Journal de la Société Française de Statistique
Vol. 155 No. 4 (2014)

Taxicab Correspondence Analysis of Ratings and Rankings

Titre: Analyse des Correspondances du Taxi de Notes et de Rangs

Vartan Choulakian[1]

Abstract: Let $\mathbf{Y}$ be an $I \times Q$ ratings data set, where Q represents the number of items, and I represents the number of rated objects or the number of individuals expressing their opinions on the Q items. This paper considers two kinds of data codings before the application of correspondence analysis (CA) or taxicab correspondence analysis (TCA), where TCA is a L_1 variant of CA: the doubled data set $\mathbf{Y}_D$ of size $I \times 2Q$, and the data set $\mathbf{Y}_{nega}$ of size $I \times (Q+1)$ where a column named nega is added representing the cumulative complementary columns. The interpretation of maps in CA of $\mathbf{Y}_D$ is based on the lever principle. We use the law of contradiction to interpret maps of CA and TCA of $\mathbf{Y}_{nega}$. We provide necessary and sufficient conditions for TCA of $\mathbf{Y}_{nega}$ or $\mathbf{Y}_D$ so that the first factor score is an affine function of the sum score of the ratings; and, if this is true for a dataset, then following Cox we suggest the use of the sum score of ratings either to reduce the Q ratings into a single index, or to summarize the underlying latent variable. This ordinal inference can be of two types: weak or strong. In the case of a rankings dataset, the proposed approach corresponds to Borda count rule or modified Borda count rule. Examples are provided.

Résumé : Soit $\mathbf{Y}$ un tableau de notes sur $I \times Q$; où I est un ensemble d'individus, et la $i$ème ligne représente les notes attribuées par l'individu i sur Q variables ou attributs. Dans cet article nous étudions deux codages du tableau $\mathbf{Y}$ avant de le traiter par analyse des correspondances (AC) ou analyse des correspondances du taxi (ACT), ACT étant une variante robuste de AC basée sur la norme L_1 : Le tableau dédoublé $\mathbf{Y}_D$ de dimension $I \times 2Q$, et le tableau $\mathbf{Y}_{nega}$ de dimension $I \times (Q+1)$ où une colonne nommée *nega* est ajoutée à $\mathbf{Y}$ representant la note complémentaire globale. L'interprétation des diagrammes du tableau $\mathbf{Y}_D$ par AC ou ACT est basée sur le principe du bras de levier. Nous utilisons la loi de contradiction pour interpréter les diagrammes du tableau $\mathbf{Y}_{nega}$ par AC ou ACT. Une condition nécessaire et suffisante pour que l'analyse du tableau $\mathbf{Y}_{nega}$ par ACT et l'analyse du tableau $\mathbf{Y}_D$ par ACT soient equivalentes est que le 1er facteur est une fonction affine du total de notes. Et si cette condition est satisfaite, suivant Cox, nous utilisons le 1er facteur comme un résumé de la variable latente. Cette inférence peut être de deux sortes, faible ou forte. Dans le cas de données des rangs représentant des préférences individuelles, la méthode correspond à la règle de Borda ou une version modifiée. Deux exemples de natures différentes sont exposés.

Keywords: Total de notes, nega, dédoublement, principe du bras de levier, loi de contradiction, équation personnelle, point aberrant, règle de Borda, codage de Nishisato, analyse des correspondances du taxi, analyse d'items
Mots-clés : Sum score, nega, doubling, lever principle, law of contradiction, personal equation, response styles, rogue items, strategic voters, Borda count, Nishisato mapping, taxicab correspondence analysis, IRT
AMS 2000 subject classifications: 62H25, 62G35

1. Introduction

Cox (2006) titled his talk "*In praise of the simple sum score*" at the International Conference on "*Statistical Latent Variables in the Health Sciences*" held in Perugia, Italy; see also Cox (2008) and Cox and Wermuth (2002). Cox assumed the following:

[1] Université de Moncton, Canada.
E-mail: vartan.choulakian@umoncton.ca

(A1) Let $\mathbf{Y} = (y_{iq})$ for $i = 1,...,I$ and $q = 1,...,Q$ be a data set, where a sample of I individuals are described by Q broadly similar measurements or scores.

(A2) The Q variables point in the same direction.

Then, he proposed the use of the simple sum score of the rows, y_{i*}, either as a summary of the data by a single index, or, as a representative of the underlying latent unobserved variable. As examples, Cox cited data sets having for subject study: true-false questions in an arithmetic test, and, items in health-related quality of life questionnaire. He concluded with the following:

(C1) When all items point in the same direction, communicability and statistical efficiency often point to simple sum score.

(C2) It is desirable to define items to make simple sum reasonably efficient.

Assumption A2 is very general but crucial; at a first glance, it seems transparent and based on the common sense; it has the important consequence that each item should contribute a positive part to the sum score. So, the major problem is how to know that (A2) is true: The approach proposed in this paper is direct and geometric based on the most robust matrix norm, named taxicab matrix norm.

The aim of this paper is to show that taxicab correspondence analysis (TCA) is a tool that can accomplish points (C1) and (C2) for a ratings data set $\mathbf{Y}$, because the first principal factor score of TCA of $\mathbf{Y}$ properly coded is intimately related to the sum score of ratings, as will be seen. This extends the main result in Choulakian et al. (2013), where they showed that for Q polytomous items, the first principal factor score of multiple TCA can always be interpreted as a sum score of Q Bernoulli random variables. TCA is a L_1 variant of correspondence analysis (CA) proposed by Choulakian (2006b).

CA is a popular and well established method to analyze questionnaires using rating scales. Cazes (1990) presented a panorama of different coding schemes of ratings for the application of correspondence analysis (CA). Essentially there are three kinds of codings of a ratings data set $\mathbf{Y} = (\mathbf{y}_{iq})$ for $i = 1,...,I$ and $q = 1,...Q$, where Q represents the number of items, and I represents the number of rated objects or the number of individuals expressing their opinion on the Q items. It is customery to represent the positive pole of an item q by $q+$ and its negative or complementary pole by $q-$. Further, let us suppose that the minimum value of $q+$ is 0 and its maximum value is m_q; thus for any item q, $(q+) + (q-) = m_q$. Let $\mathbf{1}_I$ be the vector having I components of ones and $\mathbf{m} = (m_q)$ the column vector with Q coordinates. We define $\mathbf{Y}_D = (\mathbf{Y}|\mathbf{1}_I\ \mathbf{m}\prime - \mathbf{Y})$ the doubled dataset of dimension $I \times 2Q$, and, $\mathbf{Y}_{nega} = (\mathbf{Y}|\mathbf{nega})$ of dimension $I \times (Q+1)$, where the vector $\mathbf{nega} = (\mathbf{1}_I\ \mathbf{m}\prime - \mathbf{Y})\mathbf{1}_Q$ represents the cumulative of the complementary columns. The coding $\mathbf{Y}_D$ is very well known, and it is quite old, see Benzécri (1973), page 25. The interpretation of maps produced by CA of $\mathbf{Y}_D$ are based on the well-known lever principle, see for instance, Greenacre (1984), page 175.

The coding $\mathbf{Y}_{nega}$ is the most recent, and it has been applied only twice in the statistical literature by Adames (1993) and Esmieu et al. (1993); Murtagh (2005), page 79, mentions it but provides no applications. There are some issues in the interpretation of maps produced by CA of $\mathbf{Y}_{nega}$. In this paper, we use the law of contradiction to interpret maps produced by CA and TCA of $\mathbf{Y}_{nega}$. According to Eves (1997), the method of reductio ad absurdum, which was brillantly used by Euclid, is based on two complementary principles of classical logic: the law of contradiction and the law of excluded middle. The law of contradiction shows that, in section 4, *at most* the first dimension of CA of $\mathbf{Y}_{nega}$ is interpretable.

The third is $\mathbf{Y}_{pers}$, where the data are coded according to Benzécri's personal equation mapping for Likert type data, see Benzécri and Benzécri (1989), or Cazes (1990). Dual scaling is similar to CA; within the context of dual scaling, Nishisato (1980) proposed a nonlinear mapping of ratings of data by ranking them; we consider Nishisato's coding of ratings a variant of $\mathbf{Y}_{pers}$, and its analysis by TCA is related to $\mathbf{Y}_D$ or $\mathbf{Y}_{nega}$ via the well-known Borda count or modified Borda count rule in voting theory. de Borda (1781) suggested assigning $Q-1,\ Q-2,,...,1,0$ points to the Q candidates, respectively, where the top-ranked candidate receives $Q-1$ credits, the second-ranked candidate receives $Q-2$ credits, and the last ranked candidate 0 credit. Summing the credits over the voters corresponds to the Borda count rule to rank the candidates according to the received credits from top to bottom.

This paper considers only $\mathbf{Y}_D$ and $\mathbf{Y}_{nega}$; that is why we did not define $\mathbf{Y}_{pers}$. It is well known that CA applied to these three codings of the same data set produces different numerical results, and there is no equivalence relationship between any two of them. Deniau et al. (1979) compared CA dispersion measures of two correspondence matrices, where one is obtained from the other by grouping of columns or rows; using their result we get $\lambda_\alpha^D \geq \lambda_\alpha^{nega}$ for $\alpha \geq 1$, where λ_α is the inertia (dispersion measure) of the αth principal axis.

In this paper we provide a necessary and sufficient condition for TCA of $\mathbf{Y}_{nega}$ or $\mathbf{Y}_D$ so that the first factor score is an affine function of the sum score of the ratings. And, if this is satisfied by a dataset, then following Cox, we suggest the use of the sum score of ratings either to reduce the Q ratings into a single index, or to summarize the underlying latent variable. This ordinal inference can be of two types: weak or strong, the former based on TCA of $\mathbf{Y}_{nega}$ and the latter based on TCA of $\mathbf{Y}_D$. Furthermore, if TCA of $\mathbf{Y}_{nega}$ is not related to the sum score statistic, and we want absolutely to use the sum score statistic (this is Cox's point C2 above concerning the efficiency of the sum score), then TCA can identify the rogue or incoherent items, items that do not point in the same direction as the rest, that have to be deleted so that the sum score statistic calculated on the remaining items can be used. This shows that TCA of $\mathbf{Y}_{nega}$ is completely related to the sum score statistic; while TCA of $\mathbf{Y}_D$ is related to the sum score statistic if and only if it is equivalent to TCA of $\mathbf{Y}_{nega}$. We can rephrase this otherwise in the following way: If there are rogue items, then the maps or biplots obtained by CA or TCA of $\mathbf{Y}_{nega}$ are useless and senseless by the law of contradiction; while the maps obtained by CA or TCA of $\mathbf{Y}_D$ can be useful and meaningful, because they can be interpreted by the lever principle.

