

NIASRA

NATIONAL INSTITUTE FOR APPLIED
STATISTICS RESEARCH AUSTRALIA



***National Institute for Applied Statistics Research
Australia***

University of Wollongong, Australia

Working Paper

06-22

**Correspondence Analysis and the Cressie-Read
Divergence Statistic**

Eric J. Beh and Rosaria Lombardo

*Copyright © 2022 by the National Institute for Applied Statistics Research Australia, UOW.
Work in progress, no part of this paper may be reproduced without permission from the Institute.*

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia Phone: +61 2 42215076, Fax: +61 2 42214998.

Email: karink@uow.edu.au

Correspondence Analysis and the Cressie-Read Divergence Statistic

Eric J. Beh^{1,2,3}

Rosaria Lombardo⁴

ORCID: 0000-0002-9051-9070 ORCID: 0000-0003-1746-0292

¹University of Newcastle, Australia

²University of Wollongong, Australia

³Stellenbosch University, South Africa

⁴University of Campania, Italy

Abstract

In the correspondence analysis literature, the foundations of visually and numerically summarising the association between two categorical variables rest with Pearson's chi-squared statistic. Not only is this statistic extremely popular and versatile, and yields some very useful visual and numerical properties. More recently, ties have been established that show the role that the Freeman-Tukey statistic plays in correspondence analysis and confirmed the advantages of the Hellinger distance that have long been advocated in the literature. Both Pearson's and the Freeman-Tukey statistics are special cases of the Cressie-Read divergence statistic, as are the Cressie-Read statistic, the likelihood ratio statistic and their modified versions. Therefore, this paper explores the features of correspondence analysis where the association, and the resulting low-dimensional correspondence plot, have at its foundation this divergence statistic. By doing so, this paper describes the properties of correspondence analysis for any special case of the Cressie-Read divergences statistic which includes as special cases the Hellinger distance decomposition (HDD) method and log-ratio analysis (LRA).

1 Introduction

Correspondence analysis is a very popular means of visually analysing the association between two or more categorical variables. There has been a vast amount written about it in the statistical, and allied, literature. For example, Nishisato (1980, Section 1.2), Lebart, Morineau, and Warwick (1984), Greenacre (1984), Birks, Peglar, and Austin (1996), Beh (2004) and Beh and Lombardo (2012, 2014, 2019) all provide comprehensive historical, technical, computational and bibliographical reviews of this technique. While the key feature from performing correspondence analysis is the visualisation of the nature of categorical association, a reason for its strong appeal is that the association is most commonly quantified in terms of Pearson's chi-squared statistic (Pearson, 1904). One may also refer to Plackett (1983) for a description of Karl Pearson's development of this statistic.

While Pearson's chi-squared statistic dominates much that has been written in the categorical data analysis (and related) literature, it is certainly not the only measure that one may use to assess whether there exists a statistically significant association between two or more categorical variables. Much has been discussed about some of the alternatives, including the likelihood ratio statistic (Wilks, 1938) and the Freeman-Tukey statistic (Freeman & Tukey, 1950). Despite this, very little work has been done that explores the link between measures such as these and their role in correspondence analysis. With these (and other) statistics being a special case of the Cressie-Read divergence statistic (Cressie & Read, 1984), this provides the categorical data analyst with some scope to explore further the way in which their data is associated. Therefore, rather than performing correspondence analysis by using just Pearson's statistic, this paper shows that a broader perspective can be gained by using the Cressie-Read divergence statistic to perform correspondence analysis. In doing so we can gain a meaningful understanding of distances between points depicted in a low-dimensional space, better understand the implications of using this statistic for modelling the association between the variables, and gain insight into how to improve the quality of a low-dimensional display (while also identifying when a poor quality plot arises).

Therefore, this paper will describe how correspondence analysis can be performed where the association is quantified in terms of the divergence statistic. The advantage of doing so is that correspondence analysis can be generalised for

any measure of association that belongs to the Cressie-Read divergence statistic, thereby extending the analysis beyond Pearson’s statistic. To describe this technique, we do so in the following seven sections. Section 2 gives a brief description of the Cressie-Read divergence statistic and five of the most popular special cases. We also provide some perspective of the divergence statistic as a sum-of-squares measure involving a general residual, called the *divergence residual*, that relies on a parameter δ ; changes in δ lead to special cases of the divergence statistic. Some further comments on the divergence residual are made in Section 3 while Section 4 demonstrates how correspondence analysis can be applied to this residual. In particular, we describe the application of singular value decomposition and the various properties that stem from this approach; such properties include the definition and interpretation of principal coordinates, their distance from each other (or the origin) in a low-dimensional space and their role in modelling the association between the categorical variables. In Section 5 we show how particular values of δ lead to techniques that currently exist and to new variants that have yet to be explored in any significant detail. For example, we show that when $\delta = 1$ the using the divergence residual leads to the traditional approach to correspondence analysis. We also show that when $\delta = 1/2$, our framework shares similar features with the Hellinger distance decomposition (HDD) method of Cuadras and Cuadras (2006) and leads to the correspondence analysis method based on the Freeman-Tukey statistic; see Beh, Lombardo, and Alberti (2018). We also show that when $\delta = 0$, the total inertia can be expressed using the modified likelihood ratio statistic and expands upon Greenacre’s (2009) “power family 2” form of log-ratio analysis (LRA). In Section 6 we discuss in more detail the links that exist between the correspondence analysis using the Cressie-Read divergence statistic and HDD and LRA. In doing so, we provide more insight into these two methods than was originally described by Cuadras and Cuadras (2006) and Greenacre (2009).

We demonstrate the application of this technique in Section 7 by studying the self-evaluation data collected from two consecutive Young-HUNT studies (Holmen et al., 2014). In doing so, we evaluate the benefits of using the divergence statistic as the foundation of performing correspondence analysis by comparing the configuration of points in a low-dimensional space, and other features, for specific values of δ . We also demonstrate how this technique can be used to determine the “best” and “worst” quality two-dimensional visual summary of the association. From the “best” plot that can be obtained, we examine other

aspects of the analysis including the interpretation of distances and the construction of association models. We provide some final remarks on the technique in Section 8.

2 The Cressie-Read Divergence Statistic

2.1 The Statistic

Consider an $I \times J$ two-way contingency table, \mathbf{N} , where the (i, j) th cell entry has a frequency of n_{ij} for $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$. Let the grand total of \mathbf{N} be n and let the matrix of relative frequencies be \mathbf{P} so that its (i, j) th cell entry is $p_{ij} = n_{ij}/n$ where $\sum_{i=1}^I \sum_{j=1}^J p_{ij} = 1$. Define the i th row marginal proportion by $p_{i\bullet} = \sum_{j=1}^J p_{ij}$. Similarly, define the j th column marginal proportion as $p_{\bullet j} = \sum_{i=1}^I p_{ij}$.

To determine whether there exists a statistically significant association between the row and column variables, one may calculate any number of measures. Five of the most common measures that are used for such a purpose include

$$\begin{aligned} X^2 &= n \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - p_{i\bullet}p_{\bullet j})^2}{p_{i\bullet}p_{\bullet j}} \\ G^2 &= 2n \sum_{i=1}^I \sum_{j=1}^J p_{ij} \ln \left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right) \\ T^2 &= 4n \sum_{i=1}^I \sum_{j=1}^J \left(\sqrt{p_{ij}} - \sqrt{p_{i\bullet}p_{\bullet j}} \right)^2 \\ N^2 &= n \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - p_{i\bullet}p_{\bullet j})^2}{p_{ij}} \\ M^2 &= 2n \sum_{i=1}^I \sum_{j=1}^J p_{i\bullet}p_{\bullet j} \ln \left(\frac{p_{i\bullet}p_{\bullet j}}{p_{ij}} \right) \end{aligned}$$

which are, respectively, Pearson's chi-squared statistic (Pearson, 1904), the log-likelihood ratio statistic (Wilks, 1938), the Freeman-Tukey statistic (Freeman & Tukey, 1950), the modified chi-squared statistic (Neyman, 1940, 1949) and the modified log-likelihood ratio statistic (Kullback, 1959). All these statistics are chi-squared random variables with $(I - 1)(J - 1)$ degrees of freedom and can all be expressed in terms of $p_{ij}/(p_{i\bullet}p_{\bullet j})$ which is the *Pearson ratio* of the (i, j) th cell of the contingency table. Note that it is just the ratio of the observed cell

proportion to what is expected under independence; see Goodman (1996). Beh (2004) and Beh and Lombardo (2014, pg. 123) used this ratio in the context of correspondence analysis while Greenacre (2009) refers to it as a *contingency ratio* in his description of LRA.

Another feature that these measures of association share is that they are special cases of the Cressie-Read divergence statistic (Cressie & Read, 1984) which, for a two-way contingency table, is defined as

$$\text{CR}(\delta) = \frac{2}{\delta(\delta+1)} \sum_{i=1}^I \sum_{j=1}^J n_{ij} \left\{ \left(\frac{n_{ij}}{np_{i\bullet}p_{\bullet j}} \right)^\delta - 1 \right\}. \quad (1)$$

Here, δ is the parameter of interest and lies within the interval $\delta \in (-\infty, \infty)$. The general nature of (1) ensures that specific values of δ lead to specific measures of association, all of which are chi-squared random variables. These include Pearson's chi-squared statistic, the log-likelihood ratio statistic, and the Freeman-Tukey statistic which are $X^2 = \text{CR}(\delta = 1)$, $G^2 = \text{CR}(\delta = 0)$ and $T^2 = \text{CR}(\delta = -1/2)$, respectively. The modified chi-squared statistic and the modified log-likelihood ratio statistic are also special cases such that $N^2 = \text{CR}(\delta = -2)$ and $M^2 = \text{CR}(\delta = -1)$, respectively.

At the suggestion of F. J. Anscombe in a private communication in 1985, Read and Cressie (1988, pg. 93) proposed the following amendment to their statistic

$$\text{CR}(\delta) = \frac{2}{\delta(\delta+1)} \sum_{i=1}^I \sum_{j=1}^J \left\{ n_{ij} \left[\left(\frac{n_{ij}}{np_{i\bullet}p_{\bullet j}} \right)^\delta - 1 \right] + \delta (np_{i\bullet}p_{\bullet j} - n_{ij}) \right\}. \quad (2)$$

Note that (2) leads to the same quantity given by (1) since, under the hypothesis of independence,

$$\sum_{i=1}^I \sum_{j=1}^J (np_{i\bullet}p_{\bullet j} - n_{ij}) = 0.$$

The advantage of considering (2) instead of (1) is that (2) can be expressed as a weighted sum-of-squares quantity which makes it suitable for performing correspondence analysis; we discuss this in more detail in Section 2.2. Irrespective of the choice of δ , for a given level of significance α , the statistical significance of the association may be tested by comparing $\text{CR}(\delta)$ with the $1 - \alpha$ percentile of the chi-squared distribution with $(I - 1)(J - 1)$ degrees of freedom.

In the context of goodness-of-fit testing for categorical variables, Cressie and Read (1984) suggest that the most appropriate values of δ to choose from are those that lie within the interval $[0, 3/2]$. For a two-way contingency table where

$n > 10$ and $\min(np_{i\bullet}p_{\bullet j}) > 1$ for all $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$, they advised that an appropriate choice of δ is $2/3$. Such a value leads to the chi-squared random variable

$$\text{CR} = \text{CR} \left(\delta = \frac{2}{3} \right) = \frac{9n}{5} \sum_{i=1}^I \sum_{j=1}^J p_{ij} \left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right)^{2/3} - 1 \right) \quad (3)$$

which is commonly referred to as the *Cressie-Read statistic*.

