



A New Algorithm for the Partition of Pearson's Chi-Squared Statistic for Multiway Contingency Table

Kirtee K. Kamalja¹ · Nutan V. Khangar² · Eric J. Beh^{3,4}

Accepted: 18 December 2023 / Published online: 12 February 2024
© The Indian Society for Probability and Statistics (ISPS) 2024

Abstract

Pearson's chi-squared statistic is one of the most common statistical tools used to assess the association between two or more categorical variables that have been cross-classified to form a contingency table. In many practical settings, multiple categorical variables are “paired-off” and analysed by identifying association structures between two variables only. However, there are less well-known tools that allow the analyst to explore the association structure of categorical variables that form a multi-way contingency table. This paper presents an ANOVA-like decomposition of the chi-squared statistic for four-way and five-way contingency tables and can be extended for the analysis of higher-way contingency tables. Furthermore, we propose an efficient algorithm for partitioning the statistic that leads to two-way and higher-way terms. The proposed algorithm reduces the complexity involved in the calculation of the terms of the partition and will be demonstrated by way of a simulation and practical example.

Keywords Contingency tables · Pearson's chi-squared statistic · ANOVA-like decomposition

JEL Classifications C49

✉ Kirtee K. Kamalja
kirteekamalja@gmail.com

¹ Department of Statistics, K.B.C. North Maharashtra University, Jalgaon, India

² Department of Statistics, K.R.T. Arts, B.H. Commerce and A.M. Science College, Nashik, India

³ National Institute for Applied Statistics Research Australia (NIASRA), University of Wollongong, Wollongong, Australia

⁴ Centre for Multi-Dimensional Data Visualisation (MuViSU), Stellenbosch University, Stellenbosch, South Africa

1 Introduction

There has been a great deal of attention given to the analysis of big data in the last few years (Tsai et al. 2015; Luo et al. 2016). For many areas of research—especially marketing, social, political, business and medical studies—much of their data is collected from surveys and questionnaires that comprise of questions whose responses are categorical. Understanding the association between multiple categorical variables in such cases is a big challenge for data scientists (Loughin and Scherer 1998). Central to studying these associations is Pearson’s chi-squared statistic and tests of independence; both of which are appropriate for the analysis of symmetrically associated variables where all variables are treated as explanatory variables. For the study of three categorical variables (Kroonenberg 1989; Carlier and Kroonenberg 1996) presented a decomposition of Pearson’s chi-squared statistic using Lancaster’s (1951) additive decomposition. The advantage of this decomposition is that each term reflects a specific source of association which can be tested for statistical significance.

Beh and Davy (1998) presented the partition of Pearson’s chi-squared statistic using orthogonal polynomials for three-way contingency tables where all three variables consist of completely ordered categories. Their partition isolates 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. The case when only one or two variables are ordered is studied in Beh and Davy (1999).

While Pearson’s statistic considers a symmetric association structure between categorical variables, one may also examine the partition of a measure of asymmetric association; where, for a two-way contingency table, one variable is treated as a predictor variable and the second is considered a response variable. In such cases, the Goodman–Kruskal tau index (Goodman and Kruskal 1954) may be used as an alternative to Pearson’s statistic. For ordinal two-way tables, D’Ambra et al. (2005) partitioned the Goodman–Kruskal tau index into components that reflect sources of association in terms of the location, dispersion and higher order moments for each variable. Three-way analogues of the Goodman–Kruskal tau index include the Gray–Williams statistic (Gray and Williams 1975), the delta index (Lombardo 1994, 2011) and Marcotorchino’s index (Marcotorchino 1984). Lombardo et al. (1996) presented the partition of Marcotorchino’s index that is an asymmetric analogue of Beh and Davy’s (1998, 1999) partition of Pearson’s chi-squared statistic and Lombardo (2011) presented a similar partition for the delta index.

Loisel and Takane (2016) clarified the expositions proposed by Lancaster (1951) on an ANOVA-like partition of Pearson’s statistic proposed under various representative hypotheses about the expected frequencies and further extended it to more general situations. The approach of Loisel and Takane (2016) involves transforming the diagonal matrix of hypothesized joint probabilities of the cells and the vector of observed joint probabilities from two-way/multi-way tables into a vector such that its inner product is equivalent to Pearson’s chi-squared statistic for that table. Thus, the method of partitioning the chi-squared statistic used

by Loisel and Takane (2016) involves the transformation, vectorisation and construction of fully orthogonal matrices of appropriate dimension for each factor/association term and the use of Kronecker products. The idea of the partition is based on expressing the chi-squared statistic as a quadratic form in idempotent matrix. This method is applicable to multi-way contingency tables but involves matrices of sizes almost as large as all possible combinations of categories of all variables in the multi-way table. They also compare the partition of Pearson's statistic with analogous decompositions of the log likelihood ratio statistic associated with log-linear analysis of a contingency table. Lombardo et al. (2020) developed a general framework for ANOVA-like decompositions of Pearson's statistic under the independence assumption. It also accommodates the cases of specification of theoretically driven probabilities as well as the cases in which the marginal probabilities are fixed or estimated from the data. Lombardo et al. (2021) used orthogonal polynomials and/or the singular vectors to partition the three-way symmetric and asymmetric measures of association and demonstrate this partition with examples.

Lombardo et al. (2017) provided the *chi2x3way* package in R for partitioning three-way chi-squared statistic and Marcotorchino index, under two scenarios based on Kronecker product. Lombardo and Beh (2021) developed the *CAvariants* package in R which provides six variants of two-way correspondence analysis (CA). Lombardo et al. (2022) developed the *CA3variants* package in R which performs four variants of three-way CA. Both the *chi2x3way* and *CA3variants* R-packages can only be used on three-way tables. This motivated us to develop an algorithm that will theoretically deal with the partitions for multi-way tables and easily implement in software.

The main aim of the present paper is to develop an algorithm that simplifies the calculation of the terms from the partition of Pearson's chi-squared statistic for multi-way contingency tables. Furthermore, the implementation of the new algorithm for partitions involving software that use tensors makes it more efficient than the existing ones. A Matlab function '*catanova()*' is developed which gives an ANOVA-like partition for a multi-way contingency table. Such partitions quantify two-way, three-way and higher-way terms of association of a multi-way contingency table where all the terms can be formally tested for statistical significance through this function. Thus, the generalisation of what the *chi2x3way* package does for three-way tables is incorporated into the algorithm described in this paper.

This paper is organised as follows. Section 2 briefly describes the partition of Pearson's chi-squared statistic by Carlier and Kroonenberg (1996) for three-way contingency tables. In Sect. 3, we extend their partition to four-way, five-way and multi-way contingency tables. Section 4 proposes a fast and efficient method of partitioning Pearson's statistic yielding association terms for multi-way contingency tables. Section 5 provides an overview of tensors and utilizes them for the development of the proposed algorithm. Section 6 demonstrates the speed and applicability of this algorithm through a practical and simulation study; the practical demonstration of the algorithm is undertaken by considering contingency tables that have been commonly analysed in the literature. We provide some concluding remarks in Sect. 7.

2 Partition of Pearson's Chi-Squared Statistic for a Three-way Contingency Table

Carlier and Kroonenberg (1996) presented a decomposition of Pearson's chi-squared statistic for a three-way contingency table using Lancaster's (1951) additive decomposition. Before we discuss the new algorithm (in Sect. 4), we first provide an overview of the partition of Pearson's chi-squared statistic for a three-way contingency table.

Let n individuals be classified by three categorical variables A_1, A_2 and A_3 with j_1, j_2 and j_3 number of categories respectively. Let $N = (n_{i_1 i_2 i_3})$ be the corresponding three-way contingency table where $n_{i_1 i_2 i_3}$ is the number of individuals classified into i_k^{th} category of the k^{th} variable, for $i_k = 1, 2, \dots, j_k$, and $k = 1, 2, 3$. The matrix of joint proportions is $P = \frac{1}{n}N = (p_{i_1 i_2 i_3})$ so that $0 \leq p_{i_1 i_2 i_3} < 1, \forall (i_1, i_2, i_3)$ and $\sum_{i_1, i_2, i_3} p_{i_1 i_2 i_3} = 1$. The marginal proportion of variable A_1 is $p_{i_1 \cdot \cdot} = \sum_{i_2, i_3} p_{i_1 i_2 i_3}, i_1 = 1, 2, \dots, j_1$. Define an inner product of two three-way arrays $A = (a_{i_1 i_2 i_3})$ and $B = (b_{i_1 i_2 i_3}), \mathbb{R}^{j_1 \times j_2 \times j_3}$, as

$$\langle A, B \rangle = \sum_{i_1, i_2, i_3} p_{i_1 \cdot \cdot} p_{\cdot i_2 \cdot} p_{\cdot \cdot i_3} a_{i_1 i_2 i_3} b_{i_1 i_2 i_3}.$$

By doing so, A and B are orthogonal if $\langle A, B \rangle = 0$.

The total inertia of the contingency table, which is also referred as its Pearson's mean squared contingency coefficient or, alternatively, phi-squared statistic, Φ^2 , is based on the deviation of the observed joint proportions from the three-way independence model, and is defined as

$$\Phi^2 = \sum_{i_1, i_2, i_3} \frac{(p_{i_1 i_2 i_3} - p_{i_1 \cdot \cdot} p_{\cdot i_2 \cdot} p_{\cdot \cdot i_3})^2}{p_{i_1 \cdot \cdot} p_{\cdot i_2 \cdot} p_{\cdot \cdot i_3}}.$$

We shall refer to Φ^2 as simply the *total inertia* of the contingency table. Such terminology is especially common in CA (Greenacre 1984; Beh and Lombardo 2014) and quantifies the symmetric association between categorical variables and is independent of the sample size. The total inertia, Φ^2 , can be expressed as the weighted norm of the $j_1 \times j_2 \times j_3$ array Π whose $(i_1, i_2, i_3)^{th}$ cell value is $\pi_{i_1 i_2 i_3} = \frac{p_{i_1 i_2 i_3}}{p_{i_1 \cdot \cdot} p_{\cdot i_2 \cdot} p_{\cdot \cdot i_3}} - 1$ and is a measure of departure from complete independence of that cell such that

$$\Phi^2 = \sum_{i_1, i_2, i_3} p_{i_1 \cdot \cdot} p_{\cdot i_2 \cdot} p_{\cdot \cdot i_3} \left(\frac{p_{i_1 i_2 i_3}}{p_{i_1 \cdot \cdot} p_{\cdot i_2 \cdot} p_{\cdot \cdot i_3}} - 1 \right)^2 = \Pi^2$$

Thus, Φ^2 is the weighted sum-of-squares of the deviations of the observed relative frequencies from their expected value under complete independence, where the weights are given by the product of the marginal proportions, $p_{i_1 \cdot \cdot} p_{\cdot i_2 \cdot} p_{\cdot \cdot i_3}$. Consider the following

three-way arrays and their corresponding elements that are weighted sums of the marginal proportions:

$$\begin{aligned} \mathbf{\Pi}_{\dots} &= (\pi_{\dots}), & \mathbf{\Pi}_{i_1..} &= (\pi_{i_1..} - \pi_{\dots}), \\ \mathbf{\Pi}_{.i_2.} &= (\pi_{.i_2.} - \pi_{\dots}), & \mathbf{\Pi}_{..i_3} &= (\pi_{..i_3} - \pi_{\dots}), \\ \mathbf{\Pi}_{i_1i_2.} &= (\pi_{i_1i_2.} - \pi_{i_1..} - \pi_{.i_2.} + \pi_{\dots}), & \mathbf{\Pi}_{.i_2i_3} &= (\pi_{.i_2i_3} - \pi_{.i_2.} - \pi_{..i_3} + \pi_{\dots}), \\ \mathbf{\Pi}_{i_1.i_3} &= (\pi_{i_1.i_3} - \pi_{i_1..} - \pi_{..i_3} + \pi_{\dots}), \\ \alpha \mathbf{\Pi}_{i_1i_2i_3} &= (\pi_{i_1i_2i_3} - \pi_{i_1i_2.} - \pi_{i_1.i_3} - \pi_{.i_2i_3} + \pi_{i_1..} + \pi_{.i_2.} + \pi_{..i_3} - \pi_{\dots}) \end{aligned}$$

where $\pi_{\dots} = \sum_{i_1,i_2,i_3} P_{i_1..}P_{.i_2.}P_{..i_3} \pi_{i_1i_2i_3}$, $\pi_{i_1..} = \sum_{i_2,i_3} P_{.i_2.}P_{..i_3} \pi_{i_1i_2i_3}$, $\pi_{.i_2.} = \sum_{i_1,i_3} P_{i_1..}P_{..i_3} \pi_{i_1i_2i_3}$ and so on.

Observe that $\mathbf{\Pi}$ is the summation of these eight arrays where all the arrays are orthogonal to each other with respect to the previously defined inner product on $\mathbb{R}^{i_1 \times j_2 \times j_3}$. Furthermore, since, $\mathbf{\Pi}_{i_1..} = \mathbf{\Pi}_{.i_2.} = \mathbf{\Pi}_{..i_3} = \mathbf{O}$,

$$\mathbf{\Pi} = \mathbf{\Pi}_{i_1i_2.} + \mathbf{\Pi}_{i_1.i_3} + \mathbf{\Pi}_{.i_2i_3} + \alpha \mathbf{\Pi}_{i_1i_2i_3}$$

In terms of relative frequencies, the components are,

$$\left(\frac{P_{i_1i_2i_3} - P_{i_1..}P_{.i_2.}P_{..i_3}}{P_{i_1..}P_{.i_2.}P_{..i_3}} \right) = \frac{P_{i_1i_2.} - P_{i_1..}P_{.i_2.}}{P_{i_1..}P_{.i_2.}} + \frac{P_{i_1.i_3} - P_{i_1..}P_{..i_3}}{P_{i_1..}P_{..i_3}} + \frac{P_{.i_2i_3} - P_{.i_2.}P_{..i_3}}{P_{.i_2.}P_{..i_3}} + \frac{P_{i_1i_2i_3} - \alpha P_{i_1i_2i_3}}{P_{i_1..}P_{.i_2.}P_{..i_3}},$$

where $\alpha P_{i_1i_2i_3} = P_{i_1i_2.}P_{..i_3} + P_{i_1.i_3}P_{.i_2.} + P_{.i_2i_3}P_{i_1..} - 2P_{i_1..}P_{.i_2.}P_{..i_3}$.

