

Tests of Ignoring and Eliminating in Nonsymmetric Correspondence Analysis

Yoshio Takane · Sunho Jung

The date of receipt and acceptance will be inserted by the editor

Abstract Nonsymmetric correspondence analysis (NSCA) aims to examine predictive relationships between rows and columns of a contingency table. The predictor categories of such tables are often accompanied by some auxiliary information. Constrained NSCA (CNSCA) incorporates such information as linear constraints on the predictor categories. However, imposing constraints also means that part of the predictive relationship is left unaccounted for by the constraints. A method of NSCA is proposed for analyzing the residual part along with the part accounted for by the constraints. The CATANOVA test may be invoked to test the significance of each part. The two tests parallel the distinction between tests of ignoring and eliminating, and help gain some insight into what is known as Simpson's paradox in the analysis of contingency tables. Two examples are given to illustrate the distinction.

Keywords Contingency table analysis · The Goodman-Kruskal τ index · CATANOVA · Simpson's paradox

Subject classification JEL C25, AMS 15A03 15A09

1 Introduction

Scientific research in many disciplines including psychology, biology, ecology, epidemiology, and so on, often concerns the prediction of one categorical variable from another. Nonsymmetric correspondence analysis (NSCA: Lauro and

Yoshio Takane

Department of Psychology, McGill University, 1205 Dr. Penfield Ave., Montreal, QC, H3A 1B1 Canada

Phone: 514-398-6125 Fax: 514-398-4896 E-mail: takane@psych.mcgill.ca

Sunho Jung

Department of Marketing and Management, Rotterdam School of Management, Erasmus University, PO BOX 1738, 3000 DR Rotterdam, The Netherlands

Phone: +31 10 4088822 Fax: +31 10 4089011 E-mail: sjung@rsm.nl

D’Ambra 1984) is a useful technique for analyzing a two-way contingency table in which rows and columns assume asymmetric relationships, for example, columns depend on rows, but not vice versa. (Throughout this paper, we assume that rows represent the predictive categories, and columns the criterion categories.) For such tables, the Goodman-Kruskal τ index (Goodman and Kruskal 1954) is used to indicate how well the rows predict the columns. The CATANOVA test (Light and Margolin 1971) may then be invoked to test the significance of the predictive relationship indicated by τ .

Two-way contingency tables are often accompanied by some auxiliary information on the predictor categories. Consider, for example, a table of frequencies of certain tree species (columns) in various sites (rows). The abundance of trees depends on the sites, while the converse is not true. Variables describing environmental factors may be provided that influence the population of trees in the sites. Constrained NSCA (CNSCA; Takane and Jung 2009) may be applied to incorporate such information in the prediction of abundance of various kinds of trees in various sites (see also Böckenholt and Böckenholt 1990; Böckenholt and Takane 1994; Hwang and Takane 2002; Takane, Yanai, and Mayekawa 1991; ter Braak 1986).

Higher order contingency tables can also be analyzed by CNSCA. Table 1 shows a three-way contingency table constructed by classifying 800 patients by treatment and gender (interactively coded) on the rows, and prognosis on the columns. Patient recovery is affected by the treatment and gender. We may take the prognosis as the criterion variable, and the combinations of the other two as the predictor variables. Since the rows of the table have a factorial structure, a variety of analyses can be performed by defining appropriate contrast vectors (contrasts are formal representations of constraints) representing the structure on the rows. For example, we may be specifically interested in examining how the main effect of treatment alone influences the prognosis. To achieve this goal, we supply a contrast vector representing the main effect of treatment, that is,

$$\mathbf{T}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \begin{array}{l} \text{male, medication} \\ \text{male, control} \\ \text{female, medication} \\ \text{female, control} \end{array}, \quad (1)$$

and apply CNSCA. Note that \mathbf{T}_1 takes the sum of the first and third rows (both related to the medication group), the sum of the second and fourth rows (both related to the control group), and take a difference between these two sums, thereby contrasting the two groups (the medication and control groups). (Note that \mathbf{T}_1 satisfies $\mathbf{1}'\mathbf{T}_1 = \mathbf{0}$, which is the defining property of a contrast vector.)

Imposing constraints on the predictor categories has the potential benefit of stabilizing the estimates of parameters (Takane and Jung 2009). A potential problem, on the other hand, is that part of the predictive relationship may be left unexplained by the constraints. That is, the constraints may capture some

Table 1 Cross-classification of patients in terms of Treatment, Gender, and Prognosis ($n = 800$, where n is the total sample size).