The importance of applying CA or TCA to the coded data set $\mathbf{Y}_D$ or $\mathbf{Y}_{nega}$, and not to the original data set $\mathbf{Y}$ stems from the following consideration. Suppose that all Q ratings of two distinct rows in $\mathbf{Y}$ are linearly related, that is, $y_{iq} = \alpha\, y_{i_1q}$ for $\alpha > 0$; this implies that the rows i and i_1 have identical profiles, and by CA or TCA they can be merged together into one, which will be misleading.

This paper is organized as follows: In section 2 we present a general criterion for the assumption (A2); section 3 presents an overview of TCA; section 4 presents the main theoretical results; in section 5 we present the law of contradiction and its application to CA and TCA of $\mathbf{Y}_{nega}$; in section 6 we present 2 applications; and we conclude in section 7.

We suppose that the theory of correspondence analysis (CA) is known, which can be found, among others, in Benzécri (1973), Benzécri (1992), Greenacre (1984), Gifi (1990), Nishisato (1994), Le Roux and Rouanet (2004) and Murtagh (2005).

2. A general criterion

Suppose we have a set of Q random variables $\mathbf{x}'=(X_1,X_2,...,X_Q)$ with variance covariance matrix Σ such that for 2 distinct variables $cov(X_i,X_j)=0$ with zero probability. Then it is transparent that (X_1,X_2) point in the same direction when the covariance $cov(X_1,X_2)>0$.

We consider the centered random variable $S(\varepsilon)=\varepsilon_1(X_1-EX_1)+\varepsilon_2(X_2-EX_2)+...+\varepsilon_{Q-1}(X_{Q-1}-EX_{Q-1})+(X_Q-EX_Q)=\mathbf{u}'(\mathbf{x}-E\mathbf{x})$, where $\varepsilon\in\{-1,+1\}^{Q-1}$, $\mathbf{u}'=(\varepsilon \quad 1)$, E is the averaging operator and define the centered dispersion measure

$$\lambda_\alpha(\varepsilon)=E|S(\varepsilon)|^\alpha \quad \text{for} \quad \alpha\geq 1.$$

The content of this paper is based on the following general criterion

Definition 1: The Q random variables $(X_1,X_2,...,X_Q)$ point in the same direction if $\mathbf{u}=\mathbf{1}_Q=\arg\max\lambda_\alpha(\varepsilon)$ over $\varepsilon\in\{-1,+1\}^{Q-1}$.

An important consequence of Definition 1 is: If the Q random variables $(X_1,X_2,...,X_Q)$ point in the same direction then the random variable

$$\begin{aligned} S(\varepsilon) &= S(\mathbf{1}_{Q-1}) \\ &= \sum_{i=1}^{Q}(X_i-EX_i) \end{aligned}$$

is the centered sum score statistic. Two prominent values of α are 1 and 2.

For $\alpha=2$,

$$\begin{aligned} \max\lambda_2(\varepsilon) &= variance(S(\varepsilon)) \\ &= \mathbf{u}'\Sigma\mathbf{u} \quad \text{over} \quad \varepsilon\in\{-1,+1\}^{Q-1}, \end{aligned}$$

corresponds to the centroid method of component analysis, which preceded Hotelling's PCA, see (Choulakian (2003), Choulakian (2005), Choulakian (2006a)): The centroid method of principal component analysis was used extensively in the psychometric literature before the advent of the computers. It was first proposed by Burt (1917) and developed by Thurstone (1931). The centroid method is discussed in every major book in quantitative psychology treating factor analysis of data, such as Thurstone (1947), Horst (1965) and Harman (1967). In this case, if the Q random variables $(X_1,X_2,...,X_Q)$ point in the same direction, then we have $\sum_{i<j} cov(X_i,X_j)>0$; and, in particular, for $Q=2$, we have the evident result $cov(X_1,X_2)>0$. However, note that the condition, for two distinct variables $cov(X_i,X_j)>0$, is sufficient, but not necessary, for $\sum_{i<j} cov(X_i,X_j)>0$.

For $\alpha=1$,

$$\max\lambda_1(\varepsilon)=E|S(\varepsilon)| \quad \text{over} \quad \varepsilon\in\{-1,+1\}^{Q-1}$$

corresponds to the taxicab principal components approach, which will be used in this paper.

3. Taxicab Correspondence analysis: An overview

Let $\mathbf{X}=(x_{ij})$ be a contingency table cross-classifying two nominal variables with I rows and J columns, and $\mathbf{P}=\mathbf{X}/x_{**}$ be the associated correspondence matrix with elements p_{ij}, where

$x_{**}=\sum_{j=1}^{J}\sum_{i=1}^{I}x_{ij}$ is the sample size. We define as usual $p_{i*}=\sum_{j=1}^{J}p_{ij}$, $p_{*j}=\sum_{i=1}^{I}p_{ij}$, the vector $\mathbf{r}=(p_{i*})\in\mathbb{R}^{I}$, the vector $\mathbf{c}=(p_{*j})\in\mathbb{R}^{J}$, and $\mathbf{D}_r=Diag(\mathbf{r})$ a diagonal matrix having diagonal elements p_{i*}, and similarly $\mathbf{D}_c=Diag(\mathbf{c})$. Let $k=rank(\mathbf{P})-1$. TCA is computed in two steps: In the first step we compute the taxicab singular value decomposition (TSVD) of $\mathbf{P}$, which is a stepwise matrix decomposition method based on a particular matrix norm, see below equation (3); similar to SVD, TSVD produces an ordered sequence of basic coordinates $(\mathbf{a}_\alpha,\mathbf{b}_\alpha,\lambda_\alpha)\in\mathbb{R}^{I}\times\mathbb{R}^{J}\times\mathbb{R}^{+}$ for $\alpha=1,...,k$. In the second step, by reweighting the pair of basic coordinates $(\mathbf{a}_\alpha,\mathbf{b}_\alpha)$ by respective weights of the columns, $\mathbf{D}_r$, and the rows, $\mathbf{D}_c$, we obtain TCA.

3.1. Taxicab singular value decomposition

Let $\mathbf{P}^{(\alpha)}$ be the residual data matrix at the αth iteration, where, $\mathbf{P}^{(0)}=\mathbf{P}$ for $\alpha=0$. TSVD consists of maximizing the L_1 norm of the linear combination of the columns of the matrix $\mathbf{P}^{(\alpha)}$ subject to L_∞ norm constraint, where the L_1 norm of a vector $\mathbf{v}=(v_1,...,v_m)'$ is defined to be $||\mathbf{v}||_1=\sum_{i=1}^{m}|v_i|$ and $||\mathbf{v}||_\infty=\max_i|v_i|$ is the L_∞ norm; more precisely, it is based on the following optimization problem

$$max\left|\left|\mathbf{P}^{(\alpha)}\mathbf{u}\right|\right|_1 \text{ subject to } ||\mathbf{u}||_\infty=1; \tag{1}$$

or equivalently, it can also be described as maximization of the L_1 norm of the linear combination of the rows of the matrix $\mathbf{P}^{(\alpha)}$

$$max\left|\left|\mathbf{P}^{(\alpha)\prime}\mathbf{v}\right|\right|_1 \text{ subject to } ||\mathbf{v}||_\infty=1. \tag{2}$$

Equation (1) is the dual of (2), and they can be reexpressed as matrix operator norms

$$\begin{aligned}\lambda_\alpha &= \max_{\mathbf{u}\in\mathbb{R}^J}\frac{||\mathbf{P}^{(\alpha)}\mathbf{u}||_1}{||\mathbf{u}||_\infty},\\ &= \max_{\mathbf{v}\in\mathbb{R}^I}\frac{||\mathbf{P}^{(\alpha)\prime}\mathbf{v}||_1}{||\mathbf{v}||_\infty},\\ &= \max_{\mathbf{u}\in\mathbb{R}^J,\mathbf{v}\in\mathbb{R}^I}\frac{\mathbf{v}'\mathbf{P}^{(\alpha)}\mathbf{u}}{||\mathbf{u}||_\infty\,||\mathbf{v}||_\infty},\end{aligned} \tag{3}$$

which is a well known and much discussed matrix norm related to Grothendieck problem, see for instance, Alon and Naor (2006). The solution to (3), λ_α, is a combinatorial optimization problem given by

$$\max||\mathbf{P}^{(\alpha)}\mathbf{u}||_1 \text{ subject to } \mathbf{u}\in\{-1,+1\}^J. \tag{4}$$

Equation (4) characterizes the robustness of the method, in the sense that, the weights affected to the columns (similarly to the rows by duality) are uniform ±1. The αth principal axes, $\mathbf{u}_\alpha$ and $\mathbf{v}_\alpha$, are computed by

$$\mathbf{u}_\alpha=\arg\max_{\mathbf{u}}\left|\left|\mathbf{P}^{(\alpha)}\mathbf{u}\right|\right|_1 \text{ such that } ||\mathbf{u}||_\infty=1, \tag{5}$$

and

$$\mathbf{v}_\alpha=\arg\max_{\mathbf{v}}\left|\left|\mathbf{P}^{(\alpha)\prime}\mathbf{v}\right|\right|_1 \text{ such that } ||\mathbf{v}||_\infty=1. \tag{6}$$

Let $\mathbf{a}_\alpha$ represent the αth TSVD coordinates of the rows of $\mathbf{P}^{(\alpha)}$ by projecting the rows of $\mathbf{P}^{(\alpha)}$ on the principal axis $\mathbf{u}_\alpha$, and $\mathbf{b}_\alpha$ represent the αth TSVD coordinates of the columns of $\mathbf{P}^{(\alpha)}$ by projecting the columns of $\mathbf{P}^{(\alpha)}$ on the principal axis $\mathbf{v}_\alpha$. These are given by

$$\mathbf{a}_\alpha = \mathbf{P}^{(\alpha)}\mathbf{u}_\alpha \text{ and } \mathbf{b}_\alpha = \mathbf{P}^{(\alpha)\prime}\mathbf{v}_\alpha; \tag{7}$$

and

$$||\mathbf{a}_\alpha||_1 = \mathbf{v}_\alpha'\mathbf{a}_\alpha = ||\mathbf{b}_\alpha||_1 = \mathbf{u}_\alpha'\mathbf{b}_\alpha = \lambda_\alpha. \tag{8}$$

Equations (7) are named transition formulas, because $\mathbf{v}_\alpha$ and $\mathbf{a}_\alpha$, and , $\mathbf{u}_\alpha$ and $\mathbf{b}_\alpha$, are related by

$$\mathbf{u}_\alpha = sgn(\mathbf{b}_\alpha) \quad \text{and} \quad \mathbf{v}_\alpha = sgn(\mathbf{a}_\alpha), \tag{9}$$

where $sgn(\mathbf{b}_\alpha) = (sgn(b_\alpha(1)),...,sgn(b_\alpha(J))'$, and $sgn(b_\alpha(j)) = 1$ if $b_\alpha(j) > 0$, $sgn(b_\alpha(j)) = -1$ otherwise.

