Austrian Journal of Statistics
July 2023, Volume 52, 39–70.
http://www.ajs.or.at/
doi:10.17713/ajs.v52i3.1302

Taxicab Correspondence Analysis and Taxicab Logratio Analysis: A Comparison on Contingency Tables and Compositional Data

Vartan Choulakian
Université de Moncton

Jacques Allard
Université de Moncton

Smail Mahdi
University of the West Indies

Abstract

In this paper, we attempt to see further by relating theory with practice: First, we review the principles on which three interrelated well developed methods for the analysis and visualization of contingency tables and compositional data are erected: Correspondence analysis based on Benzécri's principle of distributional equivalence, Goodman's RC association model based on Yule's principle of scale invariance, and compositional data analysis based on Aitchison's principle of subcompositional coherence. Second, we introduce a novel index named intrinsic measure of the quality of the signs of the residuals for the choice of the method. The criterion is based on taxicab singular value decomposition, on which the package TaxicabCA in R is developed. We present a minimal R script that can be executed to obtain the numerical results and the maps in this paper. Third, we introduce a flexible method based on the novel index for the choice of the constant to be added to contingency tables with zero counts so that logratio methods can be applied.

Keywords: Taxicab singular value decomposition, correspondence analysis, logratio analysis, association models, prespecified weights, quality of the signs of the residuals, zero counts.

1. Introduction

We start by citing Tukey (1977), p.400 : *the general maxim–it is a rare thing that a specific body of data tells us clearly enough how it itself should be analyzed–applies to choice of re-expression for two-way analysis.*

We consider correspondence analysis (CA) and logratio analysis (LRA) as two different popular well-developed choices of re-expression for the analysis and visualization of a contingency table (two-way frequency counts data having I rows and J columns) or a compositional data set (I individuals, also named samples, of J compositional parts).

LRA includes two independently well developed methods: RC association models for the analysis of contingency tables by Goodman (1979, 1981a,b, 1991, 1996) and, compositional data analysis (CoDA) by Aitchison (1986); see also among others, the recently published books by Filzmoser, Hron, and Templ (2018) and Greenacre (2018); and the review articles by Egozcue and Pawlowsky-Glahn (2019) , Greenacre (2020) and Greenacre, Grunsky, Bacon-Shone, Erb, and Quinn (2022). The reference book on CA is Benzécri (1973a); Beh and Lombardo (2014) present a panoramic review of CA and its variants.

CA and LRA are based on three different principles: CA on Benzécri's *distribututional equivalence principle*, RC association models on Yule's *scale invariance principle*, and CoDA on Aitchison's *subcompositional coherence principle*. We shall review these three principles when columns and rows of a table are attributed with prespecified weights. Two popular weights are the marginals of a table or uniform.

The following decomposition scheme describes the general approach of analyzing a two-way table $\mathbf{X}$ by CA or LRA re-expressed as

$$\mathbf{X_{re-expressed}} = (\mathbf{main\ effects}) + (\mathbf{bilinear\ terms}) + \mathbf{noise},$$

where *main effects* represent the scale factor of the rows and the columns of $\mathbf{X}$, the *bilinear terms* represent the shape factors or the latent variables of $\mathbf{X}$, and the last term the noise in the observations. The most important part in the above decomposition that interests us is the *bilinear terms*, which is often computed by singular value decomposition (SVD) of the pre-processed residual data set

$$\mathbf{X_1} = \mathbf{X_{re-expressed}} - \mathbf{main\ effects}.$$

So the choice of the method is crucial: Should we choose CA or LRA? And what are the effects of the weights (marginal or uniform) in LRA? These questions will be tackled within the framework of Taxicab SVD (TSVD), see Choulakian (2006, 2016).

1.1. Motivating example

Consider the contrived data sets of size 3 by 3, $\mathbf{X}_1$ and $\mathbf{X}_2$ reproduced below from Table 9, set 1 of Goodman (1996), where he compared different methods, including CA and LRA. From a mathematical point of view, the most important aspect is that the rank of each data set is 2 or 3 depending on the method used: by the parsimony principle of Occam's razor, of course one should prefer the method that reduces the rank of the data set. For,

$$\mathbf{X}_1 = \begin{pmatrix} 400 & 100 & 100 \\ 100 & 50 & 100 \\ 100 & 100 & 400 \end{pmatrix}$$

we have, rank $\mathbf{X}_1$ =2 by LRA and rank $\mathbf{X}_1$=3 by CA. And for,

$$\mathbf{X}_2 = \begin{pmatrix} 400 & 100 & 100 \\ 100 & 40 & 100 \\ 100 & 100 & 400 \end{pmatrix}$$

we have, rank $\mathbf{X}_2$ =3 by LRA and rank $\mathbf{X}_2$=2 by CA.

It is important to note that the structure of real observed data differs fundamentally from the structure of artificial data. So we need a criterion for the choice between CA or LRA.

1.2. Aims and organization

The essay entitled "On Practice" by Mao (1938) complements Tukey's above maxim, because its subtitle is "*On the Relation Between Knowledge and Practice, Between Knowing and Doing*". For a modern interpretation of Mao's cited work, see Zizek (2008).

Our aim is to relate theory with practice. Given that knowledge-knowing-theory-method start with principles, first we discuss the three basic principles, which are the starting points for each of the three methods: CA, RC association models and CoDA. Then, for data analysis we introduce and apply a simple intuitive criterion for the choice of the preprocessing, and consequently of the method. The novel criterion is the intrinsic variability measure of a principal dimension by the quality of the signs of the residuals (QSR), recently introduced by Choulakian (2021). For each principal dimension QSR is calculated via taxicab singular value decomposition (TSVD) on which the package TaxicabCA in R is developed by Allard and Choulakian (2019). Furthermore, we use the QSR for the choice of the constant to be added to zero cells in contingency tables so that LRA can be applied.

This paper is organized as follows: Section 2 discusses the fundamental differences between contingency tables and compositional data via the role of simplices; section 3 presents an overview of taxicab singular value decomposition (TSVD); section 4 presents the computational steps of the methods TCA (taxicab correspondence analysis) and TLRA (taxicab logratio analysis) for contingency tables, then for compositional data; section 5 reviews the three principles; section 6 introduces the QSR index; section 7 presents examples. Finally we conclude in section 8. The appendix presents a minimal R script to execute the computations in this paper.

2. Simplices in contingency tables and in compositional data sets

From a statistical point of view there is a fundamental difference between the structures of a two-way contingency table $\mathbf{N} = (n_{ij})$ and a compositional data set $\mathbf{X} = (x_{ij})$ for $i = 1, ..., I$ and $j = 1, ..., J$; while from a mathematical point of view the form of the resulting equations arising from different departure assumptions may be identical.

2.1. One simplex in a compositional data set

In a compositional data set, I represents the sample size, that is, the number of individuals observed; where each individual is described by J nonnegative variables (discrete or continuous), named parts measured in the same unit, such as ounce or gram, \$ or e; and the proportion of each part, $y_{ij} = x_{ij}/\sum_j x_{ij}$ is important. For example, suppose we note the expenses of one day stay in Paris and New York, and measure our expenses on three items: one-night stay in Holiday Inn, the price of one cup of coffee at Starbucks and the price of a Big Mac at McDonald's in each of the cities. So we have 2 vectors with 3 parts, where the measurement unit in Paris is in euros and in New York in dollars; x_{ij} in \$ or € are not comparable, but the proportions are comparable via logratios. So geometrically, the i-th row is found in the unit simplex $S^J = \left\{ y_{ij} | 0 \leq y_{ij} \text{ and } \sum_j y_{ij} = 1 \right\}$.

2.2. Three simplices in a contingency table

In a contingency table $\mathbf{N} = (n_{ij})$, $n = \sum_{i,j} n_{ij}$ represents the sample size, that is, the number of individuals observed; where each individual is described by two categorical variables: a row categorical variable having I levels and a column categorical variable having J levels. There are three unit simplices in a contingency table:

$$S^{IJ} = \left\{ n_{ij}/n | 0 \leq n_{ij}/n \text{ and } \sum_{i,j} n_{ij}/n = 1 \right\}$$

representing the joint probability measure on R_+^{IJ};

$$S^J = \left\{ n_{ij}/\sum_j n_{ij} | 0 \leq n_{ij} \text{ and } \sum_j n_{ij}/\sum_j n_{ij} = 1 \right\}$$

representing the conditional probability measure defined on the level i of the row variable;

$$S^I = \left\{ n_{ij} / \sum_i n_{ij} | 0 \leq n_{ij} \text{ and } \sum_i n_{ij} / \sum_i n_{ij} = 1 \right\}$$

representing the conditional probability measure defined on the level j of the column variable. CA acts simultaneously on the two simplices S^J and S^I via Benzécri's chi-squared metrics. While TCA acts simultaneously on the three simplices S^{IJ}, S^J and S^I via TSVD.

A contingency table $\mathbf{N}$ can also be represented (coded) as an indicator matrix $\mathbf{Z} = \left[\mathbf{Z}^I \;\; \mathbf{Z}^J\right] = [(z_{\alpha i}) \;\; (z_{\alpha j})]$ of size n by $(I + J)$, where $z_{\alpha i} = 0$ if individual α does not have level i of the row variable, $z_{\alpha i} = 1$ if individual α has level i of the row variable; $z_{\alpha j} = 0$ if individual α does not have level j of the column variable, $z_{\alpha j} = 1$ if individual α has level j of the column variable. Note that $\mathbf{N} = (\mathbf{Z}^I)'\mathbf{Z}^J$. CA of $\mathbf{Z}$ is mathematically equivalent to CA of $\mathbf{N}$; but the geometric structure of $\mathbf{Z}$ is not clear, see Greenacre and Hastie (1987) . TCA of $\mathbf{Z}$ is not mathematically equivalent to TCA of $\mathbf{N}$, see Choulakian (2008a).

3. An overview of taxicab singular value decomposition

Consider a matrix $\mathbf{X}$ of size $I \times J$ and $rank(\mathbf{X}) = k$. Taxicab singular value decomposition (TSVD) of $\mathbf{X}$ is a decomposition similar to SVD of $\mathbf{X}$, see Choulakian (2006, 2016).

For a vector $\mathbf{u} = (u_i)$, its taxicab or L_1 norm is $||\mathbf{u}||_1 = \sum_i |u_i|$, the Euclidean or L_2 norm is $||\mathbf{u}||_2 = (\sum_i |u_i|^2)^{1/2}$ and the L_∞ norm is $||\mathbf{u}||_\infty = \max_i \; |u_i|$.

In TSVD the calculation of the dispersion measures (δ_α), principal axes $(\mathbf{u}_\alpha, \mathbf{v}_\alpha)$ and principal scores $(\mathbf{a}_\alpha, \mathbf{b}_\alpha)$ for $\alpha = 1, ..., k$ is done in a stepwise manner. We put $\mathbf{X}_1 = \mathbf{X} = (x_{ij})$ and $\mathbf{X}_\alpha$ be the residual matrix at the α-th iteration for $\alpha = 1, ..., k$.

The variational definitions of the TSVD at the α-th iteration are

$$\begin{aligned} \delta_\alpha &= \max_{\mathbf{u} \in \mathbf{R}^J} \frac{||\mathbf{X}_\alpha \mathbf{u}||_1}{||\mathbf{u}||_\infty} = \max_{\mathbf{v} \in \mathbf{R}^I} \frac{||\mathbf{X}'_\alpha \mathbf{v}||_1}{||\mathbf{v}||_\infty} = \max_{\mathbf{u} \in \mathbf{R}^J, \mathbf{v} \in \mathbf{R}^I} \frac{\mathbf{v}' \mathbf{X}_\alpha \mathbf{u}}{||\mathbf{u}||_\infty \, ||\mathbf{v}||_\infty}, \\ &= \max \; ||\mathbf{X}_\alpha \mathbf{u}||_1 \;\; \text{subject to } \mathbf{u} \in \{-1, +1\}^J, \\ &= \max \; ||\mathbf{X}'_\alpha \mathbf{v}||_1 \;\; \text{subject to } \mathbf{v} \in \{-1, +1\}^I, \\ &= \max \mathbf{v}' \mathbf{X}_\alpha \mathbf{u} \;\; \text{subject to } \;\; \mathbf{u} \in \{-1, +1\}^J, \mathbf{v} \in \{-1, +1\}^I. \end{aligned} \tag{1}$$

The α-th principal axes are

$$\mathbf{u}_\alpha \; = \arg \max_{\mathbf{u} \in \{-1, +1\}^J} ||\mathbf{X}_\alpha \mathbf{u}||_1 \quad \text{and} \quad \mathbf{v}_\alpha \; = \arg \max_{\mathbf{v} \in \{-1, +1\}^I} ||\mathbf{X}_\alpha \mathbf{v}||_1 , \tag{2}$$

and the α-th principal projections of the rows and the columns are

$$\mathbf{a}_\alpha = \mathbf{X}_\alpha \mathbf{u}_\alpha \;\; \text{and} \;\; \mathbf{b}_\alpha = \mathbf{X}'_\alpha \mathbf{v}_\alpha. \tag{3}$$

Furthermore, the following relations are also useful

$$\mathbf{u}_\alpha = sign(\mathbf{b}_\alpha) \;\; \text{and} \;\; \mathbf{v}_\alpha = sign(\mathbf{a}_\alpha), \tag{4}$$

where $sign(.)$ is the coordinatewise sign function, $sign(x) = 1$ if $x > 0$, and $sign(x) = -1$ if $x \leq 0$.