Gender	Treatment	Prognosis		Recovery rate
		Recovery	Non-recovery	
Male	Medication	180	120	.6
	Control	70	30	.7
Female	Medication	20	80	.2
	Control	120	180	.4

portions of the relationship, but not the entire relationship. In the specific example of Table 1, if the contrast representing the main effect of treatment (\mathbf{T}_1) is imposed, the main effect of gender and the interaction between treatment and gender will be left out. The latter effects pertain to the contrast matrix

$$\mathbf{T}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ -1 & 1 \end{bmatrix} \begin{array}{l} \text{male, medication} \\ \text{male, control} \\ \text{female, medication} \\ \text{female, control} \end{array} \quad (2)$$

where the first column of \mathbf{T}_2 contrasts the two categories of gender, and the second column the interaction between treatment and gender. Note that $\mathbf{1}'\mathbf{T}_2 = \mathbf{0}$ (indicating that \mathbf{T}_2 is a contrast matrix), and $\mathbf{T}_1'\mathbf{T}_2 = \mathbf{0}$ (\mathbf{T}_1 and \mathbf{T}_2 are orthogonal). The present paper proposes a method of NSCA that allows the analysis of both parts of the predictive relationship. The two parts of the analysis parallel the distinction between tests of ignoring and eliminating other effects in multiple regression analysis and non-orthogonal ANOVA (Maxwell and Delaney 2004, pp 320-343). The method also allows analogous partitionings of the Goodman-Kruskal τ index and the CATANOVA test statistic. This example will be used in the application section to illustrate various analyses that can be carried out by the proposed method.

This paper first calls attention to the importance of analyzing the residuals along with the part accounted for by the constraints in the analysis of contingency tables. In particular, the analysis sheds further light on an old problem known as Simpson's paradox in statistics (Simpson 1951; Yule 1903). Simpson's paradox (e.g., Blyth 1972; Shapiro 1982) occurs when an apparent association between rows and columns of a table disappears or is even reversed, or when an apparent non-association between rows and columns appears as a sizable association, when a third variable is taken into consideration. These differences that occur when the effect of a third variable is ignored or eliminated can be captured by the difference between the analysis of the part explained by the constraints and that of the residuals. For example, in testing the treatment main effect in Table 1, we can either ignore or eliminate the other effects. The former (the test of ignoring) can be performed by CNSCA with \mathbf{T}_1 as the constraint, while the latter (the test of eliminating) by the analysis of residuals from CNSCA with \mathbf{T}_2 as the constraints.

This paper is organized as follows. Section 2 discusses the proposed tests of ignoring and eliminating in the analysis of contingency tables. Section 3 presents the mathematical underpinnings for the proposed method. Section 4 gives possible finer decompositions of the ignoring and eliminating effects. Section 5 discusses how the proposed method is related to other existing methods. Section 6 introduces permutation tests for determining the best dimensionality of the solution. The bootstrap method is then described for evaluating the stability of the estimates of parameters. Section 7 demonstrates applications of the proposed method to two empirical examples to illustrate its usefulness. The final section gives a summary and discussion.

2 The Proposed Method

Let \mathbf{G} and \mathbf{H} represent n by r and n by c indicator matrices of predictor and criterion variables, respectively, where n is the number of subjects (the sample size), r is the number of categories in the predictor variable, and c is the number of categories in the criterion variable. Let $\mathbf{K} = \mathbf{G}'\mathbf{G}$ and $\mathbf{D} = \mathbf{H}'\mathbf{H}$ denote diagonal matrices of row and column marginal frequencies, respectively. We assume for the moment that \mathbf{K} is nonsingular. Let $\mathbf{F} = \mathbf{G}'\mathbf{H}$ denote an r by c contingency table calculated from \mathbf{G} and \mathbf{H} . NSCA attempts to predict column categories given row categories of a contingency table. Measures of predictive power of rows on columns are given collectively by

$$\mathbf{A} = \mathbf{K}^{-1}\mathbf{Q}'_{1/K}\mathbf{F} = \mathbf{Q}_{1/K}\mathbf{K}^{-1}\mathbf{F}, \quad (3)$$

where

$$\mathbf{Q}_{1/K} = \mathbf{I} - \mathbf{1}(\mathbf{1}'\mathbf{K}\mathbf{1})^{-1}\mathbf{1}'\mathbf{K} \quad (4)$$