To obtain the $(\alpha+1)$th TSVD row and column coordinates $\mathbf{a}_{\alpha+1}$ and $\mathbf{b}_{\alpha+1}$, and corresponding principal axes $\mathbf{u}_{\alpha+1}$ and $\mathbf{v}_{\alpha+1}$, we repeat the above procedure on the residual dataset

$$\mathbf{P}^{(\alpha+1)} = \mathbf{P}^{(\alpha)} - \mathbf{a}_\alpha\mathbf{b}_\alpha'/\lambda_\alpha. \tag{10}$$

We note that the $rank(\mathbf{P}^{(\alpha+1)}) = rank(\mathbf{P}^{(\alpha)})-1$, because by (7), (8) and (9)

$$\mathbf{P}^{(\alpha+1)}\mathbf{u}_\alpha = \mathbf{0} \text{ and } \mathbf{P}^{(\alpha+1)\prime}\mathbf{v}_\alpha = \mathbf{0}; \tag{11}$$

which implies that

$$\mathbf{u}_{\alpha-1}'\mathbf{b}_\alpha = 0 \text{ and } \mathbf{v}_{\alpha-1}'\mathbf{a}_\alpha = 0 \text{ for } \alpha = 1,...,k. \tag{12}$$

From which one gets the data reconstitution formula for the correspondence matrix $\mathbf{P}$ as a function of the basic coordinates $(\mathbf{a}_\alpha, \mathbf{b}_\alpha)$ for $\alpha = 1,...,k$ associated with the dispersion measures λ_α

$$p_{ij} = p_{i.}p_{.j} + \sum_{\alpha=1}^{k} a_\alpha(i)b_\alpha(j)/\lambda_\alpha. \tag{13}$$

In TCA of $\mathbf{P}$ both basic vectors $\mathbf{a}_\alpha$ and $\mathbf{b}_\alpha$ for $\alpha = 1,...,k$ satisfy the equivariability property, see Choulakian (2008b). This means that $\mathbf{a}_\alpha$ and $\mathbf{b}_\alpha$ are balanced in the sense that

$$\begin{aligned} \frac{\lambda_\alpha}{2} &= \sum_i [a_\alpha(i)|a_\alpha(i) > 0] \\ &= -\sum_i [a_\alpha(i)|a_\alpha(i) < 0] \\ &= \sum_j [b_\alpha(j)|b_\alpha(j) > 0] \\ &= -\sum_j [b_\alpha(j)|b_\alpha(j) < 0]; \end{aligned} \tag{14}$$

this easily follows from (9) and (12).

In TSVD, the optimization problems (3), (5) or (6) can be accomplished by two algorithms. The first one is based on complete enumeration (4); this can be applied, with the present state of desktop computing power, say, when $min(I,J) \simeq 25$. The second one is based on iterating the transitional formulas (7), (8) and (9), similar to Wold (1966)'s NIPALS (nonlinear iterative partial alternating least squares) algorithm, also named criss-cross regression by Gabriel and Zamir (1979). It is easy to show that this is also an ascent algorithm. The criss-cross nonlinear algorithm can be summarized in the following way, where $\mathbf{b}$ is a starting value:

Step 1: $\mathbf{u} = sgn(\mathbf{b})$, $\mathbf{a} = \mathbf{P}^{(\alpha)}\mathbf{u}$ and $\lambda(\mathbf{a}) = ||\mathbf{a}||_1$;

Step 2: $\mathbf{v} = sgn(\mathbf{a})$, $\mathbf{b} = \mathbf{P}^{(\alpha)\prime}\mathbf{v}$ and $\lambda(\mathbf{b}) = ||\mathbf{b}||_1$;

Step 3: If $\lambda(\mathbf{b}) - \lambda(\mathbf{a}) > 0$, go to Step 1; otherwise, stop.

This is an ascent algorithm; that is, it increases the value of the objective function λ at each iteration. The convergence of the algorithm is superlinear (very fast, at most two or three iterations); however it could converge to a local maximum; so we restart the algorithm I times using each row of $\mathbf{P}^{(\alpha)}$ as a starting value. The iterative algorithm is statistically consistent in the sense that as the sample size increases there will be some observations in the direction of the principal axes, so the algorithm will find the optimal solution.

3.2. Taxicab correspondence analysis

A simple reweighting of the basic coordinates $(\mathbf{a}_\alpha, \mathbf{b}_\alpha)$ produces TCA factor scores

$$\mathbf{f}_\alpha = \mathbf{D}_r^{-1}\mathbf{a}_\alpha \text{ and } \mathbf{g}_\alpha = \mathbf{D}_c^{-1}\mathbf{b}_\alpha; \tag{15}$$

and (8) becomes

$$\mathbf{v}'_\alpha \mathbf{D}_r \mathbf{f}_\alpha = \mathbf{u}'_\alpha \mathbf{D}_c \mathbf{g}_\alpha = \lambda_\alpha. \tag{16}$$

One gets the data reconstitution formula both in TCA and CA for the correspondence matrix $\mathbf{P}$ as a function of the factor coordinates $(\mathbf{f}_\alpha, \mathbf{g}_\alpha)$ for $\alpha = 1, ..., k$ associated with the eigenvalues λ_α

$$p_{ij} = p_{i.}p_{.j}\left[1 + \sum_{\alpha=1}^{k} f_\alpha(i) g_\alpha(j)/\lambda_\alpha\right]. \tag{17}$$

The visual maps are obtained by plotting the points $(f_\alpha(i), f_\beta(i))$ for $i = 1, ..., I$ or $(g_\alpha(j), g_\beta(j))$ for $j = 1, ..., J$, for $\alpha \neq \beta$.

TCA does not admit a distance interpretation between profiles; there is no chi-square like distance in TCA. Fichet (2009) described it as a general scoring method. This paper shows that for ratings and rankings data sets TCA is related to the sum score.

More technical details about TCA and a deeper comparison between TCA and CA is done in Choulakian (2006b). Further results can be found in Choulakian et al. (2006), Choulakian (2008b), Choulakian (2008a), Choulakian and de Tibeiro (2013), Choulakian et al. (2013), Choulakian (2013) and Choulakian et al. (2014).

4. Main theoretical results

`4.1: Notation`

Let

$$\mathbf{P}_D = \frac{(\mathbf{Y} \mid \mathbf{1}_I\ \mathbf{m}\prime - \mathbf{Y})}{I\ m_*},$$

and

$$\mathbf{P}_{nega} = \frac{(\mathbf{Y} \mid (\mathbf{1}_I\ \mathbf{m}\prime - \mathbf{Y})\mathbf{1}_Q)}{I\ m_*},$$

be the correspondence matrices associated with $\mathbf{Y}_D$ and $\mathbf{Y}_{nega}$, where $m_* = \sum_{q=1}^{Q} \mathrm{m}_q = \mathbf{m}\prime\mathbf{1}_Q$.

We designate TCA residual correspondence matrices in block form by $\mathbf{P}_D^{(\alpha)} = (\mathbf{P}_{D1}^{(\alpha)}|\mathbf{P}_{D2}^{(\alpha)})$ and $\mathbf{P}_{nega}^{(\alpha)} = (\mathbf{P}_{nega1}^{(\alpha)}|\mathbf{p}_{nega}^{(\alpha)})$ for $\alpha = 0, ..., k$, where $k = rank(\mathbf{P}_D) - 1 = rank(\mathbf{P}_{nega}) - 1$, $\mathbf{P}_D^{(0)} = \mathbf{P}_D$ and $\mathbf{P}_{nega}^{(0)} = \mathbf{P}_{nega}$. Similarly we designate by $(\mathbf{f}_\alpha, \mathbf{g}_\alpha, \lambda_\alpha) \in \mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R}^+$ for $\alpha = 1, ..., k$ the ordered sequence of TCA factor scores and dispersion measures. In block notation, we write: $\mathbf{g}_\alpha^{nega} = (\mathbf{g}_{1\alpha}^{nega\prime} \mid g_{2\alpha}^{nega})'$ and $\mathbf{g}_\alpha^D = (\mathbf{g}_{1\alpha}^{D}\ ' \mid \mathbf{g}_{2\alpha}^{D}\ ')'$. As usual $y_{i*} = \sum_{q=1}^{Q} y_{iq}$ designates the sum score of the ith row, and $y_{tot} = \sum_{i=1}^{I} y_{i*}$.

The transposed vector of column masses in $\mathbf{P}_{nega}$ is

$$\begin{aligned} \mathbf{c}'_{nega} &= (\mathbf{1}'_I\ \mathbf{P}_{nega}) \\ &= (\frac{\mathbf{1}'_I\ \mathbf{Y}}{I\ m_*} \mid 1 - \frac{y_{tot}}{I\ m_*}) \\ &= (\mathbf{c}'_1 \mid c_2), \end{aligned} \tag{18}$$

and the transposed vector of column masses in $\mathbf{P}_D$ is

$$\begin{aligned} \mathbf{c}'_D &= (\mathbf{1}'_I\ \mathbf{P}_D) \\ &= (\mathbf{1}'_I\ \mathbf{Y} \mid I\ \mathbf{m}\prime - \mathbf{1}'_I\ \mathbf{Y})/(I\ m_*) \\ &= (\mathbf{c}'_{D1}|\mathbf{c}'_{D2}). \end{aligned} \tag{19}$$

So the metric matrix defined on the rows of $\mathbf{P}_D$ is

$$\begin{aligned} \mathbf{D}_c^D &= Diag(\mathbf{c}_D) \\ &= \begin{pmatrix} \mathbf{D}_{c1}| & \mathbf{0} \\ \mathbf{0} & |\ \mathbf{D}_{c2} \end{pmatrix}, \end{aligned} \tag{20}$$

and the metric matrix defined on the rows of $\mathbf{P}_{nega}$ is

$$\begin{aligned} \mathbf{D}_c^{nega} &= Diag(\mathbf{c}_{nega}) \\ &= \begin{pmatrix} \mathbf{D}_{c1}| & \mathbf{0} \\ \mathbf{0} & |\ c_2 \end{pmatrix}. \end{aligned} \tag{21}$$

Note that the submatrix $\mathbf{D}_{c1}$ is common to both $\mathbf{D}_c^D$ in (20) and $\mathbf{D}_c^{nega}$ in (21).