2.2 Approximation to the Divergence Statistic

Read and Cressie (1988, pg. 94 – 95) demonstrate that a second order Taylor series approximation of (2) around $(p_{ij}/(p_{i\bullet}p_{\bullet j}))^\delta = 1$ is

$$\text{CR}(\delta) \approx \text{CR}^*(\delta) = \frac{n}{\delta^2} \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij}^\delta - (p_{i\bullet}p_{\bullet j})^\delta)^2}{(p_{i\bullet}p_{\bullet j})^{2\delta-1}}. \quad (4)$$

Note that this approximation preserves the double centring of the the p_{ij}^δ values through its expected value under independence, $(p_{i\bullet}p_{\bullet j})^\delta$, for all δ . Note also that (4) can be alternatively, and equivalently, expressed as a power transformation of the Pearson ratio's such that

$$\text{CR}^*(\delta) = n \sum_{i=1}^I \sum_{j=1}^J p_{i\bullet}p_{\bullet j} \left[\frac{1}{\delta} \left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right)^\delta - 1 \right) \right]^2. \quad (5)$$

Equation's (4) and (5) show that the divergence statistic can be approximated by a weighted sum-of-squares of the power transformation of the ratio of the observed cell proportions compared to what is expected under the hypothesis of independence. Power transformations have been discussed in the correspondence analysis literature in the past; one may refer to Cuadras and Cuadras (2006), Cuadras, Cuadras, and Greenacre (2006) and Greenacre (2009, 2010) as key examples of this research. However, as we shall discuss further in Section 6, tying together the links between the divergence statistic and power transformations has not been discussed before, at least to the best of our knowledge.

Before we move on to the derivation of the various features of correspondence analysis using (5), the special cases that we will be considering, $\delta = 0, 1/2$ and $\delta = 1$, lead to measures of association that are equivalent to those defined using (2). That is, for $\delta = 0$, (5) gives Kullback's (1959) modified log-likelihood ratio statistic, $M^2 = \text{CR}^*(0) = \text{CR}(-1)$; this result can be obtained by determining

the limiting value of $\text{CR}^*(\delta)$ as $\delta \rightarrow 0$. Similarly, when $\delta = 1$, the resulting measure of association is Pearson's chi-squared statistic, so that $X^2 = \text{CR}^*(1) = \text{CR}(1)$. For the Freeman-Tukey statistic, T^2 we yield such a measure when $\delta = 1/2$ so that $T^2 = \text{CR}^*(1/2) = \text{CR}(-1/2)$. We shall also be looking at a fourth case, when $\delta = 2/3$, such that (4) and (5) give a second order approximation of (3).

Read and Cressie (1988, pg. 96) point out, when choosing the appropriate value of δ to use,

Departures involving large ratios of the alternative to null expected frequencies in one or two cells are best detected using large values of δ , say $\delta = 5$.

Therefore, large values of δ may be chosen to identify those cell's of a contingency table that exhibit very large, positive, Pearson ratio's; that is, when $p_{ij}/(p_{i\bullet}p_{\bullet j}) \gg 1$. Using the notation adopted here, they also advocate choosing a large negative value of δ (such as $\delta = -5$) when the Pearson ratio's are close to zero. Read and Cressie (1988, pg. 97) also point out that δ close to 0 is not recommended when assessing the overall deviation of the cell frequencies from independence, instead suggesting that δ be chosen from the interval $[1/3, 2/3]$; such an argument gives credence to their choice of $\delta = 2/3$.

3 Special Cases of the Divergence Residual

3.1 The Divergence Residual

To visualise the association between the categorical variables of a two-way contingency table, the traditional approach to correspondence analysis quantifies the *total inertia* by $\phi^2 = X^2/n$; here X^2 is Pearson's chi-squared statistic. As Section 2 shows, Pearson's chi-squared statistic is a special case of the Cressie-Read divergence statistic. Therefore, a more general framework of correspondence analysis can be adopted that quantifies the total inertia by $\phi^2(\delta) = \text{CR}^*(\delta)/n$ for any given value of δ . Therefore, from (5), a generalisation of the total inertia is defined such that

$$\phi^2(\delta) = \sum_{i=1}^I \sum_{j=1}^J p_{i\bullet}p_{\bullet j} \left[\frac{1}{\delta} \left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right)^\delta - 1 \right) \right]^2 .$$

Let

$$r_{ij}(\delta) = \frac{\sqrt{p_{i\bullet}p_{\bullet j}}}{\delta} \left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right)^\delta - 1 \right) \quad (6)$$

be referred to as the *divergence residual* of the (i, j) th cell of the contingency table. Then the sum-of-squares of these residuals gives the total inertia

$$\phi^2(\delta) = \sum_{i=1}^I \sum_{j=1}^J r_{ij}^2(\delta) .$$

We now provide an overview of the various special cases of the divergence residual that can be obtained by focusing on four specific values of δ ; $\delta = 0, 1/2, 2/3$ and 1 . Of course, other values of δ lead to other special cases which can also be considered.

3.2 The Pearson Residual: $\delta = 1$

When $\delta = 1$ the divergence residual, (6), simplifies to the *Pearson residual* of the (i, j) th cell such that

$$r_{ij}(1) = \sqrt{p_{i\bullet}p_{\bullet j}} \left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} - 1 \right) = \frac{p_{ij} - p_{i\bullet}p_{\bullet j}}{\sqrt{p_{i\bullet}p_{\bullet j}}} .$$

Note that the sum-of-squares of these residuals gives the total inertia, $X^2/n = \phi^2(1)$, which is used as the basis for quantifying the association between two categorical variables in the traditional approach to correspondence analysis.

3.3 The Modified Log-Likelihood Ratio Residual: $\delta = 0$

The divergence residual, (6), does not exist when $\delta = 0$. However, we can determine the limiting value of the residual as $\delta \rightarrow 0$ by first noting that, using the Box-Cox transformation,

$$\ln \left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} \left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right)^\delta - 1 \right) . \quad (7)$$

Therefore, the limiting value of the residual as δ approaches zero is

$$\begin{aligned} r_{ij}(0) &= \sqrt{p_{i\bullet}p_{\bullet j}} \ln \left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right) \\ &= -\sqrt{p_{i\bullet}p_{\bullet j}} \ln \left(\frac{p_{i\bullet}p_{\bullet j}}{p_{ij}} \right) \end{aligned} \quad (8)$$

and is the *modified log-likelihood ratio residual*. It is termed as such since the sum-of-squares of these residuals gives the total inertia, M^2/n , where $M^2 = CR^*(0)$. Except for the presence of the negative sign, $r_{ij}(0)$ is the same quantity whose sum-of-squares leads Greenacre (2009, equation (12)) who described the method of LRA; more will be said on the link (6) and LRA in Section 6.

The residual (8) may also be derived from the first order Taylor series expansion of the natural logarithm; $\ln x \approx x - 1$, for $0 < x \leq 2$. Therefore

$$\ln \left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right) \approx \frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} - 1$$

and is valid when $0 < p_{ij}/(p_{i\bullet}p_{\bullet j}) \leq 2$, so that the expansion applies for a Pearson ratio centred around 1 (which coincides with independence for all cells). Cuadras et al. (2006) used this approximation to compare the HDD method of Cuadras and Cuadras (2006) with the traditional approach to correspondence analysis, although no link was made between the residual and any measure of total inertia.

By applying this logarithm approximation to (8), $r_{ij}(1)$ is a first order approximation of $r_{ij}(0)$; the changes in the features of correspondence analysis as one moves from the log-transformation to the linear case (akin to δ moving from 0 to 1) is the key to the discussion of LRA made by Greenacre (2009).

3.4 “Freeman-Tukey” Residual: $\delta = 1/2$

The divergence residual, (6), simplifies to the *Freeman-Tukey residual* when $\delta = 1/2$ which is defined as

$$r_{ij} \left(\frac{1}{2} \right) = 2 \left(\sqrt{p_{ij}} - \sqrt{p_{i\bullet}p_{\bullet j}} \right).$$

Beh et al. (2018) demonstrated the utility and applicability of using this residual for performing correspondence analysis. They also showed that the sum-of-squares of these residuals gives the total inertia, T^2/n of the contingency table. Applying a singular value decomposition (SVD) to the matrix of these residuals leads to the Hellinger Distance Decomposition (HDD) method described by Cuadras and Cuadras (2006). However, no link was made between these residuals and the Freeman-Tukey statistic by the authors. Section 6 discusses some points of overlap, and some differences that exist, between the correspondence analysis technique described by Beh et al. (2018) and the HDD method of Cuadras and Cuadras (2006).

3.5 “Cressie-Read” Residual: $\delta = 2/3$

Consider now $\delta = 2/3$. Then the divergence residual (6) simplifies to the *Cressie-Read residual* defined by

$$r_{ij} \left(\frac{2}{3} \right) = \frac{3}{2} \sum_{i=1}^I \sum_{j=1}^J \sqrt{p_{i\bullet} p_{\bullet j}} \left(\left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right)^{2/3} - 1 \right).$$

In this case the total inertia of the contingency table can be expressed as the sum of squares of these residuals such that

$$\phi^2 \left(\frac{2}{3} \right) = \frac{\text{CR}}{n} = n \sum_{i=1}^I \sum_{j=1}^J r_{ij}^2 \left(\frac{2}{3} \right)$$

We shall explore the correspondence analysis of a two-way contingency table using this total inertia in Section 5.4.

4 Correspondence Analysis and the Divergence Residual

4.1 SVD of the Divergence Residual

Recall that the traditional approach to simple correspondence analysis quantifies the association between categorical variables using the total inertia based on Pearson’s chi-squared statistic. Therefore, since the X^2 is a special case of (6) when $\delta = 1$ – see Section 3.2 – this suggests that a more general framework of correspondence analysis can be considered. Such a framework involves not just applying a SVD to the matrix of Pearson residuals but it can also be made by applying it to the matrix of divergence residuals. In doing so, the correspondence analysis of a two-way contingency table can be performed using any member of the Cressie-Read divergence statistic. For the (i, j) th cell of the contingency table, SVD is applied to $r_{ij}(\delta)$ so that

$$\frac{\sqrt{p_{i\bullet} p_{\bullet j}}}{\delta} \left(\left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right)^\delta - 1 \right) = \sum_{m=1}^{M^*} a_{im}(\delta) \lambda_m(\delta) b_{jm}(\delta). \quad (9)$$

Here $a_{im}(\delta)$ and $b_{jm}(\delta)$ is the i th and j th element, respectively, of the m th left and right singular vectors of the matrix of divergence residual for some value of

δ . These elements are constrained by

$$\sum_{i=1}^I a_{im}(\delta) a_{im'}(\delta) = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}, \quad \sum_{j=1}^J b_{jm}(\delta) b_{jm'}(\delta) = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}. \quad (10)$$

Similarly, $\lambda_m(\delta)$ is the m th largest singular value of the matrix of divergence residuals for a given δ where these values are arranged in descending order so that $1 > \lambda_1(\delta) > \lambda_2(\delta) > \dots, \lambda_{M^*}(\delta) > 0$. The value of M^* depends on the choice of δ ; when $\delta = 0$ or $\delta = 1$, $M^* = \min(I, J) - 1$ while $M^* = \min(I, J)$ for all other values of δ . Such a specification of M^* is also consistent with the comments made by Cuadras and Cuadras (2006) (in their discussion of HDD) and Greenacre (2009) (in his discussion of LRA) for $\delta = 1/2$ and $\delta = 1$, respectively. We shall describe this feature further in Section 6.