is the orthogonal projector onto the null space of $\mathbf{1}$ (the vector of ones) in metric \mathbf{K} . The matrix $\mathbf{Q}_{1/K}$ eliminates the row marginal effect from the entire relationship between rows and columns. The ij^{th} element a_{ij} of \mathbf{A} is equal to

$$a_{ij} = p_{ij}/p_{i.} - p_{.j}, \quad (5)$$

where p_{ij} is the joint probability of row i and column j (the ij^{th} element of \mathbf{F}/n), $p_{i.}$ is the marginal probability of row i (the i^{th} diagonal element of \mathbf{K}/n), and $p_{.j}$ is the marginal probability of column j (the j^{th} diagonal element of \mathbf{D}/n). The first term on the right hand side of (5) is the conditional probability of column j given row i . This means that the predictive power of row i on column j is higher if the conditional probability of column j given row i is higher relative to the marginal (unconditional) probability of column j , and it is lower if $p_{ij}/p_{i.}$ is smaller than $p_{.j}$ (Takane and Jung 2009).

For dimension reduction, we may use the generalized singular value decomposition (GSVD) of (3) with metric matrices \mathbf{K} and \mathbf{I} , which is written as $\text{GSVD}(\mathbf{A})_{K, I}$ (e.g., Takane and Shibayama 1991). Let $\mathbf{A} = \mathbf{U}\mathbf{\Delta}\mathbf{V}'$ denote $\text{GSVD}(\mathbf{A})_{K, I}$, where \mathbf{U} is an r by m matrix of left singular vectors such that $\mathbf{U}'\mathbf{K}\mathbf{U} = \mathbf{I}_m$ ($m = \text{rank}(\mathbf{A})$), \mathbf{V} is a c by m matrix of right singular vectors

Table 2 The marginal table within Table 1, collapsing over gender

Treatment	Prognosis		Recovery rate
	Recovery	Non-recovery	
Medication	200	200	.5
Control	190	210	.475

such that $\mathbf{V}'\mathbf{V} = \mathbf{I}_m$, and $\mathbf{\Delta}$ is a *pd* (positive definite) diagonal matrix of order m of generalized singular values of \mathbf{A} . Then the best rank- t approximation to \mathbf{A} in the sense of $\text{SS}(\mathbf{A} - \hat{\mathbf{A}}) \rightarrow \min$, assuming that $t \leq m$, is given by $\hat{\mathbf{A}} = \mathbf{U}_t \mathbf{\Delta}_t \mathbf{V}_t'$, where the matrices with subscript t (\mathbf{U}_t , \mathbf{V}_t , and $\mathbf{\Delta}_t$) are submatrices of the respective matrices (\mathbf{U} , \mathbf{V} , and $\mathbf{\Delta}$) in the GSVD of \mathbf{A} corresponding to the t largest (generalized) singular values of \mathbf{A} . Matrix $\mathbf{U}_t \mathbf{\Delta}_t$ gives the so-called principal coordinates of row (predictor) categories, while \mathbf{V}_t the standard coordinates of column (criterion) categories (Greenacre, 1984). The method of choosing the best dimensionality t will be discussed in Section 6.

Let $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2]$ be a matrix of contrast vectors (representing predictor variables) defined on row categories such that $\mathbf{T}_1' \mathbf{T}_2 = \mathbf{0}$, and $\text{Sp}(\mathbf{T}_1) \oplus \text{Sp}(\mathbf{T}_2) = \text{Sp}(\mathbf{Q}_{1/K})$ (i.e., \mathbf{T}_2 spans the ortho-complement subspace of \mathbf{T}_1 within $\text{Sp}(\mathbf{Q}_{1/K})$, where $\text{Sp}(\mathbf{Y})$ indicates the column space of $\mathbf{Y} = \mathbf{T}_1, \mathbf{T}_2$, or $\mathbf{Q}_{1/K}$). Note that an example of \mathbf{T}_1 and \mathbf{T}_2 has been given in the introduction section, where for Table 1, \mathbf{T}_1 indicated the main effect of treatment, while \mathbf{T}_2 the main effect of gender and the interaction between treatment and gender. In CNSCA of \mathbf{F} with \mathbf{T}_1 ignoring \mathbf{T}_2 , we obtain

$$\mathbf{B} = \mathbf{Q}_{1/K} \mathbf{T}_1 (\mathbf{T}_1' \mathbf{Q}_{1/K}' \mathbf{K} \mathbf{T}_1)^{-1} \mathbf{T}_1' \mathbf{Q}_{1/K}' \mathbf{F}. \quad (6)$$