The α-th taxicab dispersion measure δ_α can be represented in many different ways

$$\begin{aligned} \delta_\alpha &= ||\mathbf{X}_\alpha \mathbf{u}_\alpha||_1 = ||\mathbf{a}_\alpha||_1 = \mathbf{a}'_\alpha \mathbf{v}_\alpha, \\ &= \left|\left|\mathbf{X}'_\alpha \mathbf{v}_\alpha\right|\right|_1 = ||\mathbf{b}_\alpha||_1 = \mathbf{b}'_\alpha \mathbf{u}_\alpha \\ &= \mathbf{v}_\alpha{}' \mathbf{X}_\alpha \mathbf{u}_\alpha = sign(\mathbf{a}_\alpha)' \mathbf{X}_\alpha sign(\mathbf{b}_\alpha). \end{aligned} \tag{5}$$

The $(\alpha+1)$-th residual matrix is

$$\mathbf{X}_{\alpha+\mathbf{1}} = \mathbf{X}_\alpha - \mathbf{a}_\alpha \mathbf{b}'_\alpha / \delta_\alpha. \tag{6}$$

An interpretation of the term $\mathbf{a}_\alpha \mathbf{b}'_\alpha/\delta_\alpha$ in (6) is that, it represents the best rank-1 approximation of the residual matrix $\mathbf{X}_\alpha$, in the sense of the taxicab matrix norm (1).

Thus TSVD of $\mathbf{X}$ corresponds to the bilinear decomposition

$$x_{ij} = \sum_{\alpha=1}^{k} a_\alpha(i) b_\alpha(j)/\delta_\alpha, \tag{7}$$

a decomposition similar to SVD, but where the vectors $(\mathbf{a}_\alpha, \mathbf{b}_\alpha)$ for $\alpha = 1, ..., k$ are conjugate, that is

$$\begin{aligned} \mathbf{a}'_\alpha \mathbf{v}_\beta &= \mathbf{a}'_\alpha sign(\mathbf{a}_\beta) \\ &= \mathbf{b}'_\alpha \mathbf{u}_\beta = \mathbf{b}'_\alpha sign(\mathbf{b}_\beta) \\ &= 0 \text{ for } \alpha \geq \beta + 1. \end{aligned} \tag{8}$$

In the package TaxicabCA in R, the calculation of the principal component weights, $\mathbf{u}_\alpha$ and $\mathbf{v}_\alpha$, are accomplished by three algorithms. The first one, based on complete enumeration equation (2), is named *exhaustive*. The second one, based on iterating the transition formulae (3,4), is named *criss-cross*. The third one, based on the genetic algorithm, is named *genetic*. The exhaustive algorithm is only feasible when I or J is small (by default, TaxicabCA does not use this algorithm if I and J are > 20). The speed of the other algorithms as a function of table's dimensions I by J has not been studied yet. In the criss-cross approach, the starting values are based on the shorter representation of the singular value decomposition (SVD) values. For the interested reader, here are few published references on the sizes of "relatively" large data sets: A health data set of size 3530 by 88 in Choulakian, Allard, and Simonetti (2013); an ecological abundance data of size 285 by 220 in Choulakian (2017); an indicator data set of size 441 by 105 in Choulakian and de Tibeiro (2013); a survey data set of size 88000 by 120 in Das, Minjares-Kyle, Wu, and Henk (2019) and Das (2021).

4. Analysis of contingency tables and compositional data

First we consider contingency tables, then relate the mathematics to compositional data sets.

Let $\mathbf{P} = \mathbf{N}/n = (p_{ij})$ of size $I \times J$ be the associated correspondence matrix (probability table) of a contingency table $\mathbf{N}$. We define as usual $p_{i+} = \sum_{j=1}^{J} p_{ij}$, $p_{+j} = \sum_{i=1}^{I} p_{ij}$, the vector $\mathbf{r} = (p_{i+}) \in \mathbf{R^I}$, the vector $\mathbf{c} = (p_{+j}) \in \mathbf{R^J}$, and $\mathbf{D}_I = Diag(\mathbf{r})$ the diagonal matrix having diagonal elements p_{i+}, and similarly $\mathbf{D}_J = Diag(\mathbf{c})$. We suppose that $\mathbf{D}_I$ and $\mathbf{D}_J$ are positive definite metric matrices of size $I \times I$ and $J \times J$, respectively; this means that the diagonal elements of $\mathbf{D}_I$ and $\mathbf{D}_J$ are strictly positive.

4.1. Independence of the row and column categories

a) The I row categories are independent of the J column categories, when

$$\sigma_{ij} = p_{ij} - p_{i+}p_{+j} = 0, \tag{9}$$

where σ_{ij} is the residual matrix of p_{ij} with respect to the independence model $p_{i+}p_{+j}$.

Remark 1. *Note that, σ_{ij} can be interpreted as the cross-covariance between the i-th row and the j-th column categories using the indicator matrix $\boldsymbol{Z}$ discussed in section 2.*

b) The independence assumption $\sigma_{ij} = 0$ can also be interpreted in another way as $\mathbf{\Delta} = (\Delta_{ij})$

$$\begin{aligned}\Delta_{ij} &= (\frac{p_{ij}}{p_{i+}p_{+j}} - 1) = 0 \\ &= \frac{1}{p_{i+}}(\frac{p_{ij}}{p_{+j}} - p_{i+}) \\ &= \frac{1}{p_{+j}}(\frac{p_{ij}}{p_{i+}} - p_{+j});\end{aligned} \tag{10}$$

this is the column and row homogeneity models. (Benzécri (1973a), p.31) named the conditional probability vector ($\frac{p_{ij}}{p_{+j}}$ for $i = 1, ..., I$ and j fixed) the profile of the j-th column; and the element $\frac{p_{ij}}{p_{i+}p_{+j}}$ the density function of the probability measure (p_{ij}) with respect to the product measure $p_{i+}p_{+j}$. The element $\frac{p_{ij}}{p_{i+}p_{+j}}$ is named Pearson ratio in Goodman (1996) and (Beh and Lombardo 2014, p.123).

c) A third way to represent the independence assumption (9) and the row and column homogeneity models (10) is via the $(w_i^R,\ w_j^C)$ weighted loglinear formulation, equation (11), assuming $p_{ij} > 0$ and defining $G_{ij} = \log(p_{ij})$,

$$\lambda_{ij} = G_{ij} - G_{i+} - G_{+j} + G_{++} = 0 \tag{11}$$

where
$G_{i+} = \sum_{j=1}^{J} G_{ij}w_j^C$, $G_{+j} = \sum_{i=1}^{I} G_{ij}w_i^R$ and $G_{++} = \sum_{j=1}^{J}\sum_{i=1}^{I} G_{ij}w_j^C w_i^R$; $w_j^C > 0$ and $w_i^R > 0$, satisfying $\sum_{j=1}^{J} w_j^C = \sum_{i=1}^{I} w_i^R = 1$, are a priori fixed or data dependent probability weights. Two popular weights are marginal ($w_j^C = p_{+j}$, $w_i^R = p_{i+}$) and uniform ($w_j^C = 1/J$,$w_i^R = 1/I$). This is implicit in equation 7 in Goodman (1996) or equation 2.2.6 in Goodman (1991); and explicit in Egozcue, Pawlowsky-Glahn, Templ, and Hron (2015).

Equation (11) is equivalent to the logratios

$$\log(\frac{p_{ij}p_{i_1j_1}}{p_{ij_1}p_{i_1j}}) = 0 \quad \text{for } i \neq i_1 \text{ and } j \neq j_1,$$

which (Goodman 1979, equation 2.2) names it "null association" model.

Equation (11) is also equivalent to

$$p_{ij} = \frac{\exp(G_{i+})\exp(G_{+j})}{\exp(G_{++})},$$

from which we deduce that : under the independence assumption the marginal row probability vector (p_{i+}) is proportional to the vector of weighted geometric means $(\exp(G_{i+}))$, and a similar property is true also for the columns; see for instance Egozcue *et al.* (2015).

4.2. Interaction factorization

Suppose that the independence-homogeneity-null association models are not true; then each of the three equivalent model formulations (9,10,11) can be generalized to explain the nonindependence-nonhomogeneity-association, named interaction, among the I rows and the J columns by adding k bilinear terms, where $k = rank(\mathbf{N})-1$. We designate any one of the interaction indices (9,10,11) by τ_{ij}.

(Benzécri 1973a, Vol.1, p. 31-32) emphasized the importance of row and column weights or metrics in multidimensional data analysis; this is the reason in the French data analysis circles any study starts with a triplet $(\mathbf{Y}, \mathbf{M}_I, \mathbf{M}_J)$, where $\mathbf{Y}$ represents the data set, $\mathbf{M}_I = Diag(m_i^r)$ is the diagonal matrix metric defined on the rows and $\mathbf{M}_J = Diag(m_j^c)$ the diagonal matrix metric defined on the columns. We follow the same procedure where:

a) In cross-covariance analysis (Cov), $\mathbf{Y} = (\tau_{ij}) = (\sigma_{ij})$ and $(\mathbf{M}_I, \mathbf{M}_J) = (Diag(1/I), Diag(1/J))$;

b) In CA, $\mathbf{Y} = (\tau_{ij}) = (\Delta_{ij})$ and $(\mathbf{M}_I, \mathbf{M}_J) = (Diag(p_{i+}), Diag(p_{+j}))$;

c) In LRA, $\mathbf{Y} = (\tau_{ij}) = (\lambda_{ij})$ and $(\mathbf{M}_I, \mathbf{M}_J) = (Diag(w_i^R), Diag(w_j^C))$ with $\sum_{j=1}^{J} w_j^C = \sum_{i=1}^{I} w_i^R = 1$.

TCov, TCA and TLRA are factorizations of the interactions in (9,10,11) in three steps:
Step 1: We double-center $\mathbf{Y} = (\tau_{ij})$

$$X(i,j) = \tau_{ij} m_i^r m_j^c \tag{12}$$

That is

$$\begin{aligned} 0 &= \sum_{i=1}^{I} X(i,j) \\ &= \sum_{j=1}^{J} X(i,j). \end{aligned}$$

This double-centering step is necessary to have the important basic equations (17,18) on which the QSR index is based.
Step 2: Calculate TSVD of $\mathbf{X} = (\tau_{ij} m_i^r m_j^c)$ as described in section 3

$$\tau_{ij} m_i^r m_j^c = \sum_{\alpha=1}^{k} a_\alpha(i) b_\alpha(j) / \delta_\alpha. \tag{13}$$

We name $(a_\alpha(i), b_\alpha(j))$ Taxicab contribution scores because they satisfy following (5)

$$\delta_\alpha = \sum_{i=1}^{I} |a_\alpha(i)| = \sum_{j=1}^{J} |b_\alpha(j)| \quad \text{for } \alpha = 1, ..., k. \tag{14}$$

Furthermore, they are centered following Step 1

$$0 = \sum_{i=1}^{I} a_\alpha(i) = \sum_{j=1}^{J} b_\alpha(j)) \quad \text{for} \quad \alpha = 1, ..., k; \tag{15}$$

And they are conjugate (in TSVD conjugacy replaces orthogonality in SVD)

$$0 = \sum_{i=1}^{I} a_\alpha(i)\ sign(a_\beta(i)) = \sum_{j=1}^{J} b_\alpha(j)\ sign(b_\beta(j)) \quad \text{for} \quad \alpha > \beta. \tag{16}$$

Let $S = \{i : a_\alpha(i) \geq 0\}$ and $T = \{j : b_\alpha(j) \geq 0\}$, so that at iteration α, $S \cup \overline{S} = I$ is an optimal partition of I and $T \cup \overline{T} = J$ is an optimal partition of J. Besides (14), the taxicab dispersion δ_α will additionally satisfy the following useful equations:

$$\begin{aligned} \delta_\alpha/2 &= \sum_{i \in S} a_\alpha(i) = -\sum_{i \in \overline{S}} a_\alpha(i) \\ &= \sum_{j \in T} b_\alpha(j) = -\sum_{j \in \overline{T}} b_\alpha(j); \end{aligned} \tag{17}$$

which tells that the Taxicab principal dimensions are *balanced*; and

$$\delta_\alpha/4 = \sum_{(i,j)\in S\times T} X_\alpha(i,j) = \sum_{(i,j)\in \overline{S}\times \overline{T}} X_\alpha(i,j) \tag{18}$$
$$= -\sum_{(i,j)\in \overline{S}\times T} X_\alpha(i,j) = -\sum_{(i,j)\in S\times \overline{T}} X_\alpha(i,j),$$

which tells that the α-th principal dimension divides the residual data matrix $\mathbf{X}_\alpha$ into 4 *balanced* quadrants, see Choulakian and Abou-Samra (2020).

Step 3: Calculate Taxicab principal factor scores $(f_\alpha(i), g_\alpha(j))$ of $\mathbf{X}$ by dividing each term in (13) by the weights $m_i^r m_j^c$

$$\tau_{ij} = \sum_{\alpha=1}^{k} f_\alpha(i) g_\alpha(j)/\delta_\alpha, \tag{19}$$

where evidently $f_\alpha(i) = a_\alpha(i)/m_i^r$ and $g_\alpha(j) = b_\alpha(j)/m_j^c$. Equation (19) is named "data reconstruction formula".