Note that by rearranging (6) we obtain the generalised SVD (GSVD) of the *adjusted divergence residual*

$$r_{ij}^*(\delta) = \frac{1}{\delta} \left(\left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right)^\delta - 1 \right) = \sum_{m=1}^{M^*} a_{im}^*(\delta) \lambda_m^*(\delta) b_{jm}^*(\delta) \quad (11)$$

where $\lambda_m^*(\delta) = \lambda_m(\delta)$, $a_{im}^*(\delta) = a_{im}(\delta) / \sqrt{p_{i\bullet}}$ and $b_{jm}^*(\delta) = b_{jm}(\delta) / \sqrt{p_{\bullet j}}$. Here

$$\sum_{i=1}^I p_{i\bullet} a_{im}^*(\delta) a_{im'}^*(\delta) = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases} \quad (12)$$

and

$$\sum_{j=1}^J p_{\bullet j} b_{jm}^*(\delta) b_{jm'}^*(\delta) = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}. \quad (13)$$

These results can help simplify our discussion of the features of correspondence analysis where the total inertia is measured in terms of δ . We now discuss some of these properties.

4.2 Principal Coordinates

To visually summarise the association between the row and column variables of \mathbf{N} , a low-dimensional correspondence plot can be constructed by plotting the principal coordinate of the i th row and j th column category along the m th axis using their principal coordinate

$$f_{im}(\delta) = a_{im}^*(\delta) \lambda_m^*(\delta) \quad (14)$$

$$g_{jm}(\delta) = b_{jm}^*(\delta) \lambda_m^*(\delta) \quad (15)$$

respectively. By jointly displaying these coordinates, one can gain an understanding of how particular row and column categories are linked. When the visual display consists of M^* dimensions it is referred to as an *optimal correspondence plot*, otherwise it is a low-dimensional space usually consisting of two dimensions (for ease of visualisation).

The constraints imposed upon $a_{im}^*(\delta)$ and $b_{jm}^*(\delta)$, defined by (12) and (13), lead to the following properties of the principal coordinates

$$\sum_{i=1}^I p_{i\bullet} f_{im}(\delta) f_{im'}(\delta) = \begin{cases} \lambda_m^2(\delta) & m = m' \\ 0 & m \neq m' \end{cases}$$

$$\sum_{j=1}^J p_{\bullet j} g_{jm}(\delta) g_{jm'}(\delta) = \begin{cases} \lambda_m^2(\delta) & m = m' \\ 0 & m \neq m' \end{cases}.$$

where $\lambda_m^2(\delta)$ is the principal inertia of the m th dimension of the correspondence plot. Therefore, irrespective of the choice of δ , the first dimension reflects (proportionally) more of the association than any other dimension in the correspondence plot. In fact, the total inertia, $\phi^2(\delta) = \text{CR}(\delta)/n$, is expressible in terms of these squared singular values and principal coordinates by

$$\phi^2(\delta) = \sum_{m=1}^{M^*} \lambda_m^2(\delta) = \sum_{m=1}^{M^*} \sum_{i=1}^I p_{i\bullet} f_{im}^2(\delta) = \sum_{m=1}^{M^*} \sum_{j=1}^J p_{\bullet j} g_{jm}^2(\delta)$$

While the principal coordinates, (14) and (15), can be used to visually summarise the association, there are other coordinates systems that can be used. These include the use of lambda-scaling described by Gower, Lubbe, and Le Roux (2011, Section 2.3.1) and Beh and Lombardo (2014, pg. 195). Alternatively, one may construct a biplot which has been described in considerable detail in the correspondence analysis literature by Greenacre (2010, Chapters 7 & 8), Gower et al. (2011) and Beh and Lombardo (2014, Section 4.5.3). We shall confine our attention here to visualising the association using principal coordinates.

4.3 Distance Measures

Now that we have defined the row and column profile coordinates, we turn our attention to their distance from the origin, and from each other, in the correspondence plot. While such distances apply to the set of row and column categories, we focus only on row point distances.

4.3.1 Distance from the Origin

Consider the power transformation of the i th centred row profile that takes the form

$$\left(\left(\frac{p_{i1}}{p_{i\bullet}} \right)^\delta - p_{\bullet 1}^\delta, \dots, \left(\frac{p_{ij}}{p_{i\bullet}} \right)^\delta - p_{\bullet j}^\delta, \dots, \left(\frac{p_{iJ}}{p_{i\bullet}} \right)^\delta - p_{\bullet J}^\delta \right).$$

The j th element of this profile can be expressed in terms of the transformed Pearson ratio such that

$$\left(\frac{p_{ij}}{p_{i\bullet}} \right)^\delta - p_{\bullet j}^\delta = p_{\bullet j}^\delta \left[\left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right)^\delta - 1 \right].$$

Therefore, its (weighted) squared Euclidean distance from zero is

$$d_I^2(i, 0; \delta) = \frac{1}{\delta^2} \sum_{j=1}^J \frac{1}{p_{\bullet j}^{2\delta-1}} \left[\left(\frac{p_{ij}}{p_{i\bullet}} \right)^\delta - p_{\bullet j}^\delta \right]^2 \quad (16)$$

$$\begin{aligned} &= \frac{1}{\delta^2} \sum_{j=1}^J \frac{1}{p_{i\bullet}} \frac{\left((p_{ij})^\delta - (p_{i\bullet} p_{\bullet j})^\delta \right)^2}{(p_{i\bullet} p_{\bullet j})^{2\delta-1}} \\ &= \sum_{j=1}^J p_{\bullet j} \left[\frac{1}{\delta} \left(\left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right)^\delta - 1 \right) \right]^2. \end{aligned} \quad (17)$$

The term $1/(\delta^2 p_{\bullet j}^{2\delta-1})$ in (16) is the weight chosen to accommodate for the value of δ under consideration and ensures that a meaningful interpretation of these distance measures can be made. To show that such a meaningful distance interpretation is gained, $d_I^2(i, 0; \delta)$ can also be expressed as the squared Euclidean distance of a row point from the origin in the correspondence plot. This can be shown by substituting the right hand side of (11) into (17) resulting in

$$\begin{aligned} d_I^2(i, 0; \delta) &= \sum_{j=1}^J p_{\bullet j} \left(\sum_{m=1}^{M^*} a_{im}^*(\delta) \lambda_m(\delta) b_{jm}^*(\delta) \right)^2 \\ &= \sum_{m=1}^{M^*} \left(\sum_{j=1}^J p_{\bullet j} (b_{jm}^*(\delta))^2 \right) (a_{im}^*(\delta) \lambda_m(\delta))^2 \\ &= \sum_{m=1}^{M^*} (f_{im}^*(\delta))^2 \end{aligned}$$

and is the squared-Euclidean distance of the i th row category from the origin in the correspondence plot. Since the Cressie-Read divergence statistic can be

expressed in terms of the Pearson ratio's - see equation (4) - then it can also be expressed in terms of this distance measure by

$$\text{CR}^*(\delta) = n \sum_{i=1}^I p_{i\bullet} d^2(i, 0; \delta) = n \sum_{i=1}^I \sum_{m=1}^{M^*} p_{i\bullet} (f_{im}^*(\delta))^2.$$

This shows that, irrespective of the choice of δ , if all of the principal coordinates along all dimensions lie at the origin of the correspondence plot then the total inertia, and hence the Cressie-Read divergence statistic, is zero. The further away from the origin a point lies, the more dominant their contribution is to the association between the categorical variables.

4.3.2 Interpoint Distances

Consider now the squared distance between the i th and i' th row profiles for some value of δ . This distance is

$$\begin{aligned} d_I^2(i, i'; \delta) &= \frac{1}{\delta^2} \sum_{j=1}^J \frac{1}{p_{\bullet j}^{2\delta-1}} \left[\left(\left(\frac{p_{ij}}{p_{i\bullet}} \right)^\delta - p_{\bullet j}^\delta \right) - \left(\left(\frac{p_{i'j}}{p_{i'\bullet}} \right)^\delta - p_{\bullet j}^\delta \right) \right]^2 \\ &= \frac{1}{\delta^2} \sum_{j=1}^J \frac{1}{p_{\bullet j}^{2\delta-1}} \left[\left(\frac{p_{ij}}{p_{i\bullet}} \right)^\delta - \left(\frac{p_{i'j}}{p_{i'\bullet}} \right)^\delta \right]^2 \end{aligned} \quad (18)$$

and may also be expressed in terms of the power transformation of the (i, j) th Pearson ratio of the i th and i' th rows such that

$$d_I^2(i, i'; \delta) = \frac{1}{\delta^2} \sum_{j=1}^J p_{\bullet j} \left(\left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right)^\delta - \left(\frac{p_{i'j}}{p_{i'\bullet} p_{\bullet j}} \right)^\delta \right)^2 \quad (19)$$

$$= \sum_{j=1}^J p_{\bullet j} \left[\frac{1}{\delta} \left(\left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right)^\delta - 1 \right) - \frac{1}{\delta} \left(\left(\frac{p_{i'j}}{p_{i'\bullet} p_{\bullet j}} \right)^\delta - 1 \right) \right]^2. \quad (20)$$

This distance measure is equivalent to (Greenacre, 2009, equation (17)) in his description of power transformations in correspondence analysis, although δ was restricted to lie within the interval $[0, 1]$.

Substituting the right hand side of (11) into (20) yields

$$\begin{aligned} d_I^2(i, i'; \delta) &= \sum_{j=1}^J p_{\bullet j} \left[\sum_{m=1}^{M^*} (a_{im}^*(\delta) \lambda_m^*(\delta) - a_{i'm}^*(\delta) \lambda_m^*(\delta)) b_{jm}^*(\delta) \right]^2 \\ &= \left[\sum_{m=1}^{M^*} (f_{im}^2(\delta) - f_{i'm}^2(\delta)) \right]^2 \left(\sum_{j=1}^J p_{\bullet j} (b_{jm}^*(\delta))^2 \right) \\ &= \sum_{m=1}^{M^*} (f_{im}^2(\delta) - f_{i'm}^2(\delta))^2 \end{aligned}$$

and is the squared-Euclidean distance between the i th and i' th row profiles in the correspondence plot. This result shows that if two row profiles (whose elements are raised to the power of δ) are equivalent then they will share identical positions in the correspondence plot. Similarly, two points that have an identical position in the correspondence plot means that their profiles have identical distributions. Such a features ensures that, irrespective of the choice of δ , the property of distributional equivalence holds.

It is worth pointing out that (19) is just equation (13) of Cuadras and Cuadras (2006) where δ is akin to their power “parameter” $1 - \alpha$. Using the notation adopted here, the study of power transformations in correspondence analysis by Cuadras and Cuadras (2006) confined their attention to values of $\delta \in [0.5, 1]$. They also showed that the property of distributional equivalence is satisfied in their framework. They also continued their discussion by proposing a measure of agreement between the traditional approach to correspondence analysis (when $\delta = 1$) and their framework (when $\delta = 1/2$). We shall not discuss this issue further but feel that it is certainly a worth investigating in the future.