The orthogonality of the arrays leads to an additive partition of the squared norm of $\mathbf{\Pi}$ as,

$$\| \mathbf{\Pi} \|^2 = \| \mathbf{\Pi}_{i_1i_2.} \|^2 + \| \mathbf{\Pi}_{i_1.i_3} \|^2 + \| \mathbf{\Pi}_{.i_2i_3} \|^2 + \| \alpha \mathbf{\Pi}_{i_1i_2i_3} \|^2. \tag{1}$$

Carlier and Kroonenberg (1996) expressed the above partition in terms of the inertia of each pair of categorical variables, and all three variables, so that

$$\begin{aligned} \Phi^2 &= \sum_{i_1,i_2} P_{i_1..}P_{.i_2.} \left(\frac{P_{i_1i_2.} - P_{i_1..}P_{.i_2.}}{P_{i_1..}P_{.i_2.}} \right)^2 + \sum_{i_1,i_3} P_{i_1..}P_{..i_3} \left(\frac{P_{i_1.i_3} - P_{i_1..}P_{..i_3}}{P_{i_1..}P_{..i_3}} \right)^2 \\ &+ \sum_{i_2,i_3} P_{.i_2.}P_{..i_3} \left(\frac{P_{.i_2i_3} - P_{.i_2.}P_{..i_3}}{P_{.i_2.}P_{..i_3}} \right)^2 + \sum_{i_1,i_2,i_3} P_{i_1..}P_{.i_2.}P_{..i_3} \left(\frac{P_{i_1i_2i_3} - \alpha P_{i_1i_2i_3}}{P_{i_1..}P_{.i_2.}P_{..i_3}} \right)^2. \end{aligned}$$

Note that this partition involves only the marginal (univariate) proportions of the three-variables A_1, A_2 and A_3 and the joint proportions of each pair of variables. Thus, the orthogonal decomposition of Pearson’s chi-squared statistic $\chi_{Total}^2 = n\Phi^2$ for N is,

$$\chi_{Total}^2 = \chi_{A_1A_2}^2 + \chi_{A_1A_3}^2 + \chi_{A_2A_3}^2 + \chi_{A_1A_2A_3}^2$$

Here $\chi_{A_1A_2}^2$ is Pearson’s chi-squared statistic of the two-way contingency table formed from the cross-classification of variables A_1 and A_2 and is a measure of deviation from complete independence of the observed cell frequencies. Similarly, $\chi_{A_1A_3}^2$ and $\chi_{A_2A_3}^2$ are

the two-variable chi-squared statistics of the tables formed by cross-classifying A_1 and A_3 , and A_2 and A_3 , respectively. The term $\chi^2_{A_1A_2A_3}$ is the Pearson’s chi-squared statistic that assesses the association between all three variables. Furthermore, the terms $\chi^2_{A_1A_2}, \chi^2_{A_1A_3}, \chi^2_{A_2A_3}$ are asymptotically chi-squared with $(j_1 - 1)(j_2 - 1), (j_1 - 1)(j_3 - 1), (j_2 - 1)(j_3 - 1)$ and $(j_1 - 1)(j_2 - 1)(j_3 - 1)$ degrees of freedom respectively. Furthermore, Carlier & Kroonenberg (1996) proposed the quantification of the global measure of dependence into four measures of dependence: the three measures for each pair-wise dependence and one measure for the three-way interaction. The following section provides a description of the ANOVA-like partition of Pearson’s chi-squared statistic for multi-way tables.

3 Partitioning Pearson’s Chi-Squared Statistic for a Multi-way Table

In this section, we provide an overview of the standard approach commonly used to partition Pearson’s chi-squared statistic. We do so by extending the additive decomposition given by Lancaster (1951) and used by Carlier and Kroonenberg (1996) to four-way and five-way tables; we note that these authors confined their attention to the analysis of a three-way contingency table. We also discuss the challenges involved in the partition of Pearson’s chi-squared statistic for multiple (more than five) categorical variables.

3.1 Partition for a Four-Way Contingency Table

Let n individuals be cross classified according to four categorical variables, A_1, A_2, A_3 and A_4 . Let j_k be the number of categories of the k^{th} variable, for $k = 1, 2, 3, 4$ and the joint frequencies be summarized into a four-way contingency table $N = (n_{i_1 i_2 i_3 i_4})$ that is of dimension $j_1 \times j_2 \times j_3 \times j_4$. Let $P = (p_{i_1 i_2 i_3 i_4})$ be the table of relative proportions such that $P = \frac{1}{n}N$. The total inertia Φ^2 is based on the deviation of the observations from what is expected under complete (four-way) independence such that,

$$\Phi^2 = \sum_{i_1, i_2, i_3, i_4} \frac{(p_{i_1 i_2 i_3 i_4} - p_{i_1 \dots} p_{i_2 \dots} p_{i_3 \dots} p_{i_4 \dots})^2}{p_{i_1 \dots} p_{i_2 \dots} p_{i_3 \dots} p_{i_4 \dots}}$$

This statistic can be expressed as the weight squared norm of Π in $\mathbb{R}^{j_1 \times j_2 \times j_3 \times j_4}$ where $\pi_{i_1 i_2 i_3 i_4} = \frac{p_{i_1 i_2 i_3 i_4}}{p_{i_1 \dots} p_{i_2 \dots} p_{i_3 \dots} p_{i_4 \dots}} - 1$ and $\|\Pi\|^2 = \sum_{i_1, i_2, i_3, i_4} p_{i_1 \dots} p_{i_2 \dots} p_{i_3 \dots} p_{i_4 \dots} (\pi_{i_1 i_2 i_3 i_4})^2$; this is apparent since this case is a four-way extension of the three-way partition described in Sect. 2. In this case, Π can be expressed as the addition of eleven orthogonal arrays that consist of ${}^4C_2=6$ two-way, ${}^4C_3=4$ three-way and ${}^4C_1=1$ four-way association terms. That is

$$\begin{aligned} \Pi &= \Pi_{i_1 i_2 \dots} + \Pi_{i_1 i_3 \dots} + \Pi_{i_1 \dots i_4} + \Pi_{i_2 i_3 \dots} + \Pi_{i_2 \dots i_4} + \Pi_{\dots i_3 i_4} \\ &+ \alpha \Pi_{i_1 i_2 i_3 \dots} + \alpha \Pi_{i_1 i_2 \dots i_4} + \alpha \Pi_{i_2 i_3 i_4 \dots} + \alpha \Pi_{i_1 i_3 i_4 \dots} + \alpha \Pi_{i_1 i_2 i_3 i_4 \dots} \end{aligned}$$

where

$$\begin{aligned}
 \pi_{i_1 i_2..} &= \frac{P_{i_1 i_2..} - P_{i_1..} P_{..i_2..}}{P_{i_1..} P_{..i_2..}}, \quad \pi_{i_1..i_3.} = \frac{P_{i_1..i_3.} - P_{i_1..} P_{..i_3.}}{P_{i_1..} P_{..i_3.}}, \quad \pi_{i_1..i_4.} = \frac{P_{i_1..i_4.} - P_{i_1..} P_{..i_4.}}{P_{i_1..} P_{..i_4.}}, \\
 \pi_{..i_2 i_3.} &= \frac{P_{..i_2 i_3.} - P_{..i_2.} P_{..i_3.}}{P_{..i_2.} P_{..i_3.}}, \quad \pi_{..i_2..i_4.} = \frac{P_{..i_2..i_4.} - P_{..i_2..} P_{..i_4.}}{P_{..i_2..} P_{..i_4.}}, \quad \pi_{..i_3 i_4.} = \frac{P_{..i_3 i_4.} - P_{..i_3.} P_{..i_4.}}{P_{..i_3.} P_{..i_4.}}, \\
 \alpha \pi_{i_1 i_2 i_3.} &= \frac{P_{i_1 i_2 i_3.} - \alpha P_{i_1 i_2 i_3.}}{P_{i_1..} P_{..i_2.} P_{..i_3.}}, \quad \alpha \pi_{i_1 i_2..i_4.} = \frac{P_{i_1 i_2..i_4.} - \alpha P_{i_1 i_2..i_4.}}{P_{i_1..} P_{..i_2..} P_{..i_4.}}, \quad \alpha \pi_{..i_2 i_3 i_4.} = \frac{P_{..i_2 i_3 i_4.} - \alpha P_{..i_2 i_3 i_4.}}{P_{..i_2..} P_{..i_3.} P_{..i_4.}}, \\
 \alpha \pi_{i_1..i_3 i_4.} &= \frac{P_{i_1..i_3 i_4.} - \alpha P_{i_1..i_3 i_4.}}{P_{i_1..} P_{..i_3.} P_{..i_4.}}, \quad \alpha \pi_{i_1 i_2 i_3 i_4.} = \frac{P_{i_1 i_2 i_3 i_4.} - \alpha P_{i_1 i_2 i_3 i_4.}}{P_{i_1..} P_{..i_2..} P_{..i_3.} P_{..i_4.}}, \\
 \alpha P_{i_1 i_2 i_3.} &= P_{i_1 i_2..} P_{..i_3.} + P_{i_1..i_3.} P_{..i_2..} + P_{..i_2 i_3.} P_{i_1..} - 2P_{i_1..} P_{..i_2..} P_{..i_3.}, \\
 \alpha P_{i_1 i_2..i_4.} &= P_{i_1 i_2..} P_{..i_4.} + P_{i_1..i_4.} P_{..i_2..} + P_{..i_2..i_4.} P_{i_1..} - 2P_{i_1..} P_{..i_2..} P_{..i_4.}, \\
 \alpha P_{..i_2 i_3 i_4.} &= P_{..i_2 i_3.} P_{..i_4.} + P_{..i_2..i_4.} P_{..i_3.} + P_{..i_3 i_4.} P_{..i_2..} - 2P_{..i_2..} P_{..i_3.} P_{..i_4.}, \\
 \alpha P_{i_1..i_3 i_4.} &= P_{i_1..i_3.} P_{..i_4.} + P_{i_1..i_4.} P_{..i_3.} + P_{..i_3 i_4.} P_{i_1..} - 2P_{i_1..} P_{..i_3.} P_{..i_4.}, \\
 \alpha P_{i_1 i_2 i_3 i_4.} &= P_{i_1 i_2..} P_{..i_3.} P_{..i_4.} + P_{i_1..i_3.} P_{..i_2..} P_{..i_4.} + P_{i_1..i_4.} P_{..i_2..} P_{..i_3.} + P_{..i_2 i_3.} P_{i_1..} P_{..i_4.} \\
 &\quad + P_{..i_2..i_4.} P_{i_1..} P_{..i_3.} + P_{..i_3 i_4.} P_{i_1..} P_{..i_2..} + P_{i_1 i_2 i_3.} P_{..i_4.} + P_{i_1 i_2..i_4.} P_{..i_3.} + P_{i_1..i_3 i_4.} P_{..i_2..} \\
 &\quad + 9P_{i_1..} P_{..i_2..} P_{..i_3.} P_{..i_4.}
 \end{aligned} \tag{2}$$

Thus, the additive partition of the squared norm of the four-way array Π can be expressed as

$$\begin{aligned}
 \|\Pi\|^2 &= \|\Pi_{i_1 i_2..}\|^2 + \|\Pi_{i_1..i_3.}\|^2 + \|\Pi_{i_1..i_4.}\|^2 + \|\Pi_{..i_2 i_3.}\|^2 + \|\Pi_{..i_2..i_4.}\|^2 + \\
 &\quad \|\Pi_{..i_3 i_4.}\|^2 + \|\alpha \Pi_{i_1 i_2 i_3.}\|^2 + \|\alpha \Pi_{i_1 i_2..i_4.}\|^2 + \|\alpha \Pi_{..i_2 i_3 i_4.}\|^2 + \|\alpha \Pi_{i_1..i_3 i_4.}\|^2 + \|\alpha \Pi_{i_1 i_2 i_3 i_4.}\|^2
 \end{aligned}$$

This partition can be alternatively, but equivalently, expressed in terms of the table’s inertia such that,

$$\begin{aligned}
 \Phi^2 &= \sum_{i_1, i_2} P_{i_1..} P_{..i_2..} \left(\frac{P_{i_1 i_2..} - P_{i_1..} P_{..i_2..}}{P_{i_1..} P_{..i_2..}} \right)^2 + \sum_{i_1, i_3} P_{i_1..} P_{..i_3.} \left(\frac{P_{i_1..i_3.} - P_{i_1..} P_{..i_3.}}{P_{i_1..} P_{..i_3.}} \right)^2 \\
 &\quad + \sum_{i_2, i_3} P_{..i_2.} P_{..i_3.} \left(\frac{P_{..i_2 i_3.} - P_{..i_2.} P_{..i_3.}}{P_{..i_2.} P_{..i_3.}} \right)^2 + \sum_{i_1, i_2, i_3} P_{i_1..} P_{..i_2.} P_{..i_3.} \left(\frac{P_{i_1 i_2 i_3.} - \alpha P_{i_1 i_2 i_3.}}{P_{i_1..} P_{..i_2.} P_{..i_3.}} \right)^2 \\
 &\quad + \sum_{i_1, i_2, i_4} P_{i_1..} P_{..i_2..} P_{..i_4.} \left(\frac{P_{i_1 i_2..i_4.} - \alpha P_{i_1 i_2..i_4.}}{P_{i_1..} P_{..i_2..} P_{..i_4.}} \right)^2 + \sum_{i_2, i_3, i_4} P_{..i_2..} P_{..i_3.} P_{..i_4.} \left(\frac{P_{..i_2 i_3 i_4.} - \alpha P_{..i_2 i_3 i_4.}}{P_{..i_2..} P_{..i_3.} P_{..i_4.}} \right)^2 \\
 &\quad + \sum_{i_1, i_3, i_4} P_{i_1..} P_{..i_3.} P_{..i_4.} \left(\frac{P_{i_1..i_3 i_4.} - \alpha P_{i_1..i_3 i_4.}}{P_{i_1..} P_{..i_3.} P_{..i_4.}} \right)^2 \\
 &\quad + \sum_{i_1, i_2, i_3, i_4} P_{i_1..} P_{..i_2..} P_{..i_3.} P_{..i_4.} \left(\frac{P_{i_1 i_2 i_3 i_4.} - \alpha P_{i_1 i_2 i_3 i_4.}}{P_{i_1..} P_{..i_2..} P_{..i_3.} P_{..i_4.}} \right)^2.
 \end{aligned}$$

Thus, the orthogonal partition of $\chi^2_{Total} = n\Phi^2$ for the four-way contingency table N is

$$\begin{aligned}\chi^2_{Total} = & \chi^2_{A_1A_2} + \chi^2_{A_1A_3} + \chi^2_{A_1A_4} + \chi^2_{A_2A_3} + \chi^2_{A_2A_4} + \chi^2_{A_3A_4} \\ & + \chi^2_{A_1A_2A_3} + \chi^2_{A_1A_2A_4} + \chi^2_{A_2A_3A_4} + \chi^2_{A_1A_3A_4} + \chi^2_{A_1A_2A_3A_4}\end{aligned}\quad (3)$$

In general, the $\chi^2_{A_{i_1}A_{i_2}\dots A_{i_j}}$ ($j \geq 2$) term of the chi-squared partition quantifies the association between A_{i_1}, A_{i_2}, \dots and A_{i_j} with $(i_1 - 1)(i_2 - 1) \dots (i_j - 1)$ degrees of freedom and can be used to test the statistical significance of the association between these variables. $\chi^2_{A_{i_1}A_{i_2}\dots A_{i_j}}$ is the aggregate of deviation of A_{i_1}, A_{i_2}, \dots and A_{i_j} from what is expected under complete independence. In fact, this association term involves information about the overall association in terms of the linear, quadratic and higher order components for each of the variables A_{i_1}, A_{i_2}, \dots and A_{i_j} when these variables have ordered categories. Beh and Davy (1999) shows that the terms from the partition of the chi-squared statistic with df larger than 1 can be further partitioned into linear, quadratic and higher order components using orthogonal polynomials. While Beh and Davy (1999) describe the link between this chi-squared partition and the ordinal log-linear model for multi-way tables, the $u_{i_1i_2\dots i_j}$ term in such a model for a multi-way table reflects the *linear* \times *linear* $\times \dots \times$ *linear* association term between the ordered categorical variables A_{i_1}, A_{i_2}, \dots and A_{i_j} and it can be obtained by further partitioning $\chi^2_{A_{i_1}A_{i_2}\dots A_{i_j}}$ using orthogonal polynomials.