Remark 2. ***1)*** *We have four particular methods:*

a) $\mathbf{Y} = (\tau_{ij}) = (\sigma_{ij})$ *and* $(\mathbf{M}_I, \mathbf{M}_J) = (Diag(1/I), Diag(1/J))$, *we get TCov analysis, known also as interbattery analysis first proposed by Tucker (1958); later on, Tenenhaus and Augendre (1996) reintroduced it within correspondence analysis circles, where they showed that the Tucker decomposition by SVD produced on some correspondence tables more interesting structure, more interpretable, than CA.*

b) $\mathbf{Y} = (\tau_{ij}) = (\Delta_{ij})$ *and* $(\mathbf{M}_I, \mathbf{M}_J) = (Diag(p_{i+}), Diag(p_{+j}))$, *we get TCA (or CA) decomposition. CA has many interpretations; essentially, for data analysis purposes Benzécri (1973a) interpreted it as weighted principal components analysis of row and column profiles. Another useful interpretation, comparable to Tucker interbattery analysis, is Hotelling (1936)'s correlation analysis, see Lancaster (1958) and Goodman (1991, 1996).*

c) $\mathbf{Y} = (\tau_{ij}) = (\lambda_{ij})$ *and* $(\mathbf{M}_I, \mathbf{M}_J) = (Diag(w_i^R), Diag(w_j^C)$, *where* (w_i^R, w_j^C) *are prespecified; we note this case by TLRA. For an example of general prespecified weights see Egozcue and Pawlowsky-Glahn (2016). For the important particular case* $(\mathbf{M}_I, \mathbf{M}_J) = (Diag(1/I), Diag(1/J))$, *we get uniformly-weighted taxicab logratio analysis (uwTLRA).*

d) $\mathbf{Y} = (\tau_{ij}) = (\lambda_{ij})$ *and* $(\mathbf{M}_I, \mathbf{M}_J) = (Diag(p_{i+}), Diag(p_{+j}))$, *where* $(w_i^R, w_j^C) = (p_{i+}, p_{+j})$ *are data-dependent, we get marginally-weighted taxicab logratio analysis (mwTLRA).*

The TCov and uwTLRA principal factor scores are uniformly weighted

$$0 = \sum_{i=1}^{I} \frac{1}{I} f_\alpha(i) = \sum_{j=1}^{J} \frac{1}{J} g_\alpha(j) \quad for \quad \alpha = 1, ..., k,$$

$$\delta_\alpha = \sum_{i=1}^{I} \frac{1}{I} |f_\alpha(i)| = \sum_{j=1}^{J} \frac{1}{J} |g_\alpha(j)| \quad for \quad \alpha = 1, ..., k.$$

TCA and mwTLRA principal factor scores are marginally weighted

$$0 = \sum_{i=1}^{I} p_{i+} f_\alpha(i) = \sum_{j=1}^{J} p_{+j} g_\alpha(j) \quad for \quad \alpha = 1, ..., k,$$

$$\delta_\alpha = \sum_{i=1}^{I} p_{i+} |f_\alpha(i)| = \sum_{j=1}^{J} p_{+j} |g_\alpha(j)| \quad for \quad \alpha = 1, ..., k.$$

What is the consequence to this? The answer is: Benzécri's principle of distributional equivalence, see Definition 4 below, which states that TCA and mwTLRA results are not changed if two proportional columns or rows are merged into one. This has the practical consequence that the effective size of sparse and large data sets can be smaller than the observed size; see Example 2 in section 7. For further details concerning sparse contingency tables see Choulakian (2017).

2) *Aitchison (1983, 1997) and Aitchison and Greenacre (2002) presented principal components analysis of a compositional data set by applying SVD to the preprocessed data of equation (11) using uniform weights. According to Greenacre and Lewi (2009), the marginally-weighted LRA, named spectral mapping, originates by Lewi (1976).*

This shows that, from a mathematical point of view, the form of the resulting equations arising from two different departure assumptions may be identical in Goodman's RC association models for contingency tables and in Aitchison's principal components analysis of a compositional data.

3) *In all methods, the symmetric maps are obtained by plotting* $(f_\alpha(i),\ f_\beta(i))$ *or* $(g_\alpha(j),\ g_\beta(j))$ *for* $\alpha \neq \beta$. *Gower (2011) enumerates nine tools of interpretation of maps-biplots, one of them being the 'nearness' of points. In case of TCA (or CA), the same idea is explicitly expressed by (Benzécri 1966, p. 56) as: "que des éléments distributionnellement proches soient proches sur le diagramme et réciproquement". For connections between TCov and TCA maps, see Choulakian (2021). For TLRA, the 'nearness' of points can be seen also in equation (20) approximating the logodds ratios. Often prior knowledge is also used for interpretation, see Example 1 in section 7.*

5. The three principles

In this section we review and discuss the basic three principles, on which CA and LRA (RC association and CoDA) are erected; but the discussion is within the framework of TSVD if needed. We observe that the aim of each method was different; and only Goodman (1996) attempted to reconcile both CA and LRA methods, because that was his principal aim.

5.1. Yule's principle of scale invariance

We start by quoting (Goodman 1996, section 10) to really understand Yule's principle of scale invariance: "Pearson's approach to the analysis of cross-classified data was based primarily on the bivariate normal. He assumed that the row and column classifications arise from underlying continuous random variables having a bivariate normal distribution, so that the sample contingency table comes from a discretized bivariate normal; and he then was concerned with the estimation of the correlation coefficient for the underlying bivariate normal. On the other hand, Yule felt that, for many kinds of contingency tables, it was not desirable in scientific work to introduce assumptions about an underlying bivariate normal in the analysis of these tables; and for such tables, he used, to a great extent, coefficients based on the odds-ratios (for example, Yule's Q and Y), coefficients that did not require any assumptions about underlying distributions. The Pearson approach and the Yule approach appear to be wholly different, but a kind of reconciliation of the two perspectives was obtained in Goodman (1981a)".

An elementary exposition of these ideas with examples can also be found in Mosteller (1968).

Goodman (1996)'s reconciliation is done in two approaches:

First, Goodman (1996) proposed "marginal-free correspondence analysis", which is discussed in Choulakian (2022).

Second, it is based on defining the a priori weights in the association index (11), $\lambda_{ij} = \lambda(p_{ij}, w_j^C, w_i^R)$, where, by its decomposition into bilinear terms, mwLRA will correspond to

Pearson's approach, while uwLRA to Yule's approach. Because log-odds

$$\log(\frac{p_{ij}p_{i_1j_1}}{p_{ij_1}p_{i_1j}}) = \lambda_{ij} + \lambda_{i_1j_1} - \lambda_{i_1j} - \lambda_{ij_1}$$
$$= \sum_{\alpha=1}^{k}(f_\alpha(i) - f_\alpha(i_1))(g_\alpha(j) - g_\alpha(j_1))/\delta_\alpha, \tag{20}$$

where the principal factor scores satisfy marginally or uniformly weighted relations, see Remark 2 (1c, 1d).

We note that, in a similar spirit see Remark 4, Kazmierczak (1985, 1987) reconciled Yule's scale invariance principle with Benzécri's principle of distributional equivalence, but in a mathematical framework in search of Euclidean metrics that satisfy both principles.

To have a clear picture of LRA with general a priori prescribed weights (w_j^C, w_i^R), we first study the properties of the association index λ_{ij}, that distinguishes it from interaction indices (9,10).

Scale invariance of an interaction index

We are concerned with the property of scale dependence or independence of the three interaction indices (9,10,11). We note that in (9,10,11), p_{ij} depends on n_{ij}, $p_{ij} = n_{ij}/\sum_{i,j} n_{ij}$. To emphasize this dependence, we express an interaction index by $\tau_{ij}(n_{ij}) = \tau(p_{ij}, m_i^R, m_j^C)$ where: in the case of the association index $\tau_{ij}(n_{ij}) = \lambda_{ij}$ is defined in (11), in the case of the nonhomogeneity index $\tau_{ij}(n_{ij}) = \Delta_{ij}$ is defined in (10), and in the case of the nonindependence index $\tau_{ij}(n_{ij}) = \sigma_{ij}$ is defined in (9). Following Yule (1912), we state the following

Definition 1. An interaction index $\tau_{ij}(n_{ij})$ is scale invariant if $\tau_{ij}(n_{ij}) = \tau_{ij}(a_i n_{ij} b_j)$ for scales $a_i > 0$ and $b_j > 0$.

It is important to note that Yule's principle of scale invariance concerns a function of four interaction terms, see equation (20); while in Definition 1 the invariance concerns either each interaction term in a contingency table discussed by Goodman or the relative values of a compositional vector discussed by Aitchison (see subsection 5.3). A similar interpretation is expressed in Kazmierczak (1985).

It is evident that neither the interaction indices (9,10) nor mwLRA are scale invariant: because they are marginal-dependent.

Concerning the association index (11) we have

Lemma 1. *The association index (11) is scale invariant.*

Proof. Let $n^* = \sum_{i,j} a_i n_{ij} b_j$, $w_j^C > 0$ and $w_i^R > 0$, satisfying $\sum_{j=1}^J w_j^C = \sum_{i=1}^I w_i^R = 1$; then

$$\begin{aligned}\tau_{ij}(a_i n_{ij} b_j) &= \lambda(a_i n_{ij} b_j/n^*, w_j^C, w_i^R)\\ &= \log(a_i n_{ij} b_j/n^*) - \sum_{j=1}^{J} w_j^C \log(a_i n_{ij} b_j/n^*)\\ &- \sum_{i=1}^{I} w_i^R \log(a_i n_{ij} b_j/n^*) + \sum_{j=1}^{J}\sum_{i=1}^{I} w_j^C w_i^R \log(a_i n_{ij} b_j/n^*)\\ &= \lambda(n_{ij}, w_j^C, w_i^R) = \lambda(p_{ij}, w_j^C, w_i^R) = \tau_{ij}(n_{ij})\\ &= \lambda(a_i p_{ij} b_j, w_j^C, w_i^R).\end{aligned} \tag{21}$$

□

Lemma 2. *To a first-order approximation,*

$$\lambda_{ij} \approx (\frac{p_{ij}}{w_j^C w_i^R} - \frac{p_{i+}}{w_i^R} - \frac{p_{+j}}{w_j^C} + 1).$$

Proof. The average value of the density function $\frac{p_{ij}}{w_j^C w_i^R}$ with respect to the product measure $w_j^C w_i^R$ is 1; so the IJ values $\frac{p_{ij}}{w_j^C w_i^R}$ are distributed around 1. By Taylor series expansion of $\log x$ in the neighborhood of $x = 1$, we have to a first-order $\log x \approx x - 1$. Putting $a_i = 1/w_i^R$ and $b_j = 1/w_j^C$ in (21), and by using $\log(\frac{p_{ij}}{w_j^C w_i^R}) \approx \frac{p_{ij}}{w_j^C w_i^R} - 1$

$$\begin{aligned}
\lambda(p_{ij}, w_j^C, w_i^R) &= \lambda(\frac{p_{ij}}{w_j^C w_i^R}, w_j^C, w_i^R) \\
&= \log(\frac{p_{ij}}{w_j^C w_i^R}) - \sum_{j=1}^{J} w_j^C \log(\frac{p_{ij}}{w_j^C w_i^R}) \\
&- \sum_{i=1}^{I} w_i^R \log(\frac{p_{ij}}{w_j^C w_i^R}) + \sum_{j=1}^{J}\sum_{i=1}^{I} w_j^C w_i^R \log(\frac{p_{ij}}{w_j^C w_i^R}) \\
&\approx \frac{p_{ij}}{w_j^C w_i^R} - 1 - (\frac{p_{i+}}{w_i^R} - 1) - (\frac{p_{+j}}{w_j^C} - 1) + 0,
\end{aligned}$$

which is the required result. □

Remark 3. *Lemma 2 provides a first order approximation to mwTLRA and uwTLRA, where we see that both first-order approximations are marginal-dependent but in different ways.*

a) In the case $(a_i, b_j) = (1/p_{i+}, 1/p_{+j})$ *and* $(w_j^C, w_i^R) = (p_{+j}, p_{i+})$ *in Lemma 2,* $\lambda_{ij} = \lambda(p_{ij}, p_{+j}, p_{i+}) \approx \frac{p_{ij}}{p_{+j}p_{i+}} - 1$; *which implies that CA (or TCA) is a first-order approximation of mwLRA (or mwTLRA), a result stated in Cuadras, Cuadras, and Greenacre (2006). Note that* $\lambda(p_{ij}, p_{+j}, p_{i+}) = \lambda(\frac{p_{ij}}{p_{+j}p_{i+}}, p_{+j}, p_{i+})$ *is stated in Goodman (1996). Furthermore mwLRA (or mwTLRA) is interpreted as prespecified TLRA by defining* $n_{ij}^* = a_i n_{ij} b_j$ *in Lemma 1 with prespecified weights* $(w_j^C, w_i^R) = (p_{+j}, p_{i+})$, *where* $p_{+j} = \sum_i n_{ij}/n$, $p_{i+} = \sum_j n_{ij}/n$ *and* $n = \sum_{i,j} n_{ij}$.

b) In the case $(a_i, b_j) = (I, J)$ *and* $(w_j^C, w_i^R) = (1/I, 1/J)$ *in Lemma 2,* $\lambda_{ij} = \lambda(p_{ij}, 1/J, 1/I) \approx IJp_{ij} - Ip_{i+} - Jp_{+j} + 1$; *which implies that the bilinear expansion of the right side by TSVD (or SVD) is a first-order approximation of uwTLRA (or uwLRA); a familiar one known as a FANOVA (factor analysis and analysis of variance), see Mandel (1971).*

5.2. Benzécri's principle of distributional equivalence

(Benzécri 1966, p. 56) presented his project of uncovering grammatical rules or patterns from texts by announcing his conceptual formulation of the principle of distributional equivalence as : "que des éléments distributionnellement proches soient proches sur le diagramme et réciproquement"; that is, "points that are near in distribution should appear near on the maps and reciprocally". And the notion of nearness of two distributions was introduced by the chi-squared distance (a weighted euclidean distance) in (Benzécri 1973a, p. 150-152). In the statistical literature, the concept of the principle of distributional equivalence is discussed in terms of distances, see for example Escofier (1978), Kazmierczak (1985, 1987), Fichet (2009) and Greenacre and Lewi (2009); except in Nishisato (1984) who named it principle of equivalent partitioning, and Choulakian (2006) who formulated it as invariance of TSVD results, see Theorem 1 below.

In the sequel, we separate Benzécri's idea of distributional equivalence property from its principle, by observing that (Aitchison 1994, section 1)'s compositional equivalence is identical to Benzécri (1966)'s distributional equivalence.