In both datasets $\mathbf{P}_D$ and $\mathbf{P}_{nega}$ the vector of row masses is

$$\mathbf{r} = \mathbf{1}_I/I, \tag{22}$$

so the metric matrix defined on the columns is

$$\mathbf{D}_r = \mathbf{I}_I/I \tag{23}$$

where $\mathbf{I}_I$ is the identity matrix of size I.

4.2 Main results

The following lemma is fundamental

Lemma 1: $\mathbf{P}_{D2}^{(\alpha)} = -\mathbf{P}_{D1}^{(\alpha)}$ for $\alpha = 1,...,k$. Further, this result implies that the law of lever or the lever principle applies; that is

$$\mathbf{D}_{c1}\mathbf{g}_{1\alpha}^{D} + \mathbf{D}_{c2}\mathbf{g}_{2\alpha}^{D} = \mathbf{0} \quad \text{for} \ \ \alpha = 1,...,k,$$

where $\mathbf{g}_{\alpha}^{D} = \binom{\mathbf{g}_{1\alpha}^{D}}{\mathbf{g}_{2\alpha}^{D}}$ contains the αth factor scores of the column categories. Or in terms of basic coordinates (15)

$$\mathbf{b}_{1\alpha}^{D} + \mathbf{b}_{2\alpha}^{D} = \mathbf{0} \quad \text{for} \ \ \alpha = 1,...,k,$$

where $\mathbf{b}_{\alpha}^{D} = \binom{\mathbf{b}_{1\alpha}^{D}}{\mathbf{b}_{2\alpha}^{D}}$ contains the αth basic vector coordinates of the column categories.

The proofs of all new results can be found in the appendix. Note that the lever principle is well known in CA, see for instance, Greenacre (1984), page 175. We have the following three theorems.

Theorem 1 ($\mathbf{P}_D$): The first principal row factor score f_{i1}^{D} is an affine function of the sum score y_{i*} iff $\mathbf{u}_1^D = (\mathbf{1}_Q \mid -\mathbf{1}_Q)$, where

$$f_{i1}^{D} = \frac{2}{m_*}(y_{i*} - \frac{y_{tot}}{I}) \quad \text{or} \quad corr(\mathbf{f}_1^{D}, \mathbf{s}) = 1,$$

where $\mathbf{s} = (y_{i*})$. Note that the way $\mathbf{u}_1^D = (\mathbf{1}_Q \mid -\mathbf{1}_Q)$ is defined in Theorem 1 eliminates the sign indeterminacy of the 1st taxicab principal axis.

Corollary 1: The first nontrivial TCA dispersion measure equals

$$\lambda_1^D = \frac{2}{I\, m_*}\sum_{i=1}^{I} |(y_{i*} - \frac{y_{tot}}{I})|.$$

If a data set satisfies Theorem 1, then $\mathbf{u}_{11}^D = \mathbf{1}_Q$; that is, the weight of each item $q+$ is 1. From which we have

Definition 2: If a dataset satisfies Theorem 1, then the Q variables strongly point in the same direction, or they are strongly coherent or strongly consistent.

Theorem 2 ($\mathbf{P}_{nega}$): Properties a, b, c are true iff $\mathbf{u}_1^{nega} = (\mathbf{1}_Q \mid -1)$.
a) The first principal row factor score f_{i1}^{nega} is an affine function of the sum score y_{i*}; that is,

$$f_{i1}^{nega} = \frac{2}{m_*}(y_{i*} - \frac{y_{tot}}{I}) \quad \text{or} \quad corr(\mathbf{f}_1^{nega}, \mathbf{s}) = 1.$$

b) The first nontrivial TCA dispersion measure equals twice taxicab norm of the vector $\mathbf{p}_{nega}^{(1)}$

$$\begin{aligned}\lambda_1^{nega} &= 2\,||\mathbf{p}_{nega}^{(1)}||_1, \\ &= \frac{2}{I\,m_*}\sum_{i=1}^{I}|(y_{i*}-\frac{y_{tot}}{I})|.\end{aligned}$$

c) The nega column $\mathbf{p}_{nega}^{(2)}$ of the residual matrix $\mathbf{P}_{nega}^{(2)}$ is zero

$$\mathbf{P}_{nega}^{(2)} = (\mathbf{P}_{nega1}^{(2)}|\mathbf{0}).$$

Property b shows that the nega column of $\mathbf{P}_{nega}^{(1)}$ accounts for 50% of the first nontrivial taxicab dispersion λ_1^{nega}. Property c shows that the residual matrix $\mathbf{P}_{nega}^{(2)}$ does not contain any information on the heavyweight nega column. Properties b and c imply that the first nontrivial factor is completely determined by the nega column, which plays a dominant heavyweight role, see Choulakian (2008b). Such a context in CA is discussed by Benzécri (1979) using asymptotic theory, and in dual scaling by Nishisato (1984), who named it forced classification.

We note that if a data set satisfies Theorem 2, then $\mathbf{u}_{11}^{nega} = \mathbf{1}_Q$; that is, the weight of each item is 1. From which we have

Definition 3: If a data set satisfies Theorem 2, then the Q variables weakly point in the same direction, or they are weakly coherent or weakly consistent.

It is evident that: If Q items are strongly coherent, then they are weakly coherent. In section 6, we present a real data set, which satisfies Theorem 2 but not Theorem 1.

Theorem 3 (equivalence): TCA of $\mathbf{Y}_D$ is equivalent to TCA of $\mathbf{Y}_{nega}$ iff $\mathbf{u}_{11}^{D} = \mathbf{u}_{11}^{nega} = \mathbf{1}_Q$; where equivalence means

a1) $\mathbf{f}_1^D = \mathbf{f}_1^{nega} = \mathbf{f}_1 = (f_{i1})$, where $f_{i1} = \frac{2}{m_*}(y_{i*} - \frac{y_{tot}}{I})$ or $corr(\mathbf{f}_1, \mathbf{s}) = 1$.
a2) $\lambda_1^{nega} = \lambda_1^D = \frac{2}{I\,m_*}\sum_{i=1}^{I}|(y_{i*} - \frac{y_{tot}}{I})|$.
a3) $\mathbf{g}_{11}^D = \mathbf{g}_{11}^{nega}$ and $\mathbf{1}_Q'\,\mathbf{D}_{c2}\,\mathbf{g}_{21}^D = c_2\,g_{21}^{nega}$.
b) For $\alpha \geq 2$, $\mathbf{P}_{D1}^{(\alpha)} = \mathbf{P}_{nega1}^{(\alpha)} = -\mathbf{P}_{D2}^{(\alpha)}$ and $\mathbf{p}_{nega}^{(\alpha)} = \mathbf{0}$.
c1) For $\alpha \geq 2$, $\mathbf{f}_\alpha^D = 2\mathbf{f}_\alpha^{nega}$.
c2) For $\alpha \geq 2$, $\lambda_\alpha^D = 2\lambda_\alpha^{nega}$.
c3) For $\alpha \geq 2$, $\mathbf{g}_{1\alpha}^D = \mathbf{g}_{1\alpha}^{nega}$ and $g_{2\alpha}^{nega} = 0$.

`4.3: TCA of a rankings dataset`

Let $\mathbf{R} = (r_{ij})$ for $i = 1,...,I$ and $j = 1,...,J$ represent a rankings data set, where J objects have been ranked by I individuals. In voting theory, J represents the number of candidates, I the number of voters, and r_{ij} is the preference ranking, usually taking values $0,...,J-1$, provided by the jth voter to the ith candidate. The well known Borda count (BC) ranking of candidates or the products is obtained by summing over the rows of $\mathbf{R}$, see for instance, Saari (1990) for some optimal properties of BC. But it is well known that the BC ranking is influenced by strategic voting, which in our case corresponds to the rankings provided by the rogue individuals; Borda names the rogue voters "dishonest": a minority of the individuals that do not rank in the same direction as the majority. Following, Torres and Greenacre (2002), we can apply TCA to the row doubled $\mathbf{R}_D$ table of size $(2I) \times J$, and of course similarly to the $\mathbf{R}_{nega}$ table of size $(I+1) \times J$. Theorems 1, 2 and 3 are also valid for TCA of $\mathbf{R}_D$ and $\mathbf{R}_{nega}$ where the roles of the columns and rows are interchanged; for instance, property a) in the 3 theorems now concerns the 1st column factor score: g_{j1} is an affine function of the jth column sum r_{*j}. So, in the case of a rankings data set, the proposed approach corresponds to the Borda count rule if all the voters point in the same direction; or modified Borda count (MBC) rule, because it applies the Borda's rule for preference ranking by using the largest number of consistent voters chosen by TCA. This is similar in spirit to Johnson (1983)'s approach.

5. The law of contradiction

Let $S+$ be a statement and and $S-$ its negation; then the law of contradiction states that $S+$ and $S-$ oppose each other: they can not both hold together, see Eves (1997). We shall use the law of contradiction as a basis for the interpretation of the maps produced by CA and TCA of $\mathbf{Y}_{nega}$ in the following way. First, we recall that there are Q items, and we represented the positive pole of an item q, for $q = 1,...Q$, by $q+$ and its negative or complementary pole by $q-$. By the law of contradiction, $q+$ and $q-$ oppose each other. Which in its turn also implies that $q+$ and $nega = \cup_{q=1}^{Q} q_-$ oppose each other or they are not associated at all, because the *nega* contains $q-$. For the interpretation of the results by CA and TCA of $\mathbf{Y}_{nega}$ we can have the following two complementary scenarios:

Scenario one happens when

$$g_{11}(q+) \geq 0 \text{ for all } q \text{ and } \quad g_1(nega) < 0; \tag{Scen1}$$

then by the law of contradiction, the first principal dimension is interpretable and it shows the opposition between the positive poles of the items to their negative poles. Conditional on Scen1, the principal factor coordinates of the higher dimensions can take two forms:

Form one given by

$$g_\alpha(nega) < 0 \quad \text{and} \quad g_{1\alpha}(q+) < 0 \text{ for some } q \quad \text{and} \quad \alpha \geq 2; \tag{Scen1F1}$$

which shows that for some items the positive and negative poles are positively associated; this contradicts the law of contradiction, which implies that, for $\alpha \geq 2$ the dimensions are not interpretable. But this, Scen1F1, happens only for CA of $\mathbf{Y}_{nega}$; for an example see subsection 6.1. For TCA of $\mathbf{Y}_{nega}$ the condition Scen1F1 never happens, because Theorem 2 states that if the condition Scen1 holds, then we will have Form two given by

$$g_2(nega) = 0 \quad \text{and} \quad g_{1\alpha}(q+) < 0 \text{ for some } q \quad \text{and} \quad \alpha \geq 2; \tag{Scen1F2}$$

which states that the *nega* column is completely eliminated from the rest of the analysis.