3.2 Partition for a Five-Way Contingency Table

Consider a five-way contingency table formed by cross-classifying the five categorical variables, A_1, A_2, A_3, A_4 and A_5 . By following the same process described in Sect. 3.1, Pearson's chi-squared statistic for such a table can be partitioned in 26 terms as follows:

$$\begin{aligned}\chi^2_{Total} = & \chi^2_{A_1A_2} + \chi^2_{A_1A_3} + \chi^2_{A_1A_4} + \chi^2_{A_1A_5} + \chi^2_{A_2A_3} + \chi^2_{A_2A_4} + \chi^2_{A_2A_5} + \chi^2_{A_3A_4} + \chi^2_{A_3A_5} + \chi^2_{A_4A_5} \\ & + \chi^2_{A_1A_2A_3} + \chi^2_{A_1A_2A_4} + \chi^2_{A_1A_2A_5} + \chi^2_{A_1A_3A_4} + \chi^2_{A_1A_3A_5} + \chi^2_{A_1A_4A_5} + \chi^2_{A_2A_3A_4} \\ & + \chi^2_{A_2A_3A_5} + \chi^2_{A_2A_4A_5} + \chi^2_{A_3A_4A_5} + \chi^2_{A_1A_2A_3A_4} + \chi^2_{A_1A_2A_3A_5} + \chi^2_{A_1A_2A_4A_5} \\ & + \chi^2_{A_1A_3A_4A_5} + \chi^2_{A_2A_3A_4A_5} + \chi^2_{A_1A_2A_3A_4A_5}.\end{aligned}\quad (4)$$

Each term of this partition can be used to test the statistical significance of the association between each pair, triplet and quartet of variables as well as all five ordinal variables. Observe that, of these 26 terms, ${}^5C_2 = 10$ terms correspond to two-way association, ${}^5C_3 = 10$ terms correspond to three-way association, ${}^5C_4 = 5$ terms correspond to the four-way association structures and one (5C_5) term reflects the association between all five variables. The formulae for each of these terms are

given in the Appendix with a purpose of showing the complexity involved in the higher order terms.

The number of terms, and their calculation, dramatically increases as the number of variables cross-classified to form a contingency table increases; for $k(> 2)$ categorical variables, the partition of Pearson’s chi-squared statistic will consist of

$$\sum_{l=2}^k \binom{k}{l} = 2^k - (k + 1)$$

terms. Therefore, for a six-way contingency table, the partition of its chi-squared statistic will consist of $2^6 - (6 + 1) = 57$ terms. In the next section, we present the general steps required to obtain the terms of the partition of the chi-squared statistic for a multi-way contingency table consisting of multiple (more than five) variables.

3.3 Partition for a Multi-way Contingency Table

Consider a sample of n individuals that is cross-classified according k categorical variables, A_1, A_2, \dots and A_k where the r^{th} variable consists of j_r categories, for $r = 1, 2, \dots, k$. Consider the following notations:

- $S = \{1, 2, \dots, k\}$
- $\underline{i}_k = (i_1 i_2 \dots i_k)$, where $i_1, i_2, \dots, i_k \in S$, is referred to as a k -tuple.
- $S_r = \{\underline{i}_r = i_1 i_2 \dots i_r \mid i_1 < i_2 < \dots < i_r, i_r \in S\}$ is a set of strictly increasing r -tuples $r = 2, 3, \dots, k$ (S_r contains ${}^k C_r$ tuples).
- $A_{\underline{i}_r} = A_{i_1} A_{i_2} \dots A_{i_r}$ is referred to as an r -variable association term, $\underline{i}_r \in S_r$.
- $n_{\underline{i}_k}$ defines the number of individuals corresponding to \underline{i}_k^{th} category of variables $\{A_{i_1}, A_{i_2}, \dots, A_{i_k}\}$.
- $p_{\underline{i}_k} = \frac{n_{\underline{i}_k}}{n}$ denotes the relative frequency of the \underline{i}_k^{th} category of the set of variables $\{A_{i_1}, A_{i_2}, \dots, A_{i_k}\}$
- $p_{i_r}^r = \sum_{i_s: s \in S - \{r\}} p_{\underline{i}_r}$ is the observed marginal relative frequency of the \underline{i}_r^{th} category of A_r , for $i_r = 1, 2, \dots, j_r$, and $r = 1, 2, \dots, k$.
- $p_{i_{r_1} i_{r_2}}^{r_1 r_2} = \sum_{i_s: s \in S - \{r_1, r_2\}} p_{\underline{i}_k}$ is the observed joint relative frequency of the two variables A_{r_1} and A_{r_2} for $i_{r_1} = 1, 2, \dots, j_{r_1}$ and $i_{r_2} = 1, 2, \dots, j_{r_2}$. (This notation can be extended to represent observed joint relative frequency of more than two variables)

Now, the k -way array $\mathbf{\Pi} = (\pi_{\underline{i}_k})$ with $\pi_{\underline{i}_k} = \frac{p_{\underline{i}_k} - \prod_{r=1}^k p_{i_r}^r}{\prod_{r=1}^k p_{i_r}^r}$ can be partitioned into weighted orthogonal components such that

$$\Pi = \sum_{i_2 \in S_2} \Pi_{i_2} + \sum_{i_3 \in S_3} \alpha \Pi_{i_3} + \dots + \sum_{i_{k-1} \in S_{k-1}} \alpha \Pi_{i_{k-1}} + \alpha \Pi_{i_k} \tag{5}$$

where $\Pi_{i_2} = (\pi_{i_2})$, $\alpha \Pi_{i_3} = (\alpha \pi_{i_3})$, ..., $\alpha \Pi_{i_k} = (\alpha \pi_{i_k})$, $\pi_{i_{r_1} i_{r_2}} = \left(\frac{p_{i_{r_1} i_{r_2}}^{r_1 r_2} - p_{i_{r_1}}^{r_1} p_{i_{r_2}}^{r_2}}{p_{i_{r_1}}^{r_1} p_{i_{r_2}}^{r_2}} \right)$, for $(i_{r_1} i_{r_2}) \in S_2$, $\alpha \pi_{i_{r_1} i_{r_2} i_{r_3}} = \left(\frac{p_{i_{r_1} i_{r_2} i_{r_3}}^{r_1 r_2 r_3} - \alpha p_{i_{r_1} i_{r_2} i_{r_3}}^{r_1 r_2 r_3}}{p_{i_{r_1}}^{r_1} p_{i_{r_2}}^{r_2} p_{i_{r_3}}^{r_3}} \right)$ for $(i_{r_1} i_{r_2} i_{r_3}) \in S_3$ and so on.

In general,

$$\alpha \pi_{i_j} = \left(\frac{p_{i_j}^{r_j} - \alpha p_{i_j}^{r_j}}{\prod_{u=1}^j p_{i_u}^{r_u}} \right), i_j = (i_{r_1} i_{r_2} \dots i_{r_j}) \in S_j, j = 3, 4, \dots, k$$

The weighted norm of Π_{i_2} is the inertia corresponding to the two-way association terms of A_{i_2} , $i_2 \in S_2$ while the weighted norm of $\alpha \Pi_{i_3}$ is the inertia corresponding to three-way association terms of A_{i_3} , $i_3 \in S_3$. We refer to the above method as the traditional method. We demonstrate the complexity in this method through the step-wise algorithm for partitioning the chi-squared statistic for four-way table. This will also help to make the notations more familiar.

Let $N = (n_{i_1 i_2 i_3 i_4})$ and $p_{i_1 i_2 i_3 i_4} = n_{i_1 i_2 i_3 i_4} / n$ be a $j_1 \times j_2 \times j_3 \times j_4$ contingency table and array of relative frequencies, respectively. Here $S = \{1, 2, 3, 4\}$ and S_2, S_3, S_4 represent the set of tuples corresponding to two-way, three-way and four-way association terms respectively:

- $S_2 = \{i_1 i_2, i_1 i_3, i_1 i_4, i_2 i_3, i_2 i_4, i_3 i_4\}$ corresponds to $\{A_1 A_2, A_1 A_3, A_1 A_4, A_2 A_3, A_2 A_4, A_3 A_4\}$ to
- $S_3 = \{i_1 i_2 i_3, i_1 i_2 i_4, i_1 i_3 i_4, i_2 i_3 i_4\}$ corresponds to $\{A_1 A_2 A_3, A_1 A_2 A_4, A_1 A_3 A_4, A_2 A_3 A_4\}$ to
- $S_4 = \{i_4 = i_1 i_2 i_3 i_4\}$ corresponds to $\{A_1 A_2 A_3 A_4\}$

The marginal relative frequencies for A_1, A_2, A_3 and A_4 are,

$$p_{i_1}^1 = p_{i_1 \dots}, i_1 = 1, 2, \dots, j_1 \quad p_{i_2}^2 = p_{i_2 \dots}, i_2 = 1, 2, \dots, j_2$$

$$p_{i_3}^3 = p_{i_3 \dots}, i_3 = 1, 2, \dots, j_3 \quad p_{i_4}^4 = p_{i_4 \dots}, i_4 = 1, 2, \dots, j_4.$$

The marginal relative frequencies for all pairs of variables are,

$$p_{i_1 i_2}^{12} = p_{i_1 i_2 \dots}, \quad i_1 = 1, 2, \dots, j_1, \quad i_2 = 1, 2, \dots, j_2,$$

$$p_{i_1 i_3}^{13} = p_{i_1 i_3 \dots}, \quad i_1 = 1, 2, \dots, j_1, \quad i_3 = 1, 2, \dots, j_3,$$

$$\dots$$

$$p_{i_3 i_4}^{34} = p_{i_3 i_4 \dots}, \quad i_3 = 1, 2, \dots, j_3, \quad i_4 = 1, 2, \dots, j_4,$$

The marginal relative frequencies for all triplets of variables are

$$\begin{aligned}
 p_{i_1 i_2 i_3}^{123} &= p_{i_1 i_2 i_3}, & i_1 &= 1, 2, \dots, j_1, & i_2 &= 1, 2, \dots, j_2, & i_3 &= 1, 2, \dots, j_3, \\
 p_{i_1 i_2 i_4}^{124} &= p_{i_1 i_2 i_4}, & i_1 &= 1, 2, \dots, j_1, & i_2 &= 1, 2, \dots, j_2, & i_4 &= 1, 2, \dots, j_4, \\
 &\dots & & & & & & \\
 p_{i_2 i_3 i_4}^{234} &= p_{i_2 i_3 i_4}, & i_2 &= 1, 2, \dots, j_2, & i_3 &= 1, 2, \dots, j_3, & i_4 &= 1, 2, \dots, j_4,
 \end{aligned}$$

The steps to obtain the partition of Pearson’s chi-squared statistic using traditional method are:

- (i) Calculate the $j_1 \times j_2 \times j_3 \times j_4$ array $\mathbf{\Pi} = (\pi_{i_4})$ where $\pi_{i_4} = \frac{p_{i_4}}{p_{i_1}^2 p_{i_2}^2 p_{i_3}^2 p_{i_4}^4} - 1$.
- (ii) Calculate the 4-way arrays $\mathbf{\Pi}_{i_2}$ for $i_2 \in S_2$ as follows:

$$\begin{aligned}
 (\mathbf{\Pi}_{i_2})_{i_4} &= \frac{p_{i_1 i_2}^{12}}{p_{i_1}^1 p_{i_2}^2} - 1, & (\mathbf{\Pi}_{i_3})_{i_4} &= \frac{p_{i_1 i_3}^{13}}{p_{i_1}^1 p_{i_3}^3} - 1, \dots \\
 (\mathbf{\Pi}_{i_3 i_4})_{i_4} &= \frac{p_{i_3 i_4}^{34}}{p_{i_3}^3 p_{i_4}^4} - 1, & i_r &= 1, 2, \dots, j_r, & r &= 1, 2, \dots, 4.
 \end{aligned}$$

- (iii) Calculate the 4-way arrays $\alpha \mathbf{\Pi}_{i_3}$ for $i_3 \in S_3$ as follows:

$$\begin{aligned}
 (\alpha \mathbf{\Pi}_{i_1 i_2 i_3})_{i_4} &= \frac{p_{i_1 i_2 i_3}^{123} - \alpha p_{i_1 i_2 i_3}^{123}}{p_{i_1}^1 p_{i_2}^2 p_{i_3}^3}, & (\alpha \mathbf{\Pi}_{i_1 i_2 i_4})_{i_4} &= \frac{p_{i_1 i_2 i_4}^{124} - \alpha p_{i_1 i_2 i_4}^{124}}{p_{i_1}^1 p_{i_2}^2 p_{i_4}^4}, \\
 &\dots & & \\
 (\alpha \mathbf{\Pi}_{i_2 i_3 i_4})_{i_4} &= \frac{p_{i_2 i_3 i_4}^{234} - \alpha p_{i_2 i_3 i_4}^{234}}{p_{i_2}^2 p_{i_3}^3 p_{i_4}^4}, & i_r &= 1, 2, \dots, j_r, & r &= 1, 2, \dots, 4.
 \end{aligned}$$

where

$$\begin{aligned}
 \alpha p_{i_1 i_2 i_3}^{123} &= p_{i_1 i_2}^{12} p_{i_3}^3 + p_{i_1 i_3}^{13} p_{i_2}^2 + p_{i_2 i_3}^{23} p_{i_1}^1 - 2 p_{i_1}^1 p_{i_2}^2 p_{i_3}^3, \\
 \alpha p_{i_1 i_2 i_4}^{124} &= p_{i_1 i_2}^{12} p_{i_4}^4 + p_{i_1 i_4}^{14} p_{i_2}^2 + p_{i_2 i_4}^{24} p_{i_1}^1 - 2 p_{i_1}^1 p_{i_2}^2 p_{i_4}^4, \\
 \alpha p_{i_2 i_3 i_4}^{234} &= p_{i_2 i_3}^{23} p_{i_4}^4 + p_{i_2 i_4}^{24} p_{i_3}^3 + p_{i_3 i_4}^{34} p_{i_2}^2 - 2 p_{i_2}^2 p_{i_3}^3 p_{i_4}^4, \\
 \alpha p_{i_1 i_3 i_4}^{134} &= p_{i_1 i_3}^{13} p_{i_4}^4 + p_{i_1 i_4}^{14} p_{i_3}^3 + p_{i_3 i_4}^{34} p_{i_1}^1 - 2 p_{i_1}^1 p_{i_3}^3 p_{i_4}^4.
 \end{aligned}$$