Note that $\mathbf{Q}_{1/K}' \mathbf{K} \mathbf{Q}_{1/K} = \mathbf{Q}_{1/K}' \mathbf{K} = \mathbf{K} \mathbf{Q}_{1/K}$. Note also that $\mathbf{1}' \mathbf{K} \mathbf{Q}_{1/K} \mathbf{T}_1 = \mathbf{0}$. We say that \mathbf{B} pertains to the effect of \mathbf{T}_1 ignoring \mathbf{T}_2 because it is essentially the same as analyzing the total predictive relationship between rows (treatment) and columns (prognosis) of a contingency table obtained by collapsing Table 1 across gender categories (i.e., ignoring gender), as presented in Table 2. For dimension reduction, we obtain $\text{GSVD}(\mathbf{B})_{K,I}$. Let $\mathbf{B} = \mathbf{U} \mathbf{\Delta} \mathbf{V}'$ denote this GSVD. As before, $\hat{\mathbf{B}} = \mathbf{U}_t \mathbf{\Delta}_t \mathbf{V}_t'$ gives the best rank- t approximation of \mathbf{B} , $\mathbf{U}_t \mathbf{\Delta}_t$ gives the principal coordinates of the predictive categories, and \mathbf{V}_t gives the standard coordinates of the criterion categories.

In some cases, each column of $\mathbf{Q}_{1/K} \mathbf{T}_1$ has a specific meaning, and its effect may be of interest individually. In such cases, the weights applied to $\mathbf{Q}_{1/K} \mathbf{T}_1$ to obtain the row representation $\mathbf{U}_t \mathbf{\Delta}_t$ may be in order. They can be calculated by

$$\mathbf{W}_t = (\mathbf{T}_1' \mathbf{Q}_{1/K}' \mathbf{K} \mathbf{T}_1)^{-1} \mathbf{T}_1' \mathbf{Q}_{1/K}' \mathbf{K} \mathbf{U}_t \mathbf{\Delta}_t, \quad (7)$$

or more directly by $\text{GSVD}((\mathbf{T}_1' \mathbf{Q}_{1/K}' \mathbf{K} \mathbf{T}_1)^{-1} \mathbf{T}_1' \mathbf{Q}_{1/K}' \mathbf{F})_{T_1' \mathbf{Q}_{1/K}' \mathbf{K} \mathbf{T}_1, I}$. In the latter, the row representation is recovered by $\mathbf{Q}_{1/K} \mathbf{T}_1 \mathbf{W}_t = \mathbf{P}_{\mathbf{Q}_{1/K} \mathbf{T}_1 / K} \mathbf{U}_t \mathbf{\Delta}_t = \mathbf{U}_t \mathbf{\Delta}_t$.

The residuals from \mathbf{B} (i.e., $\mathbf{E} = \mathbf{A} - \mathbf{B}$) can be expressed as

$$\mathbf{E} = \mathbf{K}^{-1}\mathbf{T}_2(\mathbf{T}'_2\mathbf{K}^{-1}\mathbf{T}_2)^{-1}\mathbf{T}'_2\mathbf{K}^{-1}\mathbf{F}. \quad (8)$$

Note that $\mathbf{T}'_1\mathbf{Q}'_{1/K}\mathbf{T}_2 = \mathbf{0}$, and that $\mathbf{1}'\mathbf{K}\mathbf{K}^{-1}\mathbf{E} = \mathbf{0}$. The matrix \mathbf{E} may also be subjected to GSVD (i.e., $\text{GSVD}(\mathbf{E})_{K,I}$) for dimension reduction. Let $\mathbf{U}\mathbf{\Delta}\mathbf{V}'$ denote $\text{GSVD}(\mathbf{E})_{K,I}$. As before, $\hat{\mathbf{E}} = \mathbf{U}_t\mathbf{\Delta}\mathbf{V}'_t$ gives the best rank- t approximation to \mathbf{E} , $\mathbf{U}_t\mathbf{\Delta}_t$ the principal coordinates of the predictor categories, and \mathbf{V}_t the standard coordinates of the criterion categories. The weights applied to $\mathbf{K}^{-1}\mathbf{T}_2$ to obtain the row representation can be obtained by $\mathbf{W}_t = (\mathbf{T}'_2\mathbf{K}\mathbf{T}_2)^{-1}\mathbf{T}'_2\mathbf{U}_t\mathbf{\Delta}_t$, or by $\text{GSVD}((\mathbf{T}'_2\mathbf{K}^{-1}\mathbf{T}_2)^{-1}\mathbf{T}'_2\mathbf{K}^{-1}\mathbf{F})_{T'_2K^{-1}T_2,I}$. In the latter, $\mathbf{U}_t\mathbf{\Delta}_t$ can be obtained from \mathbf{W}_t by $\mathbf{K}^{-1}\mathbf{T}_2\mathbf{W}_t = \mathbf{P}_{K^{-1}T_2/K}\mathbf{U}_t\mathbf{\Delta}_t = \mathbf{U}_t\mathbf{\Delta}_t$.