Definition 2. Two non negative vectors $\mathbf{x}$ and $\mathbf{y}$ of the same size are distributionally/compositionally equivalent if $\mathbf{x} = C\mathbf{y}$, where C is a positive constant.

Definition 3. A method of analysis satisfies distributional/compositional equivalence property, if it is applied to a table of nonnegative values where two rows (or two columns) are proportional, then the method will produce maps where the two proportional rows (or the two proportional columns) coincide.

Remark 4. *a) Definition 3 is implicitly stated by Benzécri (1966, page 56) as: que des éléments distributionnellement proches soient proches sur le diagramme et réciproquement.*

b) Definition 3 is a nonmetric variant of Kazmierczak (1985, 1987)'s definition of "generalized distributional equivalence property", which states: Let $\mathbf{x}$ *and* $\mathbf{y}$ *be two proportional rows (or columns) in a nonnegative table of size* I *by* J*; if we replace* $\mathbf{x}$ *and* $\mathbf{y}$ *by two proportional vectors* $\mathbf{x}_1$ *and* $\mathbf{y}_1$ *such that* $\mathbf{x}+\mathbf{y} = \mathbf{x}_1+\mathbf{y}_1$*, then the distances between the columns (or rows) are not changed. Proposition 1 is discussed by Kazmierczak (1985, 1987) within the framework of Euclidean and Riemannian geometries. The proof within Taxicab framework is very simple.*

Proposition 1. *TCA, mwTLRA and TLRA satisfy the distributional/compositional equivalence property. That is, if the 1st row of a non negative table is proportional to the 2nd row, and* $\lambda_{ij} = \sum_{\alpha=1}^{k} f_\alpha(i)g_\alpha(j)/\delta_\alpha$*, then* $f_\alpha(1) = f_\alpha(2)$*.*
The proof is found in Appendix 2.

Remark 5. *The TCov method does not satisfy the distributional/compositional equivalence property in Definition 3.*

Definition 4. A method of analysis satisfies Benzécri's principle of distributional equivalence, when results are not changed if two proportional columns (or rows) are merged into one.

Benzécri's principle in Definition 4 is very basic and widely used in elementary statistics. For instance: consider a batch of 4 numbers x_i associated with its prespecified weights $w_i : (2, 1/4), (3, 1/4), (7, 1/4)$ and $(2, 1/4)$; then the batch is equivalent to the merged batch $(2, 2/4), (3, 1/4), (7, 1/4)$.

Theorem 1. *Let* $\mathbf{X} = (x_{ij})$ *for* $i = 1, 2, ..., I$ *and* $j = 1, ..., J$ *be a contingency table or a compositional data set. Suppose the first two rows of* $\boldsymbol{X}$ *are proportional,* $x_{1j} = Cx_{2j}$ *for* $j = 1, ..., J$*, where* C *is a strictly positive constant. TLRA of* $\boldsymbol{X}$ *with a priori weights* (w_i^R, w_j^C) *is equivalent to TLRA of* $X^{merge} = (x_{ij}^m)$ *for* $i = (1 + 2), 3, ..., I$ *and* $j = 1, ..., J$ *with weights* (w_i^{Rm}, w_j^C)*, where* $x^m_{(1+2)j} = x_{1j} + x_{2j}$ *and* $w^{Rm}_{1+2} = w_1^R + w_2^R$*, and* $x_{ij}^m = x_{ij}$ *and* $w_i^{Rm} = w_i^R$ *for* $i = 3, ..., I$*.*

Theorem 1 means the following:

Suppose the first two rows of the correspondence table $(p_{ij} = x_{ij}/\sum_{i,j} x_{ij})$ are proportional, $p_{1j} = Cp_{2j}$ for $j = 1, ..., J$, where C is a strictly positive constant; and (p_{ij}^{merge}) is the merged correspondence table of size $(I - 1)$ by J, such that

$$p^{merge}_{(1+2)j} = p_{1j} + p_{2j} \quad \text{with weights } (w^{Rm}_{(1+2)j}, w_j^C)$$
$$p_{ij}^{merge} = p_{ij} \quad \text{for} \quad i = 3, ..., I \text{ with weights } (w_i^R, w_j^C)$$

and consider the bilinear decompositions

$$\lambda_{ij} = \sum_{\alpha=1}^{k} f_\alpha(i)g_\alpha(j)/\delta_\alpha \qquad \text{for} \quad i = 1, ..., I \quad \text{and} \quad j = 1, ..., J$$

$$\lambda_{ij}^{merge} = \sum_{\alpha=1}^{k} f_\alpha^m(i) g_\alpha^m(j)/\delta_\alpha^m \quad \text{for} \quad i = (1+2), ..., I \quad \text{and} \quad j = 1, ..., J.$$

Then

$$\begin{aligned} f_\alpha^m(1+2) &= f_\alpha(1) = f_\alpha(2) \\ f_\alpha^m(i) &= f_\alpha(i) \text{ for } i = 3 \cdots I \\ g_\alpha^m(j) &= g_\alpha(j) \\ \delta_\alpha^m &= \delta_\alpha. \end{aligned}$$

The proof is found in the Appendix.

Corollary 1. *Marginal-dependent TLRA (mwTLRA) satisfies Benzécri's principle of distributional equivalence.*

Corollary 2. *TLRA with prespecified weights, and in particular uwTLRA, does not satisfy Benzécri's principle of distributional equivalence.*

5.3. Aitchison's principle of subcompositional coherence

In a paper entitled *Principles of compositional data analysis*, Aitchison (1994) summarized his main ideas.

The first is the compositional /distributional equivalence of two proportional compositional vectors that we stated in Definition 2; which is quantified as a point in unit simplex, see subsection 2.1.

The second is scale invariance of units for each compositional vector of J parts as we provided an example in section 2.

The third, which seems to be the most important, is the principle of subcompositional coherence. We think the scale invariance of the odds ratios in Yule's principle, is intimately related to the principle of subcompositional coherence in the following way. Let $\mathbf{N}$ be a count data set which can be interpreted both as a contingency table (the row variable has I levels and the column variable has J columns) and a compositional data set (I individuals, also named samples, of J compositional parts); see for instance the *rodent* data set in Example 2. As a contingency table, we consider the scale invariant odds ratios of the rows (i, i_1) and the columns (j, j_1)

$$\frac{p_{ij}p_{i_1j_1}}{p_{ij_1}p_{i_1j}} = \frac{p_{ij}/p_{ij_1}}{p_{i_1j}/p_{i_1j_1}}, \tag{22}$$

which compares the ratio of two odds: the odd of (j, j_1) given level i and the odd of (j, j_1) given level i_1. Now suppose that the data set is compositional, then the odds ratio (22) will compare the relative variation of the odds (j, j_1) of two individuals (i, i_1); for an individual i the scale independent representation of a compositional vector will be the odds p_{ij}/p_{ij_1}. So the set of distinct but redundant odds of size $J(J-1)/2$, will be a scale invariant representation of the compositional vector. That is the relative values of the components are the basic building blocks in a composition; in addition these relative values of components are unchanged for a subset of parts. This evident and important observation by Aitchison is named principle of subcompositional coherence; in fact it can be considered a corollary to Yule's odds ratio.

Another consequence in considering the ratio p_{ij}/p_{ij_1} is, by fixing , without loss of generality, $j_1 = J$, one gets $(J-1)$ distinct but basic odds p_{ij}/p_{iJ} of size $(J-1)$, named (alr) transformation. The clr transformation is $p_{ij}/\exp(G_{i+})$ of size J, where $\exp(G_{i+})$ is the geometric mean defined in subsection 4.1.

Another important contribution of CoDA is Aitchison's geometry, which is the use of Euclidean geometry restricted in the simplex. This has the practical-useful consequence that a lot of classical multivariate statistical theory can easily be applied in CoDA. But other

geometries in the simplex have been studied, see for instance Nielsen and Sun (2017), which to our knowledge have not been applied in CoDA; TSVD falls in this category.

We emphasize the fact that only the principal components analysis of the covariances of the clr tansformed compositional data is mathematically identical to Goodman's RC association model.

Isometric logratio transformation

Another active topic in CoDA is isometric logratio (ilr) representation, introduced by Egozcue and Pawlowsky-Glahn (2005), and further developed by Fiserova and Hron (2011) and Hron, Filzmoser, and De Caritat (2017); it is developed within the Euclidean geometry. Isometry is the study of group transformations that keeps distances invariant between any two points in a metric space. Given that each metric space can be characterized by its algebraic group isometries, isometries in the Euclidean space are characterized by the construction of an orthonormal basis, which can be topic based or data based. In CoDA, ilr transformation is implicitly considered topic based, similar to the use of discrete wavelet transform used in image compression, see Strang (2019). Egozcue and Pawlowsky-Glahn (2005, 2019) presented *topic based* "easily interpretable" ilr transformation named sequential binary partition (SBP) of the J parts (almost identical to the discrete Haar wavelet transform); the topic based SBP can be replaced by or compared with a *data based* SBP by TSVD.

A particular kind of ilr is the pivot logratios (plr), where the orthonormal basis has a triangular shape apart from the first row. For example for $J = 5$ parts, the orthonormal basis is

$$\begin{bmatrix} 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} \\ 4/\sqrt{20} & -1/\sqrt{20} & -1/\sqrt{20} & -1/\sqrt{20} & -1/\sqrt{20} \\ 0 & 3/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} \\ 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

This has very simple interpretation displaying the importance of the log parts in the following sense. We designate the 5 columns by the 5 ordered log parts $x_{(i)}$: $x_{(1)} \succ x_{(2)} \succ x_{(3)} \succ x_{(4)} \succ x_{(5)}$ according to their importance. Then, the second row shows that the first log part, $x_{(1)}$, being the most important opposes to the rest and it is eliminated; the third row shows that the second most important log part $x_{(2)}$ opposes to $\{x_{(3)}, x_{(4)}, x_{(5)}\}$ and it is eliminated; and so on. The underlying idea is that each row starting with the second is made up by a heavyweight log-part. In CA heavyweights are discussed by Benzécri (1979) and Lebart (1979), in TCA by Choulakian (2008b).

Finally, we mention Greenacre (2020)'s amalgamated or summated logratios (slr).

6. Quantifying the intrinsic quality of a taxicab principal axis

We briefly review the quality of measures of a principal dimension in the Euclidean framework, then within the Taxicab framework. We think this comparison will be insightful, see Choulakian (2021).

6.1. Euclidean framework

Within the Euclidean framework a common used measure of the quality of a principal dimension α of the residual matrix $\mathbf{X}_1$ described in (13), is the proportion of variance explained (or

inertia in the case of CA)

$$\begin{aligned}\tau_1(\alpha) &= \textit{proportion}(\text{explained variance of dimension } \alpha)\\ &= \frac{\sigma_\alpha^2}{\sum_{\beta=1}^{k}\sigma_\beta^2} \quad \text{for} \quad \alpha = 1,...,k\\ &= \frac{\sigma_\alpha^2}{\sum_{(i,j)}|X_1(i,j)|^2}.\end{aligned}$$

Another variant is

$$\begin{aligned}\tau_2(\alpha) &= \textit{proportion}(\text{explained residual variance of dimension } \alpha)\\ &= \frac{\sigma_\alpha^2}{\sum_{\beta=\alpha}^{k}\sigma_\beta^2} \quad \text{for} \quad \alpha = 1,...,k\\ &= \frac{\sigma_\alpha^2}{\sum_{(i,j)}|X_\alpha(i,j)|^2}.\end{aligned}$$

Note that $\tau_1(\alpha)$ and $\tau_2(\alpha)$ are extrinsic measures of quality of the residuals in the residual matrix $\mathbf{X}_\alpha$, because they compare the intrinsic dispersion of a principal axis σ_α^2 to the total dispersion $\sum_{\alpha=1}^{k}\sigma_\alpha^2$ or to the partial residual dispersion $\sum_{\beta=\alpha}^{k}\sigma_\beta^2$. Furthermore we have the following evident result lemma 3 that should be compared with Lemma 4. In Lemma 3a (upper bound of 1 is not attained) while in Lemma 4a (upper bound of 1 is attained).

Lemma 3.

a) $1 > \tau_i(\alpha)$ *for* $i = 1, 2$ *and for* $\alpha = 1, ..., k-1$.

b) For $\alpha = k$, $1 = \tau_2(\alpha)$.

6.2. Taxicab framework

The Taxicab variant of τ_2 is particularly adapted in TSVD

$$QSR_\alpha = \frac{\delta_\alpha}{\sum_{(i,j)}|X_\alpha(i,j)|},$$

which we will interpret as a new intrinsic measure of quality of the signs of the residuals in the residual matrix $\mathbf{X}_\alpha$ for $\alpha = 1, ..., k$.

Let $S \cup \overline{S} = I_1$ be the optimal principal axis partition of I_1, and similarly $T \cup \overline{T} = J_1$ be the optimal principal axis partition of J_1, such that $S = \{i : a_\alpha(i) > 0\} = \{i : v_\alpha(i) > 0\}$ and $T = \{j : b_\alpha(j) > 0\} = \{j : u_\alpha(j) > 0\}$ by (4). Thus the data set is divided into 4 quadrants. Based on the equations (18), we define a new index quantifying the quality of the signs of the residuals in each quadrant of the αth residual matrix $\mathbf{X}_\alpha$ for $\alpha = 1, ..., k$.