- (iv) Calculate the 4-way array $\alpha \mathbf{\Pi}_{i_4}$ for $i_4 \in S_4$ as follows:

$$(\alpha \mathbf{\Pi}_{i_1 i_2 i_3 i_4})_{i_4} = \frac{p_{i_4} - \alpha p_{i_1 i_2 i_3 i_4}^{1234}}{p_{i_1}^1 p_{i_2}^2 p_{i_3}^3 p_{i_4}^4},$$

where

$$\begin{aligned}
 \alpha p_{i_1 i_2 i_3 i_4}^{1234} = & p_{i_1 i_2}^{12} p_{i_3}^3 p_{i_4}^4 + p_{i_1 i_3}^{13} p_{i_2}^2 p_{i_4}^4 + p_{i_1 i_4}^{14} p_{i_2}^2 p_{i_3}^3 + p_{i_2 i_3}^{23} p_{i_1}^1 p_{i_4}^4 \\
 & + p_{i_2 i_4}^{24} p_{i_1}^1 p_{i_3}^3 + p_{i_3 i_4}^{34} p_{i_1}^1 p_{i_2}^2 + p_{i_1 i_2 i_3}^{123} p_{i_4}^4 + p_{i_1 i_2 i_4}^{124} p_{i_3}^3 \\
 & + p_{i_2 i_3 i_4}^{234} p_{i_1}^1 + p_{i_1 i_3 i_4}^{134} p_{i_2}^2 + 9 p_{i_1}^1 p_{i_2}^2 p_{i_3}^3 p_{i_4}^4.
 \end{aligned}$$

(xxii) Calculate the chi-squared partition as follows:

$$\begin{aligned}
 \chi_{Total}^2 &= n \|\mathbf{\Pi}\|^2 = n \sum_{i_4} p_{i_1}^1 p_{i_2}^2 p_{i_3}^3 p_{i_4}^4 \mathbf{\Pi}_{i_4}^2, \\
 \chi_{A_1 A_2}^2 &= n \|\mathbf{\Pi}_{i_1 i_2}\|^2 = n \sum_{i_4} p_{i_1}^1 p_{i_2}^2 (\mathbf{\Pi}_{i_1 i_2})_{i_4}^2, \\
 \chi_{A_1 A_3}^2 &= n \|\mathbf{\Pi}_{i_1 i_3}\|^2 = n \sum_{i_4} p_{i_1}^1 p_{i_3}^3 (\mathbf{\Pi}_{i_1 i_3})_{i_4}^2, \dots \\
 \chi_{A_3 A_4}^2 &= n \|\mathbf{\Pi}_{i_3 i_4}\|^2 = n \sum_{i_4} p_{i_3}^3 p_{i_4}^4 (\mathbf{\Pi}_{i_3 i_4})_{i_4}^2, \\
 \chi_{A_1 A_2 A_3}^2 &= n \|\alpha \mathbf{\Pi}_{i_1 i_2 i_3}\|^2 = n \sum_{i_4} p_{i_1}^1 p_{i_2}^2 p_{i_3}^3 (\alpha \mathbf{\Pi}_{i_1 i_2 i_3})_{i_4}^2, \\
 \dots & \\
 \chi_{A_1 A_3 A_4}^2 &= n \|\alpha \mathbf{\Pi}_{i_1 i_3 i_4}\|^2 = n \sum_{i_4} p_{i_1}^1 p_{i_3}^3 p_{i_4}^4 (\alpha \mathbf{\Pi}_{i_1 i_3 i_4})_{i_4}^2 \\
 \text{and } \chi_{A_1 A_2 A_3 A_4}^2 &= n \|\alpha \mathbf{\Pi}_{i_1 i_2 i_3 i_4}\|^2 = n \sum_{i_4} p_{i_1}^1 p_{i_2}^2 p_{i_3}^3 p_{i_4}^4 (\alpha \mathbf{\Pi}_{i_1 i_2 i_3 i_4})_{i_4}^2.
 \end{aligned}$$

□

Observe that the term $p_{i_1 i_2 i_3}^{r_1 r_2 r_3}$ in ${}_{\alpha} \pi_{i_1 i_2 i_3}$ is simply a marginal of relative frequencies while $\alpha p_{i_1 i_2 i_3}^{r_1 r_2 r_3}$ is a function of all univariate and bivariate marginals. In general, the term $\alpha p_{i_1 i_2 \dots i_j}^{r_j} = \alpha p_{i_1 i_2 \dots i_j}^{r_1 r_2 \dots r_j}$ in ${}_{\alpha} \pi_j = \alpha \pi_{i_1 i_2 \dots i_j}$ is based on the sum-of-products of j univariate, ${}^j C_2$ bivariate up to ${}^j C_{j-1}$ ($j - 1$)-variate marginal proportions. The explicit form of $\alpha p_{i_1 i_2 \dots i_j}^{r_j}$ for $j \geq 6$ will be far more complex; this is apparent by just observing the case where $j = 4, 5$. A natural question then arises – if one is not able to formulate an expression for $\alpha p_{i_1 i_2 \dots i_j}^{r_j}$ for $j \geq 6$ then how can one quantify the terms and implement the calculation of the partition using software? Another issue is the rapidly increasing number of association terms as the number of

categorical variables increases. The partitioning of Pearson's chi-squared statistic for a multi-way table is theoretically and computationally challenging by the method described above. Section 4 will present a new algorithm for dealing with this complexity.

4 A New Algorithm for Partitioning Pearson Chi-squared Statistic

In this section, we present a new and efficient method of partitioning Pearson's chi-squared statistic for multi-way tables which overcomes the theoretical complexities and simplifies the computational difficulties described in Sect. 3. The proposed method gives all components of the partition without the need to evaluate the terms of ${}_a p_{i_r}^j$ for $j \geq 3$. We describe this method for a four-way table and demonstrate how it can be extended to multi-way tables formed from the cross-classification of at least five categorical variables. We present the terms of the partition in the form of categorical ANOVA and show that they can be used for testing the significance of association.

4.1 An Algorithm for a Four-Way Contingency Table

Evaluating the terms of the partition of Pearson's chi-squared statistic for a four-way contingency table involves calculating all 3-way and 4-way association terms of ${}_a p_{i_1 i_2 i_3}$, ${}_a p_{i_1 i_2 i_4}$, ${}_a p_{i_2 i_3 i_4}$, ${}_a p_{i_1 i_3 i_4}$, ${}_a p_{i_1 i_2 i_3 i_4}$. These terms are functions of the marginal relative frequencies of all univariate, bivariate and trivariate subsets of categorical variables. The calculation of the relative marginal proportions for a given set of variables is a feasible task. However, the calculation of the ${}_a p$'s is computationally intractable nor is there a mathematical link between ${}_a p_{i_1 i_2 i_3}$, ${}_a p_{i_1 i_2 i_4}$, ${}_a p_{i_2 i_3 i_4}$, ${}_a p_{i_1 i_3 i_4}$ and ${}_a p_{i_1 i_2 i_3 i_4}$. We now present an algorithm that will provide an easy and fast means of calculating these terms, thereby providing a more efficient way of partitioning Pearson's chi-squared statistic.

Calculate the total inertia χ_{Total}^2 for a four-way contingency table as

$$\chi_{T,A_1 A_2 A_3 A_4}^2 = n \sum_{i_1, i_2, i_3, i_4} \frac{\left(p_{i_1 i_2 i_3 i_4} - \prod_{j=1}^4 p_{i_j}^j \right)^2}{\prod_{j=1}^4 p_{i_j}^j}.$$

Calculate the inertia values corresponding to all of the two-way association terms $\chi_{A_1 A_2}^2$, $\chi_{A_1 A_3}^2$, $\chi_{A_1 A_4}^2$, $\chi_{A_2 A_3}^2$, $\chi_{A_2 A_4}^2$ and $\chi_{A_3 A_4}^2$ of the partition using

$$\chi_{A_{r_1} A_{r_2}}^2 = n \sum_{i_{r_1}, i_{r_2}} \frac{\left(p_{i_{r_1} i_{r_2}}^{r_1 r_2} - p_{i_{r_1}}^{r_1} p_{i_{r_2}}^{r_2} \right)^2}{p_{i_{r_1}}^{r_1} p_{i_{r_2}}^{r_2}}, \quad (r_1, r_2) \in S_2;$$

- Rather than calculating $\alpha P_{i_1 i_2 i_3}$, $\alpha P_{i_1 i_2 \cdot i_4}$, $\alpha P_{\cdot i_2 i_3 i_4}$ and $\alpha P_{i_1 \cdot i_3 i_4}$ when computing $\chi_{A_1 A_2 A_3}^2$, $\chi_{A_1 A_2 A_4}^2$, $\chi_{A_2 A_3 A_4}^2$ and $\chi_{A_1 A_3 A_4}^2$, calculate the total inertia values $\chi_{T, A_1 A_2 A_3}^2$, $\chi_{T, A_2 A_3 A_4}^2$, $\chi_{T, A_1 A_3 A_4}^2$ and $\chi_{T, A_1 A_2 A_4}^2$ corresponding to all of the three-way association terms where

$$\chi_{T, A_1 A_2 A_3}^2 = n \sum_{i_1, i_2, i_3} \frac{(p_{i_1 i_2 i_3} - p_{i_1}^1 p_{i_2}^2 p_{i_3}^3)^2}{p_{i_1}^1 p_{i_2}^2 p_{i_3}^3},$$

$$\chi_{T, A_2 A_3 A_4}^2 = n \sum_{i_2, i_3, i_4} \frac{(p_{\cdot i_2 i_3 i_4} - p_{i_2}^2 p_{i_3}^3 p_{i_4}^4)^2}{p_{i_2}^2 p_{i_3}^3 p_{i_4}^4}$$
 and so on.
- Now calculate the inertia values corresponding to all the three-way association terms using

$$\begin{aligned} \chi_{A_1 A_2 A_3}^2 &= \chi_{T, A_1 A_2 A_3}^2 - \chi_{A_1 A_2}^2 - \chi_{A_1 A_3}^2 - \chi_{A_2 A_3}^2, \\ \chi_{A_2 A_3 A_4}^2 &= \chi_{T, A_2 A_3 A_4}^2 - \chi_{A_2 A_3}^2 - \chi_{A_2 A_4}^2 - \chi_{A_3 A_4}^2, \\ \chi_{A_1 A_3 A_4}^2 &= \chi_{T, A_1 A_3 A_4}^2 - \chi_{A_1 A_3}^2 - \chi_{A_1 A_4}^2 - \chi_{A_3 A_4}^2, \\ \chi_{A_1 A_2 A_4}^2 &= \chi_{T, A_1 A_2 A_4}^2 - \chi_{A_1 A_2}^2 - \chi_{A_1 A_4}^2 - \chi_{A_2 A_4}^2. \end{aligned}$$

- Calculate the total inertia corresponding to the four-way association term as,

$$\begin{aligned} \chi_{A_1 A_2 A_3 A_4}^2 &= \chi_{T, A_1 A_2 A_3 A_4}^2 - \left(\chi_{A_1 A_2 A_3}^2 + \chi_{A_1 A_2 A_4}^2 + \chi_{A_2 A_3 A_4}^2 + \chi_{A_1 A_3 A_4}^2 \right) \\ &\quad - \left(\chi_{A_1 A_2}^2 + \chi_{A_1 A_3}^2 + \chi_{A_1 A_4}^2 + \chi_{A_2 A_3}^2 + \chi_{A_2 A_4}^2 + \chi_{A_3 A_4}^2 \right) \end{aligned}$$

This algorithm avoids the complexity in the evaluation of the chi-squared partition. We now present the structure of the algorithm for a multi-way contingency table.

4.2 An Algorithm for a Multi-way Contingency Table

Suppose we now consider a k -way contingency table formed from the cross-classification of the categorical variables A_1, A_2, \dots, A_k . An ANOVA-like partition of the chi-squared statistic for such a table is given as follows:

- (i) Calculate Pearson’s chi-squared statistic $\chi_T^2 = n\Phi^2$ where $\Phi^2 = \sum_{i_k} \frac{(p_{i_k} - \prod_{j=1}^k p_{i_j}^j)^2}{\prod_{j=1}^k p_{i_j}^j}$
- (ii) Calculate the chi-squared statistic, $\chi_{A_{r_1} A_{r_2}}^2$ corresponding to all the ${}^k C_2$ two-way contingency tables by considering each pair of variables, such that $\chi_{A_{r_1} A_{r_2}}^2 = n \sum_{i_{r_1}, i_{r_2}} \frac{(p_{i_{r_1} i_{r_2}}^{r_1 r_2} - p_{i_{r_1}}^{r_1} p_{i_{r_2}}^{r_2})^2}{p_{i_{r_1}}^{r_1} p_{i_{r_2}}^{r_2}}, (r_1, r_2) \in S_2$.