Scenario two happens when

$$g_1(nega) < 0 \quad \text{and} \quad g_{1\alpha}(q+) < 0 \text{ for some } q \quad \text{and} \quad \alpha \geq 1; \tag{Scen2}$$

then for both methods CA and TCA of $\mathbf{Y}_{nega}$, the results are not interpretable by the law of contradiction.

The above discussion shows that the nega coding is perfectly suited for TCA; but it is very restrictive for CA, because at most the first dimension will be interpretable. Indeed, Adames (1993) interpreted only the first dimension of CA of the Foie Canard dataset; in the next section, we will be able to interpret the first two dimensions of TCA of the Foie Canard dataset. In the next section, we consider only TCA of $\mathbf{Y}_{nega}$.

6. Examples

We present two examples of different kinds.

6.1. Example 1: Foie de Canard dataset

6.1.1: Introduction

Table 1 presents a three-way data set of ratings, $\mathbf{X} = (x_{qji})$ for $q = 1,,,,4$, $j = 1,...,5$ and $i = 1,...,10$, where five judges have rated ten foie de canard (duck liver) products on four items *general appearance (A), cut look (C), odour (O)* and *taste (T).* The data are bipolar ranging from a negative extreme score of minimum value of 0 to a positive extreme score of m_q for $q = 1,...,4$, where $m_1 = 20$, $m_2 = 40$, $m_3 = 40$ and $m_4 = 150$, obviously taste being the most important. Originally, a weighted sum score statistic based on Table 3, where each judge had uniform weight, was used to order the products from the best to the worst

$$f4 \succeq f8 \succeq f6 \succeq f3 \succeq f7 \succeq f10 \succeq f1 \succeq f9 \succeq f2 \succeq f5.$$

However, Lavialle et al. (1990) provided five classifications of the ten products using different optimization criteria based on pairwise comparisons with thresholds; these are grouped in two classes. The first class is composed of two preference orderings

$$f8 \succeq f6 \succeq f4 \succeq f10 \succeq f7 \succeq f1 \succeq f3 \succeq f9 \succeq f2 \succeq f5; \tag{LQV1}$$

$$f6 \succeq f8 \succeq f4 \succeq f10 \succeq f7 \succeq f1 \succeq f3 \succeq f9 \succeq f2 \succeq f5, \tag{LQV2}$$

where the difference between the two orderings LQV1 and LQV2 is that the first two products have interchanged their positions. The second class, composed of three preference orderings, is based on LQV4

$$f4 \succeq f6 \succeq f8 \succeq f10 \succeq f7 \succeq f1 \succeq f3 \succeq f9 \succeq f2 \succeq f5, \tag{LQV3}$$

$$f4 \succeq f6 \succeq f8 \succeq f7 \succeq f10 \succeq f1 \succeq f3 \succeq f9 \succeq f2 \succeq f5, \tag{LQV4}$$

TABLE 1. *Foie de Canard ratings data set. Sscore − is the sum score without the ratings A3 and T3.*

Item	Product									
Judge	*f1*	*f2*	*f3*	*f4*	*f5*	*f6*	*f7*	*f8*	*f9*	*f10*
A1	20	10	10	20	10	20	10	20	10	10
C1	20	20	30	40	10	40	20	30	10	20
O1	40	20	30	40	10	40	30	30	20	20
T1	120	60	120	130	100	120	100	100	90	100
A2	10	20	0	20	0	15	20	10	0	20
C2	20	30	20	40	5	10	15	30	0	10
O2	15	20	30	30	5	30	40	5	40	20
T2	50	30	70	80	10	50	70	40	50	40
A3	15	10	5	10	10	15	5	15	5	15
C3	25	20	20	30	20	35	10	25	20	25
O3	20	15	30	15	15	20	15	30	30	20
T3	80	50	110	70	90	30	70	60	110	85
A4	5	15	5	15	5	15	10	15	0	10
C4	10	15	30	30	30	20	20	35	10	10
O4	0	10	10	30	0	30	30	35	10	0
T4	50	80	20	50	0	80	50	100	50	50
A5	5	5	10	15	0	10	15	15	10	15
C5	10	10	20	20	0	20	15	25	10	20
O5	20	10	15	20	0	30	20	25	20	20
T5	40	40	65	50	40	80	70	70	60	85
Sscore−	480	430	535	675	260	665	560	640	440	495

$$f4 \succeq f6 \succeq f8 \succeq f7 \succeq f10 \succeq f3 \succeq f1 \succeq f9 \succeq f2 \succeq f5, \qquad \text{(LQV5)}$$

because only two products permute their positions in (LQV3 and LQV4) and in (LQV4 and LQV5).

6.1.2: TCA

By TCA, we shall discuss the following 3 points concerning this data set:

First, using two different codings provide preference orderings of the products based on Theorems 1, 2 and 3.

Second, reveal the existence of response styles or personal equations.

Third, see whether the judges are individually coherent (consistent) or not.

Let $\mathbf{Y} = (y_{i(qj)} = x_{qji}) \in \mathbb{R}^{10\times 20}$ for $q = 1,,,,4$, $j = 1,...,5$ and $i = 1,...,10$ be the flattened form of the three-way dataset $\mathbf{X}$; to be precise, $\mathbf{Y}$ corresponds to the transpose of the data set of ratings in Table 1. We shall consider only the analysis of $\mathbf{Y}_{nega}$. Figure 1 displays the principal map obtained by TCA, where we see that two items by Judge 3, T3 and A3, are associated with the nega column: By the law of contradiction, this is senseless, absurd and not interpretable, because the opposing poles $T3+$ and $T3-$ will be associated with each other, given that the nega column contains the negative pole $T3-$. Further, we have the following Pearson correlation values, and for completeness we also present the corresponding values for CA:

$$Corr(\mathbf{f}_1^{TCA}, sum\ score) = 0.9851 \quad \text{and} \quad Corr(\mathbf{f}_1^{CA}, sum\ score) = 0.9497.$$

We repeat the analysis of the data set by deleting only item T3; this produces

$$Corr(\mathbf{f}_1^{TCA}, sum\ score) = 0.9994 \quad \text{and} \quad Corr(\mathbf{f}_1^{CA}, sum\ score) = 0.9903.$$

By deleting both items T3 and A3 from the analysis we get a perfect TCA correlation:

$$Corr(\mathbf{f}_1^{TCA}, sum\ score) = 1 \quad \text{and} \quad Corr(\mathbf{f}_1^{CA}, sum\ score) = 0.9903;$$

consequently using the sum score statistic, Sscore-, shown in the last row of Table 1, we get the following preference ordering of the products grouped into four classes according to the Sscore-values

$$(f4 \succeq f6 \succeq f8) \succeq (f7 \succeq f3) \succeq (f10 \succeq f1 \succeq f9 \succeq f2) \succeq f5. \qquad \text{(TCA1)}$$

The ordering is very similar to the ordering LQV5, where only two products, f10 and f3, have interchanged their positions.