4.4 The Divergence Correlation Model

For the (i, j) th cell of a contingency table, any departure from independence between the variables can be assessed by quantifying the difference between the observed cell proportion and what is expected under the hypothesis of independence. Such a departure may also be determined by quantifying how different p_{ij}^δ is to $(p_{i\bullet}p_{\bullet j})^\delta$ or, equivalently, by assessing how different the divergence residual is from zero. To do so, rearranging (11) leads to the saturated correlation model for the (i, j) th cell proportion, $p_{ij} \equiv p_{ij}(\delta)$,

$$p_{ij}(\delta) = p_{i\bullet}p_{\bullet j} \left(1 + \delta \sum_{m=1}^{M^*} a_{im}^*(\delta) \lambda_m^*(\delta) b_{jm}^*(\delta) \right)^{1/\delta} \quad (21)$$

where $a_{im}^*(\delta)$ and $b_{jm}^*(\delta)$ are constrained by (12) and (13) respectively.

Equation (21) shows that, irrespective of δ , when the row and column variables are completely independent (so that $\lambda_m(\delta) = 0$ for $m = 1, \dots, M^*$) then $p_{ij}(\delta) = p_{i\bullet}p_{\bullet j}$ as expected, for $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$.

By using the row and column principal coordinates defined by (14) and (15)

p_{ij} may be estimated, for any δ , from

$$p_{ij}(\delta) = p_{i\bullet}p_{\bullet j} \left(1 + \delta \sum_{m=1}^{M^*} \frac{f_{im}(\delta) g_{jm}(\delta)}{\lambda_m^*(\delta)} \right)^{1/\delta} \quad (22)$$

and is the saturated correspondence model of the Cressie-Read divergence statistic. This model confirms that the origin of the correspondence plot coincides with independence between the row and column variables, irrespective of δ . It also shows that points far from the origin reflect a deviation from what we expect under independence for those row and/or column categories. How far from the origin a point needs to be before we conclude that it plays a statistically significant role in the association structure between the variables can be determined using confidence regions. We shall not discuss this issue here but we direct the interested reader to, for example, Lebart et al. (1984), Greenacre (2007), Beh (2001, 2010), Ringrose (2012), Beh and Lombardo (2015), (Beh & Lombardo, 2014, Chapter 8) and the references mentioned therein.

While Cressie and Read (1984) recommend using values of δ such that $\delta \in [0, 3/2]$ for inferential purposes, when exploring the structure of the association between two categorical variables using correspondence analysis, it is also interesting to see what impact δ has at its extremes. For example, since

$$\lim_{\delta \rightarrow \pm\infty} \left(1 + \delta \sum_{m=1}^{M^*} a_{im}^*(\delta) \lambda_m^*(\delta) b_{jm}^*(\delta) \right)^{1/\delta} = 1$$

then $p_{ij}(\pm\infty) \rightarrow p_{i\bullet}p_{\bullet j}$. Therefore, the row and column points in an optimal (or even a low) dimensional correspondence plot will get closer to the origin as $\delta \rightarrow \pm\infty$. The (relative) exception to this are those row and/or column categories that play a very dominant role in defining the association structure between the variables. The position of these categories will still approach the origin for increasing values of δ but do so more slowly than other categories.

One can also see that since

$$\lim_{\delta \rightarrow 0} \left(1 + \delta \sum_{m=1}^{M^*} a_{im}^*(\delta) \lambda_m^*(\delta) b_{jm}^*(\delta) \right)^{1/\delta} = \exp \left\{ \sum_{m=1}^{M^*} a_{im}^*(\delta) \lambda_m^*(\delta) b_{jm}^*(\delta) \right\}$$

then the limiting value of p_{ij} as $\delta \rightarrow 0$ is

$$p_{ij}(0) = p_{i\bullet}p_{\bullet j} \exp \left\{ \sum_{m=1}^{M^*} a_{im}^*(0) \lambda_m^*(0) b_{jm}^*(0) \right\} \quad (23)$$

and is akin (but not identical) to the RC model of Goodman (1979) and Rom and Sarkar (1992, equation (2.4)).

For all values of δ , (22) reduces to the independence model when $\lambda_m(\delta) = 0$ for $m = 1, \dots, M^*$. However, the interpretation of $\lambda_m(\delta)$ can differ substantially for different values of δ . For example, $\lambda_m(1)$ is a correlation term lying within the $[0, 1]$ interval, while $\lambda_m(0)$ is an association term whose value is bounded by $[0, \infty)$; see, for example, Gilula, Krieger, and Ritov (1988) for more on this point. Therefore, some care must be taken when assigning an interpretation to $\lambda_m(\delta)$.

5 Special Cases

Now that we have described the key features of correspondence analysis using the Cressie-Read divergence statistic, we turn our attention to examining some special cases. Rather than providing a comprehensive overview of each of these cases we shall focus instead on the structure of the SVD of the matrix of divergence residuals, (9), when $\delta = 0, 1/2, 2/3$ and 1. We shall also confine our attention to describing the various distances measures in these four cases.

5.1 CA using Pearson's Chi-squared Statistic

5.1.1 SVD of Divergence Residuals when $\delta = 1$

The traditional approach to correspondence analysis can be performed by substituting $\delta = 1$ into (9) so that

$$r_{ij}(1) = \frac{p_{ij} - p_{i\bullet}p_{\bullet j}}{\sqrt{p_{i\bullet}p_{\bullet j}}} = \sum_{m=1}^{M^*} a_{im}\lambda_m b_{jm} \quad (24)$$

where $a_{im} = a_{im}(1)$, $b_{jm} = a_{im}(1)$ and $\lambda_m = \lambda_m(1)$. By substituting $\delta = 1$ into (10), a_{im} and b_{jm} are constrained by

$$\sum_{i=1}^I a_{im}a_{im'} = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}, \quad \sum_{j=1}^J b_{jm}b_{jm'} = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}.$$

An equivalent approach to performing this version of correspondence analysis is to substitute $\delta = 1$ into (11) yielding the GSVD of the matrix of centred Pearson ratio's such that

$$r^*(1) = \frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} - 1 = \sum_{m=1}^{M^*} a_{im}^* \lambda_m^* b_{jm}^* \quad (25)$$

where $\lambda_m^* = \lambda_m(1) (= \lambda_m)$. Here $a_{im}^* = a_{im}(1) / \sqrt{p_{i\bullet}}$ and $b_{jm}^* = b_{jm}(1) / \sqrt{p_{\bullet j}}$ are constrained by (12) and (13), respectively, for $\delta = 1$. This approach to correspondence analysis was discussed by, for example, Beh (2004) and Beh and Lombardo (2014). In (24), λ_m (or, equivalently, in (25), λ_m^*) is the m th largest singular value of the $I \times J$ matrix of Pearson residuals (centred ratio's) and are arranged in descending order.

5.1.2 Distances when $\delta = 1$

When $\delta = 1$ is substituted into (18), the difference between the elements of i th and i' th row profiles can be assessed using the chi-squared distance so that

$$d_I^2(i, i'; 1) = \sum_{j=1}^J \frac{1}{p_{\bullet j}} \left(\frac{p_{ij}}{p_{i\bullet}} - \frac{p_{i'j}}{p_{i'\bullet}} \right)^2 = \sum_{i=1}^I \sum_{m=1}^{M^*} (f_{im}^* - f_{i'm}^*)^2.$$

Therefore, two identically distributed row profiles share identical positions in the optimal correspondence plot. This is the well known distance measure of two row points (or profiles) in the traditional approach to correspondence analysis.

5.2 CA using the Freeman-Tukey Statistic

5.2.1 SVD of Divergence Residuals when $\delta = 1/2$

By substituting $\delta = 1/2$ into (9) yields the SVD of the matrix of Freeman-Tukey residuals where the (i, j) th element is

$$r_{ij} \left(\frac{1}{2} \right) = 2 \left(\sqrt{p_{ij}} - \sqrt{p_{i\bullet} p_{\bullet j}} \right) = \sum_{m=1}^{M^*} \check{a}_{im} \check{\lambda}_m \check{b}_{jm}.$$

Here $\check{\lambda}_m = \lambda_m(1/2)$ is the m th singular value of the matrix of Freeman-Tukey residuals. Also, $\check{a}_{im} = a_{im}(1/2)$ and $\check{b}_{jm} = b_{im}(1/2)$ is the m th element of the i th left and j th right singular vectors of this matrix, respectively, and are constrained by

$$\sum_{i=1}^I \check{a}_{im} \check{a}_{im'} = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}, \quad \sum_{j=1}^J \check{b}_{jm} \check{b}_{jm'} = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}.$$

An equivalent approach to performing this variant of correspondence analysis is to substitute $\delta = 1/2$ into (11) leaving us with the GSVD of the centred square root of the Pearson ratio's so that

$$r_{ij}^* \left(\frac{1}{2} \right) = 2 \left(\sqrt{\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}}} - 1 \right) = \sum_{m=1}^{M^*} \check{a}_{im}^* \check{\lambda}_m^* \check{b}_{jm}^*.$$

where $\check{\lambda}_m^* = \check{\lambda}_m$, $\check{a}_{im}^* = \check{a}_{im}/\sqrt{p_{i\bullet}}$ and $\check{b}_{jm}^* = \check{b}_{jm}/\sqrt{p_{\bullet j}}$. In this case

$$\sum_{i=1}^I p_{i\bullet} \check{a}_{im}^* \check{a}_{im'}^* = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}, \quad \sum_{j=1}^J p_{\bullet j} \check{b}_{jm}^* \check{b}_{jm'}^* = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}.$$

Beh et al. (2018) described the correspondence analysis of a two-way contingency table using this residual and showed its link to the Freeman-Tukey statistic. Cuadras and Cuadras (2006) and Cuadras et al. (2006) also considered a rescaled version of this residual (to accommodate for a different scaling of the matrix of singular vectors) but did not link their approach to the Freeman-Tukey statistic.

5.2.2 Distances when $\delta = 1/2$

When centring the square-root transformation of a row profile, we can assess its distance from another square-root transformed row profile by substituting $\delta = 1/2$ into (18). Doing so yields

$$d_I^2\left(i, i'; \frac{1}{2}\right) = 4 \sum_{j=1}^J \left(\sqrt{\frac{p_{ij}}{p_{i\bullet}}} - \sqrt{\frac{p_{i'j}}{p_{i'\bullet}}} \right)^2 \quad (26)$$

and is the squared Hellinger distance between the profile of the i th and i' th rows. This distance can also be expressed in terms of their principal coordinates by

$$d_I^2\left(i, i'; \frac{1}{2}\right) = \sum_{m=1}^{M^*} (\check{f}_{im} - \check{f}_{i'm})^2,$$

where $\check{f}_{im} = f_{im}(1/2)$, which is the squared-Euclidean distance between these points in the optimal space. This distance measure was described by Greenacre (2009, equation (17)) and examined with reference to the Freeman-Tukey statistic by Beh et al. (2018). Beh et al. (2018) also discussed the role of the Freeman-Tukey statistic and residual in terms of row (and column) Hellinger distances in a low-dimensional correspondence plot; their discussion supported the comments of Domenges and Volle (1979), Rao (1995) and Cuadras and Cuadras (2006) who strongly advocated the use of Hellinger distances in correspondence analysis. We shall highlight these distance measures in Section 5.3.2 where they are described in the context of a special case of the divergence residual.