The above derivation assumed that both \mathbf{T}_1 and \mathbf{T}_2 were known explicitly, which may not be true in general. However, using their complementarity, we can derive one from the other. That is, \mathbf{E} can be derived from \mathbf{B} by $\mathbf{E} = \mathbf{A} - \mathbf{B}$, and \mathbf{B} from \mathbf{E} by $\mathbf{B} = \mathbf{A} - \mathbf{E}$. It is thus sufficient to know only one of them. That the complementarity relation $\mathbf{A} = \mathbf{B} + \mathbf{E}$ holds with \mathbf{A} defined in (3), \mathbf{B} defined in (6), and \mathbf{E} defined in (8) will be shown in the next section.

Obviously, we can interchange the roles of \mathbf{T}_1 and \mathbf{T}_2 above. We may analyze

$$\mathbf{J} = \mathbf{Q}_{1/K}\mathbf{T}_2(\mathbf{T}'_2\mathbf{Q}'_{1/K}\mathbf{K}\mathbf{T}_2)^{-1}\mathbf{T}'_2\mathbf{Q}'_{1/K}\mathbf{F}, \quad (9)$$

which gives the CNSCA of \mathbf{F} with \mathbf{T}_2 ignoring \mathbf{T}_1 . We may also analyze

$$\mathbf{L} = \mathbf{K}^{-1}\mathbf{T}_1(\mathbf{T}'_1\mathbf{K}^{-1}\mathbf{T}_1)^{-1}\mathbf{T}'_1\mathbf{K}^{-1}\mathbf{F}, \quad (10)$$

which gives the CNSCA of \mathbf{F} with \mathbf{T}_1 eliminating \mathbf{T}_2 . Again, we have $\mathbf{A} = \mathbf{J} + \mathbf{L}$.

The significance of the overall predictability (\mathbf{A}) as well as the effects of ignoring (\mathbf{B} and \mathbf{J}) and eliminating (\mathbf{E} and \mathbf{L}) can be tested using the CATANOVA C -statistic (Light and Margolin 1971; see also Anderson and Landis (1980), and D'Ambra, Beh, and Amenta (2005)). Let

$$\text{TSS} = 1 - \text{tr}(\mathbf{D}^2)/n^2 \quad (11)$$

denote the total variance in the contingency table \mathbf{F} , where n is the sample size (i.e., $n = \mathbf{1}'_C\mathbf{D}\mathbf{1}_C$). Let

$$\text{BSS}_A = \frac{1}{n}\text{SS}(\mathbf{A})_{K,I} = \text{tr}(\mathbf{A}'\mathbf{K}\mathbf{A})/n \quad (12)$$

indicate the explained variance due to the predictability of rows on columns. The Goodman-Kruskal τ index, which indicates the overall predictability of rows on columns, is defined by

$$\tau_A = \text{BSS}_A/\text{TSS}. \quad (13)$$

(τ_A is sometimes called multiple τ (Gray and Williams (1981) because it represents the total predictability of two or more predictor variables.) The

CATANOVA test may be used to test the significance of the τ index. The test statistic known as the C -statistic (Light and Margolin 1971) is defined by

$$C_A = (n - 1)(c - 1)\tau_A. \quad (14)$$

This statistic follows the asymptotic chi-square distribution with $(c - 1)(r - 1)$ degrees of freedom as $n \rightarrow \infty$ under the null hypothesis of no predictability of the rows on the columns.