Definition 5. For $\alpha = 1, ..., k$, the measure of the quality of the signs of the residuals in the quadrant $E \times F \subseteq I_1 \times J_1$ is

$$\begin{aligned}QSR_\alpha(E,F) &= \frac{\sum_{(i,j)\in E\times F} p_{(i,j)}^{(\alpha)}}{\sum_{(i,j)\in E\times F}|p_{(i,j)}^{(\alpha)}|} \text{ and by (21)}\\ &= \frac{\delta_\alpha/4}{\sum_{(i,j)\in E\times F}|p_{(i,j)}^{(\alpha)}|} \text{ for } (E,F) = (S,T) \text{ or } (\bar{S},\bar{T})\\ &= \frac{-\delta_\alpha/4}{\sum_{(i,j)\in E\times F}|p_{(i,j)}^{(\alpha)}|} \text{ for } (E,F) = (\bar{S},T) \text{ or } (S,\bar{T})\end{aligned}$$

for $E = S$ and $\overline{S}$, and, $F = T$ and $\overline{T}$. Sometimes we express it also in %.

We have the following easily proved

Lemma 4.

a) $-1 \leq QSR_\alpha(E,F) \leq 1$*; furthermore,* $QSR_\alpha = 1$ *if and only if* $QSR_\alpha(S,T) = QSR_\alpha(\overline{S},\overline{T}) = -QSR_\alpha(S,\overline{T}) = -QSR_\alpha(\overline{S},T) = 1$ *for* $\alpha = 1,...,k-1$.

b) For $\alpha = k$, $QSR_\alpha = 1$.

The interpretation of $QSR_\alpha(E,F) = \pm 1$ is that in the quadrant $E \times F$ the residuals have one sign; and this is a signal for very influential cells or columns or rows; for an example see Choulakian (2021). So Lemma 4 provides a necessary and sufficient condition for $QSR_\alpha = 1$, which is not true for $\tau_1(\alpha)$ and $\tau_2(\alpha)$. Geometry plays its unique role.

Notation: $QSR_\alpha(+) = \left\{QSR_\alpha(S,T), QSR_\alpha(\overline{S},\overline{T})\right\}$ and $QSR_\alpha(-) = \left\{QSR_\alpha(S,\overline{T}), QSR_\alpha(\overline{S},T)\right\}$.

Remark 6. *The computation of the elements of* $QSR_\alpha(+)$ *and* $QSR_\alpha(-)$ *are done easily in the following way. We note that the* α*-th principal axis can be written as*

$$\mathbf{u}_\alpha = \mathbf{u}_{\alpha+} + \mathbf{u}_{\alpha-},$$

where $\mathbf{u}_{\alpha+} = (\mathbf{u}_\alpha + \mathbf{1}_J)/2$ *and* $\mathbf{u}_{\alpha-} = (\mathbf{u}_\alpha - \mathbf{1}_J)/2$; *similarly*

$$\mathbf{v}_\alpha = \mathbf{v}_{\alpha+} + \mathbf{v}_{\alpha-},$$

where $\mathbf{v}_{\alpha+} = (\mathbf{v}_\alpha + \mathbf{1}_I)/2$ *and* $\mathbf{v}_{\alpha-} = (\mathbf{v}_\alpha - \mathbf{1}_I)/2$, *where* $\mathbf{1}_I$ *designates a column vector of 1's of size* I. *So*

$$QSR_\alpha(S,T) = \frac{\delta_\alpha/4}{\mathbf{v}'_{\alpha+} abs(\mathbf{X}_\alpha)\mathbf{u}_{\alpha+}},$$

$$QSR_\alpha(\overline{S},\overline{T}) = \frac{\delta_\alpha/4}{\mathbf{v}'_{\alpha-} abs(\mathbf{X}_\alpha)\mathbf{u}_{\alpha-}},$$

$$QSR_\alpha(S,\overline{T}) = \frac{\delta_\alpha/4}{\mathbf{v}'_{\alpha-} abs(\mathbf{X}_\alpha)\mathbf{u}_{\alpha+}},$$

$$QSR_\alpha(\overline{S},T) = \frac{\delta_\alpha/4}{\mathbf{v}'_{\alpha+} abs(\mathbf{X}_\alpha)\mathbf{u}_{\alpha-}},$$

where $abs(\mathbf{X}_\alpha) = (|X_\alpha(i,j)|)$.

6.3. Zero counts

In CoDA and RC association models, data with zero values are changed into positive values, so that the log transformation can be applied. Lubbe, Filzmoser, and Templ (2021) compare different strategies proposed in the literature. Researchers often distinguish compositional data that are counts (discrete) from continuous; an interesting discussion on counts compositional data, named lattice compositions, is discussed in Lovell, Chua, and McGrath (2020). Here we present a flexible method that includes three different but related Bayesian estimators of relative frequencies in the presence of zero counts in contingency tables.

First, for a contingency table with zero counts (n_{ij}), Egozcue *et al.* (2015) consider the modified table $(n_{ij} + 1/(IJ))$ named Perks point estimator of the cell counts. This is a particular instance of a Bayesian point estimator of the relative frequency of a cell with Dirichlet prior having the general form of $(n_{ij} + \omega_j)/(n + t)$, see (Bernard 2005, equation 4).

Second, another solution is to consider the modified table $(n_{ij}+1)$ based on arguments in (Benzécri 1973b, p. 218) and Emerson and Stoto (1983), that we describe. We consider the power family of transformations

$$\begin{aligned} T_\alpha(n_{ij}) &= n_{ij}^\alpha \quad \text{for } \alpha > 0 \\ &= \log_b(n_{ij}+c) \quad \text{for } \alpha = 0. \end{aligned}$$

Find the values of the constants b and c such that the family of curves $T_\alpha(n_{ij})$ passes through two common points $T_\alpha(n_{ij}=0)=0$ and $T_\alpha(n_{ij}=1)=1$ for $\alpha \geq 0$. The solution is $b=2$ and $c=1$, and we call it the BES approach; which is equivalent to the Bayes-Laplace estimator, see (Bernard 2005, subsection 3.2).

A third estimator mentioned by (Bernard 2005, subsection 3.2) is Jeffreys estimator when $c=1/2$.

How can we decide the best, if it exists, among these three estimators? A flexible approach that includes the three estimators is to consider the transformation

$$T(c) = \log(n_{ij}+c) \quad \text{for } c = 1, 1/2, 1/(IJ), \tag{23}$$

and find the value of c that maximizes QSR_1 or $QSR_1 + QSR_2$ measures.

We note that for lattice compositions, it is sufficient to replace IJ by J above.

Remark 7. *Another interesting observation is: Suppose* $\mathbf{Z} = (z_{ij})$ *is an incidence matrix, a presence-absence data set, where* $z_{ij} = 0$ *means level* j *is absent in the* i*-th individual,* $z_{ij} = 1$ *means level* j *is present in the* i*-th individual. CA (or TCA) is a popular method for the analysis of such tables, see for an example Choulakian and Abou-Samra (2020). Putting* $p_{ij} = z_{ij}/\sum_{i,j} z_{ij}$ *and supposing that the marginals* $p_{+j} > 0$ *and* $p_{i+} > 0$*, CA (or TCA) data reconstruction formula is*

$$p_{ij} = p_{+j}p_{i+}(1 + \sum_{\alpha=1}^{k} f_\alpha(i)g_\alpha(j)/\delta_\alpha).$$

Now suppose we apply uwLRA (or uwTLRA); observing $\log_2(z_{ij}+1) = z_{ij}$*, then the data reconstruction formula is*

$$p_{ij} = p_{+j}/I + p_{i+}/J - 1/(IJ) + \sum_{\alpha=1}^{k} f_\alpha(i)g_\alpha(j)/\delta_\alpha,$$

which is the first-order approximation of uwTLRA (or uwLRA) found in Remark 3b: a familiar one known as a FANOVA (factor analysis and analysis of variance), see Mandel (1971).

Remark 8. *Here, we summarize how to analyze a contingency table or a lattice distribution by CA, mwTLRA and uwTLRA: we can have 4 cases.*

Case 1a) There are not zero counts and there are not 2 proportional rows or columns: Then we can analyze $\boldsymbol{N}$ *by CA, mwTLRA or uwTLRA.*

Case 1b) There are not zero counts but there are at least 2 proportional rows or columns: Then we can analyze $\boldsymbol{N}$ *or* $\boldsymbol{N}_{merged}$ *by CA or mwTLRA; and only* $\boldsymbol{N}$ *by uwTLRA.*

Case 2a) There are zero counts and there are not 2 proportional rows or columns: Then we can analyze $\boldsymbol{N}$ *by CA,* $(\mathbf{N} + c\mathbf{1}_I\mathbf{1}'_J)$ *by mwTLRA or uwTLRA.*

Case 2b) There are zero counts but there are at least 2 proportional rows or columns: Then we can analyze $\boldsymbol{N}$ *or* $\boldsymbol{N}_{merged}$ *by CA,* $(\boldsymbol{N}_{merged} + c\mathbf{1}_I\mathbf{1}'_J)$ *by mwTLRA; and* $(\mathbf{N}_{modified} + c\mathbf{1}_I\mathbf{1}'_J)$ *by uwTLRA: where* $\mathbf{N}_{modified} = (n^*_{ij})$ *is computed in two steps to satisfy Lemma 1 and Proposition 2. In Step 1,* $n^1_{ij} = n_{ij}/\gcd_i$*, where* $\gcd_i$ *is the greatest common divisor of the strictly positive counts of the* i*-th row* $(n_{ij} > 0$ *for* $j = 1, ..., J$ *and* i *fixed); in Step 2,* $n^*_{ij} = n^1_{ij}/\gcd(n^1_{ij} > 0$

for $i = 1, ..., I$ and j fixed). Step 1 is needed in case there are at least 2 proportional rows; Step 2 is needed in case there are at least 2 proportional columns. Rodent data set in Example 2 is such an example where there are groups of 7 proportional rows, so Step 2 is not needed, so $n^*_{ij} = n^1_{ij}$.

7. Examples

Here we consider the analysis of two tables.

7.1. Food compositional data

The food compositional data set, displayed in Appendix 3, is of size 25 by 9 and analyzed quite in Smail detail by uwLRA in Pawlowsky-Glahn and Egozcue (2011). These data are percentages of consumption of 9 different kinds of food in 25 countries in Europe in the early eighties. The 9 different kinds of food are: red meat (RM); white meat (WM); fish (F); eggs (E); milk (M); cereals (C); starch (S); nuts (N); fruit and vegetables (FV). The 25 countries are divided into 16 western (w) and 9 eastern (e) countries. It is evident that in Table 1, TCA $QSR_1 = 77.89\%$ and mwTLRA $QSR_1 = 80.1\%$ values are significantly higher than the uwTLRA $QSR_1 = 68.69\%$, so we choose either TCA or mwTLRA. On the maps displayed in Figures 1, 2 and 3, the 9 food kinds are represented by their symbols and the 25 countries by their symbols eastern (e) or western (w). Applying the R code in Appendices 2 and 3, one sees that mwTLRA and TCA maps are very similar and they discriminate much better the eastern and the western countries than the uwTLRA map: All eastern countries are clustered in the third quadrant, except one located in the first quadrant. We also note that uwTLRA map is very similar to LRA map in Pawlowsky-Glahn and Egozcue (2011).

In Table 1, we also presented the first three taxicab dispersion measures, δ_α for $\alpha = 1, 2, 3$. δ_1 is of the same order for the three methods; however QSR_1 values are different, which shows that QSR_α are not related to δ_α in general.

Table 1: QSR in (%) of Food data for the first 3 dimensions

		TCA		
α	$QSR_\alpha(+)$	$QSR_\alpha(-)$	QSR_α	δ_α
1	**(86.58, 71.16)**	**(-96.04, -65.21)**	**77.89**	0.2524
2	(56.01, **61.84)**	**(-64.99, -46.48)**	**56.40**	0.1041
3	(83.11, 41.49)	(-68.57, -54.06)	57.79	0.0848
		mwTLRA		
α	$QSR_\alpha(+)$	$QSR_\alpha(-)$	QSR_α	δ_α
1	**(90.19, 78.02)**	**(-94.91, -64.55)**	**80.11**	0.2773
2	(54.77, **67.23)**	**(-66.49, -43.51)**	**56.21**	0.1086
3	(81.41, 48.68)	(-69.18, -55.59)	61.29	0.0939
		uwTLRA		
α	$QSR_\alpha(+)$	$QSR_\alpha(-)$	QSR_α	δ_α
1	**(87.43,63.19)**	**(-89.99,-50.36)**	**68.69**	0.2755
2	(47.57,**68.06)**	**(-62.51,-47.07)**	**54.83**	0.1541
3	(71.49,61.51)	(-62.47,-62.94)	64.37	0.1443

7.2. Rodent abundance data

We consider the *rodent* data set of size 28 by 9 found in TaxicabCA in R package. This is an abundance data set of 9 kinds of rats in 28 cities in California. Choulakian (2017) analyzed it by comparing the CA and TCA maps; Choulakian (2021) showed that it has quasi-2-blocks

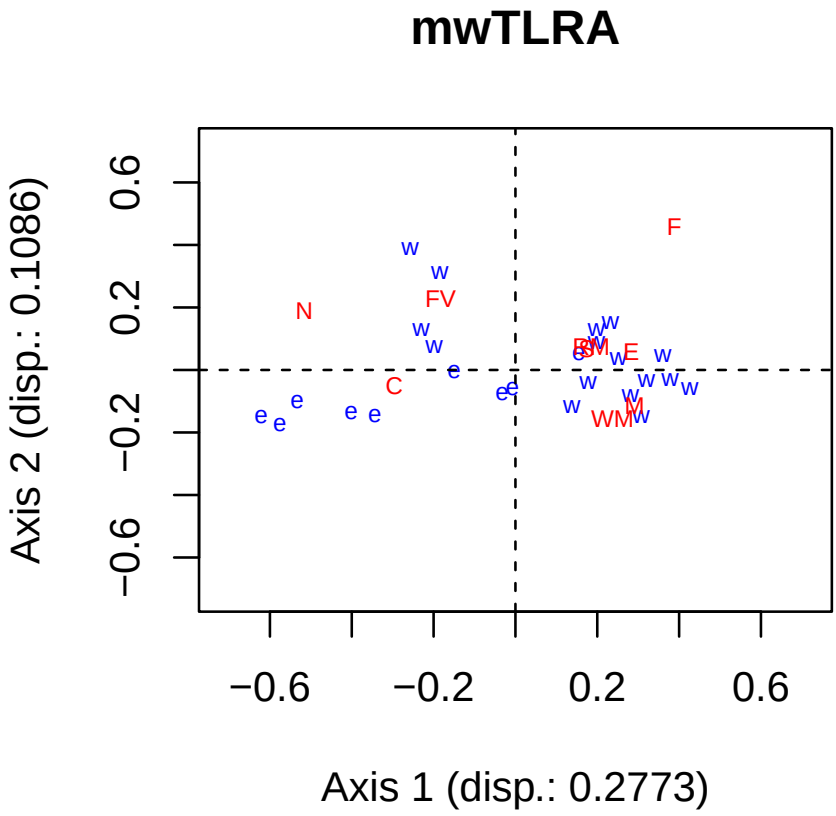

Figure 1: mwTLRA plot of Food Compositional data

diagonal structure; furthermore Choulakian (2022) analyzed it by Goodman's marginal-free CA and marginal-free TCA methods.