5.3 CA using the Modified Log-Likelihood Ratio Statistic

5.3.1 SVD of Divergence Residuals when $\delta = 0$

As we described in Section 3.1, (6) clearly does not exist for $\delta = 0$. Although, from (7), the SVD of $r_{ij}(0)$ is just the SVD of the modified log-likelihood ratio residual, (8), such that

$$-\sqrt{p_{i\bullet}p_{\bullet j}} \ln \left(\frac{p_{i\bullet}p_{\bullet j}}{p_{ij}} \right) = \sum_{m=1}^{M^*} a_{im}(0) \lambda_m(0) b_{jm}(0) .$$

Equivalently, a correspondence analysis can be performed on a two-way contingency table by applying a GSVD on the matrix where the (i, j) th element the natural logarithm of the Pearson ratio so that

$$\ln \left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right) = \sum_{m=1}^{M^*} \tilde{a}_{im}^* \tilde{\lambda}_m^* \tilde{b}_{jm}^* . \quad (27)$$

where $\tilde{\lambda}_m^* = \lambda_m(0)$. Here, $\tilde{a}_{im}^* = a_{im}(0) / \sqrt{p_{i\bullet}}$ and $\tilde{b}_{jm}^* = b_{jm}(0) / \sqrt{p_{\bullet j}}$ are constrained by

$$\sum_{i=1}^I p_{i\bullet} \tilde{a}_{im}^* \tilde{a}_{im'}^* = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases} , \quad \sum_{j=1}^J p_{\bullet j} \tilde{b}_{jm}^* \tilde{b}_{jm'}^* = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases} ,$$

respectively. In the event that a zero-cell frequency is observed there are a range of strategies that can be considered, such as the formal procedures described by Ishii-Kuntz (1994). One may also simply replace the zero cell frequency with a small but positive value such as 0.2 (Evers & Namboodiri, 1979), 0.05 (Beh & Farver, 2009) or 10^{-8} (Clogg and Eliason, 1987, p. 13). For the sake of simplicity, we shall replace the zero cell frequency with 0.05.

Since the first order Taylor series expansion of the natural logarithm function is $\ln x \approx x - 1$, for $0 < x \leq 2$, (25) is a first order Taylor series approximation of (27) and is used as the basis for comparing traditional correspondence analysis with the LRA of a contingency table; see Cuadras and Cuadras (2006) and Greenacre (2009).

5.3.2 Distances when $\delta = 0$

The squared distance between the i th and i' row profiles when $\delta = 0$ is

$$d_I^2(i, i'; 0) = \lim_{\delta \rightarrow 0} d_I^2(i, i'; \delta)$$

$$\begin{aligned}
&= \sum_{j=1}^J p_{\bullet j} \left[\lim_{\delta \rightarrow 0} \frac{1}{\delta} \left(\left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right)^\delta - 1 \right) - \lim_{\delta \rightarrow 0} \frac{1}{\delta} \left(\left(\frac{p_{i'j}}{p_{i'\bullet} p_{\bullet j}} \right)^\delta - 1 \right) \right]^2 \\
&= \sum_{j=1}^J p_{\bullet j} \left(\ln \left(\frac{p_{ij}}{p_{i\bullet} p_{\bullet j}} \right) - \ln \left(\frac{p_{i'j}}{p_{i'\bullet} p_{\bullet j}} \right) \right)^2 \\
&= 2 \sum_{j=1}^J p_{\bullet j} \ln \left(\frac{p_{ij}/p_{i\bullet}}{p_{i'j}/p_{i'\bullet}} \right) \tag{28}
\end{aligned}$$

so that if two row profiles have identical distributions, they will share an identical position in the correspondence plot where the total inertia is measured in terms of the modified log-likelihood ratio statistic, M^2 . Equation (28) was briefly discussed by Greenacre (2009, pg. 3113) in the context of LRA. However, from Section 4.3.2, it follows that this distance is equivalent to the squared-Euclidean distance between the i th and i' th row points in the correspondence plot so that

$$d_I^2(i, i'; 0) = \sum_{m=1}^{M^*} (\tilde{f}_{im}^* - \tilde{f}_{i'm}^*)^2$$

where $\tilde{f}_{im}^* = f_{im}(0)$ is the principal coordinate of the i th row category along the m th principal axis.

Note that unlike other measures of squared distance that involve the difference between two transformed profile elements, the presence of the natural logarithm function means that such a distance is expressed as the ratio of the two elements. Of course, this distance can be expressed as the difference between two logarithm terms such that

$$d_I^2(i, i'; 0) = \sum_{j=1}^J p_{\bullet j} \left\{ \ln \left(\frac{p_{ij}}{p_{i\bullet}} \right) - \ln \left(\frac{p_{i'j}}{p_{i'\bullet}} \right) \right\}^2$$

and was a focus of the recent discussion by Gzyl (2019) who studied logarithmic distances in the context of positive values from two random variables.

5.4 CA using Cressie-Read Residual

Now that we have examined the features of correspondence analysis when using the Freeman-Tukey, modified log-likelihood and Pearson's statistics as the basis for quantifying the total inertia, we now do so for $\delta = 2/3$.

5.4.1 SVD of Divergence Residuals when $\delta = 2/3$

When $\delta = 2/3$, a correspondence analysis can be performed by applying the SVD to the matrix of these Cressie-Read residuals such that

$$\frac{3}{2}\sqrt{p_{i\bullet}p_{\bullet j}} \left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right)^{2/3} - 1 \right) = \sum_{m=1}^{M^*} \dot{a}_{im} \dot{\lambda}_m \dot{b}_{im}$$

where $\dot{a}_{im} = a_{im}(2/3)$, $\dot{b}_{jm} = b_{jm}(2/3)$ and $\dot{\lambda}_m = \lambda_m(2/3)$. This variant of correspondence analysis is equivalent to performing the GSVD to the matrix of centred power (of $2/3$) transformed Pearson ratio's so that

$$\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}} \right)^{2/3} - 1 = \frac{2}{3} \sum_{m=1}^{M^*} \dot{a}_{im}^* \dot{\lambda}_m^* \dot{b}_{im}^*$$

where $\dot{a}_{im}^* = \dot{a}_{im}/\sqrt{p_{i\bullet}}$, $\dot{b}_{jm}^* = \dot{b}_{jm}/\sqrt{p_{\bullet j}}$ and $\dot{\lambda}_m^* = \dot{\lambda}_m$. Therefore, the constraints imposed upon \dot{a}_{im}^* and \dot{b}_{jm}^* are

$$\sum_{i=1}^I p_{i\bullet} \dot{a}_{im}^* \dot{a}_{im'}^* = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases}, \quad \sum_{j=1}^J p_{\bullet j} \dot{b}_{jm}^* \dot{b}_{jm'}^* = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases},$$

respectively.

5.4.2 When $\delta = 2/3$

In the special case when $\delta = 2/3$ - which as we saw above concerns the Cressie-Read statistic - the weighted squared Euclidean distance of the i th transformed row profile to the origin is

$$d_I^2 \left(i, i'; \frac{2}{3} \right) = \frac{9}{4} \sum_{j=1}^J p_{\bullet j}^{-1/3} \left[\left(\frac{p_{ij}}{p_{i\bullet}} \right)^{2/3} - \left(\frac{p_{i'j}}{p_{i'\bullet}} \right)^{2/3} \right]^2$$

so that such a distances relies less on the column marginal proportions than if $\delta = 0$ or $\delta = 1$. More will be said on the impact of the column marginal proportions on the distance between two row profiles in the following section.

6 Links with LRA & HDD

Throughout our discussion, we have directly compared the results that stem from the correspondence analysis using the Cressie-Read divergence statistic with

other methods that have been proposed in the past. In particular, some comparisons have been made with this technique and the LRA of Greenacre (2009) and the Hellinger Distance Decomposition (HDD) method described by Cuadras and Cuadras (2006). In doing so, these two methods can be viewed as special cases of the technique proposed in this paper. However, there are other points of discussion that are worth highlighting and so we now turn our attention to these.

6.1 Benefits of $\delta = 1/2$

Rao (1995) points out that there are benefits of using the Hellinger distance (when $\delta = 1/2$) instead of the chi-squared distance ($\delta = 1$) when comparing the difference in profiles of two intra-variable categories. One of the key benefits is that the Hellinger distance between the i th and i' th row profiles (say) does not rely on any information contained in the column marginals. Such a feature is observed in our framework as well and this can be verified by considering (18). It can be expressed in terms of the square root of the elements of the profiles by

$$d_I^2(i, i'; 2\delta) = \frac{4}{(2\delta)^2} \sum_{j=1}^J \frac{1}{p_{\bullet j}^{2\delta-1}} \left[\left(\sqrt{\frac{p_{ij}}{p_{i\bullet}}} \right)^{2\delta} - \left(\sqrt{\frac{p_{i'j}}{p_{i'\bullet}}} \right)^{2\delta} \right]^2$$

so that the Hellinger distance of the row profiles is observed only when $\delta = 1/2$ and a chi-squared distance arises only when $\delta = 1$. Note that as $\delta \rightarrow \infty$ the impact on the column marginal proportions becomes more important. However, their dominance on this distance measure also diminishes as $\delta \rightarrow -\infty$ and we would consider these values when identifying the row and column pairs whose expected values are less than what would be expected (in the extreme case) under the hypothesis of independence.

6.2 On the value of M^*

Cuadras and Cuadras (2006) and Cuadras et al. (2006) point out that in the traditional approach to correspondence analysis, $M^* = \min(I, J) - 1$ while for HDD $M^* = \min(I, J)$. We have also identified this feature when $\delta = 1$ and $\delta = 1/2$, respectively; we have also found that $M^* = \min(I, J)$ for Greenacre's (2009) LRA (when $\delta = 0$) while for all values of δ (except 0 or 1) then $M^* = \min(I, J) - 1$. Cuadras and Cuadras (2006) and Cuadras et al. (2006) also state that the additional dimension when using the Hellinger distance is a drawback. The inclusion

of an additional dimension certainly appears to be a problem unless one focuses instead on the quality of the two- (or three-) dimensional correspondence plot used to visualise the association. While the Hellinger distance (and the distance in LRA, see (28)) requires one more dimension than the traditional approach to correspondence analysis, the focus should instead be on the relative size of the principal inertia of a two- (or three-) dimensional display for some value of δ . That is, for a two-dimensional correspondence plot, say, the inclusion of another dimension may have a negligible effect on $(\lambda_1^2(\delta) + \lambda_2^2(\delta)) / \phi^2(\delta)$ since the total inertia $\phi^2(\delta)$ may be significantly smaller than $\phi^2(1)$. As Cuadras et al. (2006) show in their example, as we do in ours in Section 7, the extra dimension needed for $\delta = 0$ or $\delta = 1/2$ is negated by the improvements in the quality of the two-dimensional correspondence plot yielded from performing a correspondence analysis using these δ values.

6.3 On modelling the association

While it has long been established that the traditional approach to correspondence analysis has links to log-linear models Van der Heijden, de Falguerolles, and de Leeuw (1989); Goodman (1986), Cuadras et al. (2006) note that HDD (when $\delta = 1/2$) has no links to such models. This is not necessarily a problem since, as Beh et al. (2018) point out, there are some advantages in considering the resulting association model which is

$$p_{ij} = p_{i\bullet} p_{\bullet j} \left(1 + \frac{1}{2} \sum_{m=1}^{M^*} \check{a}_{im} \check{\lambda}_m \check{b}_{jm} \right)^2$$

which can be derived from (21) when $\delta = 1/2$. From this result complete independence arises when all $\lambda_m^* = 0$, $m = 1, 2, \dots, M^* = \min(I, J)$. It is also clear that this result will yield a positive cell proportion for the (i, j) th cell, a property that is not guaranteed with the traditional approach to correspondence analysis. In fact, from (21) a positive cell proportion will only be guaranteed when δ is an even number. This includes $\delta = 0$ where the resulting association model is given by (23).

