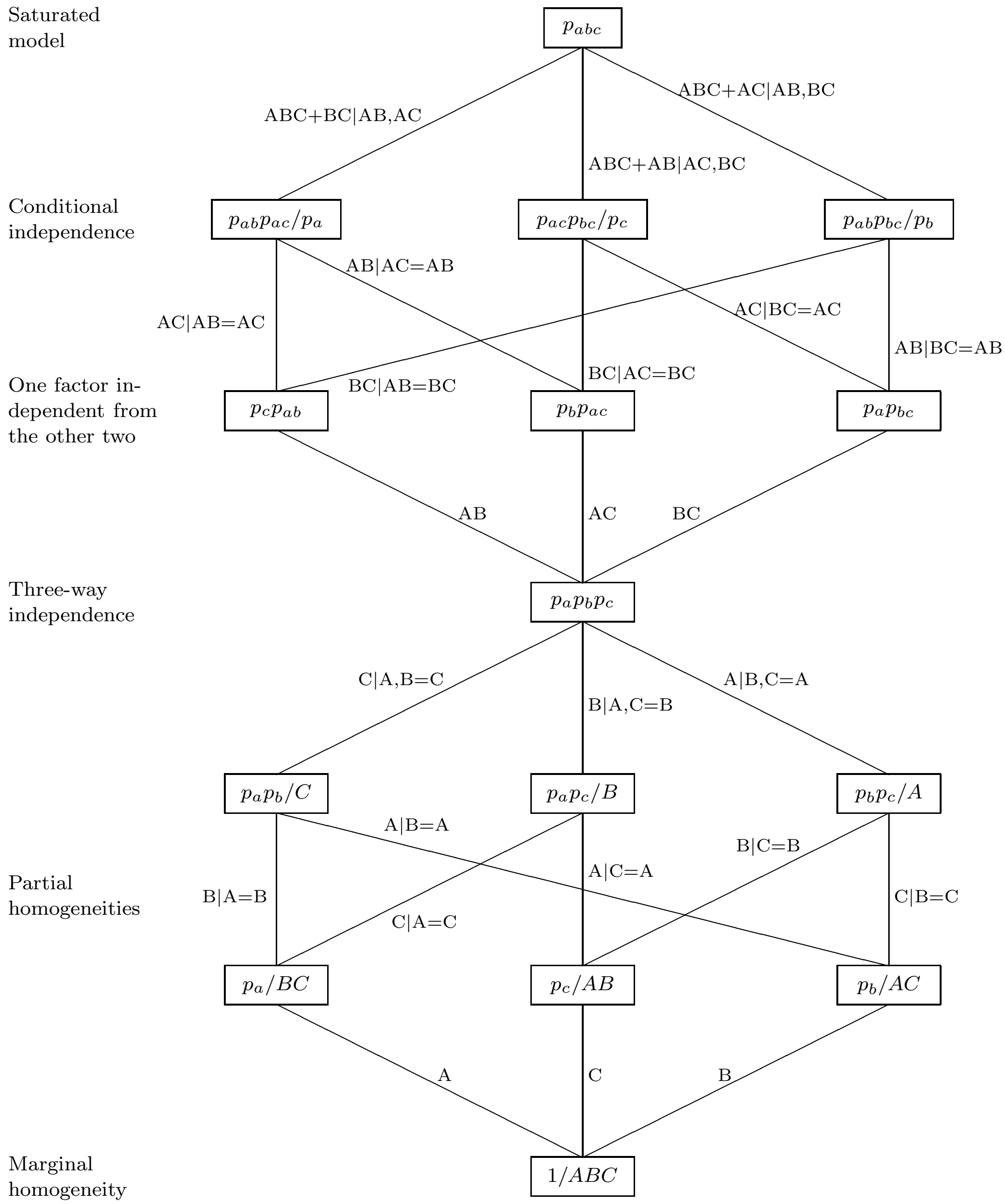

Fig. 1 A hierarchy of hypotheses in log-linear analysis of three-way contingency tables