Let $\mathbf{N}$ be the original data set of size 28 by 9; the following function in R calculates the proportion of zero counts in $\mathbf{N}$

- sum(N == 0)/length(N)

  [1] 0.6626984

  The function "*CombineCollinearRowsCols*" in the package TaxicabCA in R merges the rows and the columns of $\mathbf{N}$, which are proportional; we see that the size of the Nmerged is 21 by 9.

- Nmerged ← CombineCollinearRowsCols(N, rows = T, cols = T)
- dim(Nmerged)

  [1] 21 9

Here, we are in the Case 2b of Remark 8: We can analyze $\mathbf{N}$ or $\mathbf{N}_{merged}$ by CA, $(\mathbf{N}_{merged} + c\mathbf{1}_I\mathbf{1}'_J)$ by mwTLRA; and $(\mathbf{N}_{\text{modified}} + c\mathbf{1}_I\mathbf{1}'_J)$ by uwTLRA. $\mathbf{N}_{\text{modified}}$ can be calculated using the package "numbers" by Borchers (2021) in the following way:

install.packages("numbers")

library(numbers)

Nmodified ← N

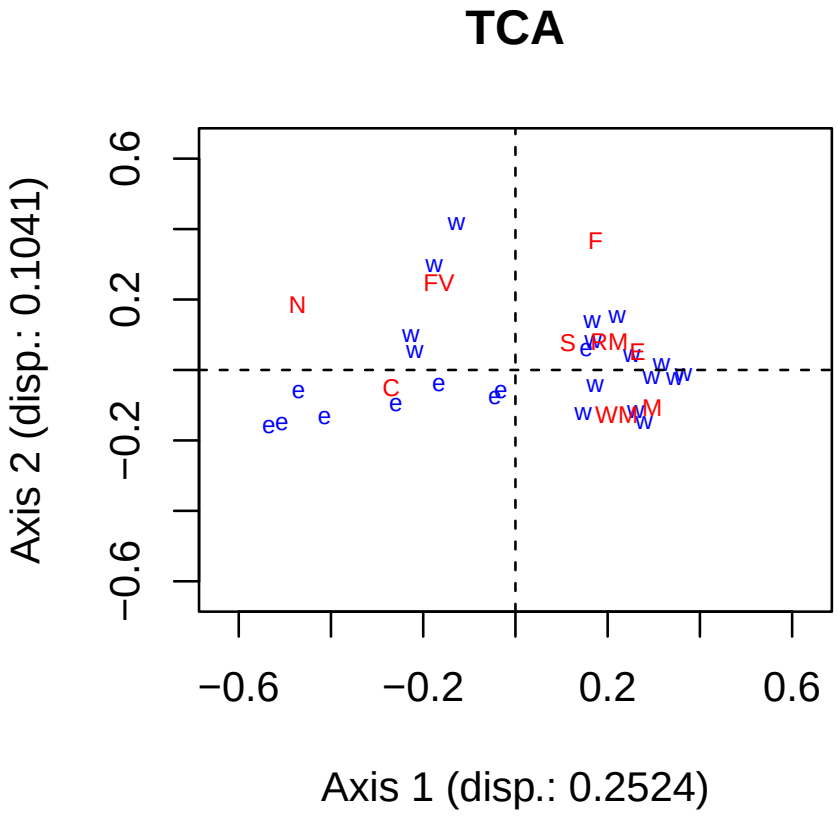

Figure 2: TCA plot of Food Composition data

```
nRow ← nrow(N)
nCol ←ncol(N)
   for (ii in 1:nRow) {
    vpositive ← filter(function(i) i > 0, N[ii,1:nCol])
    numb ← length(vpositive)
      if(numb < 2){gcd=vpositive}
     else {gcd=mGCD(vpositive)}
    Nmodified[ii,1:nCol] ← N[ii,1:nCol]/gcd
}
```

Table 2 displays the numerical values of QSR_α and δ_α for $\alpha = 1, 2$ for TCA, uwTLRA and mwTLRA methods. For the last two methods, the optimal choice of the constant is $c = 1$. We summarize the numerical results in Table 1:

First, it makes sense, that the dispersion values of the uwTLRA and mwTLRA methods using BES Bayes-Laplace estimator, $c = 1$, are the smallest among the three methods BES, Jeffreys and Perks.

Second, for the choice among the three methods TCA, uwTLRA or mwTLRA, QSR_1 is not informative, while QSR_2 is informative: The choice clearly is TCA or mwTLRA.

In conclusion, we prefer TCA, because it is unaffected by the presence of zeros. This example shows that, the addition of a constant $c > 0$ to a count data with zero cells complicates the application of logratio methods. We completely agree with Egozcue, Pawlowsky-Glahn, and Gloor (2018) who state "*the presence of zeroes introduces a new source of contamination*".

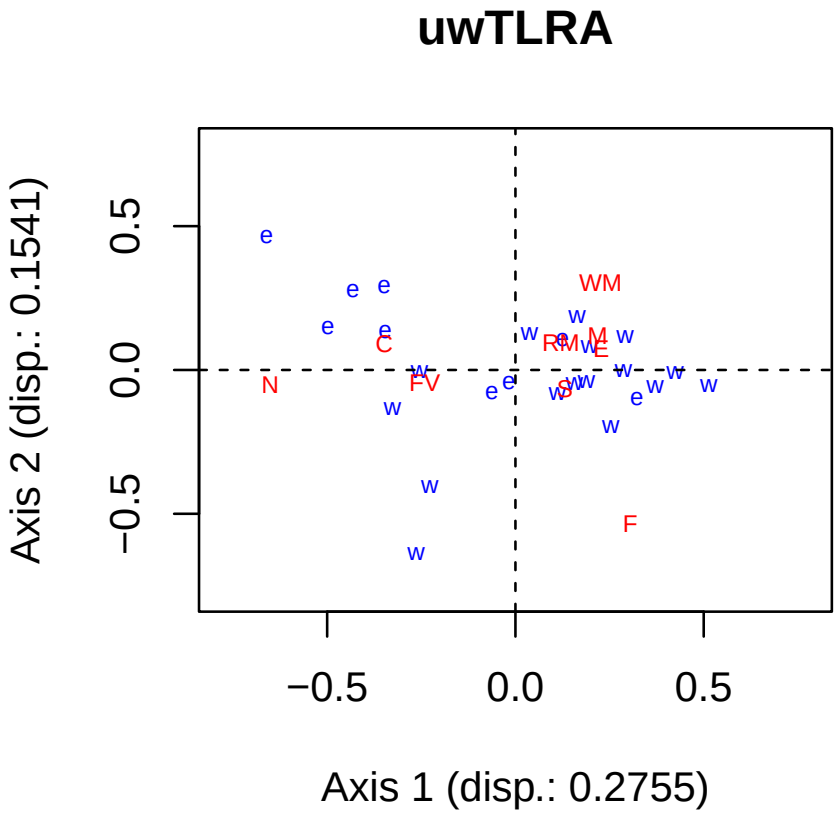

Figure 3: uwTLRA plot of Food Compositional data

8. Conclusion

In this paper, we, as dwarfs on the shoulders of three giants Benzécri-Goodman-Aitchison, attempted to see further by relating theory with practice. First, we reviewed the principles on which three interrelated well developed methods, for the analysis and visualization of contingency tables and compositional data, are erected: CA based on Benzécri's principle of distributional equivalence, Goodman's RC association model based on Yule's principle of scale invariance, and CoDA based on Aitchison's principle of subcompositional coherence.

Second, we introduced a novel index named the intrinsic measure of the quality of the signs of the residuals (QSR) for the choice of the preprocessing (double-centering), and consequently of the method among TCA, mwTLRA and uwTLRA. The criterion is based on taxicab singular value decomposition (TSVD) on which the package TaxicabCA in R is developed. We presented a minimal R script that can be executed to obtain the numerical results and the maps in this paper.

Third, we introduced a flexible method based on the QSR index for the choice of the constant to be added to contingency tables with zero counts so that TLRA methods can be applied.

We conclude by re-citing (Tukey 1977, p. 400): "*the general maxim–it is a rare thing that a specific body of data tells us clearly enough how it itself should be analyzed–applies to choice of re-expression for two-way analysis*". In this paper we studied three choices of re-expression for contingency tables and compositional data: TCA, mwTLRA and uwTLRA.

Table 2: QSR in (%) and dispersions of Rodent data for the first 2 dimensions

		TCA		
	QSR_1	QSR_2	δ_1	δ_2
	67.8	83.9	0.478	0.422
		uwTLRA		
Estimator	QSR_1	QSR_2	δ_1	δ_2
BES	69.1	57.5	0.403	0.198
Jeffreys	67.4	53.9	0.493	0.245
Perks	67.2	51.0	1.356	0.682
		mwTLRA		
Estimator	QSR_1	QSR_2	δ_1	δ_2
BES	66.0	85.2	0.404	0.373
Jeffreys	65.2	83.8	0.466	0.425
Perks	64.4	74.6	0.891	0.735

Do all roads lead to Rome? In unipolar world, yes; but in multipolar world, no. In the analysis of contingency tables and compositional data, the world is multipolar.

Acknowledgements

Choulakian's research has been supported by NSERC of Canada. The authors thank M. Greenacre for providing recent references of his works. The authors thank two reviewers and Pr. M. Templ, editor of the Austrian Journal of Statistics, for their constructive comments.

Appendix 1

Proposition 2. *TCA and TLRA satisfy the distributional/compositional equivalence property. That is,*

$$\lambda_{ij} = \sum_{\alpha=1}^{k} f_\alpha(i) g_\alpha(j)/\delta_\alpha \quad \textit{with } f_\alpha(1) = f_\alpha(2).$$

Proof. We prove it for TLRA, the proof for TCA being similar. Without loss of generality, suppose that the first two rows of the correspondence table (p_{ij}) are proportional; that is, $p_{1j} = Cp_{2j}$ for $j = 1, ..., J$, where C is a strictly positive constant. The proof is composed in 2 steps.

First, we show that the interaction terms described in (11) satisfy

$$\lambda_{1j} = \lambda_{2j}.$$

We have

$$\begin{aligned}
\lambda_{1j} &= \log(p_{1j}) - \sum_{j=1}^{J} w_j^C \log(p_{1j}) - G_{+j} + G_{++} \\
&= \log(Cp_{2j}) - \sum_{j=1}^{J} w_j^C \log(Cp_{2j}) - G_{+j} + G_{++} \\
&= \log(C) + \log(p_{2j}) - \sum_{j=1}^{J} w_j^C (\log(C) + \log(p_{2j})) - G_{+j} + G_{++} \\
&= \lambda_{2j}, \text{ because } \sum_{j=1}^{J} w_j^C = 1.
\end{aligned}$$

Second, we follow the Steps (1,2,3) described by equations (12,13,14,19) and the required result follows easily. □

Theorem 2. *Let* $\mathbf{X} = (x_{ij})$ *for* $i = 1, 2, ..., I$ *and* $j = 1, ..., J$ *be a contingency table or a compositional data set. Suppose the first two rows of* $\boldsymbol{X}$ *are proportional,* $x_{1j} = Cx_{2j}$ *for* $j = 1, ..., J$*, where* C *is a strictly positive constant. TLRA of* $\boldsymbol{X}$ *with a priori weights* (w_i^R, w_j^C) *is equivalent to TLRA of* $X^{merge} = (x_{ij}^m)$ *for* $i = (1+2), 3, ..., I$ *and* $j = 1, ..., J$ *with weights* (w_i^{Rm}, w_j^C)*, where* $x_{(1+2)j}^m = x_{1j} + w_{2j}$ *and* $w_{1+2}^{Rm} = w_1^R + w_2^R$*, and* $x_{ij}^m = x_{ij}$ *and* $w_i^{Rm} = w_i^R$ *for* $i = 3, ..., I$*.*

Theorem 2 means the following:

Suppose the first two rows of the correspondence table $(p_{ij} = x_{ij}/\sum_{i,j} x_{ij})$ are proportional, $p_{1j} = Cp_{2j}$ for $j = 1, ..., J$, where C is a strictly positive constant; and (p_{ij}^{merge}) is the merged correspondence table of size $(I-1)$ by J, such that

$$p_{(1+2)j}^{merge} = p_{1j} + p_{2j} \quad \text{with weights } (w_{(1+2)}^{Rm}, w_j^C)$$
$$p_{ij}^{merge} = p_{ij} \quad \text{for} \quad i = 3, ..., I \text{ with weights } (w_i^R, w_j^C)$$

and consider the bilinear decompositions

$$\lambda_{ij} = \sum_{\alpha=1}^{k} f_\alpha(i) g_\alpha(j)/\delta_\alpha \qquad \text{for} \quad i = 1, ..., I \quad \text{and} \quad j = 1, ..., J$$

$$\lambda_{ij}^{merge} = \sum_{\alpha=1}^{k} f_\alpha^m(i) g_\alpha^m(j)/\delta_\alpha^m \qquad \text{for} \quad i = (1+2), 3, ..., I \quad \text{and} \quad j = 1, ..., J.$$