6.4 On the choice of δ

By performing a correspondence analysis where the Cressie-Read divergence statistic, (2), rests at its foundations, virtually any value of δ can be consid-

ered. While, for inferential purposes, $\delta \in [0, 3/2]$ is recommended, values of δ outside of this range can be considered for identifying row/column pairs whose expected value greatly differs from what is expected under the hypothesis of independence. For example, the traditional approach to correspondence analysis and LRA are very much related through the divergence residual; the former occurs when $\delta = 1$ while the latter when δ is asymptotically zero. Thus, as the title of Greenacre (2010) says, “Log-ratio analysis is a limiting case of correspondence analysis”. However, in the LRA of Greenacre (2009), δ was confined to lie within the interval $[0, 1]$ so that comparisons could be made of the configuration of points (defined by their principal coordinates) when comparing LRA (when $\delta = 0$) with the traditional correspondence analysis approach ($\delta = 1$); Greenacre (2009) considered the values $\delta = 0, 0.25, 0.5, 0.75$ and 1 in his study of LRA. However, in his description of LRA, no discussion was given on how to interpret the total inertia for a given δ value, or what δ value gave the “best” visualisation of the association. Further insight into LRA was provided by Sarnacchiaro, D’Ambra, and Camminatiello (2015) who highlighted its link to Altham’s index (Altham, 1970) that is used to quantify the association between the categorical variables of a two-way contingency table using log-odds ratio’s. An extension of this link for three-way contingency tables was described by Lombardo, Camminatiello, D’Ambra, and Beh (2021). In the framework of association models, de Rooij and Heiser (2005) discussed an alternative way to graphically depict log-odds ratio’s based on distances between principal coordinates.

By using the notation we have adopted in this paper, the HDD method described in Cuadras and Cuadras (2006) considered those values of δ that lie within the interval $[1/2, 1]$. This was done so that comparisons could be made of the Hellinger and chi-squared distances between two intra-variable profiles projected onto a low-dimensional (Euclidean) space. However, no discussion was given on the implications of the change in the total inertia or, in the case of $\delta = 1/2$, how it relates to the Freeman-Tukey statistic. More information on this can be found in Beh et al. (2018).

The technique outlined in this paper therefore shows that, not only is HDD and LRA special cases of the method, but that virtually any value of δ can be considered and the association is quantified in terms of the Cressie-Read divergence statistic. A caveat to this statement though is that, depending on the value, negative δ ’s can lead to some computational problems due to the SVD of (9).

7 Application

7.1 The HUNT Study, Norway

The Trøndelag Health Study (HUNT Study) is a large population study undertaken in the Nord-Trøndelag county of Norway and its database is used for medical and health-related research. For this study, participants are aged 20 years or older, and it was first undertaken between 1984 and 1986 (Krokstad et al., 2013). For those who are aged 13 – 19 years there is the Young-HUNT study where the first survey was held between 1995 and 1997 (and is referred to as Young-HUNT1) while a follow-up study was undertaken during the 2000 – 2001 period and is referred to as the Young-HUNT2 study (Holmen et al., 2014). One of the questions that was asked in these two studies was “How is your overall health at the moment?” and participants were asked to self-report using a four point scale with responses “Poor”, “Not very good”, “Good” and “Very good”. Table 1 is a cross-classification of the responses of 2348 adolescents who participated in Young-HUNT1 and Young-HUNT2. This two-way contingency table was studied by Breidablik, Meland, and Lydersen (2008) and Fagerland, Lydersen, and Laake (2017).

Table 1: Contingency table of 2348 adolescents who participated in Young-HUNT1 and Young-HUNT2 and self-reported on their overall health

Young-HUNT2	Young-HUNT1				Total
	Poor	Not very good	Good	Very Good	
Poor	2	3	3	3	11
Not very good	2	58	98	14	172
Good	8	162	949	252	1371
Very good	4	48	373	369	794
Total	16	271	1423	638	2348

7.2 Measuring the Association Between the Variables

The Pearson chi-squared test of independence yields a chi-squared statistic of 361.42 for Table 1. One may also determine the modified log-likelihood ratio test statistic, the Freeman-Tukey statistic and the Cressie-Read statistic for this

contingency table; these are 301.55, 286.33 and 297.66, respectively. Since these are all special cases of the Cressie-Read divergence statistic which is a chi-squared random variable with $(I - 1)(J - 1) = 9$ degrees of freedom, their p-value is less than 0.0001. Therefore, for all four special cases of the divergence statistic, there is a statistically significant association between the adolescent responses to the question asked in Young-HUNT1 and Young-HUNT2.

7.3 A Correspondence Analysis: $\delta = 0, 0.5, 2/3, 1$

Since there exists a statistically significant association between the two variables of Table 1, one can perform a correspondence analysis to visually identify how these variables are associated. When studying this association, we shall focus our attention on using the modified log-likelihood statistic ($\delta = 0$), the Freeman-Tukey statistic ($\delta = 1/2$), the Cressie-Read statistic ($\delta = 2/3$) and Pearson's statistic ($\delta = 1$) as the measures of association. Doing so produces a two-dimensional correspondence plot which is given by Figure 1; note that the (row) responses to the Young-HUNT1 study have the suffix “-1” while the (column) responses to the Young-HUNT2 study have the suffix “-2”. For these plots, Figures 1 a), b), c) and d) are constructed where the modified log-likelihood statistic, Freeman-Tukey statistic, Cressie-Read and Pearson statistics are used to determine the total inertia of the table. For these four values of δ , all two-dimensional correspondence plots display at least 90% of the association between the variables and give similar (but not identical) configurations of points. All four plots suggest that the “Good” and “Very good” (VG) responses across the two time periods are very similar while there is some variation in the “Not very good” (NVG) responses and even more variation in the “Poor” responses. There are also some obvious differences in the four configurations. For example, when $\delta = 0$, the configuration appears to be flipped along the first principal axis, but is otherwise similar to the other three plots. Also, the “Poor” response in both studies is located much further away from the origin when $\delta = 1$ than for any of the other values of δ .

There are also some similar, and very different, comparisons that can be made when observing the numerical summaries from the analysis. Table 2 summarises the total inertia, the first two principal inertia values for each of the four plots of Figure 1 and the percentage contribution of these two inertia values to the total inertia. We now discuss these summaries in the context of Figure 1.

Table 2: Total inertia, principal inertia values, and percentage contribution of principal inertia values to the total inertia of the two dimensions for and $\delta = 0, 0.5, 2/3$ and 1 of Table 1

Summary	$\delta = 0$	$\delta = 0.5$	$\delta = 2/3$	$\delta = 1$
$\lambda_1^2(\delta)$	0.111	0.105	0.105	0.109
$\lambda_2^2(\delta)$	0.011	0.015	0.018	0.031
% Cont.	95.36%	98.75%	97.39%	90.73%
$CR^*(\delta)$	$M^2 = 301.55$	$T^2 = 286.33$	$CR^2 = 297.65$	$X^2 = 361.42$

Suppose we consider the case where the traditional approach to correspondence analysis is applied to Table 1; that is, when $\delta = 1$. In doing so, Figure 1d) shows that the two-dimensional correspondence plot visually describes 90.73% of the association between the two variables of Table 1; see also the last column of Table 2. While this is a very good quality plot and is the most popular display to use in correspondence analysis, it is the poorest visual representation of the association among the four δ values considered. The association between the two variables can be better captured when either $\delta = 0$, $\delta = 0.5$ or $\delta = 2/3$ is used. Of these three values, the “best” two-dimensional display is obtained when $\delta = 0.5$ and summarises 98.75% of the association; see Figure 1b) (we shall talk more about the use of the term “best” in the next section). Therefore, rather than using Pearson’s chi-squared statistic as the measure of association, a much better visual display of the association can be obtained using the Freeman-Tukey statistic. In this case, rather than using the chi-squared distance to assess the difference between the profiles of the rows (and columns), Hellinger distances provide a more optimal view of their difference when projected onto a two-dimensional display. Calculating the total inertia using the Cressie-Read statistic ($\delta = 2/3$) yields a two-dimensional correspondence plot that depicts 97.39% of the association (Figure 1c)). Its quality and configuration of points is similar to that of Figure 1b) which should be of no surprise since the power of the transformation, δ , used to calculate the Freeman-Tukey statistic and the Cressie-Read statistic are quite similar. Using the modified log-likelihood ratio test ($\delta = 0$) statistic also provides a better two-dimensional display (95.36%) than when using Pearson’s statistic and the difference between the row profiles can be quantified using (28).

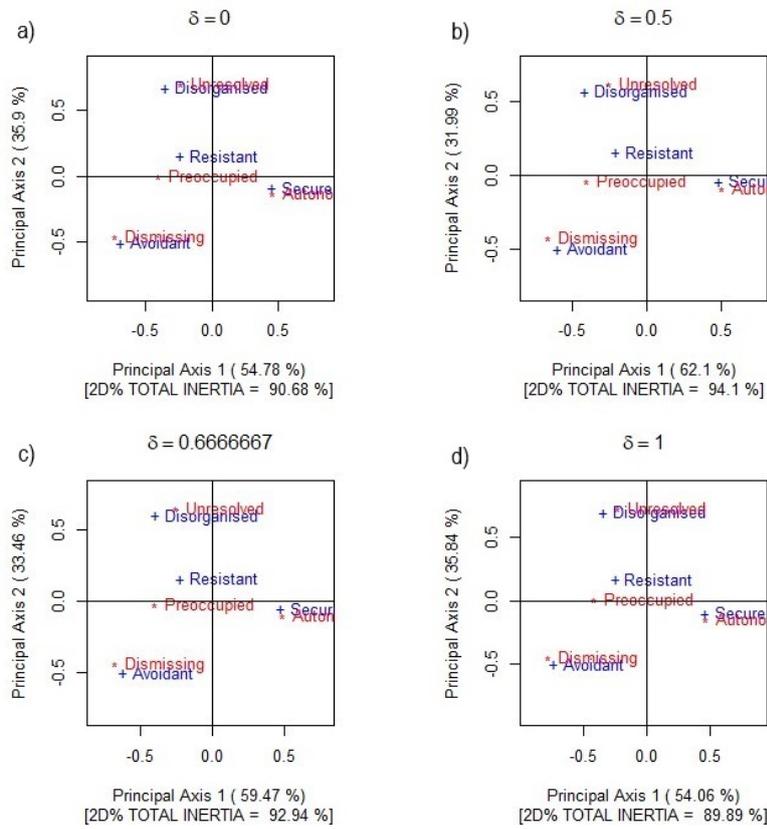


Figure 1: Correspondence plot of Table 1 using the Cressie-Read divergence statistic with a) $\delta = 0$, b) $\delta = 1/2$ c) $\delta = 2/3$ d) $\delta = 1$

7.4 Finding the “Best” & “Worst” Two-Dimensional Correspondence Plot

To gain an understanding of what impact changing δ has on the two-dimensional correspondence plot of Table 1, consider Figure 2 which shows the following. Figure 2a) gives the principal inertia values for the first principal axis (dashed line), second principal axis (dotted line) and the two axes combined (solid line) for $\delta \in [0, 3/2]$. Figure 2b) visually shows the percentage contribution of the first principal axis (dashed line), second principal axis (dotted line) and the combined inertia of these axes to the total inertia. Figure 2a) shows that the inertia of the first principal axis remains fairly stable at about 0.110 for $0 < \delta < 1.3$ then increases quite sharply for $\delta = 1.3$ to 1.5, while the second principal inertia increases steadily for $0 < \delta < 1$ then increases quite quickly for $\delta > 1$. While this might suggest that the quality of a two-dimensional display for values of δ ranging from 0 to 1.0 should be fairly similar, this is not the case since the total inertia rises steadily for values of δ greater than about 0.8 (not shown). As a result of this feature, the impact of the axes of a two-dimensional correspondence plot is variable for $\delta \in [0, 3/2]$; see Figure 2b). This plot shows that there is an increase in the percentage contribution to the total inertia of the first dimension for $\delta = 0$