ships among the three (families of) two-way interaction effects (all eliminating the main effects) are, however, much more complicated (Cheng et al., 2006). First of all, AB = AB|AC = AB|BC, but AB|AC,BC is different from any of the former three. Similar relations also hold for AC and BC interaction effects (i.e., AC = AC|AB = AC|BC $\neq$ AC|AC,BC, and BC = BC|AB = BC|AC $\neq$ BC|AB,AC). As noted above, two-way interaction effects ignoring other two-way interaction effects represent the differences between the three-way independence hypothesis and one-factor independence hypotheses. The two-way interaction effects eliminating only one of the other two families of two-way interaction effects represent differences between one-factor independence hypotheses and conditional independence hypotheses (e.g., AB|AC represents the difference between $p_{abc} = p_b p_{ac}$ and $p_{abc} = p_{ab}p_{ac}/p_a$, and AB|BC represents the difference between $p_{abc} = p_a p_{bc}$ and $p_{abc} = p_{ab}p_{bc}/p_b$, but these two differences are identical, both pertaining to the hypothesis of $p_{ab} = p_a p_b$). However, the two-way interaction effects eliminating both of the other two families of two-way interaction effects are part of the differences between conditional independence hypotheses and the saturated model in which no constraints are imposed on p_{abc} (e.g., AB|AC,BC represents part of the difference between $p_{abc} = p_{ac}p_{bc}/p_c$ and unconstrained p_{abc}). The two-way interaction effects eliminating both of the other two-way interaction effects (e.g., AB|AC,BC) represent homogeneous parts of simple two-way interaction effects (e.g., AB interaction effects at different levels of C), and differ from the other two-way interaction effects (e.g., AB, AB|AC, AB|BC). The remaining portion of the differences between the conditional independence hypotheses and the saturated model pertains to the three-way interaction effects (i.e., ABC|AB,AC,BC), representing nonuniform parts of simple two-way interaction effects. There are no explicit expressions of no three-way interaction effects that separate the saturated model (with the three-way interaction effects) at the very top of the hierarchy and the one without the three-way interaction effects. Some iterative fitting procedure is necessary to fit the model of no three-way interaction effects. There are three conditional independence hypotheses, and correspondingly three connections between these hypotheses and the saturated model. The differences associated with these connections all represent two-way interaction effects eliminating the other two families of two-way interaction effects plus the three-way interaction effects. This means that the joint effects of the three families of two-way interaction effects can be partitioned in three different ways: AB,AC,BC = AB + AC + BC|AB,AC = AB + BC + AC|AB,BC = AC + BC + AB|AC,BC. We may add A + B + C + ABC|AB,AC,BC to each of the three partitions of AB,AC,BC, and obtain three alternative partitions of the total LR chi-square under the complete marginal homogeneity hypothesis.

The values of part LR chi-squares obtained from the example data set (Table 1) are presented in Table 4 for the three possible partitions of the total LR chi-square corresponding to three possible partitions of the joint two-way interaction effects. The values reported in the table were obtained by Hiloglinear and Loglinear procedures in SPSS by Takane and Zhou (2013). Cheng et al. (2006) obtained essentially the same results. As discussed above, the main effects are independent from each other. Ignoring and eliminating other effects are identical, so that only the ignoring effects are shown in the table. Two-way interaction effects, on the other hand, are not independent from each other. Instead, there are three alternative partitions. Two of the three families of two-way interaction effects ignoring other families of two-way interaction effects can be added as part of the joint effects of all three families of two-way interaction effects. However, the third two-way interaction effect to be added to this sum to obtain the joint effects of all two-way interaction effects must be the remaining two-way interaction effects eliminating those already added. There are three ways to do this, resulting in three alternative partitions of the joint two-way interactions. It is also observed that two-way interaction effects remain invariant even if another family of two-way interaction effects are eliminated, but not if both of the two remaining families are simultaneously eliminated. As noted above, the three-way interaction effects in log-linear analysis cannot be obtained in closed form, but should be estimated iteratively. Note that the values of part LR chi-squares are in all cases very similar to the corresponding Pearsonian part chi-squares.