Tests involved in the two parts of the proposed analysis (i.e., tests of ignoring and eliminating) entail decomposing the Goodman-Kruskal τ index and the CATANOVA test statistic in a straightforward manner. Let BSS_B and BSS_E be $\frac{1}{n}SS(\mathbf{B})_{K, I}$ and $\frac{1}{n}SS(\mathbf{E})_{K, I}$, respectively. We have $BSS_A = BSS_B + BSS_E$. Part of the overall τ index pertaining to the effect of \mathbf{T}_1 ignoring \mathbf{T}_2 is defined as

$$\tau_B = BSS_B/TSS, \quad (15)$$

while the complementary part pertaining to the effect of \mathbf{T}_2 eliminating \mathbf{T}_1 is defined as

$$\tau_E = BSS_E/TSS. \quad (16)$$

(τ_B is simply the basic τ index for the marginal table (Table 2), while τ_E is sometimes called partial τ (Gray and Williams 1981).) The CATANOVA test statistic can also be analogously decomposed into two parts. Let r_1 and r_2 denote $\text{rank}(\mathbf{T}_1)$ and $\text{rank}(\mathbf{T}_2)$, respectively, such that $r_1 + r_2 = \text{rank}(\mathbf{T})$. Then, the portion of the overall C -statistic pertaining to the test of \mathbf{T}_1 ignoring \mathbf{T}_2 is obtained by

$$C_B = (n - 1)(c - 1)BSS_B/TSS. \quad (17)$$

Under the hypothesis that \mathbf{T}_1 ignoring \mathbf{T}_2 has no predictability on columns, this statistic has an asymptotic chi-square distribution with $(c - 1)r_1$ degrees of freedom. The complementary part of the C -statistic pertaining to the test of \mathbf{T}_2 eliminating \mathbf{T}_1 , on the other hand, is obtained by

$$C_E = (n - 1)(c - 1)BSS_E/TSS. \quad (18)$$

This statistic follows the asymptotic chi-square distribution with $(c - 1)r_2$ degrees of freedom under the hypothesis that \mathbf{T}_2 eliminating \mathbf{T}_1 has no predictability on columns. Obviously, analogous arguments hold for \mathbf{J} and \mathbf{L} , where we test the effects of \mathbf{T}_2 ignoring \mathbf{T}_1 , and of \mathbf{T}_1 eliminating \mathbf{T}_2 .

3 Mathematical Underpinnings

In this section, we show the complementarity of \mathbf{B} and \mathbf{E} , and of \mathbf{J} and \mathbf{L} . Let $\mathbf{X} = \mathbf{Q}_{1/K}\mathbf{T}_1$ and $\mathbf{R} = \mathbf{Q}'_{1/K}\mathbf{T}_2 = \mathbf{T}_2$, or $\mathbf{X} = \mathbf{Q}_{1/K}\mathbf{T}_2$ and $\mathbf{R} = \mathbf{Q}'_{1/K}\mathbf{T}_1 = \mathbf{T}_1$. Note that with these definitions of \mathbf{X} and \mathbf{R}

$$\mathbf{X}'\mathbf{R} = \mathbf{0}, \quad (19)$$

and

$$\text{Sp}([\mathbf{X}, \mathbf{R}]) = \text{Sp}(\mathbf{Q}_{1/K}). \quad (20)$$

The analyses discussed in the previous section suggest the following decomposition of matrix $\mathbf{K}^{-1}\mathbf{Q}'_{1/K}$:

$$\mathbf{K}^{-1}\mathbf{Q}'_{1/K} = \mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}' + \mathbf{K}^{-1}\mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}^{-1}. \quad (21)$$

Since $\mathbf{K}^{-1}\mathbf{Q}'_{1/K} = \mathbf{Q}_{1/K}\mathbf{K}^{-1}$, and because

$$\mathbf{Q}_{1/K}\mathbf{X} = \mathbf{X}, \quad (22)$$

and

$$\mathbf{Q}_{1/K}\mathbf{K}^{-1}\mathbf{R} = \mathbf{K}^{-1}\mathbf{R}, \quad (23)$$

(21) can be rewritten as

$$\mathbf{Q}_{1/K}\mathbf{K}^{-1} = \mathbf{Q}_{1/K}\mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}' + \mathbf{Q}_{1/K}\mathbf{K}^{-1}\mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}^{-1}. \quad (24)$$