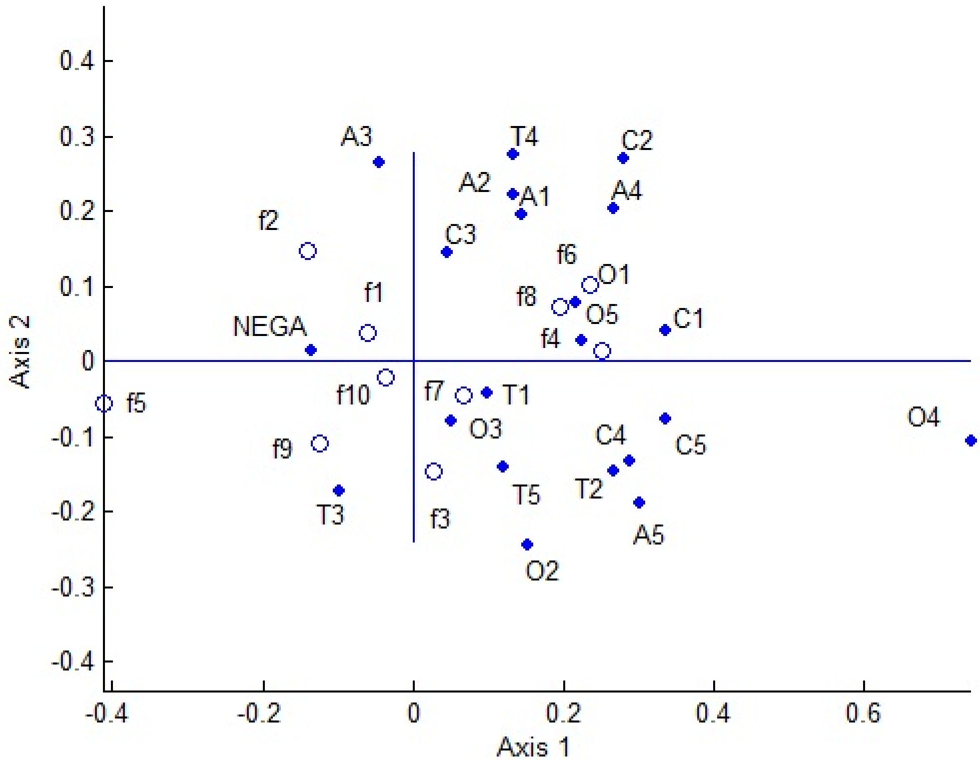

FIGURE 1. *TCA biplot of Foie Gras*

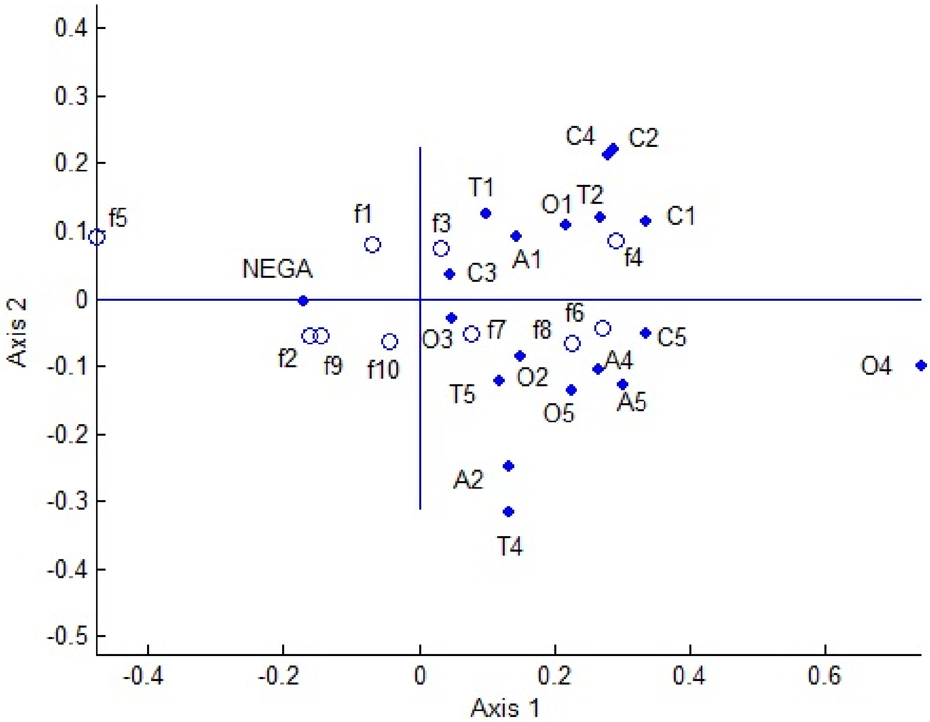

FIGURE 2. *TCA biplot of Foie Gras without A3 and T3*

Figure 2 displays the principal map obtained from TCA of $\mathbf{Y}_{nega-}$, where $-$ designates the fact that items T3 and A3 are completely eliminated from the $\mathbf{Y}$ dataset. The first factor represents the sum score statistic, and the above preference ordering, TCA1, can be seen in Figure 2.

The second TCA factor is also interpretable and brings further insight on the judges in the following way: The points (T5, O5, C5 and A5), which represent Judge 5, are found in the fourth quadrant of Figure 2; while the points (T1, O1, C1 and A1), which represent Judge 1, are found in the first quadrant of Figure 2. So the second factor opposes Judge 5 to Judge 1. What does this mean? Let $y_{iqj} = x_{(qj)i}$ for $i = 1,...,10,\ q = 1,...,4$ and $j = 1,...,5$ represent the score of judge j of product i on item q. We note that in Table 1, $x_{(q1)i} \geq x_{(q5)i}$ for all $i = 1,...,10,\ q = 1,...,4$ except for $(q,i) = (1,7)$ and $(1,10)$; that is, 38 out of 40 ratings of Judge 1 are superior or equal to the corresponding ratings of Judge 5. This means that the Judges 1 and 5 have different response styles, or different personal equations. Judge 1 has acquiscence response style-he is liberal and provides high scores; while Judge 5 has disacquiscence response style-he is conservative and provides low scores. This provides an answer to the second question.

Table 2 displays TCA dispersion measures, where we see that TCA of the coded Foie de Canard complete data sets $\mathbf{Y}_D$ and $\mathbf{Y}_{nega}$ are not equivalent, because $\lambda_1^D = 0.1763 > \lambda_1^{nega} = 0.1552$; while TCA of the coded Foie de Canard data sets without (T3, A3), $\mathbf{Y}_{D-}$ and $\mathbf{Y}_{nega-}$, are equivalent, because properties a2 and c2 of Theorem 2 are satisfied; in particular $\lambda_1^{D-} = \lambda_1^{nega-} = 0.1796$.

http://www.sfds.asso.fr/journal

TABLE 2. *Dispersion measures of Foie de Canard data set.*

	Whole data set		Data set without **(T3, A3)**	
	TCA of $\mathbf{Y}_D$	TCA of $\mathbf{Y}_{nega}$	TCA of $\mathbf{Y}_{D-}$	TCA of $\mathbf{Y}_{nega-}$
α	λ_α	λ_α	λ_α	λ_α
1	0.1763	0.1552	0.1796	0.1796
2	0.1187	0.0752	0.1334	0.0667
3	0.1108	0.0631	0.1268	0.0634
4	0.0908	0.0523	0.0878	0.0439
5	0.0807	0.0461	0.0824	0.0412
6	0.0741	0.0381	0.0800	0.0400
7	0.0580	0.0330	0.0526	0.0263
8	0.0336	0.0167	0.0306	0.0153
9	0.0222	0.0113	0.0178	0.0089

Now we discuss the third point concerning the coherence or the consistency of the judges' ratings by considering the five subtables $\mathbf{Y}_j = (y_{iq(j)})$ of size 10×4 separately. Based on Theorems 1, 2 and 3, the calculations show that:

Judges 1, 4 and 5 are strongly consistent, because

$$Corr(\mathbf{f}_1^D, sum\ score) = Corr(\mathbf{f}_1^{nega}, sum\ score) = 1.$$

Judge 2 is weakly consistent, because

$$Corr(\mathbf{f}_1^D, sum\ score) = 0.449, \text{ but } Corr(\mathbf{f}_1^{nega}, sum\ score) = 1.$$

Judge 3 is inconsistent, because

$$Corr(\mathbf{f}_1^D, sum\ score) = 0.935 \text{ and } Corr(\mathbf{f}_1^{nega}, sum\ score) = 0.809.$$

6.3 TCA of rankings via Nishisato's mapping of the ratings

Table 3 shows the total ratings given by the five judges to the ten products. We note that the sums are quite different: The total score of Judge 1 is 1570 which is around 1.5 times larger than the total scores provided individually by the Judges 2, 4 and 5. This means that the results obtained in the previous subsection do not weigh uniformly the judges. Nishisato (1980)'s mapping of the data transforms the initial ratings into rankings displayed in Table 4. This mapping is in the same spirit as Benzécri's personal equation mapping. In particular note that in Table 3, the ratings of Judge 1 for the ten products are strictly superior to the corresponding ratings of Judges 2 and 5; but this fact is no more true in Table 4. This means that the bias in the personal equations have been reduced or completely eliminated.

TABLE 3. *Foie de Canard marginal ratings.*

judge	Product										Sum
	f1	*f2*	*f3*	*f4*	*f5*	*f6*	*f7*	*f8*	*f9*	*f10*	
1	200	110	190	230	130	220	160	180	130	150	1570
2	95	100	120	170	20	105	145	85	90	90	1020
3	140	95	165	125	135	100	100	130	165	145	1300
4	65	120	65	125	35	145	110	185	70	70	990
5	75	65	110	105	40	140	120	135	100	140	1030

TABLE 4. *Foie de Canard rankings.*

judge	Product										Sum
	f1	*f2*	*f3*	*f4*	*f5*	*f6*	*f7*	*f8*	*f9*	*f10*	
1	7	0	6	9	1.5	8	4	5	1.5	3	45
2	4	5	7	9	0	6	8	1	2.5	2.5	45
3	6	0	8.5	3	5	1.5	1.5	4	8.5	7	45
4	1.5	6	1.5	7	0	8	5	9	3.5	3.5	45
5	2	1	5	4	0	8.5	6	7	3	8.5	45
BC	20.5	12	28	32	6.5	32	24.5	26	19	24.5	
MBC	14.5	12	19.5	29	1.5	30.5	23	22	10.5	17.5	

In Table 4 each judge has identical weight. Let $\mathbf{R} = (r_{ij})$ for $i = 1,...,5$ and $j = 1,...,10$ represent the dataset of rankings in Table 4. The well known Borda count (BC) ranking of the products, displayed in Table 4, is obtained by summing over the columns of $\mathbf{R}$, see for instance, Saari (1990). But it is well known that the BC ranking is influenced by strategic voting, which in our case corresponds to the ratings provided by the rogue judges: a minority of the judges that do not rate in the same direction as the majority of the judges. Following, Torres and Greenacre (2002), we can apply TCA to the row doubled $\mathbf{R}_D$ table of size 10×10, and of course similarly to the $\mathbf{R}_{nega}$ table of size 6×10. Theorems 1, 2 and 3 are also valid for TCA of $\mathbf{R}_D$ and $\mathbf{R}_{nega}$ where the roles of the columns and rows are interchanged. TCA of $\mathbf{R}_D$ and $\mathbf{R}_{nega}$ reveal that Judge 3 is deviant; so, we eliminate Judge 3, and denote the resulting data set of size 4×10 by $\mathbf{R}_-$. We find that TCA of $\mathbf{R}_{D-}$ and $\mathbf{R}_{nega-}$ are equivalent; so we can use its column sums to rank the products. We name these column sums modified Borda count (MBC) displayed in Table 4, because it applies the Borda's rule for preference ranking by using the largest number of consistent voters chosen by TCA. Using MBC we obtain the following robust ranking of the products:

$$(f6 \succeq f4) \succeq (f7 \succeq f8) \succeq (f3 \succeq f10) \succeq (f1 \succeq f2 \succeq f9) \succeq f5. \qquad \text{(Robust)}$$

Among the seven preference rankings of the products displayed above, we prefer the last one because of its robustness. Comparing the seven preference rankings listed above, we conclude that the products, $f4$ and $f6$, are the best, and the product $f5$ is the worst.