Table 4 A summary of the log LR chi-squares associated with log-linear analysis for the example data set in Table 1

Effects	Decomposition 1	Decomposition 2	Decomposition 3
A	6.21	6.21	6.21
B	0.31	0.31	0.31
C	2.11	2.11	2.11
AB	24.23		24.23
AB|AC	(24.23)		(24.23)
AB|BC	(24.23)		(24.23)
AB|BC,AC		12.22	
AC	69.54	69.54	
AC|AB	(69.54)	(69.54)	
AC|BC	(69.54)	(69.54)	
AC|AB,BC			57.53
BC		32.01	32.01
BC|AB		(32.01)	(32.01)
BC|AC		(32.01)	(32.01)
BC|AB,AC	20.00		
ABC|AB,BC,AC	6.82	6.82	6.82
AB,BC,AB	113.77	113.77	113.77
AB,BC,AC,ABC	120.59	120.59	120.59
+A,B,C (Total)	129.22	129.22	129.22

6 Concluding Remarks

In this paper, we gave a comprehensive account of what is involved in partitioning Pearson's chi-square statistic. We began this endeavor by first developing a basic tool for partitioning chi-squares for one-way tables, which was then gradually expanded to higher order tables. We have found that partitioning Pearson's statistic depends on the hypotheses about the expected frequencies (probabilities). Under the complete independence conditions, part chi-squares in ANOVA-like partitions of the total chi-square are asymptotically independent. However, entire partitions are affected whenever a new hypothesis is adopted.

Derived partitions of Pearson's statistic are compared with analogous partitions of the log LR chi-square statistic. The LR chi-square statistic admits invariant partitions, e.g., the part LR chi-square due to the AB interaction effects, i.e., LR(AB), remains the same whether the three-way independence hypothesis is assumed, or the complete marginal homogeneity is assumed. (Remember that the effect AB here is in fact AB|A,B,C.) This makes easier to perform so-called step-up tests. We may, for example, start with the test of the three-way independence hypothesis as the null hypothesis against one of the one-factor independence hypotheses. If the three-way independence hypothesis is rejected, we may "step up" the test by adopting the former alternative hypothesis (one of the one-factor independence hypotheses) as the null hypothesis against one of the conditional independence hypotheses as the alternative hypothesis, and so on.

This is in marked contrast with Pearson's chi-square statistic. While step-up tests are not impossible with Pearson's statistic, a new partition results every time we adopt a new null hypothesis. Suppose, for example, we start with complete marginal homogeneity and derive a partition of the total chi-square given in the second column of Table 2. Suppose further that this null hypothesis is rejected against the three-way independence hypothesis. Then, a new partition of the total chi-square has to be derived under the three-way independence condition to perform step-up tests. Of course, one can stick with one common null hypothesis in Pearson's statistic for all the tests (based on a single partition of the total chi-square). It should be kept in mind, however, that all these tests involve the same null hypothesis (*albeit* against different alternatives). A bit of comfort is that values of part chi-squares which are not zeroed out are in most cases very similar across different partitions generated by different null hypotheses, and their significance/nonsignificance tends to remain unchanged unless they are close to the borderline.

Refrences

Cheng PE, Liou JW, Liou M, Aston JAD. Data information in contingency tables: A fallacy of hierarchical loglinear models. J Data Sci 4:387–398 (2006).

Greenacre MJ. Theory and applications of correspondence analysis. Academic Press, London (1984).

Irwin JO. A note on the subdivision of χ^2 into components. Biometrika 36:130–134 (1949).

Lancaster HO. The derivation and partition of χ^2 in certain discrete distributions. Biometrika 36:117–129 (1949).

Lancaster HO. Complex contingency tables treated by the partition of chi^2. J ROY STAT SOC B 13: 242–249 (1951).

Nishisato S. Analysis of categorical data: Dual scaling and its applications. University of Toronto Press, Toronto (1980).

Pearson K. On the criterion that a given system of deviation from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophy Magazine 50:157–172 (1900).

Snedecor GW. Chi-squares of Bartlett, Mood and Lancaster in a 2^3 contingency table. Biometrics 14:560–562 (1958).

Takane Y, Zhou L. Anatomy of Pearson's chi-square statistic in three-way contingency tables. In: Millsap RE, van der Ark LA, Bolt DM, Woods CM. (eds.) New developments in quantitative psychology. Springer, New York, pp 41–57 (2013).