If there is no $\mathbf{Q}_{1/K}$ in front of each term in this identity, this is nothing but Khatri's (1966) lemma used to show the equivalence of CCA (canonical correspondence analysis; ter Braak, 1986) and CALC (canonical analysis with linear constraints; Böckenholt and Böckenholt 1990) in Takane, Yanai, and Mayekawa (1991). The matrix $\mathbf{Q}_{1/K}$ has the effect of eliminating the trivial solution corresponding to the most dominant singular value in NSCA, since $\mathbf{1}'\mathbf{K}\mathbf{Q}_{1/K} = \mathbf{0}$. The identity (21) can also be rewritten as

$$\mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}' = \mathbf{K}^{-1}(\mathbf{Q}'_{1/K} - \mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}^{-1}), \quad (25)$$

or

$$\begin{aligned} & \mathbf{K}^{-1}\mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{K}\mathbf{K}^{-1}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}^{-1} \\ &= \mathbf{K}^{-1}(\mathbf{Q}'_{1/K} - \mathbf{K}\mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{K}^{-1}\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{K}^{-1}). \end{aligned} \quad (26)$$

Eq. (25) shows that a nonsymmetric version of CCA (NSCCA), which is a special case of CNSCA of \mathbf{F} with the predictor matrix \mathbf{X} (written as NSCCA(\mathbf{X})), and which obtains GSVD($\mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{F}$) $_{K, I}$, is equivalent to obtaining:

$$\text{GSVD}(\mathbf{K}^{-1}(\mathbf{Q}'_{1/K} - \mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}^{-1})\mathbf{F})_{K, I}.$$

The latter is a nonsymmetric version of CALC (NSCALC), which is written as NSCALC(\mathbf{R}). The original CALC used \mathbf{I} for $\mathbf{Q}'_{1/K}$ and an \mathbf{R} which included $\mathbf{K}\mathbf{1}$ as part of it in the above formula. The difference in the present case arose from the fact that we *a priori* eliminated the effect of $\mathbf{K}\mathbf{1}$. Eq. (26) shows NSCCA($\mathbf{K}^{-1}\mathbf{R}$) is equivalent to NSCALC($\mathbf{K}\mathbf{X}$); both analyze the part of $\mathbf{K}^{-1}\mathbf{Q}'_{1/K}\mathbf{F}$ complementary to NSCCA(\mathbf{X}) = NSCALC(\mathbf{R}).

It is interesting to note that decomposition (21) has a one-to-one correspondence with the following decomposition of projector $\mathbf{P}_{GQ_{1/K}}$:

$$\mathbf{P}_{GQ_{1/K}} = \mathbf{P}_{GQ_{1/K}X} + \mathbf{P}_{GQ_{1/K}(Q'_{1/K}KQ_{1/K})^{-1}R}, \quad (27)$$

where $\mathbf{P}_{GQ_{1/K}}$ can further be rewritten as

$$\begin{aligned}\mathbf{P}_{GQ_{1/K}} &= \mathbf{GQ}_{1/K}(\mathbf{Q}'_{1/K}\mathbf{KQ}_{1/K})^{-}\mathbf{Q}'_{1/K}\mathbf{G}' \\ &= \mathbf{GQ}_{1/K}\mathbf{K}^{-1}\mathbf{Q}'_{1/K}\mathbf{G}' = \mathbf{GK}^{-1}\mathbf{Q}'_{1/K}\mathbf{G}',\end{aligned}\quad (28)$$

since $\mathbf{K}^{-1}\mathbf{Q}'_{1/K} \subset \{(\mathbf{Q}'_{1/K}\mathbf{KQ}_{1/K})^{-}\}$, where $^{-}$ indicates a g-inverse. Because of (22) and (23), we have $\mathbf{GQ}_{1/K}\mathbf{X} = \mathbf{GX}$, and $\mathbf{GQ}_{1/K}(\mathbf{Q}'_{1/K}\mathbf{KQ}_{1/K})^{-}\mathbf{R} = \mathbf{GQ}_{1/K}\mathbf{K}^{-1}\mathbf{R} = \mathbf{GK}^{-1}\mathbf{R}$. Thus, the above decomposition can be rewritten as:

$$\begin{aligned}\mathbf{GK}^{-1}\mathbf{Q}'_{1/K}\mathbf{G}' \\ = \mathbf{GX}(\mathbf{X}'\mathbf{KX})^{-1}\mathbf{X}'\mathbf{G}' + \mathbf{GK}^{-1}\mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}^{-1}\mathbf{G}'.\end{aligned}\quad (29)$$

By pre- and postmultiplying both sides of (29) by $\mathbf{K}^{-1}\mathbf{G}'$ and \mathbf{GK}^{-1} , respectively, we obtain (26), establishing that (27) implies (21). Showing the reverse is straightforward by tracing the above process backward.