6.2. *Example 2: Sum score in IRT*

Evaluating student achievement is an important task in education, and sum score of items scores play an important role in estimating a student's position on the latent trait ability scale, denoted by θ. There is a clear distinction between the observed sum score statistic and the unobservable latent trait θ, in parametric item response theory (IRT), such as the Rasch (1PL) and the two parameter logistic (2PL) models, and in nonparametric IRT, such as the Mokken model; see for instance Meijer et al. (1990). Three general assumptions of modern IRT, known under the name of monotone homogeneity models, are local independence of the items, monotonicity of item response functions, and unidimensionality of the latent trait, refer to Sijtsma and Junker (2006) for a comprehensive review. If our aim, say, is only to order the students according to their ability, no such assumptions are needed in TCA approach proposed in this paper: If the items point in the same direction, then the sum score is the latent ability θ. Here, we consider three binary datasets two of them simulated and one real. The two simulated datasets, each of size 1000×100, are from Pr. C. Anderson's website http: //faculty. ed. uiuc. edu/ cja/ lma_as_irt/ index. html: The first is generated from the Rash model, the second from the 2PL model. The third is the Fractions dataset found at *http: //www. blackwell publishers. co. uk/ rss/,* and was used by Tatsuoka (2002); it consists of correct-incorrect responses of 536 fifth grade students to 20 fraction addition and subtraction problems; for instance item 11 is: $4\frac{1}{3} - 2\frac{4}{3}$. For the three data sets, we find

$$Corr(\mathbf{f}_1^D, sum\ score) = Corr(\mathbf{f}_1^{nega}, sum\ score) = 1,$$

which shows that the sum score is the unobservable latent variable in the three datasets, and all the items in each dataset point in the same direction. Note that for random data

$$Pr\{Corr(\mathbf{f}_1^{nega}, sum\ score) = 1\} = \frac{1}{2^{Q-1}}.$$

7. Conclusion

Cazes (2011) described Data Analysis as an Experimental Science the way it was practiced at Benzécri's Laboratory in Paris during the seventies and eighties: All kinds of data were analyzed using different kinds of codings; and it is remarkably unique in its kind that almost all the activities were reported in Benzécri's journal *Les Cahiers de L'Analyse Des Données*. The nega coding was one of these codings experimented twice, and which did not take off in CA. The law of contradiction showed that at most the first dimension of CA of $\mathbf{Y}_{nega}$ is interpretable. This paper shows that the nega coding is perfectly fit for TCA.

TCA of $\mathbf{Y}_{nega}$ is completely related to the sum score statistic; while TCA of $\mathbf{Y}_D$ is related to the sum score statistic if and only if it is equivalent to TCA of $\mathbf{Y}_{nega}$. In this case it is preferable to use the coding $\mathbf{Y}_{nega}$, because the TCA maps will be less cluttered. We can rephrase this in the following way: If there are rogue items, then the maps or biplots obtained by CA or TCA of $\mathbf{Y}_{nega}$ are useless and senseless; while the maps obtained by CA or TCA of $\mathbf{Y}_D$ can be useful and meaningful. Moreover, for a rankings dataset the proposed method corresponds to the famous Borda's count rule or to its modified version

Further, the following conjecture seems to be true: $\lambda_\alpha^D(TCA) \geq \lambda_\alpha^{nega}(TCA)$ for $\alpha \geq 1$. As mentioned in the introduction, it is true for $\lambda_\alpha^D(CA) \geq \lambda_\alpha^{nega}(CA)$ for $\alpha \geq 1$.

7.1. Acknowleldgements:

Choulakian's research is financed by NSERC of Canada. The author thanks the editor Pr. G. Celeux, the associate editor, two anonymous reviewers and Pr. D. R. Cox for their constructive comments.

References

Adames, G. (1993). Duck liver and tradition: analysis of ratings of a competition. *Les Cahiers de L'Analyse des Données*, XVIII:389–398.

Alon, N. and Naor, A. (2006). Approximating the cut-norm via grothendieck's inequality. *SIAM Journal on Computing*, 35:787–803.

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

Benzécri, J. (1979). On the analysis of a table with one heavyweight column (in french). *Les Cahiers de L'Analyse des Données*, IV:413–416.

Benzécri, J. (1992). *Correspondance Analysis Handbook*. N.Y:Marcel Dekker.

Benzécri, J. and Benzécri, F. (1989). Codage linéaire par morceaux et équation personnelle. *Les Cahiers de L'Analyse des Données*, XIV:203–210.

Burt, C. (1917). *The Distribution and Relations of Educational Abilities.* London:P.S. King and Son.

Cazes, P. (1990). Codage d'une variable continue en vue de l'analyse des correspondances. *Revue de Statistique Appliquée*, 38(3):35–51.

Cazes, P. (2011). Some comments on correspondence analysis. *www.youtube.com/watch?v=cisfaltVBTI*.

Choulakian, V. (2003). The optimality of the centroid method. *Psychometrika*, 68:473–475.

Choulakian, V. (2005). Transposition invariant principal component analysis in l1 for long tailed data. *Statistics and Probability Letters*, 71:23–31.

Choulakian, V. (2006a). L1 norm projection pursuit principal component analysis. *Computational Statistics and Data Analysis*, 50:1441–1451.

Choulakian, V. (2006b). 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*, 73:309–319.

Choulakian, V. (2013). *Advances in Latent Variables*, chapter The simple sum score statistic in taxicab correspondence analysis, page 6. Vita e Pensiero, Milan, Italy.

Choulakian, V., Allard, J., and Simonetti, B. (2013). Multiple taxicab correspondence analysis of a survey related to health services. *Journal of Data Science*, 11(2):205–229.

Choulakian, V. and de Tibeiro, J. (2013). Graph partitionong by correspondence analysis and taxicab correspondence analysis. *Journal of Classification*, 30:397–427.

Choulakian, V., Kasparian, S., Miyake, M., Akama, H., Makoshi, N., and Nakagawa, M. (2006). A statistical analysis of synoptic gospels. *JADT'2006*, pages 281–288.

Choulakian, V., Simonetti, B., and Gia, T. (2014). Some new aspects of taxicab correspondence analysis. *Statistical Methods and Applications*, 23:401–406.

Cox, D. (2006). In praise of the simple sum score. *www.stat.unpg.it/forcina/shlav/.../Cox2.pdf*.

Cox, D. (2008). On an internal method for deriving a summary measure. *Biometrika*, 95:1002–1005.

Cox, D. and Wermuth, N. (2002). On some models for binary variables parallel in compexity with the multivariate gaussian distribution. *Biometrika*, 89:462–469.

de Borda, J. (1781). Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 102:657–665.

Deniau, C., Oppenheim, G., and Benzécri, J. (1979). An effect of the refining of a partition on the eigenvalues arising from a correspondence table (in french). *Les Cahiers de l'Analyse des Données*, IV(3):289–297.

Esmieu, D., Gopalan, T., and Maiti, G. (1993). On the use of ratings in marketing studies for the introduction of a new product (in french). *Les Cahiers de L'Analyse des Données*, XVIII:399–426.

Eves, H. (1997). *Foundations and Fundamental Concepts of Mathematics*. N.Y. : Dover.

Fichet, B. (2009). Metrics of lp-type and distributional equivalence principle. *Advances in Data Analaysis and Classification*, 3:305–314.

Gabriel, K. and Zamir, S. (1979). Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*, 21:489–498.

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. N.Y:Wiley.

Greenacre, M. (1984). *Theory and Applications of Correspondence Analysis.* London:Academic Press.

Harman, H. (1967). *Modern Factor Analysis*. Chicago:The University of Chicago Press.

Horst, P. (1965). *Factor Analysis of Data Matrices*. Holt Rinehart and Winston.

Johnson, C. (1983). A characterization of borda's rule via optimization. *IMA preprint*, 41.
Lavialle, O., Qannari, E., and Vidal, C. (1990). Order aggregation under constraints: Ordering of products by sensory ratings (in french). *Revue de Statistique Appliquée*, 38(4):61–73.
Le Roux, B. and Rouanet, H. (2004). *Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis*. Dodrecht:Kluewer-Springer.
Meijer, R., Sijtsma, K., and Smid, N. (1990). Theoretical and empirical comparison of the mokken and the rasch approach to irt. *Applied Psychological Measurement*, 14:283–298.
Murtagh, F. (2005). *Correspondence Analysis and Data Coding with Java and R.* London:Chapman & Hall/CRC.
Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and Its Applications.* Toronto:University of Toronto Press.
Nishisato, S. (1984). Forced classification: A simple application of quantification method. *Psychometrika*, 49:25–36.
Nishisato, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis.* Hillsdale NJ: Lawrence Erlbaum.
Saari, D. (1990). The borda dictionary. *Social Choice and Welfare*, 7:279–317.
Sijtsma, K. and Junker, B. (2006). Item response theory: Past performance, present developments, and future expactations. *Behaviormetrika*, 33(1):75–102.
Tatsuoka, C. (2002). Data-analytic methods for latent partially ordered classification models. *Jornal of the Royal Statistical Society Series C (Applied Statistics)*, 51:337–350.
Thurstone, L. (1931). Multiple factor analysis. *Psychological Review*, 31:406–427.
Thurstone, L. (1947). *Multiple factor analysis.* Chicago: The University of Chicago Press.
Torres, A. and Greenacre, M. (2002). Dual scaling and correspondence analysis of preferences, paired comparisons and ratings,. *International Journal of Research in Marketing*, 19(4):401–405.
Wold, H. (1966). *Multivariate Analysis*, chapter Estimation of principal components and related models by iterative least squares, pages 391–420. N.Y:Academic Press.

Appendix

We apply TCA to the correspondence matrice

$$\mathbf{P}_D = \frac{(\mathbf{Y} \mid (\mathbf{1}_I\, \mathbf{m}\prime - \mathbf{Y}))}{I\, m_*}, \tag{24}$$

where $m_* = \sum_{q=1}^{Q} m_q = \mathbf{m}\prime\mathbf{1}_Q$. Note that $\min_q = 0$ and $\max_q > 0$ for $q = 1,...,Q$. Let $\mathbf{P}_D = (\mathbf{P}_{D1}|\mathbf{P}_{D2})$, and designate TCA of residual correspondence matrices by $\mathbf{P}_D^{(\alpha)} = (\mathbf{P}_{D1}^{(\alpha)}|\mathbf{D}_{D2}^{(\alpha)})$ for $\alpha = 0,...,k$, where $k = rank(\mathbf{P}_D) - 1$.