- (iii) Calculate Pearson's total chi-squared statistic corresponding to a set of ${}^k C_r$ variables, $\{A_{i_1}, A_{i_2}, \dots, A_{i_r}\}$ such that $\chi_{T, A_{i_1} A_{i_2} \dots A_{i_r}}^2 = n \sum_{i_r} \frac{(p_{i_r} - \prod_{j=1}^r p_{i_j}^j)^2}{\prod_{j=1}^r p_{i_j}^j}$ for $i_r \in S_r, r = 3, 4, \dots, k - 1$.
- (iv) Calculate the contribution of j -way association term $A_{i_1} A_{i_2} \dots A_{i_j}$ in the partition of Pearson's chi-squared statistic is given by,

$$\chi_{A_{i_1} A_{i_2} \dots A_{i_j}}^2 = \chi_{T, A_{i_1} A_{i_2} \dots A_{i_j}}^2 - \sum_{r=2}^{j-1} \sum_{i_r \in S_r} \chi_{A_{i_1} A_{i_2} \dots A_{i_r}}^2, j = 3, 4, \dots, k - 1.$$

Thus, to calculate the contribution of any of the association terms from the partition of Pearson's chi-squared statistic, we need to calculate all two-way and higher-way association terms. This task is much simpler than calculating all components of $\mathbf{\Pi}$ separately through ${}_{\alpha} \pi_{i_j}$ and ${}_{\alpha} p_{i_j}^r$ for $i_j \in S_j$ for $j = 3, 4, \dots, k$.

4.3 Categorical ANOVA for a Multi-way Table

Consider a k -way contingency table consisting of ordered variables. The partition of Pearson's chi-squared statistic can be presented in the form of categorical ANOVA where the statistical significance of each association term can be tested using a chi-squared test of independence with degrees of freedom dependent on the term under consideration. Table 1 gives the general ANOVA table for the analysis of association for k categorical variables.

The practical implementation of the proposed algorithm can be further simplified using matrix-based software packages such as Matlab and Mathematica. Furthermore, the evaluation of the univariate and all multivariate association terms can be determined using tensors, which are akin to a multi-dimensional array. As we shall see, a n -mode product of a tensor with a matrix can be evaluated to obtain the $\mathbf{\Pi}$ -arrays and their norms. The tensor product simplifies the implementation of the partitioning algorithm described earlier in this section. In Sect. 5, we shall briefly introduce tensors and tensor products and briefly describe how they can be used to obtain the partition of Pearson's chi-squared statistic for a multi-way contingency table.

5 Tensor-Matrix Product and the Partition for a Multi-way Table

A tensor decomposition has been widely used in analysis of multi-way data; see, for example Kolda and Bader (2009) and the references mentioned therein. Lathauwer et al. (2000) investigated the problem of generalising the singular value decomposition (SVD) of matrices to higher-order tensors. They do this by defining scalar product of tensors, Frobenius-norm of a tensor and tensor product

Table 1 Categorical ANOVA for k variables

Order of association	Association term	χ^2 value	Degrees of freedom
Two-way	A_1A_2	$\chi^2_{A_1A_2}$	$(j_1 - 1)(j_2 - 1)$
	A_1A_3	$\chi^2_{A_1A_3}$	$(j_1 - 1)(j_3 - 1)$
	\vdots	\vdots	\vdots
	A_1A_k	$\chi^2_{A_1A_k}$	$(j_1 - 1)(j_k - 1)$
	A_2A_3	$\chi^2_{A_2A_3}$	$(j_2 - 1)(j_3 - 1)$
	\vdots	\vdots	\vdots
	A_2A_k	$\chi^2_{A_2A_k}$	$(j_2 - 1)(j_k - 1)$
	\vdots	\vdots	\vdots
	$A_{k-1}A_k$	$\chi^2_{A_{k-1}A_k}$	$(j_{k-1} - 1)(j_k - 1)$
Three-way	$A_1A_2A_3$	$\chi^2_{A_1A_2A_3}$	$\prod_{k=1}^3 (j_r - 1)$
	\vdots	\vdots	\vdots
	$A_{k-2}A_{k-1}A_k$	$\chi^2_{A_{k-2}A_{k-1}A_k}$	$\prod_{r=k-2}^k (j_r - 1)$
	\vdots	\vdots	\vdots
k -way	$A_1A_2 \dots A_k$	$\chi^2_{A_1A_2 \dots A_k}$	$\prod_{r=1}^k (j_r - 1)$
	Total	χ^2_T	$\prod_{r=1}^k j_r - \sum_{r=1}^k (j_r - 1) - 1$

with a matrix. The n -mode product of N -way tensor with a matrix is defined as follows.

Definition. The n -mode product of tensor $\mathcal{A} = (a_{i_1 i_2 \dots i_n \dots i_N}) \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_n \times \dots \times I_N}$ (i.e. a tensor of dimension $I_1 \times I_2 \times \dots \times I_n \times \dots \times I_N$) by a matrix $\mathbf{B} = (b_{j_n}) \in \mathbb{C}^{J \times I_n}$ is denoted by $\mathcal{A} \times_n \mathbf{B}$, and is an $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$ -tensor $\mathcal{C} = (c_{i_1 i_2 \dots i_{n-1} j_n i_{n+1} \dots i_N})$ such that

$$c_{i_1 i_2 \dots i_{n-1} j_n i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} a_{i_1 i_2 \dots i_n \dots i_N} b_{j_n i_n}.$$

This n -mode product has two interesting properties that are pertinent to the development of our algorithm:

(a) Given a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ and the matrices $\mathbf{B} \in \mathbb{R}^{J_n \times I_n}$ and $\mathbf{C} \in \mathbb{R}^{J_m \times I_m}$, ($n \neq m$)

$$(\mathcal{A} \times_n \mathbf{B}) \times_m \mathbf{C} = (\mathcal{A} \times_m \mathbf{C}) \times_n \mathbf{B} = \mathcal{A} \times_n \mathbf{B} \times_m \mathbf{C}.$$

(b) Given a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ and matrices $\mathbf{B} \in \mathbb{R}^{J_n \times I_n}$ and $\mathbf{C} \in \mathbb{R}^{K_n \times J_n}$,

$$(\mathcal{A} \times_n \mathbf{B}) \times_n \mathbf{C} = \mathcal{A} \times_n (\mathbf{CB})$$

We discuss the tensor, the n -mode product of a tensor with a matrix and demonstrate how they can be used to evaluate all of the association terms of the partition of Pearson’s chi-squared statistic through $\mathbf{\Pi}$ and $\mathbf{\Pi}_{i_j}$, $i_j \in S_j, j = 2, 3, \dots, k - 1$. The

(unweighted) norm of \mathcal{A} is $\|\mathcal{A}\| = (\sum_{i_1, i_2, \dots, i_k} a_{i_1 i_2 \dots i_k}^2)^{1/2}$ while, for a $j_1 \times j_2 \times \dots \times j_k$ tensor \mathcal{A} , we define its weighted norm as

$$\|\mathcal{A}\|_w = \left(\sum_{i_1, i_2, \dots, i_k} (a_{i_1 i_2 \dots i_k}^2) p_{i_1}^1 p_{i_2}^2 \dots p_{i_k}^k \right)^{\frac{1}{2}}$$

where $p_{i_1}^1, p_{i_2}^2, \dots, p_{i_k}^k$ are the expected values of the joint cell proportions under independence between the variables $A_{i_1}, A_{i_2}, \dots, A_{i_k}$.

For k ordered categorical variables, N, \mathbf{P} and $\hat{\mathbf{p}} = (\prod_{j=1}^k p_{i_j}^j)$ can all be expressed as a $j_1 \times j_2 \times \dots \times j_k$ dimensional tensor while $\mathbf{D}_r = \text{diag}(p_{i_1}^r, p_{i_2}^r, \dots, p_{i_{j_r}}^r)$ is a $j_r \times j_r$ diagonal matrix for the r^{th} variable A_{i_r} .

To obtain an additive partition of Pearson’s chi-squared statistic, we provide a simple method to obtain $\mathbf{\Pi}$ and $\mathbf{\Pi}_{i_j}$, for $i_j \in S_j, j = 2, 3, \dots, k - 1$ when they can be considered as tensors. The weighted norm of each $\mathbf{\Pi}_{i_j}$ gives the total inertia corresponding to the association A_{i_j} which can be further used to obtain inertia corresponding to the association between the variables $A_{i_1}, A_{i_2}, \dots, A_{i_j}$. The algorithm can be expressed in the form of tensors, and its properties, as follows:

i) Calculate $\mathbf{\Pi} = (\mathbf{P} - \hat{\mathbf{P}}) \times_1 \mathbf{D}_1^{-1} \times_2 \mathbf{D}_2^{-1} \times_3 \dots \times_k \mathbf{D}_k^{-1}$. Then the total inertia of the contingency table can be expressed as the weighted norm.

$$\Phi^2 = \|\mathbf{\Pi}\|_w^2 = \|\mathbf{\Pi} \times_{r=1}^k \mathbf{D}_r^{1/2}\|^2.$$

ii) At this step we determine the two-way, three-way, ..., $(j - 1)$ -way association terms from the partition. Therefore, set $j = k - 1$. Then a set S_j consists of ${}^k C_j$ tuples and $A_{i_1} A_{i_2} \dots A_{i_j}, i_j \in S_j$ is the corresponding association term. Let $S_{k-j} = \{i_1, i_2, \dots, i_{k-j} \mid i_r \in S_k \text{ but not in } S_j\}$. For each of the $i_j \in S_j$, determine the j -dimensional tensor $N_{i_j} = N \times_{i_r \in S_{k-j}} E_{1 i_r}$ (tensor of cell frequencies), P_{i_j} (tensor of relative cell frequencies) and \hat{p}_{i_j} (tensor of expected relative frequencies under complete independence) where E_{mn} is $m \times n$ matrix with all elements 1. Then

$$\mathbf{\Pi}_{i_j} = (P_{i_j} - \hat{p}_{i_j}) \times_{i_r \in S_j} \mathbf{D}_i^{-1}.$$

iii) From step ii), calculate $\chi_{T, A_{i_1} A_{i_2} \dots A_{i_j}}^2 = \|\mathbf{\Pi}_{i_j}\|_w^2 = \|\mathbf{\Pi}_{i_j} \times_{i_r \in S_j} \mathbf{D}_i^{1/2}\|^2$

iv) Repeat the steps (ii) and (iii) for $j = 2, 3, \dots, k - 2$ and set $\chi_{T, A_{i_1} A_{i_2}}^2 = \chi_{A_{i_1} A_{i_2}}^2$ for $i_1 i_2 \in S_2$.

v) Now that each two-way, three-way, ..., $(j - 1)$ -way association term is calculated using the above steps, the j -way association term, $A_{i_1} A_{i_2} \dots A_{i_j}$, from the partition of Pearson’s chi-squared statistic can be calculated by

$$\chi_{A_{i_1} A_{i_2} \dots A_{i_j}}^2 = \chi_{T, A_{i_1} A_{i_2} \dots A_{i_j}}^2 - \sum_{r=2}^{j-1} \sum_{i_r \in S_r} \chi_{A_{i_1} A_{i_2} \dots A_{i_r}}^2, \quad j = 3, 4, \dots, k - 1$$

The next section demonstrates the applicability of this algorithm.

6 Numerical Results

In this section, we present a practical demonstration of the partition of Pearson's chi-squared statistic for a four-way table and perform tests on the statistical significance of each of the association terms. Matlab code (not given here) is used to execute the tensor based algorithm described in Sect. 5. We also develop Matlab code for the analysis of a four-way and five-way table using the traditional method of partitioning discussed in Sect. 3. Furthermore, we compare the execution time (in seconds) of the traditional method and the new algorithm through empirical and simulated multi-way contingency tables.

6.1 PISA Study

Consider the data summarised in Table 2. It was obtained from the 2003 PISA (Programme for International Student Assessment) study of a student's gender (A), their country of origin (B) and their response to two survey questions (C, D). The two survey questions are as follows.

Q1. "I am just not good at math" and

Q2. "In math class the teacher continues teaching until the students understand."

This data was also analysed by Azen and Walker (2011) and consists of the following categorical variables:

Variable label	Variable	No. of categories	Categories of the variable
A	Gender	2	Female, male
B	Country	3	Australia, France, United States
C	Response to Q1	4	Strongly agree, agree, disagree, strongly disagree
D	Response to Q2	4	Every lesson, most lessons, some lessons, hardly/never

Consider the four-way contingency table of Table 2 which is of dimension $2 \times 3 \times 4 \times 4$. We assume that all the categorical variables are ordered variables.

Pearson's chi-squared statistic of this table is 1918.2484 and, with 86 degrees of freedom, has an extremely small p -value (< 0.0001). This suggests that there exists an association between at least two of the four variables, although it does not reveal

Table 2 Four-way contingency table of student responses from the PISA study

Gender	Country	Self-efficacy for Q1	No. of students whose response to Q2 is				Total
			Every lesson	Most lessons	Some lessons	Hardly/never	
Female	Australia	Strongly agree	208	211	223	116	758
		Agree	462	579	418	135	1594
		Disagree	1165	1139	552	150	3006
		Strongly Disagree	350	209	86	40	685
	France	Strongly agree	110	128	137	84	459
		Agree	142	160	136	71	509
		Disagree	283	296	217	83	879
		Strongly Disagree	130	88	86	28	332
	United States	Strongly agree	123	108	108	64	403
		Agree	210	213	171	58	652
		Disagree	464	396	198	63	1121
		Strongly disagree	228	133	72	18	451
Male	Australia	Strongly agree	91	136	103	77	407
		Agree	282	495	331	106	1214
		Disagree	963	1446	694	203	3306
		Strongly Disagree	524	391	197	53	1165
	France	Strongly Agree	60	69	62	56	247
		Agree	95	139	102	51	387
		Disagree	233	327	216	102	878
		Strongly Disagree	153	138	89	38	418
	United States	Strongly Agree	83	66	80	47	276
		Agree	155	177	150	57	539
		Disagree	409	474	241	68	1192
		Strongly Disagree	272	186	104	33	595

which bivariate, trivariate or four-way association terms are dominant or statistically significant. To examine further the structure of this association, we partition Pearson's statistic in accordance with (3).

Table 3 Categorical ANOVA for contingency table of student responses from the PISA study

Association term	X^2 value	% of X^2 value	DF	p value
AB	13.0503	0.68	2	0.0015
AC	413.1086	21.54	3	<0.0001
AD	60.0572	3.13	3	<0.0001
BC	290.3186	15.13	6	<0.0001
BD	202.8416	10.57	6	<0.0001
CD	813.8252	42.43	9	<0.0001
ABC	16.2139	0.85	6	0.0127
ABD	9.4708	0.49	6	0.1488
ACD	48.5608	2.53	9	<0.0001
BCD	34.5523	1.80	18	0.0108
ABCD	16.2491	0.85	18	0.5752
Total	1918.2484	100	86	<0.0001

As we discussed in Sect. 5, we use the n -mode products to obtain all Π -arrays and calculate their weighted norms and do so using Matlab. The 11 terms of (3) are evaluated using the traditional method described in Sect. 3 and the new algorithm described in Sect. 4. Both approaches yield the same terms of the partition (which are summarised in Table 3), however the new algorithm performed the calculations in 0.0320 s while the traditional method took twice as long – 0.0631 s—such time-scales may be small, and on the surface insignificant, but it does highlight improvements in the efficiency of the partition of Pearson’s statistic when compared with the procedure in Sect. 3.