(where the contribution is 76.7%) to a δ value slightly larger than zero (where the percentage contribution of this axis to the total inertia is 86.7%). However, there is a steady decline in quality of the first dimension as $\delta \rightarrow 1.34$ where its contribution to the total inertia reduces to about 40%, and then steadily increases again. On the other hand, the second principal axis becomes more dominant in the two-dimensional display as $\delta \rightarrow 1.34$ (increasing from about 8.9% to 40.5%) and then it reduces in quality. While not shown, for $3/2 < \delta < \infty$, the quality of the two-dimensional correspondence plot approaches 100% and does so very quickly for $\delta > 2$. For these values of δ , the impact of the second principal axis reduces to near zero so that all of the association can be depicted using a single (the first) principal axis. As a result of this feature, all categories lie relatively close to the origin except for the “Poor” category for each Young–HUNT survey. Thus, these two categories are the most dominant source of association between the two variables. In fact, the Pearson ratio for the (1, 1)th cell of Table 1 is $p_{11}/(p_{1\bullet}p_{\bullet 1}) = 26.68$ and is by far the largest of the 16 ratios in the contingency table, with the next largest only being $p_{22}/(p_{2\bullet}p_{\bullet 2}) = 2.92$ while the ratio closest to zero is $p_{24}/(p_{2\bullet}p_{\bullet 4}) = 0.30$. Therefore, choosing large values of δ (beyond 5 in this case) with the Cressie-Read divergence statistic in the context of correspondence analysis helps to visually identify cells that deviate greatly from independence.

By confining our attention to those values of δ that lie within $[0, 3/2]$, Figure 2b) shows that the poorest quality (or perhaps “worst” but we use this term with some caution) two-dimensional correspondence plot for Table 1 occurs when $\delta = 1.23$ so that only 87.47% of the association is visually depicted. For this value of δ , the first principal axis contributes to 54.58% of the association while 32.89% is depicted using the second axis. Figure 3a) gives the resulting two-dimensional correspondence plot and shows that the “Poor” responses are separated more clearly from the rest of the configuration than the plots in Figure 1; this feature is not surprising since δ is relatively large (in comparison with those considered in Figure 1) and so the dominance of the “Poor” category in both studies becomes more prominent. In fact, for this contingency table, the configuration in Figure 3a) matches most closely to Figure 1d) which is obtained using the traditional approach to correspondence analysis. For $\delta = 1.23$, the chi-squared

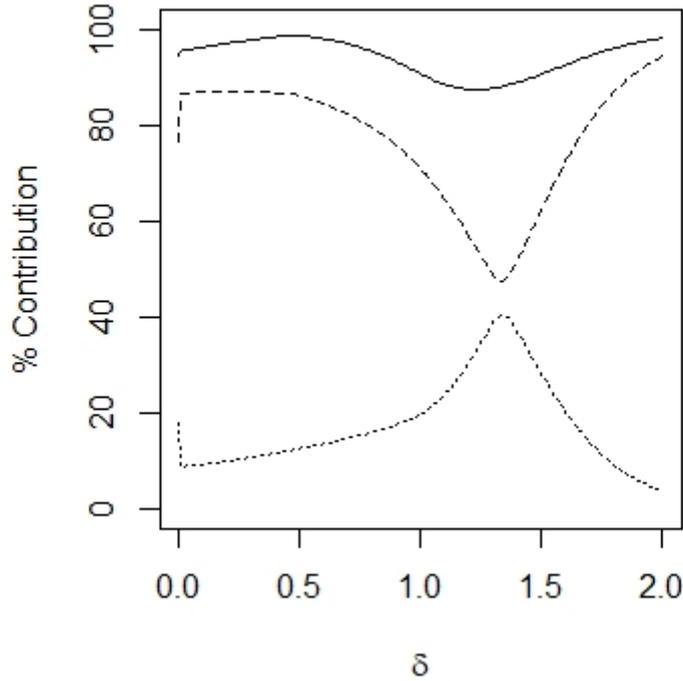


Figure 2: Solid line is the a) percentage contribution to the total inertia and b) Principal inertia values of a two-dimensional correspondence plot for $\delta \in [0, 3/2]$. The dashed lines refer to the first principal axis and the dotted line refers to the second principal axis.

statistic from which the total inertia of Figure 3a) is determined is

$$CR^*(1.23) = 1551.986 \sum_{i=1}^4 \sum_{j=1}^4 \frac{(p_{ij}^{1.23} - (p_{i\bullet} p_{\bullet j})^{1.23})^2}{(p_{i\bullet} p_{\bullet j})^{1.46}} = 496.105.$$

For values of $\delta \in [0, 3/2]$ the “best” quality display is achieved when δ is 0.479 and explains 98.86% percent of the total inertia; see Figure 3b). We use the term “best” here cautiously but mean for it to be the value of δ within the interval that produces a two-dimensional correspondence plot that explains as close to 100% of the association as possible. We do concede that the “best” quality display may be assessed using other criteria depending on what the analyst wishes to focus on and it is for this reason that the term is given in quotation marks. For example, one may defined “best” to be the value of δ that maximises the divergence statistic for $\delta \in [0, 3/2]$.

For Figure 3b), the first principal axis contributes to 86.43% of the association while the second axis accounts for 12.43%. When compared with the four figures in Figure 1, Figure 3b) most closely resembles the configuration of points in Figure 1b); this should not be surprising since this value of δ is closest to 0.5 than any of the three other values considered above. In this case, the chi-squared

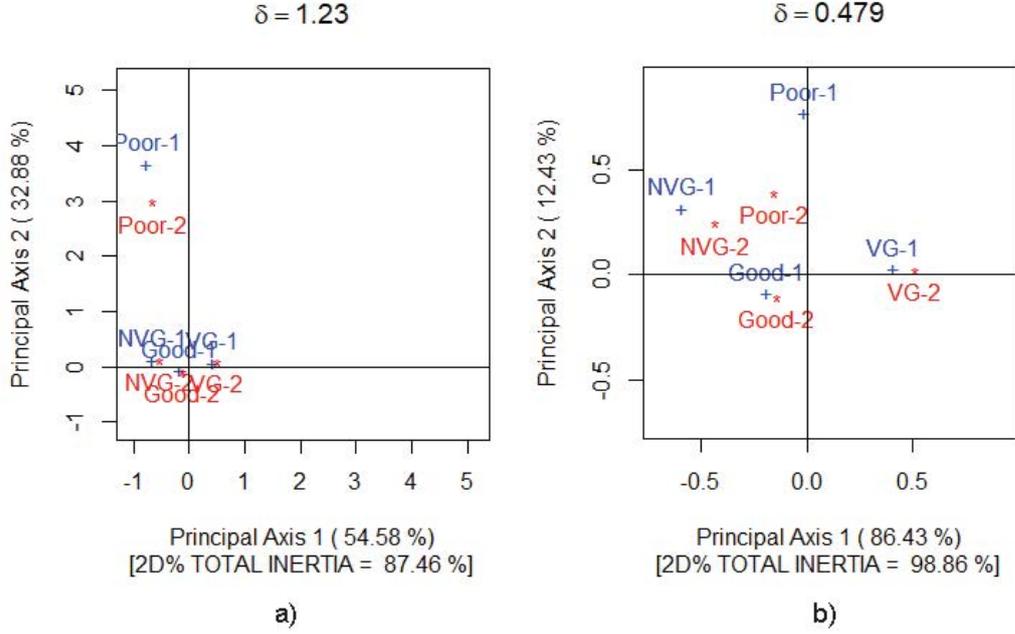


Figure 3: The a) “worst” and b) “best” two-dimensional correspondence plot of Table 1 for $\delta \in [0, 3/2]$

statistic that is used as the measure of association is

$$\begin{aligned} CR^*(0.479) &= 10233.67 \sum_{i=1}^4 \sum_{j=1}^4 (p_{i\bullet} p_{\bullet j})^{0.042} \left(p_{ij}^{0.479} - (p_{i\bullet} p_{\bullet j})^{0.479} \right)^2 \\ &= 285.461 \end{aligned}$$

and is very close to the Freeman-Tukey statistic of $T^2 = 286.33$.

The squared distance of two points (from the same variable) in Figure 1b) obtained using $\delta = 0.479$ can be assessed by observing the differences in the row, and column, profiles where the elements of the profiles are raised to the power of 0.479. When doing so, this squared distance of the i th and i' th row categories, say, in the optimal correspondence plot, can be expressed as

$$d_I^2(i, i'; \delta = 0.479) = 4.36 \sum_{j=1}^4 p_{\bullet j}^{0.042} \left(\left(\frac{p_{ij}}{p_{i\bullet}} \right)^{0.479} - \left(\frac{p_{i'j}}{p_{i'\bullet}} \right)^{0.479} \right)^2$$

and is an approximation of the Hellinger distance between two points; see (26). Therefore, the two-dimensional correspondence analysis that displays the maximal amount of the total inertia (regardless of the magnitude of the total inertia)

that exists in Table 1 is when $\delta = 0.479$. Thus, any differences between two row, or column, profiles can be better approximated using the Hellinger distance than the chi-squared distance and virtually all of this difference is captured in the two dimensions of Figure 3b). In fact, since $\delta \approx 0.5$ here, the distance between the row points in Figure 3b) are virtually independent of the column marginal proportions and this is reflected by the small contribution of $p_{\bullet j}^{0.042}$ in the calculation of CR^* (0.479). One can also make the same conclusion of the distance of the column points in this correspondence plot and the small contribution that $p_{i\bullet}^{0.042}$ makes in calculating this chi-squared statistic.

8 Discussion

The discussion of power transformations of profiles in correspondence analysis is not new. One only has to consult Greenacre (2009), Cuadras and Cuadras (2006) and Cuadras et al. (2006), for example, for a general discussion of this issue. Related to these discussions is the importance of the role of Hellinger distances in a correspondence plot, a feature advocated by Rao (1995) and Cuadras and Cuadras (2006) and explored in the context of the Freeman-Tukey statistic by Beh et al. (2018). However, these features can all be incorporated into the one framework by using the Cressie-Read divergence statistic as the primary measure of the association between the variables. Doing so means that the Pearson and Freeman-Tukey statistics can be considered as special cases of the divergence statistic, thereby providing a generalised approach to assess the nature of the association between the variables. By applying a SVD to the matrix of divergence residuals defined by (3) leads to a power transformation of the profiles where the power need not be constrained to lie within $[0, 1]$ but can take on any value that makes the Cressie-Read divergence statistic interpretable; by following the recommendation of Cressie and Read (1984) we have considered $\delta \in [0, 3/2]$ and discussed the implications of performing a correspondence analysis for values of $\delta > 3/2$, especially for identifying the most dominant categories that contribute to the association. As a result of the properties that stem from the SVD of (3), distance measures of two row, or column, points in the optimal correspondence plot can be assessed in terms of the power transformation applied to the row, or column, profiles. A special case is where the power is 0.5 and leads to Hellinger distances while chi-squared distances can be obtained for $\delta = 1$.