`Proof of Lemma 1`

By induction. For $\alpha = 1$, we have

$$\begin{aligned}
\mathbf{P}_D^{(1)} &= \mathbf{P}_D - \mathbf{r}\,\mathbf{c}_D' \\
&= \frac{(\mathbf{Y} - \frac{\mathbf{1}_I\mathbf{1}_I'\mathbf{Y}}{I} \mid -(\mathbf{Y} - \frac{\mathbf{1}_I\mathbf{1}_I'\mathbf{Y}}{I}))}{I\, m_*} \quad \text{by (24, 22 and 19)} && (25) \\
&= (\mathbf{P}_{D1}^{(1)} \mid \mathbf{P}_{D2}^{(1)}) \\
&= (\mathbf{P}_{D1}^{(1)} \mid -\mathbf{P}_{D1}^{(1)}). && (26)
\end{aligned}$$

Equation (26) implies that the lever principle (28) applies, because let:

$$\mathbf{v}_1^D = \arg \max_{\mathbf{v}\in\{-1,+1\}^I} \left|\left|\mathbf{P}_D^{(1)\prime}\mathbf{v}\right|\right|_1,$$

then the first basic coordinates of the columns (7) will be

$$\begin{aligned} \mathbf{b}_1^D &= \mathbf{P}_D^{(1)\prime}\mathbf{v}_1^D \\ &= \binom{\mathbf{P}_{D1}^{(1)\prime}\ \mathbf{v}_1^D}{-\mathbf{P}_{D1}^{(1)\prime}\ \mathbf{v}_1^D} \text{ by (26)} \\ &= \binom{\mathbf{b}_{11}^D}{\mathbf{b}_{21}^D}. \end{aligned} \tag{27}$$

But (27) implies

$$\mathbf{b}_{11}^D + \mathbf{b}_{21}^D = \mathbf{0}. \tag{28}$$

Let us show $\mathbf{P}_{D2}^{(\alpha)} = -\mathbf{P}_{D1}^{(\alpha)}$ for $\alpha = 2$ and by induction it will be true for all α values. By (10) the residual correspondence matrix is

$$\begin{aligned} \mathbf{P}_D^{(2)} &= \mathbf{P}_D^{(1)} - \mathbf{a}_1^D\mathbf{b}_1^{D\prime}/\lambda_1^D \\ &= (\mathbf{P}_{D1}^{(1)} \mid \mathbf{P}_{D2}^{(1)}) - (\mathbf{a}_1^D\mathbf{b}_{11}^{D\prime}\ |\mathbf{a}_1^D\mathbf{b}_{21}^{D\prime})/\lambda_1^D \text{ by (27)} \\ &= (\mathbf{P}_{D1}^{(1)} \mid -\mathbf{P}_{D1}^{(1)}) - (\mathbf{a}_1^D\mathbf{b}_{11}^{D\prime} \mid -\mathbf{a}_1^D\mathbf{b}_{11}^{D\prime})/\lambda_1^D \text{ by (28)} \\ &= (\mathbf{P}_{D1}^{(2)} \mid -\mathbf{P}_{D1}^{(2)}). \end{aligned}$$

Using the same calculations as in the previous paragraph when $\alpha = 1$, we get for $\alpha = 2$

$$\mathbf{b}_{1\alpha}^D + \mathbf{b}_{2\alpha}^D = \mathbf{0}. \tag{29}$$

By (15) and (20), equation (29) can be expressed as

$$\mathbf{D}_{c1}\mathbf{g}_{1\alpha}^D + \mathbf{D}_{c2}\mathbf{g}_{2\alpha}^D = \mathbf{0} \quad \text{for } \alpha = 1...k,$$

which is the lever principle.

`Proof of Theorem 1 and Corollary 1`

In equation (26), we note that the second block matrix in $\mathbf{P}_D^{(1)}$ is the negative of the first block matrix, so $\mathbf{u}_1^D = \binom{\mathbf{u}_{11}^D}{\mathbf{u}_{21}^D} = \binom{\mathbf{u}_{11}^D}{-\mathbf{u}_{11}^D}$ that maximizes λ_1 in (3). By (7)

$$\begin{aligned} \mathbf{a}_1^D &= \mathbf{P}_D^{(1)}\mathbf{u}_1^D \\ &= 2\mathbf{P}_{D1}^{(1)}\mathbf{u}_{11}^D \quad \text{by (26)} \\ &= \frac{2}{I\,m_*}(\mathbf{Y}\mathbf{u}_{11}^D - \frac{\mathbf{1}_I\mathbf{1}_I'\mathbf{Y}\mathbf{u}_{11}^D}{I}) \quad \text{by (25)} && (30) \\ &= \frac{2}{I\,m_*}(\mathbf{Y}\mathbf{1}_Q - \frac{\mathbf{1}_I\, y_{tot}}{I}) \quad \text{iff } \mathbf{u}_{11}^D = \mathbf{1}_Q, && (31) \end{aligned}$$

where $y_{tot} = \mathbf{1}_I'\mathbf{Y}\mathbf{1}_Q$. By (15) and (31) we have

$$\begin{aligned} \mathbf{f}_1^D &= \mathbf{D}_r^{-1}\mathbf{a}_1^D \\ &= \frac{2}{m_*}(\mathbf{Y}\mathbf{1}_Q - \frac{\mathbf{1}_I\, y_{tot}}{I}) \text{ by (23) and (31),} \end{aligned}$$

which coordinatewise is $f^D_{i1} = \frac{2}{m_*}(y_{i*} - \frac{y_{tot}}{I})$; that is, the 1st row factor score, f_{i1}, is an affine function of the sum score y_{i*} iff $\mathbf{u}^D_{11} = \mathbf{1}_Q$.

To see what happens if some $u^D_{11j} = -1$, we consider the case when only one, say, $u^D_{11Q} = -1$. Then by (30), we have

$$\begin{aligned} f^D_{i1} &= \frac{2}{m_*}\left[\sum_{j=1}^{Q-1}(y_{ij} - \frac{y_{*j}}{I}) - (y_{iQ} - \frac{y_{*Q}}{I})\right] \\ &= \frac{2}{m_*}\left[\sum_{j=1}^{Q}(y_{ij} - \frac{y_{*j}}{I}) - 2(y_{iQ} - \frac{y_{*Q}}{I})\right] \\ &= \frac{2}{m_*}\left[(y_{i*} - \frac{y_{tot}}{I}) - 2(y_{iQ} - \frac{y_{*Q}}{I})\right]. \end{aligned} \quad (32)$$

Equation (32) shows that the points (f^D_{i1}, y_{i*}) will locate on many parallel lines defined by the rating of the ith respondent on item Q.

By (8)

$$\begin{aligned} \lambda^D_1 &= ||\mathbf{a}^D_1||_1 \\ &= \frac{2}{I\, m_*}\sum_{i=1}^{I}|(y_{i*} - y_{tot}/I)| \quad \text{by (31)}, \end{aligned}$$

which is Corollary 1.

`Proof of Theorem 2`

The proof is similar as in Theorem 1. First, we note that $\mathbf{P}^{(1)}_D = (\mathbf{P}^{(1)}_{D1} \mid -\mathbf{P}^{(1)}_{D1})$ in (25) and $\mathbf{P}^{(1)}_{nega}$ are related by

$$\begin{aligned} \mathbf{P}^{(1)}_{nega} &= (\mathbf{P}^{(1)}_{nega1} \mid \mathbf{p}^{(1)}_{nega}) \\ & \quad (\mathbf{P}^{(1)}_{D1} \mid -\mathbf{P}^{(1)}_{D1}\mathbf{1}_Q). \end{aligned} \quad (33)$$

Let $\mathbf{u}^{nega}_1 = \binom{\mathbf{u}^{nega}_{11}}{u^{nega}_{21}}$ that maximizes λ^{nega}_1 in (3). By (7)

$$\begin{aligned} \mathbf{a}^{nega}_1 &= \mathbf{P}^{(1)}_{nega}\mathbf{u}^{nega}_1 \\ &= \mathbf{P}^{(1)}_{D1}\mathbf{u}^{nega}_{11} - \mathbf{P}^{(1)}_{D1}\mathbf{1}_Q u^{nega}_{21} \quad \text{by (33)} \quad (34) \\ &= 2\mathbf{P}^{(1)}_{D1}\mathbf{1}_Q \quad \text{iff } \mathbf{u}^{nega}_{11} = \mathbf{1}_Q \text{ and } u^{nega}_{21} = -1 \\ &= -2\mathbf{p}^{(1)}_{nega} \quad \text{iff } \mathbf{u}^{nega}_{11} = \mathbf{1}_Q \text{ and } u^{nega}_{21} = -1 \quad \text{by (33)} \quad (35) \\ &= \frac{2}{I\, m_*}(\mathbf{Y}\mathbf{1}_Q - \frac{\mathbf{1}_I\, y_{tot}}{I}) \quad \text{iff } \mathbf{u}^{nega}_{11} = \mathbf{1}_Q \text{ and } u^{nega}_{21} = -1. \quad (36) \end{aligned}$$

From (35, 36) properties a and b follow easily.

To see what happens if some $u^{nega}_{11j} = -1$, we consider the case when only one, say, $u^{nega}_{11Q} = -1$. Then by (34), we have

$$
\begin{aligned}
f_{i1}^{nega} &= \frac{1}{m_*}\left[\sum_{j=1}^{Q-1}(y_{ij}-\frac{y_{*j}}{I})-(y_{iQ}-\frac{y_{*Q}}{I})\right]+\frac{1}{m_*}\left[(y_{i*}-\frac{y_{tot}}{I})\right] \\
&= \frac{1}{m_*}\left[\sum_{j=1}^{Q}(y_{ij}-\frac{y_{*j}}{I})-2(y_{iQ}-\frac{y_{*Q}}{I})\right]+\frac{1}{m_*}\left[(y_{i*}-\frac{y_{tot}}{I})\right] \\
&= \frac{2}{m_*}\left[(y_{i*}-\frac{y_{tot}}{I})-(y_{iQ}-\frac{y_{*Q}}{I})\right]. \qquad (37)
\end{aligned}
$$

Equation (37) shows that the points (f_{i1}^{D}, y_{i*}) will locate on many parallel lines defined by the rating of the ith respondent on item Q. However, (37) is different from (32).

The proof of property c follows from the following 2 facts:

Fact 1: We have

$$
\begin{aligned}
\lambda_1^{nega} &= ||\mathbf{a}_1^{nega}|| \quad \text{by} \quad (8) \\
&= 2||\mathbf{p}_{nega}^{(1)}|| \quad \text{by} \quad (35). \qquad (38)
\end{aligned}
$$

Fact 2: Consider the basic vector $\mathbf{b}_1^{nega} = (\begin{smallmatrix}\mathbf{b}_{11}^{nega} \\ b_{21}^{nega}\end{smallmatrix})$. By (9)

$$
\begin{aligned}
sign(b_{21}^{nega}) &= u_{21}^{nega} \\
&= -1 \quad \text{by} \ (36),
\end{aligned}
$$

and

$$\lambda_1^{nega} = -2b_{21}^{nega} \quad \text{by}(14). \qquad (39)$$

We have

$$
\begin{aligned}
\mathbf{p}_{nega}^{(2)} &= \mathbf{p}_{nega}^{(1)} - \frac{\mathbf{a}_1^{nega} b_{21}^{nega}}{\lambda_1^{nega}} \quad \text{by} \ (10) \\
&= \mathbf{0} \quad \text{by (35, 38, 39)}.
\end{aligned}
$$

`Proof of Theorem 3`

The proof follows easily from Theorems 1 and 2 and Lemma 1.