Table 3 also summarises the percentage contribution each term makes to the chi-squared statistic, the degrees of freedom (DF) of the term and its p -value. It shows that, for the variables associated with Table 2, all pair-wise association terms from the partition of its chi-squared statistic are statistically significant, the largest of which is the term CD which describes the association between the responses to Question 1 and Question 2. Such an association contributes to nearly half (42.43%) of the association structure in the contingency table. In fact, the participants’ response to Question 2 is not significantly associated with their gender or country of origin (term ABD) while the only statistically significant trivariate association term is between the gender of those who participated and their response to Question 1 and Question 2 (term ACD). Gaining more insight into how gender and the responses to the questions are associated can be further investigated using CA and, more generally, multi-way data analysis (Greenacre 1984; Lebart et al. 1984; Kroonenberg 2008; Beh and Lombardo 2014).

6.2 Computational Speed

We now turn our attention to assessing the computational speed of the new algorithm when compared to the approach outlined in Sect. 3 and the approach by Loisel and Takane (2016). This is achieved through the analysis of real and simulated

contingency tables (or high-dimensional arrays for the purposes of using tensors as described in Sect. 5) that are simulated from a discrete uniform distribution on the interval [7, 100]. Such a range is chosen to ensure that each of the expected cell frequencies of the randomly generated contingency table exceeds 5 thereby satisfying Cochran's rules (Cochran 1952, 1954). These tables were randomly generated using the `unidrnd()` function in Matlab. For the sake of brevity, we do not summarise the ANOVA output of these partitions (we consider the example in Sect. 6.1 as providing the empirical evidence on their numerical equivalence) but instead record the execution time it takes for the partition to take place using the two approaches. The execution times are given in Table 4 for five real data sets (that are either three-way, four-way and five-way) ranging in dimension from $3 \times 3 \times 3$ to $3 \times 5 \times 2 \times 2 \times 2$. We also summarise in Table 4 the execution times of seven randomly generated contingency tables consisting of the cross-classification of five, six and seven categorical variables.

Table 4 shows that the new algorithm described in Sect. 5 provides a quicker execution time of the partition of Pearson's chi-squared statistic than the traditional method of partition and that by Loisel and Takane (2016). This is apparent when comparing the execution times for the larger contingency tables, including the four-way table of Azen and Walker (2011), the five-way table given in Jarausch and Arminger (1989) and the randomly simulated five-way contingency tables. As the execution times for method described by Loisel and Takane (2016) are longer when compared with the new algorithm and traditional method, we present a comparison of a new algorithm with that of the traditional one. In these cases, the new algorithm at least halves the execution time needed for the partition to be performed. The new algorithm was also able to partition the Pearson chi-squared statistic for a six-way and seven-way contingency table; albeit, the calculation time took a little longer for these large tables than it did to partition a three-, four- or five-way contingency table.

The reason behind the longer execution times taken by Loisel and Takane (2016) method lies in the methodology used in the partition. Every chi-squared term is expressed as a quadratic form in an idempotent matrix of size which is product of categories of all variables (i.e. all possible combinations of categories of all variables). For example, in the case of the simulated contingency table of size $10 \times 5 \times 5 \times 5 \times 5$, every chi-squared statistic corresponding to the two-way, three-way, four-way and five-way association terms (in all 26 such terms) is calculated as a quadratic form in a 6250 sized matrix. It may be due to this reason that the program runs out of memory. The methodology given by Loisel and Takane (2016) has a strong theoretical basis but software implementation proves to be very challenging in this case. This problem can be seen to be overcome using the new algorithm presented in this paper.

For the traditional method of partition, calculating the association terms in a six-way, seven-way, or "higher"-way contingency tables are very complex and are not generalisable. As a result, there is no software available to perform a generic partition of Pearson's chi-squared statistic for multi-way contingency tables. Instead, the analyst is required to develop separate code for the analysis of a three-, four- and "higher"-way contingency table. It is for this reason that no execution time is included in Table 4 for the randomly generated six-way and seven-way tables.

Table 4 Execution time to obtain the partition using traditional method, method by Loisel and Takane (2016) and the new method

References	Data summary	Dimension	Execution time (in seconds)		Improvement in execution time by M3 over M1
			Traditional method (M1)	By Loisel and Takane (2016) (M2)	
Carlier and Kroonenberg (1996)	Behavior of 30 infant-mother pairs during the first six months of life	7 × 6 × 6	0.0450	0.137	0.0390 13.3%
Beh and Davy (1998, 1999)	Happiness, schooling and number of sibling's data on 1517 people	3 × 3 × 3	0.0471	0.056	0.0320 32.1%
D'Ambra and Amenta (2011), D'Ambra et al. (2012)	Religion, years of formal education and attitude towards abortion for 3181 individuals	3 × 3 × 3	0.0462	0.059	0.0319 31.0%
Azen and Walker (2011)	Responses of the students to two survey questions	4 × 4 × 3 × 2	0.0631	0.098	0.0320 49.3%
Jarausch and Armingier (1989)	Count of members and non-members of Nazi party for teachers	3 × 5 × 2 × 2 × 2	0.0784	0.181	0.0310 60.5%
Simulated data (using <code>unidirnd()</code>)	Simulated data from discrete <i>Uniform</i> [7, 100]	3 × 5 × 2 × 2 × 2	0.0772	0.197	0.0319 58.7%
		3 × 5 × 3 × 3 × 3	0.0769	0.433	0.0309 59.8%
		5 × 5 × 4 × 3 × 5	0.0789	13.422	0.0320 59.4%
		5 × 5 × 5 × 5 × 5	0.0652	114.89	0.0329 49.5%
		10 × 5 × 5 × 5 × 5	0.0790	*	0.0389 50.8%
	3 × 3 × 4 × 3 × 3 × 3	–	–	–	–
	5 × 5 × 4 × 2 × 2 × 3 × 3	–	–	–	–

*Out of memory message appears and further execution stopped

However, the new algorithm overcomes this problem; the new algorithm calculates the higher-order terms of the partition of the chi-squared statistic using lower-order (eg two-, or three-, way) association terms.

7 Discussion

This paper develops an algorithm for partitioning Pearson's chi-squared statistic for four-way and five-way contingency tables and extends to the multi-way case. The traditional approach to partitioning Pearson's chi-squared, or equivalently Φ^2 , statistic for multi-way tables is computationally complex and has rarely been performed in practice or implemented into current statistical packages. This paper has proposed a new method of partition using tensors that provide the analyst with a means of calculating multi-way sources of association based on the prior calculation of two-, three-, and other "lower"-order association terms. Furthermore, the ANOVA-like partition can be presented in the form of categorical ANOVA which allows for one to test the statistical significance of each term of the partition. Our demonstration of the new algorithm shows that using tensors and tensor products allows one to simply, and quickly, partition Pearson's statistic. We do acknowledge though that the execution times between the two approaches is in the order of a fraction of a second and therefore may appear, on the surface, insignificant (we alluded to this point in Sect. 6.1). However, we must keep in mind that, in practice, a study generally will consist of more than four or five variables and in the commercial and industrial sectors data is collected on 100 s and even 1000 s of categorical variables. This is certainly true for the analysis of questionnaires and surveys where responders are asked to "tick a box". For example, the 2016 Australian Census consisted of 51 questions, most of which may be treated as categorical variables. Simultaneously exploring the association between all, or even a subset, of such a large number of categorical variables is almost never performed due, for example, to the computational complexities in doing so. The computational efficiency of the algorithm outlined in Sect. 4 is apparent since the execution time of the partition can be generally at least halved when compared with the traditional method of partition described in Sect. 3. While such efficiencies gained appear small, our demonstrations here suggest that better efficiencies can be gained for the analysis of truly large number of variables.

Further developments of this algorithm can be extended beyond the case of studying symmetric association between categorical variables. That is, rather than just partitioning Pearson's chi-squared statistic for multi-way contingency tables, this algorithm can be extended to situations where a partition of multi-way version of asymmetric association is required. These may include the partitions of the delta index (Lombardo 2011), the Marcotorchino index (Marcotorchino 1984) and the Gray–Williams index (Gray & Williams 1975) which have been confined to the analysis of asymmetric association among only three categorical variables. Other areas of further investigation include utilising the algorithm in the context of

modelling categorical variables using association models, correspondence models and log linear models. We shall leave the development of the algorithm to these cases for future consideration.

Appendix

Consider a classification of n individuals based on five categorical variables A_1, A_2, A_3, A_4 and A_5 . Let j_k be the number of categories of the k^{th} variable, for $k = 1, 2, \dots, 5$ and the joint frequencies be summarised into a five-way contingency table $N = (n_{i_1 i_2 i_3 i_4 i_5})$ of dimension $j_1 \times j_2 \times j_3 \times j_4 \times j_5$. Let $P = (p_{i_1 i_2 i_3 i_4 i_5})$ be the table of relative proportions such that $p = \frac{1}{n}N$.

The total inertia Φ^2 based on the deviations from the five-way complete independence model is

$$\Phi^2 = \sum_{i_1, i_2, i_3, i_4, i_5} \frac{(p_{i_1 i_2 i_3 i_4 i_5} - p_{i_1 \dots} p_{i_2 \dots} p_{i_3 \dots} p_{i_4 \dots} p_{i_5 \dots})^2}{p_{i_1 \dots} p_{i_2 \dots} p_{i_3 \dots} p_{i_4 \dots} p_{i_5 \dots}}$$

The total inertia Φ^2 can be expressed as a squared norm of $\Pi = (\pi_{i_1 i_2 i_3 i_4 i_5})$ in $\mathbb{R}^{j_1 \times j_2 \times j_3 \times j_4 \times j_5}$ where

$$\pi_{i_1 i_2 i_3 i_4 i_5} = \frac{p_{i_1 i_2 i_3 i_4 i_5}}{p_{i_1 \dots} p_{i_2 \dots} p_{i_3 \dots} p_{i_4 \dots} p_{i_5 \dots}} - 1,$$

And $\|\Pi\|^2 = \sum_{i_1, i_2, i_3, i_4, i_5} p_{i_1 \dots} p_{i_2 \dots} p_{i_3 \dots} p_{i_4 \dots} p_{i_5 \dots} (\pi_{i_1 i_2 i_3 i_4 i_5})^2$.

Now, Π can be expressed as the addition of twenty-six orthogonal arrays that consist of ${}^5C_2 = 10$ two-way, ${}^5C_3 = 10$ three-way, ${}^5C_4 = 5$ four-way and ${}^5C_5 = 1$ five-way association terms. That is

$$\begin{aligned} \Pi = & \Pi_{i_1 i_2 \dots} + \Pi_{i_1 i_3 \dots} + \Pi_{i_1 \dots i_4} + \Pi_{i_1 \dots i_5} + \Pi_{i_2 i_3 \dots} + \Pi_{i_2 i_4 \dots} + \Pi_{i_2 \dots i_5} + \Pi_{i_3 i_4 \dots} + \Pi_{i_3 i_5 \dots} \\ & + \Pi_{i_4 i_5} + \alpha \Pi_{i_1 i_2 i_3 \dots} + \alpha \Pi_{i_1 i_2 i_4 \dots} + \alpha \Pi_{i_1 i_2 \dots i_5} + \alpha \Pi_{i_1 i_3 i_4 \dots} + \alpha \Pi_{i_1 i_3 \dots i_5} + \alpha \Pi_{i_1 \dots i_4 i_5} \\ & + \alpha \Pi_{i_2 i_3 i_4 \dots} + \alpha \Pi_{i_2 i_3 \dots i_5} + \alpha \Pi_{i_2 i_4 i_5} + \alpha \Pi_{i_3 i_4 i_5} + \alpha \Pi_{i_1 i_2 i_3 i_4 \dots} \\ & + \alpha \Pi_{i_1 i_2 i_3 \dots i_5} + \alpha \Pi_{i_1 i_2 i_4 \dots i_5} + \alpha \Pi_{i_1 i_3 i_4 \dots i_5} + \alpha \Pi_{i_2 i_3 i_4 \dots i_5} + \alpha \Pi_{i_1 i_2 i_3 i_4 i_5} \end{aligned}$$

where

$$\begin{aligned}
 \pi_{i_1 i_2 \dots} &= \frac{P_{i_1 i_2 \dots} - P_{i_1 \dots} P_{i_2 \dots}}{P_{i_1 \dots} P_{i_2 \dots}}, & \pi_{i_1 i_3 \dots} &= \frac{P_{i_1 \dots i_3 \dots} - P_{i_1 \dots} P_{i_3 \dots}}{P_{i_1 \dots} P_{i_3 \dots}}, & \pi_{i_1 i_4} &= \frac{P_{i_1 \dots i_4} - P_{i_1 \dots} P_{i_4}}{P_{i_1 \dots} P_{i_4}}, \\
 \pi_{i_1 \dots i_5} &= \frac{P_{i_1 \dots i_5} - P_{i_1 \dots} P_{i_5}}{P_{i_1 \dots} P_{i_5}}, & \pi_{i_2 i_3 \dots} &= \frac{P_{i_2 i_3 \dots} - P_{i_2 \dots} P_{i_3 \dots}}{P_{i_2 \dots} P_{i_3 \dots}}, & \pi_{i_2 i_4} &= \frac{P_{i_2 i_4} - P_{i_2 \dots} P_{i_4}}{P_{i_2 \dots} P_{i_4}}, \\
 \pi_{i_2 \dots i_5} &= \frac{P_{i_2 \dots i_5} - P_{i_2 \dots} P_{i_5}}{P_{i_2 \dots} P_{i_5}}, & \pi_{i_3 i_4} &= \frac{P_{i_3 i_4} - P_{i_3 \dots} P_{i_4}}{P_{i_3 \dots} P_{i_4}}, & \pi_{i_3 i_5} &= \frac{P_{i_3 i_5} - P_{i_3 \dots} P_{i_5}}{P_{i_3 \dots} P_{i_5}}, \\
 \pi_{\dots i_4 i_5} &= \frac{P_{\dots i_4 i_5} - P_{\dots i_4} P_{i_5}}{P_{\dots i_4} P_{i_5}},
 \end{aligned}$$