Then

$$f_\alpha^m(1+2) = f_\alpha(1) = f_\alpha(2)$$
$$g_\alpha^m(j) = g_\alpha(j)$$
$$\delta_\alpha^m = \delta_\alpha.$$

Proof. We sketch the steps for a proof:

a) We show $\lambda_{(1+2)j}^{merge} = \lambda_{2j}$ and we use 1 2.

$$\lambda_{(1+2)j}^{merge} = \log(p_{(1+2)j}^{merge}) - \sum_{j=1}^{J} w_j^C \log(p_{(1+2)j}^{merge}) \tag{A1}$$
$$- \sum_{i=1}^{I} w_i^R \log(p_{(1+2)j}^{merge}) + \sum_{j=1}^{J} \sum_{i=1}^{I} w_j^C w_i^R \log(p_{(1+2)j}^{merge}).$$

Let us consider separately the four terms, Ti for $i = 1, 2, 3, 4$, on the right-side of equation (A1).

$$T1 = \log(p_{(1+2)j}^{merge})$$
$$= \log(C+1) + \log(p_{2j}).$$

$$T2 = \sum_{j=1}^{J} w_j^C T1$$
$$= \log(C+1) + \sum_{j=1}^{J} w_j^C \log(p_{2j}).$$

$$
\begin{aligned}
T3 &= \sum_{i=1}^{I} w_i^R \log(p_{ij}^{merge}) \\
&= \sum_{i\geq 3}^{I} w_i^R \log(p_{ij}) + w_{(1+2)}^{Rm} \log(p_{(1+2)j}^{merge}) \\
&= \sum_{i\geq 3}^{I} w_i^R \log(p_{ij}) + (w_1^R + w_2^R)\,[\log(C+1) + \log(p_{2j})] \\
&= \sum_{i\geq 3}^{I} w_i^R \log(p_{ij}) + (w_1^R + w_2^R)\log(C+1) + w_1^R \log(p_{1j}/C) + w_2^R \log(p_{2j}) \\
&= \sum_{i=1}^{I} w_i^R \log(p_{ij}) + (w_1^R + w_2^R)\log(C+1) - w_1^R \log(C).
\end{aligned}
$$

$$
\begin{aligned}
T4 &= \sum_{j=1}^{J} w_j^C T3 \\
&= \sum_{j=1}^{J} w_j^C \sum_{i=1}^{I} w_i^R \log(p_{ij}) + (w_1^R + w_2^R)\log(C+1) + w_1^R \log(C).
\end{aligned}
$$

Now

$$
\begin{aligned}
\lambda_{ij}^{merge} &= T1 - T2 - T3 + T4 \\
&= \lambda_{2j} = \lambda_{1j}.
\end{aligned}
$$

b) We show $\lambda_{ij}^{merge} = \lambda_{ij}$ for $i = 3, 4, ..., I$.

The proof is similar to a) above.

c) We use the three steps of section 4.2, where Step 1 is double-centering.

In Step 2: $w_i^R \lambda_{ij} w_j^C = \sum_{\alpha=1}^{k} a_\alpha(i) b_\alpha(j)/\delta_\alpha$ and $w_i^{Rm} \lambda_{ij}^{merge} w_j^C = \sum_{\alpha=1}^{k} a_\alpha^m(i) b_\alpha^m(j)/\delta_\alpha^m$, where $\frac{a_\alpha(1)}{w_1^R} = \frac{a_\alpha(2)}{w_2^R} = \frac{a_\alpha(1+2)}{w_1^R + w_2^R}$, because of the equalities in a) and b). Furthermore, $\delta_\alpha = \delta_\alpha^m$ and $b_\alpha(j) = b_\alpha^m(j)$.

In Step 3: $\lambda_{ij} = \sum_{\alpha=1}^{k} f_\alpha(i) g_\alpha(j)/\delta_\alpha$ and $\lambda_{ij}^{merge} = \sum_{\alpha=1}^{k} f_\alpha^m(i) g_\alpha^m(j)/\delta_\alpha^m$, where by c) $f_\alpha(1) = f_\alpha(2) = f_\alpha^m(1+2)$. □

Appendix 2: The R code

```
#  dataMatrix holds the raw data
library(TaxicabCA)
dataMatrix = as.matrix(rodent)
rownames(dataMatrix) <- paste('␣',1:28,sep='␣')
colnames(dataMatrix) <- paste('C',1:9,sep='␣')
# TCA visualization
tca.Data <- tca(dataMatrix, nAxes=2,algorithm = 'exhaustive')
tca.Data$dataName <- 'TCA'
plot(
tca.Data,
axes = c(1, 2),
labels.rc = c(1, 1),
col.rc = c('blue','red'),
```

```
pch.rc = c(3, 3, 0.1, 0.1),
mass.rc = c(F, F),
cex.rc = c(0.6, 0.6),
jitter = c(F, T)
)
rowProba <- apply(dataMatrix,1,sum)/sum(dataMatrix)
colProba <- apply(dataMatrix,2,sum)/sum(dataMatrix)
nRow <- nrow(dataMatrix)
nCol <- ncol(dataMatrix)
#### TCOV, TLRA=uwTLRA and TWLRA=mwTLRA visualizations
and QSR computation
# Uncomment ONE to choose the method depending on the centering
centeringMethod <- 'TCA-TCOV'
centeringMethod <- 'uwTRA'
centeringMethod <- 'mwTLRA'
### Step 1 - Centering the data matrix:
#  According to the centering method chosen ABOVE
# TCA-TCOV Centering
if (centeringMethod == 'TCA-TCOV') {
Proba <- dataMatrix/sum(dataMatrix)
centeredDataMatrix <- Proba - rowProba %*% t(colProba)
}
37
## mwTLRA or uwTRA Centering
# Add 1 to each cell if there is a cell with 0 count
if (sum(dataMatrix == 0) > 0) {
#dataMatrix1 = dataMatrix + 1
#dataMatrix1 = dataMatrix + 1/2
dataMatrix1 = dataMatrix + 1/(nRow*nCol)
#rowS <- apply(dataMatrix,1,sum)
#dataMatrix1 = 10000*dataMatrix/rowS + 1
Proba <- dataMatrix1/sum(dataMatrix1)
}
logTaxa <- log(Proba)
# uwTRA centering
if (centeringMethod == 'uwTRA') {
Gj <- t(rep(1/nRow, nRow)) %*% logTaxa
Gi <- logTaxa %*% rep(1/nCol, nCol)
Gt <- t(rep(1/nRow, nRow)) %*% logTaxa %*% rep(1/nCol, nCol)
Lambda <- logTaxa + rep(1, nRow) %*% Gt %*% t(rep(1, nCol)) -
((rep(1, nRow) %*% Gj) +(Gi %*% t(rep(1, nCol))))
centeredDataMatrix <- Lambda*(rep(1/nRow, nRow) %*% t(rep(1/nCol,
nCol)))
}
# mwTLRA Centering
if (centeringMethod == 'mwTLRA') {
Gjtilda <- t(rowProba) %*% logTaxa
Gitilda <- logTaxa %*% colProba
Gtilda <- t(rowProba) %*% logTaxa %*% colProba
Lambdatilda <- logTaxa + rep(1, nRow) %*% Gtilda %*% t(rep(1,
nCol)) -
((rep(1, nRow) %*% Gjtilda) +(Gitilda %*% t(rep(1, nCol))))
centeredDataMatrix <- Lambdatilda*(rowProba %*% t(colProba))
}
```

```
### Step 2 - Compute the Taxicab SVD for the centered matrix
# Common to the 3 methods
nAxes <- 4 # Must have nAxes < min(nRow, nCol)
axesNames <- paste('Axis',1:nAxes,sep='␣')
# Create the matrices required to receive the results
rowScores <- matrix(NA, nrow = nRow, ncol = nAxes)
rownames(rowScores) <- rownames(centeredDataMatrix)
38
colScores <- matrix(NA, ncol = nCol, nrow = nAxes)
colnames(colScores) <- colnames(centeredDataMatrix)
dispersion <- rep(NA, nAxes) # matrix(0, nrow = nAxes, ncol = 1)
QSR <- matrix(NA, nrow = nAxes, ncol = 5)
colnames(QSR) <- c('V+U+','V-U-','V-U+','V+U-','All')
rownames(QSR) <- colnames(rowScores) <- rownames(colScores) <-
names(dispersion) <- axesNames
residuals <- centeredDataMatrix
for (iiAxis in 1:nAxes) {
# The search functions come from TaxicabCA
# Uncomment ONE search method - As of 2020, on a desktop computer,
# Exhaustive is only feasible for nRow < 22
axisResult <- SearchExhaustive(residuals)
# axisResult <- SearchCrissCross(residuals)
# axisResult <- SearchGenteticAlgorithm(residuals)
U <- axisResult$uMax
dispersion[iiAxis] <- axisResult$L1Max
rowScores[, iiAxis] <- residuals %*% t(axisResult$uMax)
V <- sign(rowScores[, iiAxis, drop = F])
colScores[iiAxis, ] <- t(V) %*% residuals
# Compute the quality of the signs of the residuals (QSR) for each UV
'quadrant'
QSR[iiAxis,1] <- 0.25*axisResult$L1Max/sum(abs(residuals[V >= 0,U
>= 0]))
QSR[iiAxis,2] <- 0.25*axisResult$L1Max/sum(abs(residuals[V < 0,U <
0]))
QSR[iiAxis,4] <- -0.25*axisResult$L1Max/sum(abs(residuals[V >= 0,U
< 0]))
QSR[iiAxis,3] <- -0.25*axisResult$L1Max/sum(abs(residuals[V < 0,U >=
0]))
# Compute the overall quality of the signs of the residuals
QSR[iiAxis,5] <- axisResult$L1Max/sum(abs(residuals))
# Update the residuals for the next iteration
residuals <- residuals - rowScores[, iiAxis, drop = F] %*%
(colScores[iiAxis,,drop = F]/dispersion[iiAxis])
}
###  Step 3: Visualisation
# TCOV Visualisation
39
if (centeringMethod == 'TCA-TCOV') {
tcov.Data <- list(rowScores = rowScores, colScores = colScores,
dispersion = dispersion, dataName = 'TCOV')
# Add class 'tca' to the class of the list in order to
# call plot.tca from TaxicabCA automatically
class(tcov.Data) <- c(class(tcov.Data),'tca')
# Open a graphics window outside of RStudio (if RStudio is used)
```

```
# dev.new(noRStudioGD = TRUE)
# Call plot.tca from TaxicabCA (automatically: tlra.Data is class 'tca')
# Use labels.rc = c(1,1) only if the data has rownames and/or colnames.
# Otherwise, use labels.rc = c(0,0), c(1,0) or c(0,1)
plot(
tcov.Data,
axes = c(1, 2),
labels.rc = c(1, 1),
col.rc = c('blue', 'red'),
pch.rc = c(5, 5, 0.1, 0.1),
mass.rc = c(F, F),
cex.rc = c(0.6, 0.6),
jitter = c(F, T)
)
}
# uwTLRA Visualisation
if (centeringMethod == 'uwTRA') {
tlra.Data <- list(rowScores = rowScores*nRow, colScores = colScores*nCol,
dispersion = dispersion, dataName = 'uwTLRA')
class(tlra.Data) <- c(class(tlra.Data),'tca')
plot(
tlra.Data,
axes = c(1, 2),
labels.rc = c(1, 1),
col.rc = c('blue', 'red'),
pch.rc = c(5, 5, 0.1, 0.1),
mass.rc = c(F, F),
cex.rc = c(0.6, 0.6),
jitter = c(F, F)
)
}
40
# mwTLRA Visualisation
if (centeringMethod == 'mwTLRA') {
twlra.Data <- list(rowScores = rowScores/(rowProba %*% t(rep(1,nAxes))),
colScores = colScores/(rep(1,nAxes) %*% t(colProba)),
dispersion = dispersion, dataName = 'mwTLRA')
class(twlra.Data) <- c(class(twlra.Data),'tca')
plot(
twlra.Data,
axes = c(1, 2),
labels.rc = c(1, 1),
col.rc = c('blue', 'red'),
pch.rc = c(5, 5, 0.1, 0.1),
mass.rc = c(F, F),
cex.rc = c(0.6, 0.6),
jitter = c(F, T)
)
}
# Print the numerical results
options(digits=3)
print(centeringMethod)
print(dispersion)
print(QSR)
```