What is unique about this approach to correspondence analysis is that we can determine the value of δ that gives the “best” quality low-dimensional correspondence plot while still maintaining interpretability of principal coordinates, distances, association models and other important features. We can also use the same technique for identifying the “worst” quality plot. While we have restricted our attention to the case where such a plot consists of two-dimensions, a similar argument can be made for a correspondence plot consisting of three (or more) dimensions. By “best”, we have confined δ to lie within $[0, 3/2]$ and our application in Section 7 shows that $\delta = 1$ does not always provide the most optimal two-dimensional correspondence plot. As we have described in this paper, the “best” quality two-dimensional plot of Table 1 is when $\delta \approx 0.5$ so that the Freeman-Tukey statistic is a more appropriate measure of association that leads to such a plot, not Pearson’s chi-squared statistic which is commonly used in applications of correspondence analysis. Visually studying the association using the Freeman-Tukey statistic has a few advantages over using the Pearson statistic. Firstly, the Freeman-Tukey statistic can address any overdispersion issues that may be present in the contingency table; a topic of discussion raised by Agresti (2002), Haberman (1973) and studied from a correspondence analysis perspective by Beh et al. (2018). Secondly, when $\delta = 1/2$ Hellinger distances can be used to assess differences between two row profiles, say, leads to a low-dimensional display where the column marginal information plays no part in the configuration of the row points. Other applications of this technique will yield different values of δ that give the “best” visual representation of the association using the correspondence plot, all depending the nature of the data being analysed.

There are further issues that can be explored that are beyond the scope of this paper; we have alluded to a few throughout our discussion. One such issue is to determine a routine to determine the value of δ that leads to the “best” quality display. In this paper, this value has been determined by simply narrowing in on an interval where the appropriate value of δ lies. However, a more efficient approach would be to determine a more direct method of determining this value.

Another avenue for further investigation is to generalise this method for more than two categorical variables so that it can be adopted in the context of multiple correspondence analysis. One could certainly use the method outline above by applying it to the Burt matrix of a multi-way contingency table, or when stacking is used (Weller & Romney, 1990; Greenacre, 2007). While joint correspondence analysis (JCA) (Greenacre, 1988) is designed to improve the quality of a corre-

spondence plot by removing redundant information from the Burt matrix, further improvements could be made by incorporating the approach outlined into this paper into JCA. Alternatively, rather than transforming a multi-way contingency table into a two-way form that captures only bivariate association terms, the procedure could be adapted to analyse the association in a multi-way contingency table. This is certainly possible for the analysis of a three-way contingency table described by, for example, Carlier and Kroonenberg (1996) and Lombardo, Beh, and Kroonenberg (2021). This could be achieved by incorporating the extensions made to the Cressie-Read divergence statistic outlined in M. C. Pardo (1996), L. Pardo and Pardo (2003) and J. A. Pardo (2010) into the framework described in this paper.

References

- Agresti, A. (2002). *Categorical data analysis (2nd ed)*. New York: Wiley.
- Altham, P. M. E. (1970). The measurement of association of rows and columns for an $r \times s$ contingency table. *Journal of the Royal Statistical Society, Series B*, 32, 63 – 73.
- Beh, E. J. (2001). Confidence circles for correspondence analysis using orthogonal polynomials. *Journal of Applied Mathematics and Decision Sciences*, 5, 35 – 45.
- Beh, E. J. (2004). Simple correspondence analysis: A bibliographic review. *International Statistical Review*, 72, 257 – 284.
- Beh, E. J. (2010). Elliptical confidence regions for simple correspondence analysis. *Journal of Statistical Planning and Inference*, 140, 2582 – 2588.
- Beh, E. J., & Farver, T. B. (2009). An evaluation of non-iterative methods for estimating the linear-by-linear parameter of ordinal log-linear models. *Australian and New Zealand Journal of Statistics*, 51, 335 - 352.
- Beh, E. J., & Lombardo, R. (2012). A genealogy of correspondence analysis. *Australian and New Zealand Journal of Statistics*, 54, 137 – 168.
- Beh, E. J., & Lombardo, R. (2014). *Correspondence analysis: Theory, practice and new strategies*. Chichester: Wiley.
- Beh, E. J., & Lombardo, R. (2015). Confidence regions and p-values for classical and non-symmetric correspondence analysis. *Communications in Statistics Theory and Methods*, 95 - 114.

- Beh, E. J., & Lombardo, R. (2019). A genealogy of correspondence analysis: Part 2 - the variants. *Electronic Journal of Applied Statistical Analysis*, *12*, 552 - 603.
- Beh, E. J., Lombardo, R., & Alberti, G. (2018). Correspondence analysis and the freeman-tukey statistic: A study of archaeological data. *Computational Statistics and Data Analysis*, *128*, 73 - 86.
- Birks, H., Peglar, S., & Austin, H. (1996). An annotated bibliography of canonical correspondence analysis and related constrained ordination methods 1986 - 1993. *Abstracta Botanica*, *20*, 17 - 36.
- Breidablik, H.-J., Meland, E., & Lydersen, S. (2008). Self-rated health during adolescence: stability and predictors of change (young-HUNT study, norway). *European Journal of Public Health*, *19*, 73 - 78.
- Carrier, A., & Kroonenberg, P. M. (1996). Decompositions and biplots in three-way correspondence analysis. *Psychometrika*, *61*, 355-373.
- Clogg, C. C., & Eliason, S. R. (n.d.). Some common problems in log-linear analysis. *Sociological Methods & Research*.
- Cressie, N. A. C., & Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, *46*, 440 - 464.
- Cuadras, C. M., & Cuadras, D. (2006). A parametric approach to correspondence analysis. *Linear Algebra and its Applications*, *417*, 64 - 74.
- Cuadras, C. M., Cuadras, D., & Greenacre, M. (2006). A comparison of different methods for representing categorical data. *Communications in Statistics - Simulation and Computation*, *35*, 447 - 459.
- de Rooij, M., & Heiser, W. J. (2005). Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data. *Psychometrika*, *70*, 99 - 122.
- Domenges, D., & Volle, M. (1979). Analyse factorielle sphérique: une exploration. *Annales de l'inséé*, *35*, 3 - 84.
- Evers, M., & Namboodiri, N. (1979). On the design matrix strategy in the analysis of categorical data. In K. Shuessler (Ed.), *Sociological methodology* (pp. 86 - 111). San Francisco, CA: Jossey-Bass.
- Fagerland, M., Lydersen, S., & Laake, P. (2017). *Statistical analysis of contingency tables*. CRC Press.
- Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular and square root. *The Annals of Mathematical Statistics*, *21*, 607 - 611.
- Gilula, Z., Krieger, M., & Ritov, Y. (1988). Ordinal association in contingency

- tables: Some interpretative aspects. *Journal of the American Statistical Association*, *83*, 540 – 545.
- Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, *74*, 537 – 552.
- Goodman, L. A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review*, *54*, 243 – 270.
- Goodman, L. A. (1996). A single general method for the analysis of cross-classified data: reconciliation and synthesis of some methods of Pearson, Yule, and Fisher and also some methods of correspondence analysis and association analysis. *Journal of the American Statistical Association*, *91*, 408 – 428.
- Gower, J. C., Lubbe, S., & Le Roux, N. (2011). *Understanding biplots*. Wiley.
- Greenacre, M. (1984). *Theory and application of correspondence analysis*. London: Academic Press.
- Greenacre, M. (1988). Correspondence analysis of multivariate categorical data by weighted least-squares. *Biometrika*, *75*, 457 – 467.
- Greenacre, M. (2007). *Correspondence analysis in practice (2nd ed)*. London: Chapman & Hall/CRC.
- Greenacre, M. (2009). Power transformations in correspondence analysis. *Computational Statistics and Data Analysis*, *53*, 3107 – 3116.
- Greenacre, M. (2010). Log-ratio analysis is a limiting case of correspondence analysis. *Mathematical Geosciences*, *42*, 129 – 134.
- Gzyl, H. (2019). Best predictors in logarithmic distance between random variables. *Journal of Applied Mathematics, Statistics and Informatics*, *15*(2), 2 – 28.
- Haberman, S. (1973). The analysis of residuals in cross-classified tables. *Biometrics*, *75*, 457 – 467.
- Holmen, T., Bratberg, G., Krokstad, S., Langhammer, A., Hveem, K., Midthjell, K., ... Holmen, J. (2014). Cohort profile of the young-hunt study, norway: A population-based study of adolescents. *International Journal of Epidemiology*, *43*, 536 – 544.
- Ishii-Kuntz, M. (1994). *Ordinal log-linear models*. Newbury Park, CA: Sage University Series on Quantitative Applications in the Social Sciences.
- Krokstad, S., Langhammer, A., Hveem, K., Holmen, T., Midthjell, K., Stene,

- T., ... Holmen, J. (2013). Cohort profile: The HUNT study, Norway. *International Journal of Epidemiology*, 42, 968 – 977.
- Kullback, S. (1959). *Information theory and statistics*. Wiley.
- Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis*. Wiley.
- Lombardo, R., Beh, E. J., & Kroonenberg, P. M. (2021). Symmetrical and non-symmetrical variants of three-way correspondence analysis for ordered variables. *Statistical Science*, (in press).
- Lombardo, R., Camminatiello, I., D’Ambra, A., & Beh, E. J. (2021). Assessing the italian tax courts sytem by weighted three-way log-ratio analysis. *Socio-Economic Planning Sciences*, (to appear).
- Neyman, J. (1940). Contribution to the theory of certain test criteria. *Bulletin de L’Institut International de Statistique*, 24, 44 – 86.
- Neyman, J. (1949). Contributions to the theory of the χ^2 test. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, 239 – 273.
- Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. University of Toronto Press.
- Pardo, J. A. (2010). An approach to multiway contingency tables based on the ϕ -divergence test statistics. *Journal of Multivariate Analysis*, 101, 2305 – 2319.
- Pardo, L., & Pardo, M. C. (2003). Minimum power-divergence estimator in three-way contingency tables. *Journal of Statistical Computation and Simulation*, 73, 819 – 831.
- Pardo, M. C. (1996). An empirical investigation of cressie and read tests for the hypothesis of independence in three-way contingency tables. *Kybernetika*, 32, 175 – 183.
- Pearson, K. (1904). On the theory of contingency and its relation to association and normal correlation. *Drapers Memoirs*, Biometric Series, Vol 1, London.
- Plackett, R. (1983). Karl Pearson and the chi-squared test. *International Statistical Review*, 51, 59 – 72.
- Rao, C. R. (1995). A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Qüestió*, 19, 23 – 63.
- Read, T. R. C., & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. Springer-Verlag.

- Ringrose, T. J. (2012). Bootstrap confidence regions for correspondence analysis. *Journal of Computational Statistics and Simulation*, *82*, 1397 – 1413.
- Rom, D., & Sarkar, S. K. (1992). A generalized model for the analysis of association in ordinal contingency tables. *Journal of Statistical Planning and Inference*, *33*, 205 – 212.
- Sarnacchiaro, P., D’Ambra, L., & Camminatiello, I. (2015). Measures of association and visualization of log odds ratio structure for a two way contingency table. *Australian & New Zealand Journal of Statistics*, *57*, 363 – 376.
- Van der Heijden, P., de Falguerolles, A., & de Leeuw, J. (1989). A combined approach to contingency table analysis using correspondence analysis and log-linear analysis. *Journal of the Royal Statistical Society, Series C*, *38*, 249 – 292.
- Weller, S. C., & Romney, A. K. (1990). *Metric scaling: Correspondence analysis*. Newbury Park, CA: Sage: Sage University Paper Series on Quantitative Applications in the Social Sciences 07-075.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, *9*, 60 – 62.