$$\begin{aligned}
 \alpha \pi_{i_1 i_2 i_3 \dots} &= \frac{P_{i_1 i_2 i_3 \dots} - \alpha P_{i_1 i_2 i_3 \dots}}{P_{i_1 \dots} P_{i_2 \dots} P_{i_3 \dots}}, & \alpha \pi_{i_1 i_2 i_4} &= \frac{P_{i_1 i_2 i_4} - \alpha P_{i_1 i_2 i_4}}{P_{i_1 \dots} P_{i_2 \dots} P_{i_4}}, \\
 \alpha \pi_{i_1 i_2 i_5} &= \frac{P_{i_1 i_2 i_5} - \alpha P_{i_1 i_2 i_5}}{P_{i_1 \dots} P_{i_2 \dots} P_{i_5}}, & \alpha \pi_{i_1 i_3 i_4} &= \frac{P_{i_1 i_3 i_4} - \alpha P_{i_1 i_3 i_4}}{P_{i_1 \dots} P_{i_3 \dots} P_{i_4}}, \\
 \alpha \pi_{i_1 i_3 i_5} &= \frac{P_{i_1 i_3 i_5} - \alpha P_{i_1 i_3 i_5}}{P_{i_1 \dots} P_{i_3 \dots} P_{i_5}}, & \alpha \pi_{i_1 i_4 i_5} &= \frac{P_{i_1 i_4 i_5} - \alpha P_{i_1 i_4 i_5}}{P_{i_1 \dots} P_{i_4 \dots} P_{i_5}}, \\
 \alpha \pi_{i_2 i_3 i_4} &= \frac{P_{i_2 i_3 i_4} - \alpha P_{i_2 i_3 i_4}}{P_{i_2 \dots} P_{i_3 \dots} P_{i_4}}, & \alpha \pi_{i_2 i_3 i_5} &= \frac{P_{i_2 i_3 i_5} - \alpha P_{i_2 i_3 i_5}}{P_{i_2 \dots} P_{i_3 \dots} P_{i_5}}, \\
 \alpha \pi_{i_2 i_4 i_5} &= \frac{P_{i_2 i_4 i_5} - \alpha P_{i_2 i_4 i_5}}{P_{i_2 \dots} P_{i_4 \dots} P_{i_5}}, & \alpha \pi_{i_3 i_4 i_5} &= \frac{P_{i_3 i_4 i_5} - \alpha P_{i_3 i_4 i_5}}{P_{i_3 \dots} P_{i_4 \dots} P_{i_5}}, \\
 \alpha \Pi_{i_1 i_2 i_3 i_4} &= \frac{P_{i_1 i_2 i_3 i_4} - \alpha P_{i_1 i_2 i_3 i_4}}{P_{i_1 \dots} P_{i_2 \dots} P_{i_3 \dots} P_{i_4}}, & \alpha \Pi_{i_1 i_2 i_3 i_5} &= \frac{P_{i_1 i_2 i_3 i_5} - \alpha P_{i_1 i_2 i_3 i_5}}{P_{i_1 \dots} P_{i_2 \dots} P_{i_3 \dots} P_{i_5}}, \\
 \alpha \Pi_{i_1 i_2 i_4 i_5} &= \frac{P_{i_1 i_2 i_4 i_5} - \alpha P_{i_1 i_2 i_4 i_5}}{P_{i_1 \dots} P_{i_2 \dots} P_{i_4 \dots} P_{i_5}}, & \Pi_{i_1 i_3 i_4 i_5} &= \frac{P_{i_1 i_3 i_4 i_5} - \alpha P_{i_1 i_3 i_4 i_5}}{P_{i_1 \dots} P_{i_3 \dots} P_{i_4 \dots} P_{i_5}}, \\
 \alpha \Pi_{i_2 i_3 i_4 i_5} &= \frac{P_{i_2 i_3 i_4 i_5} - \alpha P_{i_2 i_3 i_4 i_5}}{P_{i_2 \dots} P_{i_3 \dots} P_{i_4 \dots} P_{i_5}}, & \alpha \pi_{i_1 i_2 i_3 i_4 i_5} &= \frac{P_{i_1 i_2 i_3 i_4 i_5} - \alpha P_{i_1 i_2 i_3 i_4 i_5}}{P_{i_1 \dots} P_{i_2 \dots} P_{i_3 \dots} P_{i_4 \dots} P_{i_5}},
 \end{aligned}$$

$$\begin{aligned}
 \alpha P_{i_1 i_2 i_3 i_4} &= P_{i_1 i_2 \dots P \dots i_3 \dots P \dots i_4} + P_{i_1 i_3 \dots P \dots i_2 \dots P \dots i_4} + P_{i_1 i_4 \dots P \dots i_2 \dots P \dots i_3} + P_{i_2 i_3 \dots P \dots i_1 \dots P \dots i_4} \\
 &\quad + P_{i_2 i_4 \dots P \dots i_1 \dots P \dots i_3} + P_{i_3 i_4 \dots P \dots i_1 \dots P \dots i_2} + P_{i_1 i_2 i_3 \dots P \dots i_4} + P_{i_1 i_2 i_4 \dots P \dots i_3} \\
 &\quad + P_{i_2 i_3 i_4 \dots P \dots i_1} + P_{i_1 i_3 i_4 \dots P \dots i_2} + 9P_{i_1 \dots P \dots i_2 \dots P \dots i_3 \dots P \dots i_4}, \\
 \alpha P_{i_1 i_2 i_3 i_5} &= P_{i_1 i_2 \dots P \dots i_3 \dots P \dots i_5} + P_{i_1 i_3 \dots P \dots i_2 \dots P \dots i_5} + P_{i_1 i_5 \dots P \dots i_2 \dots P \dots i_3} + P_{i_2 i_3 \dots P \dots i_1 \dots P \dots i_5} \\
 &\quad + P_{i_2 i_5 \dots P \dots i_1 \dots P \dots i_3} + P_{i_3 i_5 \dots P \dots i_1 \dots P \dots i_2} + P_{i_1 i_2 i_3 \dots P \dots i_5} + P_{i_1 i_2 i_5 \dots P \dots i_3} \\
 &\quad + P_{i_2 i_3 i_5 \dots P \dots i_1} + P_{i_1 i_3 i_5 \dots P \dots i_2} + 9P_{i_1 \dots P \dots i_2 \dots P \dots i_3 \dots P \dots i_5}, \\
 \alpha P_{i_1 i_2 i_4 i_5} &= P_{i_1 i_2 \dots P \dots i_4 \dots P \dots i_5} + P_{i_1 i_4 \dots P \dots i_2 \dots P \dots i_5} + P_{i_1 i_5 \dots P \dots i_2 \dots P \dots i_4} + P_{i_2 i_4 \dots P \dots i_1 \dots P \dots i_5} \\
 &\quad + P_{i_2 i_5 \dots P \dots i_1 \dots P \dots i_4} + P_{i_3 i_4 \dots P \dots i_1 \dots P \dots i_2} + P_{i_1 i_2 i_4 \dots P \dots i_5} + P_{i_1 i_2 i_5 \dots P \dots i_4} \\
 &\quad + P_{i_2 i_4 i_5 \dots P \dots i_1} + P_{i_1 i_4 i_5 \dots P \dots i_2} + 9P_{i_1 \dots P \dots i_2 \dots P \dots i_4 \dots P \dots i_5}, \\
 \alpha P_{i_1 i_3 i_4 i_5} &= P_{i_1 i_3 \dots P \dots i_4 \dots P \dots i_5} + P_{i_1 i_4 \dots P \dots i_3 \dots P \dots i_5} + P_{i_1 i_5 \dots P \dots i_3 \dots P \dots i_4} + P_{i_2 i_4 \dots P \dots i_1 \dots P \dots i_5} \\
 &\quad + P_{i_3 i_5 \dots P \dots i_1 \dots P \dots i_4} + P_{i_3 i_4 \dots P \dots i_1 \dots P \dots i_2} + P_{i_1 i_3 i_4 \dots P \dots i_5} + P_{i_1 i_3 i_5 \dots P \dots i_4} \\
 &\quad + P_{i_2 i_4 i_5 \dots P \dots i_1} + P_{i_1 i_4 i_5 \dots P \dots i_3} + 9P_{i_1 \dots P \dots i_3 \dots P \dots i_4 \dots P \dots i_5}, \\
 \alpha P_{i_2 i_3 i_4 i_5} &= P_{i_2 i_3 \dots P \dots i_4 \dots P \dots i_5} + P_{i_2 i_4 \dots P \dots i_3 \dots P \dots i_5} + P_{i_2 i_5 \dots P \dots i_3 \dots P \dots i_4} + P_{i_3 i_4 \dots P \dots i_2 \dots P \dots i_5} \\
 &\quad + P_{i_3 i_5 \dots P \dots i_2 \dots P \dots i_4} + P_{i_4 i_5 \dots P \dots i_2 \dots P \dots i_3} + P_{i_2 i_3 i_4 \dots P \dots i_5} + P_{i_2 i_3 i_5 \dots P \dots i_4} \\
 &\quad + P_{i_3 i_4 i_5 \dots P \dots i_2} + P_{i_2 i_4 i_5 \dots P \dots i_3} + 9P_{i_2 \dots P \dots i_3 \dots P \dots i_4 \dots P \dots i_5}, \\
 \text{and } \alpha P_{i_1 i_2 i_3 i_4 i_5} &= P_{i_1 i_2 \dots P \dots i_3 \dots P \dots i_4 \dots P \dots i_5} + P_{i_1 i_3 \dots P \dots i_2 \dots P \dots i_4 \dots P \dots i_5} + P_{i_1 i_4 \dots P \dots i_2 \dots P \dots i_3 \dots P \dots i_5} \\
 &\quad + P_{i_1 i_5 \dots P \dots i_2 \dots P \dots i_3 \dots P \dots i_4} + P_{i_2 i_3 \dots P \dots i_1 \dots P \dots i_4 \dots P \dots i_5} + P_{i_2 i_4 \dots P \dots i_1 \dots P \dots i_3 \dots P \dots i_5} \\
 &\quad + P_{i_2 i_5 \dots P \dots i_1 \dots P \dots i_3 \dots P \dots i_4} + P_{i_3 i_4 \dots P \dots i_1 \dots P \dots i_2 \dots P \dots i_5} + P_{i_3 i_5 \dots P \dots i_1 \dots P \dots i_2 \dots P \dots i_4} \\
 &\quad + P_{i_4 i_5 \dots P \dots i_1 \dots P \dots i_2 \dots P \dots i_3} + P_{i_1 i_2 i_3 \dots P \dots i_4 \dots P \dots i_5} + P_{i_1 i_2 i_4 \dots P \dots i_3 \dots P \dots i_5} \\
 &\quad + P_{i_1 i_2 i_5 \dots P \dots i_3 \dots P \dots i_4} + P_{i_1 i_3 i_4 \dots P \dots i_2 \dots P \dots i_5} + P_{i_1 i_3 i_5 \dots P \dots i_2 \dots P \dots i_4} \\
 &\quad + P_{i_1 i_4 i_5 \dots P \dots i_2 \dots P \dots i_3} + P_{i_2 i_3 i_4 \dots P \dots i_1 \dots P \dots i_5} + P_{i_2 i_3 i_5 \dots P \dots i_1 \dots P \dots i_4} \\
 &\quad + P_{i_2 i_4 i_5 \dots P \dots i_1 \dots P \dots i_3} + P_{i_3 i_4 i_5 \dots P \dots i_1 \dots P \dots i_2} + P_{i_1 i_2 i_3 i_4 \dots P \dots i_5} + P_{i_1 i_2 i_3 i_5 \dots P \dots i_4} \\
 &\quad + P_{i_1 i_2 i_4 i_5 \dots P \dots i_3} + P_{i_1 i_3 i_4 i_5 \dots P \dots i_2} + P_{i_2 i_3 i_4 i_5 \dots P \dots i_1} \\
 &\quad + 24P_{i_1 \dots P \dots i_2 \dots P \dots i_3 \dots P \dots i_4 \dots P \dots i_5}.
 \end{aligned}$$

Thus, an additive partition of the squared norm of a five-way array Π can be expressed as

$$\begin{aligned}
 \Pi^2 &= \Pi_{i_1 i_2 \dots}^2 + \Pi_{i_1 i_3 \dots}^2 + \Pi_{i_1 i_4 \dots}^2 + \Pi_{i_1 i_5 \dots}^2 + \Pi_{i_2 i_3 \dots}^2 + \Pi_{i_2 i_4 \dots}^2 \\
 &\quad + \Pi_{i_2 i_5 \dots}^2 + \Pi_{i_3 i_4 \dots}^2 + \Pi_{i_3 i_5 \dots}^2 + \Pi_{i_4 i_5 \dots}^2 + \alpha \Pi_{i_1 i_2 i_3 \dots}^2 \\
 &\quad + \alpha \Pi_{i_1 i_2 i_4 \dots}^2 + \alpha \Pi_{i_1 i_2 i_5 \dots}^2 + \alpha \Pi_{i_1 i_3 i_4 \dots}^2 + \alpha \Pi_{i_1 i_3 i_5 \dots}^2 \\
 &\quad + \alpha \Pi_{i_1 i_4 i_5 \dots}^2 + \alpha \Pi_{i_2 i_3 i_4 \dots}^2 + \alpha \Pi_{i_2 i_3 i_5 \dots}^2 + \alpha \Pi_{i_2 i_4 i_5 \dots}^2 \\
 &\quad + \alpha \Pi_{i_3 i_4 i_5 \dots}^2 + \alpha \Pi_{i_1 i_2 i_3 i_4 \dots}^2 + \alpha \Pi_{i_1 i_2 i_3 i_5 \dots}^2 + \alpha \Pi_{i_1 i_2 i_4 i_5 \dots}^2 \\
 &\quad + \alpha \Pi_{i_1 i_3 i_4 i_5 \dots}^2 + \alpha \Pi_{i_2 i_3 i_4 i_5 \dots}^2 + \alpha \Pi_{i_1 i_2 i_3 i_4 i_5 \dots}^2.
 \end{aligned}$$