Appendix 3: FOOD compositional data in R

```
# FOOD data set AJS
dataMatrix = matrix(c(10.1, 1.4, 0.5, 8.9, 0.2,
 42.3, 0.6, 5.5, 1.7, 8.9, 14.0, 4.3, 19.9, 2.1,
 28.0, 3.6, 1.3, 4.3, 13.5, 9.3, 4.1, 17.5, 4.5,
 26.6,  5.7, 2.1, 4.0, 7.8,  6.0, 1.6, 8.3, 1.2,
 56.7, 1.1, 3.7, 4.2, 9.7, 11.4, 2.8, 12.5, 2.0,
 34.3,5.0, 1.1, 4.0, 10.6, 10.8, 3.7, 25.0, 9.9,
 21.9, 4.8, 0.7, 2.4, 9.5, 4.9, 2.7,  33.7, 5.8,
 26.3, 5.1, 1.0, 1.4, 18.0, 9.9, 3.3, 19.5, 5.7,
 28.1,  4.8, 2.4, 6.5, 9.3, 4.6, 2.1, 16.6, 3.0,
 43.6, 6.4, 3.4, 2.9, 8.4, 11.6, 3.7, 11.1, 5.4,
 24.6, 6.5, 0.8, 3.6,  41,11.4, 12.5, 4.1, 18.8,
 3.4, 18.6, 5.2, 1.5, 3.8, 10.2, 3.0, 2.8, 17.6,
 5.9, 41.7, 2.2, 7.8, 6.5,  5.3, 12.4, 2.9, 9.7,
 0.3, 40.1 4.0, 5.4, 4.2, 13.9, 10.0, 4.7, 25.8,
 2.2, 24.0, 6.2, 1.6, 2.9, 9.0,  5.1, 2.9, 13.7,
 3.4, 36.8, 2.1, 4.3, 6.7, 9.4, 4.7, 2.7,  23.3,
 9.7, 23.0, 4.6, 1.6, 2.7, 6.9, 10.2, 2.7, 19.3,
 3.0, 36.1, 5.9, 2.0, 6.6,  6.2, 3.7, 1.1,  4.9,
 14.2, 27.0, 5.9, 4.7, 7.9, 6.2, 6.3, 1.5, 11.1,
 1.0, 49.6, 3.1, 5.3,  2.8, 7.1, 3.4, 3.1, 8.6,
 7.0, 29.2,  5.7, 5.9, 7.2,9.9,  7.8, 3.5, 24.7,
 7.5, 19.5, 3.7, 1.4, 2.0,13.1, 10.1, 3.1, 23.8,
 2.3, 25.6, 2.8, 2.4, 4.9, 9.5, 13.6, 3.6, 23.4,
 2.5, 22.4, 4.2, 1.8, 3.7, 17.4, 5.7, 4.7, 20.6,
 4.3,  24.3, 4.7, 3.4, 3.3,  4.4, 5.0, 1.2, 9.5,
 0.6, 55.9, 3.0, 5.7, 3.2),nrow=25,byrow=T)
colnames(dataMatrix) <- c('RM','WM','E','M','F','C','S','N','FV')
rownames(dataMatrix) <- c('e','w','w','e','e','w','w','w','e','e',
'w','w','e','w','w','w','e','w','e','w','w','w','w','w','e')
```

References

Aitchison J (1983). "Principal Component Analysis of Compositional Data." *Biometrica*, **1**(70), 57–65.

Aitchison J (1986). *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

Aitchison J (1994). *Principles of Compositional Data Analysis. A Chapter in Multivariate Analysis and Its Applications*. Institute of Mathematical Statistics, Hayward.

Aitchison J (1997). "The One-hour Course in Compositional Data Analysis or Compositional Data Analysis Is Simple." **24**, 3–35. ISBN 978-84-87867-76-7.

Aitchison J, Greenacre M (2002). "Biplots for Compositional Data." *Journal of the Royal Statistical Society. Series C*, **51**(4), 375–392.

Allard J, Choulakian V (2019). "Taxicab Analysis in R."

Beh E, Lombardo R (2014). *Correspondence Analysis: Theory, Practice and New Strategies*. N.Y: Wiley.

Benzécri J (1966). "Linguistique et Mathématique." *Revue Philosophique de la France et de l'Étranger*, **156**(3), 309–374.

Benzécri JP (1973a). *L'Analyse des Données*, volume 2. Paris:Dunod.

Benzécri JP (1973b). *L'Analyse des Données: La Taxinomie*, volume 1. Paris:Dunod.

Benzécri JP (1979). "On the Analysis of a Table, One of Whose Columns Has a Predominant Weight (in French)." *Les Cahiers de L'Analyse des Données*, **4**, 413–416.

Bernard JM (2005). "An Introduction to the Imprecise Dirichlet Model for Multinomial Data." *International Journal of Approximate Reasoning*, **39**(2), 123–150.

Borchers HW (2021). "Numbers."

Choulakian V (2006). "Taxicab Correspondence Analysis." *Psychometrika*, **71**, 333–345.

Choulakian V (2008a). "Multiple Taxicab Correspondence Analysis." *Advances in Data Analysis and Classification*, **2**, 177–206.

Choulakian V (2008b). "Taxicab Correspondence Analysis of Contingency Tables with One Heavyweight Column." *Psychometrika*, **73**(2), 309–319.

Choulakian V (2016). "Matrix Factorizations Based on Induced Norms." *Statistics, Optimization and Information Computing*, **4**, 1–14.

Choulakian V (2017). "Taxicab Correspondence Analysis of Sparse Contingency Tables." *Italian Journal of Applied Statistics*, **29**(2), 153–179.

Choulakian V (2021). "Quantification of Intrinsic Quality of a Principal Dimension in Correspondence Analysis and Taxicab Correspondence Analysis." *https://arxiv.org/pdf/arXiv:2108.10685.*

Choulakian V (2022). "Some Notes on Goodman's Marginal-free Correspondence Analysis." *https://arxiv.org/pdf/2202.01620.pdf.*

Choulakian V, Abou-Samra J (2020). "Mean Absolute Deviations about the Mean, the Cut Norm and Taxicab Correspondence Analysis." *Open Journal of Statistics*, **10**(1), 97–112.

Choulakian V, Allard J, Simonetti B (2013). "Multiple Taxicab Correspondence Analysis of a Survey Related to Health Service." *Journal of Data Science*, **11**, 205–229.

Choulakian V, de Tibeiro J (2013). "Graph Partitioning by Correspondence Analysis and Taxicab Correspondence Analysis." *Journal of Classification*, **30**, 397–427.

Cuadras CM, Cuadras D, Greenacre M (2006). "A Comparison of Different Methods for Representing Categorical Data." *Communications in Statistics - Simulation and Computation*, **35**(2), 447–459.

Das S (2021). "Identifying Key Patterns in Motorcycle Crashes: Findings from Taxicab Correspondence Analysis." *Transportmetrica A: Transport*, **17**(4), 593–614.

Das S, Minjares-Kyle L, Wu L, Henk R (2019). "Understanding Crash Potential Associated with Teen Driving: Survey Analysis Using Multivariate Graphical Method." *Journal of Safety Research*, **70**, 213–222.

Egozcue JJ, Pawlowsky-Glahn V (2005). "Groups of Parts and Their Balances in Compositional Data Analysis." *Mathematical Geology*, **37**, 795–828.

Egozcue JJ, Pawlowsky-Glahn V (2016). "Changing the Reference Measure in the Simplex and Its Weighting Effects." *Austrian Journal of Statistics*, **45**(4), 25–44.

Egozcue JJ, Pawlowsky-Glahn V (2019). "Compositional Data: The Sample Space and Its Structure." *TEST*, **28**, 599–638.

Egozcue JJ, Pawlowsky-Glahn V, Gloor GB (2018). "Linear Association in Compositional Data Analysis." *Austrian Journal of Statistics*, **47**(1), 3–31.

Egozcue JJ, Pawlowsky-Glahn V, Templ M, Hron K (2015). "Independence in Contingency Tables Using Simplicial Geometry." *Communications in Statistics - Theory and Methods*, **18**(44), 3978–3996.

Emerson JD, Stoto MA (1983). "Transforming data. Chapter 4 in Hoaglin DC, Mosteller F, Tukey JW (ed)." *Understanding Robust and Exploratory Data Analysis, Wiley: NY*.

Escofier B (1978). "Analyse Factorielle et Distances Répondant au Principe d'Équivalence Distributionnelle." *Revue de Statistique Appliquée*, **26**(4), 29–37.

Fichet B (2009). "Metrics of Lp -type and Distributional Equivalence principle." *Advances in Data Analysis and Classification*, **3**(article 305).

Filzmoser P, Hron K, Templ M (2018). *Applied Compositional Data Analysis*. Springer, Switzerland.

Fiserova E, Hron K (2011). "On the Interpretation of Orthonormal Coordinates for Compositional Data." *Mathematical Geosciences*, **43**, 455–468.

Goodman LA (1979). "Simple Models for the Analysis of Association in Crossclassifications Having Ordered Categories." *Journal of the American Statistical Association*, **74**, 537–555.

Goodman LA (1981a). "Association Models and Canonical Correlation in the Analysis of Cross-classifications Having Ordered Categories." *Journal of the American Statistical Association*, **76**, 320–334.

Goodman LA (1981b). "Association Models and the Bivariate Normal for Contingency Tables with Ordered Categories." *Biometrika*, **68**, 347–355.

Goodman LA (1991). "Measures, Models, and Graphical Displays in the Analysis of Cross-classified Data." *Journal of the American Statistical Association*, **86**(4), 1085–1111.

Goodman LA (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 (2011). "Biplots: Taking stock." URL https://www.youtube.com/watch-v=Tm5933EuI-Q.

Greenacre M (2018). *Compositional Data Analysis in Practice*. Chapman and Hall/CRC Press, Boca Raton.

Greenacre M (2020). "Amalgamations Are Valid in Compositional Data Analysis, Can Be Used in Agglomerative Clustering, and Their Logratios Have an Inverse Transformation." *Applied Computing and Geosciences*, **5**, 100017.

Greenacre M, Grunsky E, Bacon-Shone J, Erb I, Quinn T (2022). "Aitchison's Compositional Data Analysis 40 Years on: A Reappraisal." URL https://arxiv.org/pdf/2201.05197.pdf.

Greenacre M, Hastie T (1987). "The Geometric Interpretation of Correspondence Analysis." *Journal of the American Statistical Association*, **5**(100017), 437–447.

Greenacre M, Lewi P (2009). "Distributional Equivalence and Subcompositional Coherence in the Analysis of Compositional Data, Contingency Tables and Ratio-scale Measurements." *Journal of Classification*, **26**, 29–54.

Hotelling H (1936). "Relations between Two Sets of Variables." *Biometrika*, **28**, 321–377.

Hron K, Filzmoser P, De Caritat PEA (2017). "Weighted Pivot Coordinates for Compositional Data and Their Application to Geochemical Mapping." *Mathematical Geosciences*, **49**(6), 797–814.

Kazmierczak JB (1985). "Analyse Logarithmique: Deux Exemples d'Application." *Revue de Statistique Appliquée*, **33**(1), 13–24.

Kazmierczak JB (1987). "Sur l'Usage d'un Principe d'Invariance pour Aider au Choix d'une Métrique." *Statistique et Analyse des Données*, **12**(3), 37–57.

Lancaster HO (1958). "The Structure of Bivariate Distributions." *The Annals of Mathematical Statistics*, **29**, 719–736.

Lebart L (1979). "Example of the Analysis of a Table with a Column of Predominant Weight (in French)." *Les Cahiers de L'Analyse des Données*, **4**, 417–422.

Lewi P (1976). "Spectral Mapping, a Technique for Classifying Biological Activity Profiles of Chemical Compounds." *Arzneimittelforschung*, **4**(26), 1295–1300.

Lovell DR, Chua XY, McGrath A (2020). "Counts: An Outstanding Challenge for Log-ratio Analysis of Compositional Data in the Molecular Biosciences." *NAR Genomics and Bioinformatics*, **2**(2).

Lubbe S, Filzmoser P, Templ M (2021). "Comparison of Zero Replacement Strategies for Compositional Data with Large Numbers of Zeros." *Chemometrics and Intelligent Laboratory Systems*, **248**(210:104).

Mandel J (1971). "A New Analysis of Variance Model for Non-additive Data." *Technometrics*, **13**(1), 1–8.

Mao Z (1938). "On Practice." URL https://www.marxists.org/reference/archive/mao/selected-works/volume-1/mswv1-16.htm.

Mosteller F (1968). "Association and Estimation in Contingency Tables." *Journal of the American Statistical Association*, **63**(321), 1–28.

Nielsen F, Sun K (2017). "Clustering in Hilbert Simplex Geometry." In *Geometric Structures of Information*, pp. 297–331. Springer. URL arXiv:1704.00454.

Nishisato S (1984). "Forced Classification: A Simple Application of a Quantification Method." *Psychometrika*, **49**(1), 25–36.

Pawlowsky-Glahn V, Egozcue JJ (2011). "Exploring Compositional Data with the CoDa-Dendrogram." *Austrian Journal of Statistics*, **40**(1-2), 25–36.

Strang G (2019). "Change of Basis; Image Compression." URL https://www.youtube.com/watch-v=0h43aV4aH7Iandt=1270s.

Tenenhaus M, Augendre H (1996). "Analyse Factorielle Inter-batteries de Tucker et Analyse Canonique aux Moindres Carrés Partiels." *In Recueil des Résumés des Communications des 28ème Journées de Statistique*, pp. 693–697.

Tucker LR (1958). "An Inter-battery Method of Factor Analysis." *Psychometrika*, **23**, 111–136.

Tukey JW (1977). *Exploratory Data Analysis*. Addison-Wesley: Reading, Massachusetts.

Yule GU (1912). "On the Methods of Measuring Association between Two Attributes." *Journal of the Royal Statistical Society*, **75**(6), 579–652.

Zizek S (2008). *On Paractice and Contradiction*. Verso Books, London, UK.

Affiliation:

Vartan Choulakian
Département de mathematiques et de statistique
Université de Moncton
Moncton, Canada
E-mail: vartan.choulakian@umoncton.ca

Jacques Allard
Département de mathematiques et de statistique
Université de Moncton
Moncton, Canada
E-mail: jacques.allard@umoncton.ca

Smail Mahdi
Department of Computer Science, Mathematics and Physics
University of the West Indies
Cave Hill Campus
Barbados
E-mail: smail.mahdi@cavehill.uwi.edu

Austrian Journal of Statistics
published by the Austrian Society of Statistics
Volume 52
July 2023

http://www.ajs.or.at/
http://www.osg.or.at/

Submitted: 2021-02-14
Accepted: 2022-07-26