This partition can be equivalently expressed in terms of table’s inertia such that

$$\begin{aligned}
 \Phi^2 = & \sum_{i_1, i_2} p_{i_1 \dots p_{i_2 \dots}} \left(\frac{p_{i_1 i_2 \dots} - p_{i_1 \dots p_{i_2 \dots}}}{p_{i_1 \dots p_{i_2 \dots}}} \right)^2 + \sum_{i_1, i_3} p_{i_1 \dots p_{i_3 \dots}} \left(\frac{p_{i_1 i_3 \dots} - p_{i_1 \dots p_{i_3 \dots}}}{p_{i_1 \dots p_{i_3 \dots}}} \right)^2 \\
 & + \sum_{i_1, i_4} p_{i_1 \dots p_{i_4 \dots}} \left(\frac{p_{i_1 i_4 \dots} - p_{i_1 \dots p_{i_4 \dots}}}{p_{i_1 \dots p_{i_4 \dots}}} \right)^2 + \sum_{i_1, i_5} p_{i_1 \dots p_{i_5 \dots}} \left(\frac{p_{i_1 i_5 \dots} - p_{i_1 \dots p_{i_5 \dots}}}{p_{i_1 \dots p_{i_5 \dots}}} \right)^2 \\
 & + \sum_{i_2, i_3} p_{i_2 \dots p_{i_3 \dots}} \left(\frac{p_{i_2 i_3 \dots} - p_{i_2 \dots p_{i_3 \dots}}}{p_{i_2 \dots p_{i_3 \dots}}} \right)^2 + \sum_{i_2, i_4} p_{i_2 \dots p_{i_4 \dots}} \left(\frac{p_{i_2 i_4 \dots} - p_{i_2 \dots p_{i_4 \dots}}}{p_{i_2 \dots p_{i_4 \dots}}} \right)^2 \\
 & + \sum_{i_2, i_5} p_{i_2 \dots p_{i_5 \dots}} \left(\frac{p_{i_2 i_5 \dots} - p_{i_2 \dots p_{i_5 \dots}}}{p_{i_2 \dots p_{i_5 \dots}}} \right)^2 + \sum_{i_3, i_4} p_{i_3 \dots p_{i_4 \dots}} \left(\frac{p_{i_3 i_4 \dots} - p_{i_3 \dots p_{i_4 \dots}}}{p_{i_3 \dots p_{i_4 \dots}}} \right)^2 \\
 & + \sum_{i_3, i_5} p_{i_3 \dots p_{i_5 \dots}} \left(\frac{p_{i_3 i_5 \dots} - p_{i_3 \dots p_{i_5 \dots}}}{p_{i_3 \dots p_{i_5 \dots}}} \right)^2 + \sum_{i_4, i_5} p_{i_4 \dots p_{i_5 \dots}} \left(\frac{p_{i_4 i_5 \dots} - p_{i_4 \dots p_{i_5 \dots}}}{p_{i_4 \dots p_{i_5 \dots}}} \right)^2 \\
 & + \sum_{i_1, i_2, i_3} p_{i_1 \dots p_{i_2 \dots p_{i_3 \dots}}} \left(\frac{p_{i_1 i_2 i_3 \dots} - \alpha p_{i_1 i_2 i_3 \dots}}{p_{i_1 \dots p_{i_2 \dots p_{i_3 \dots}}}} \right)^2 + \sum_{i_1, i_2, i_4} p_{i_1 \dots p_{i_2 \dots p_{i_4 \dots}}} \left(\frac{p_{i_1 i_2 i_4 \dots} - \alpha p_{i_1 i_2 i_4 \dots}}{p_{i_1 \dots p_{i_2 \dots p_{i_4 \dots}}}} \right)^2 \\
 & + \sum_{i_1, i_2, i_5} p_{i_1 \dots p_{i_2 \dots p_{i_5 \dots}}} \left(\frac{p_{i_1 i_2 i_5 \dots} - \alpha p_{i_1 i_2 i_5 \dots}}{p_{i_1 \dots p_{i_2 \dots p_{i_5 \dots}}}} \right)^2 + \sum_{i_1, i_3, i_4} p_{i_1 \dots p_{i_3 \dots p_{i_4 \dots}}} \left(\frac{p_{i_1 i_3 i_4 \dots} - \alpha p_{i_1 i_3 i_4 \dots}}{p_{i_1 \dots p_{i_3 \dots p_{i_4 \dots}}}} \right)^2 \\
 & + \sum_{i_1, i_3, i_5} p_{i_1 \dots p_{i_3 \dots p_{i_5 \dots}}} \left(\frac{p_{i_1 i_3 i_5 \dots} - \alpha p_{i_1 i_3 i_5 \dots}}{p_{i_1 \dots p_{i_3 \dots p_{i_5 \dots}}}} \right)^2 + \sum_{i_1, i_4, i_5} p_{i_1 \dots p_{i_4 \dots p_{i_5 \dots}}} \left(\frac{p_{i_1 i_4 i_5 \dots} - \alpha p_{i_1 i_4 i_5 \dots}}{p_{i_1 \dots p_{i_4 \dots p_{i_5 \dots}}}} \right)^2 \\
 & + \sum_{i_2, i_3, i_4} p_{i_2 \dots p_{i_3 \dots p_{i_4 \dots}}} \left(\frac{p_{i_2 i_3 i_4 \dots} - \alpha p_{i_2 i_3 i_4 \dots}}{p_{i_2 \dots p_{i_3 \dots p_{i_4 \dots}}}} \right)^2 + \sum_{i_2, i_3, i_5} p_{i_2 \dots p_{i_3 \dots p_{i_5 \dots}}} \left(\frac{p_{i_2 i_3 i_5 \dots} - \alpha p_{i_2 i_3 i_5 \dots}}{p_{i_2 \dots p_{i_3 \dots p_{i_5 \dots}}}} \right)^2 \\
 & + \sum_{i_2, i_4, i_5} p_{i_2 \dots p_{i_4 \dots p_{i_5 \dots}}} \left(\frac{p_{i_2 i_4 i_5 \dots} - \alpha p_{i_2 i_4 i_5 \dots}}{p_{i_2 \dots p_{i_4 \dots p_{i_5 \dots}}}} \right)^2 + \sum_{i_3, i_4, i_5} p_{i_3 \dots p_{i_4 \dots p_{i_5 \dots}}} \left(\frac{p_{i_3 i_4 i_5 \dots} - \alpha p_{i_3 i_4 i_5 \dots}}{p_{i_3 \dots p_{i_4 \dots p_{i_5 \dots}}}} \right)^2 \\
 & + \sum_{i_1, i_2, i_3, i_4} p_{i_1 \dots p_{i_2 \dots p_{i_3 \dots p_{i_4 \dots}}} \left(\frac{p_{i_1 i_2 i_3 i_4 \dots} - \alpha p_{i_1 i_2 i_3 i_4 \dots}}{p_{i_1 \dots p_{i_2 \dots p_{i_3 \dots p_{i_4 \dots}}}} \right)^2 \\
 & + \sum_{i_1, i_2, i_3, i_5} p_{i_1 \dots p_{i_2 \dots p_{i_3 \dots p_{i_5 \dots}}} \left(\frac{p_{i_1 i_2 i_3 i_5 \dots} - \alpha p_{i_1 i_2 i_3 i_5 \dots}}{p_{i_1 \dots p_{i_2 \dots p_{i_3 \dots p_{i_5 \dots}}}} \right)^2 \\
 & + \sum_{i_1, i_2, i_4, i_5} p_{i_1 \dots p_{i_2 \dots p_{i_4 \dots p_{i_5 \dots}}} \left(\frac{p_{i_1 i_2 i_4 i_5 \dots} - \alpha p_{i_1 i_2 i_4 i_5 \dots}}{p_{i_1 \dots p_{i_2 \dots p_{i_4 \dots p_{i_5 \dots}}}} \right)^2 \\
 & + \sum_{i_1, i_3, i_4, i_5} p_{i_1 \dots p_{i_3 \dots p_{i_4 \dots p_{i_5 \dots}}} \left(\frac{p_{i_1 i_3 i_4 i_5 \dots} - \alpha p_{i_1 i_3 i_4 i_5 \dots}}{p_{i_1 \dots p_{i_3 \dots p_{i_4 \dots p_{i_5 \dots}}}} \right)^2 \\
 & + \sum_{i_2, i_3, i_4, i_5} p_{i_2 \dots p_{i_3 \dots p_{i_4 \dots p_{i_5 \dots}}} \left(\frac{p_{i_2 i_3 i_4 i_5 \dots} - \alpha p_{i_2 i_3 i_4 i_5 \dots}}{p_{i_2 \dots p_{i_3 \dots p_{i_4 \dots p_{i_5 \dots}}}} \right)^2 \\
 & + \sum_{i_1, i_2, i_3, i_4, i_5} p_{i_1 \dots p_{i_2 \dots p_{i_3 \dots p_{i_4 \dots p_{i_5 \dots}}} \left(\frac{p_{i_1 i_2 i_3 i_4 i_5 \dots} - \alpha p_{i_1 i_2 i_3 i_4 i_5 \dots}}{p_{i_1 \dots p_{i_2 \dots p_{i_3 \dots p_{i_4 \dots p_{i_5 \dots}}}} \right)^2.
 \end{aligned}$$

Thus, the orthogonal partition of $n\Phi^2 = \chi_{Total}^2$ for the five-way contingency table N is therefore

$$\begin{aligned}\chi_{Total}^2 = & \chi_{A_1A_2}^2 + \chi_{A_1A_3}^2 + \chi_{A_1A_4}^2 + \chi_{A_1A_5}^2 + \chi_{A_2A_3}^2 + \chi_{A_2A_4}^2 + \chi_{A_2A_5}^2 + \chi_{A_3A_4}^2 + \chi_{A_3A_5}^2 + \chi_{A_4A_5}^2 \\ & + \chi_{A_1A_2A_3}^2 + \chi_{A_1A_2A_4}^2 + \chi_{A_1A_2A_5}^2 + \chi_{A_1A_3A_4}^2 + \chi_{A_1A_3A_5}^2 + \chi_{A_1A_4A_5}^2 + \chi_{A_2A_3A_4}^2 \\ & + \chi_{A_2A_3A_5}^2 + \chi_{A_2A_4A_5}^2 + \chi_{A_3A_4A_5}^2 + \chi_{A_1A_2A_3A_4}^2 + \chi_{A_1A_2A_3A_5}^2 + \chi_{A_1A_2A_4A_5}^2 \\ & + \chi_{A_1A_3A_4A_5}^2 + \chi_{A_2A_3A_4A_5}^2 + \chi_{A_1A_2A_3A_4A_5}^2\end{aligned}$$

Acknowledgements We would like to thank the Referees for their constructive comments and suggestions, which improved the manuscript significantly.

Funding The second author would like to thank University Grants Commission, New Delhi, India for awarding Rajiv Gandhi National Fellowship (Award No.: F1-17.1/2011-12/RGNF-SC-MAH-5331) for pursuing a Ph.D.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

- Azen R, Walker CM (2011) *Categorical data analysis for the behavioral and social sciences*. Routledge, London
- Beh EJ, Davy PJ (1998) Partitioning Pearson's chi-squared Statistic for a completely ordered three-way contingency table. *Aust N Z J Stat* 40:465–477
- Beh EJ, Davy PJ (1999) Partitioning Pearson's chi-squared statistic for a partially ordered three-way contingency table. *Aust N Z J Stat* 41:233–246
- Beh EJ, Lombardo R (2014) *Correspondence analysis: theory practice and new strategies*. Wiley, Chichester
- Carlier A, Kroonenberg PM (1996) Decompositions and biplots in three-way correspondence analysis. *Psychometrika* 61:355–373
- Cochran WG (1952) The χ^2 Test of goodness of fit. *Ann Math Stat* 23:315–345
- Cochran WG (1954) Some methods for strengthening the common χ^2 tests. *Biometrics* 10:417–451
- D'Ambra A, Amenta P (2011) Correspondence analysis with linear constraints of ordinal cross-classifications. *J Classif* 28:70–92
- D'Ambra L, Beh EJ, Amenta P (2005) CATANOVA for two-way contingency tables with ordinal variables using orthogonal polynomials. *Commun Stat-Theor Methods* 34:1755–1769
- D'Ambra L, D'Ambra A, Sarnacchiaro P (2012) Visualizing main effects and interaction in multiple non-symmetric correspondence analysis. *J Appl Stat* 39:2165–2175
- De Lathauwer L, De Moor B, Vandewalle J (2000) A multilinear singular value decomposition. *SIAM J Matrix Anal Appl* 21(4):1253–1278
- Goodman L, Kruskal W (1954) Measures of association for cross-classifications. *J Am Stat Assoc* 49:732–764
- Gray LN, Williams JS (1975) Goodman and Kruskal's tau b: multiple and partial analogs, *American Statistical Association (Proceedings of the Social Statistics Section)*, p 444–448
- Greenacre MJ (1984) *Theory and applications of correspondence analysis*. Academic Press, London
- Jarausch KH, Arminger G (1989) The German teaching profession and Nazi party membership: a demographic logit model. *J Interdiscip Hist* 20:197–225

- Kolda TG, Bader BW (2009) Tensor decompositions and applications. *SIAM Rev* 51:455–500
- Kroonenberg PM (1989) Singular value decomposition of interactions in three-way contingency tables. In: Coppi R, Bolasco S (eds) *Multiway data analysis*. North Holland, Amsterdam, pp 169–184
- Kroonenberg PM (2008) *Applied multiway data analysis*. Wiley, NJ
- Lancaster HO (1951) Complex contingency tables treated by the partition of the chi-square. *J R Stat Soc, Ser B* 13:242–249
- Lebart L, Morineau A, Warwick KM (1984) *Multivariate descriptive statistical analysis*. Wiley
- Loisel S, Takane Y (2016) Partitions of Pearson's chi-square statistic for frequency tables: a comprehensive account. *Comput Stat* 31:1429–1452
- Lombardo R, Beh EJ (2021) *CAvariants package*. <https://cran.r-project.org/web/packages/CAvariants/index.html>
- Lombardo R (1994) Models of decomposition for the analysis of the dependence in three-way contingency tables (in Italian). Unpublished PhD Thesis. Department of Mathematics and Statistics, University of Naples Federico II
- Lombardo R (2011) Three-way association measure decompositions: the delta index. *J Stat Plan Inference* 141:1789–1799
- Lombardo R, Beh EJ, Kroonenberg PM (2021) Symmetrical and non-symmetrical variants of three-way correspondence analysis for ordered variables. *Stat Sci* 36(4):542–561
- Lombardo R, Beh EJ, Michel van de V (2022) *CA3variants package*. <https://cran.r-project.org/web/packages/CA3variants/index.html>
- Lombardo R, Carlier A, D'Ambrà L (1996) Non-symmetric correspondence analysis for three-way contingency tables. *Methodologica* 4:59–80
- Lombardo R, Takane Y, Beh EJ (2017) *Chi2x3way package*. <https://cran.r-project.org/web/packages/chi2x3way/index.html>
- Lombardo R, Takane Y, Beh E (2020) Familywise decompositions of Pearson's chi-square statistic in the analysis of contingency tables. *Adv Data Anal Classif* 14:629–649
- Loughin TM, Scherer PN (1998) Testing for association in contingency tables with multiple column responses. *Biometrics* 54:630–637
- Luo J, Wu M, Gopukumar D, Zhao Y (2016) Big data application in biomedical research and health care: a literature review. *Biomed Inform Insights* 8:1–10
- Marcotorchino F (1984) Utilisation des Comparaisons par Paires en Statistique des Contingencies Partie II, Report # F071, Etude du Centre Scientifique, IBM, France
- Tsai CW, Lai CF, Chao HC, Vasilakos AV (2015) Big data analytics: a survey. *J Big Data* 2(21):1–31

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.