Thus far we have assumed that \mathbf{K} is nonsingular (i.e., of full rank). Consider, however, the situation in which we allow predictor categories with 0 frequencies in the diagonal matrix of row totals. In practice, we may discard those categories before the data analysis is conducted, but this strategy may not work in the bootstrap method (see Section 6).

The case of singular \mathbf{K} is quite challenging because $\text{Sp}(\mathbf{R})$ is not necessarily in $\text{Sp}(\mathbf{Q}'_{1/K}\mathbf{G}')$, and \mathbf{GX} and $\mathbf{GK}^{-1}\mathbf{R}$ in (27) are not necessarily mutually orthogonal. There are two possible orthogonal decompositions in this case, depending on which of \mathbf{X} and \mathbf{R} we put more emphasis on. Both are based on the following well-known decomposition of the orthogonal projector defined by a row block matrix $[\mathbf{M}, \mathbf{N}]$ (e.g., Takane and Yanai 1999):

$$\mathbf{P}_{[\mathbf{M}, \mathbf{N}]} = \mathbf{P}_M + \mathbf{P}_{Q_M N}.\quad (30)$$

One is

$$\mathbf{P}_{GQ_{1/K}} = \mathbf{P}_{GX} + \mathbf{P}_{Q_{GX}GQ_{1/K}},\quad (31)$$

where $\mathbf{Q}_{GX}\mathbf{GQ}_{1/K}$ can further be rewritten as

$$\mathbf{Q}_{GX}\mathbf{GQ}_{1/K} = \mathbf{GQ}_{1/K}\mathbf{Q}_{X/K} = \mathbf{GQ}_{X^*/K},\quad (32)$$

where $\mathbf{X}^* = [\mathbf{1}, \mathbf{X}]$. (The matrix $\mathbf{Q}_{GX} = \mathbf{I} - \mathbf{GX}(\mathbf{X}'\mathbf{KX})^{-}\mathbf{X}'\mathbf{G}'$ is the orthogonal projector onto the null space of \mathbf{GX} ($\text{Ker}(\mathbf{X}'\mathbf{G}')$), whereas $\mathbf{Q}_{X^*/K} = \mathbf{I} - \mathbf{X}^*(\mathbf{X}^*\mathbf{KX}^*)^{-}\mathbf{X}^*\mathbf{K}$ is the orthogonal projector onto the null space of \mathbf{X}^* ($\text{Ker}(\mathbf{X}^*)$) in metric \mathbf{K} .) The other is

$$\mathbf{P}_{GQ_{1/K}} = \mathbf{P}_{GQ_{1/K}(\mathbf{Q}'_{1/K}\mathbf{KQ}_{1/K})^{-}\mathbf{R}} + \mathbf{P}_{Q_{GQ_{1/K}\mathbf{K}^+\mathbf{R}}GQ_{1/K}},\quad (33)$$

where the matrices that define the two projectors on the right hand side of (33) can further be written as

$$\mathbf{GQ}_{1/K}(\mathbf{Q}'_{1/K}\mathbf{KQ}_{1/K})^{-}\mathbf{R} = \mathbf{GQ}_{1/K}\mathbf{K}^+\mathbf{R},\quad (34)$$

and

$$\mathbf{Q}_{GQ_{1/K}K^+R}GQ_{1/K} = GQ_{1/K}Q_{Q_{1/K}K^+R/K} = GQ_{R^*/K}, \quad (35)$$

where $\mathbf{R}^* = [\mathbf{1}, \mathbf{Q}_{1/K}\mathbf{K}^+\mathbf{R}]$, and $^+$ indicates a Moore-Penrose inverse.

In the previous section, we set $\mathbf{X} = \mathbf{Q}_{1/K}\mathbf{T}_1$ and $\mathbf{R} = \mathbf{T}_2$ (respectively, $\mathbf{X} = \mathbf{Q}_{1/K}\mathbf{T}_2$ and $\mathbf{R} = \mathbf{T}_1$) to obtain the CNSCA of \mathbf{F} with \mathbf{T}_1 ignoring \mathbf{T}_2 , and \mathbf{T}_2 eliminating \mathbf{T}_1 (respectively, \mathbf{T}_2 ignoring \mathbf{T}_1 and \mathbf{T}_1 eliminating \mathbf{T}_2). The role of \mathbf{X} and \mathbf{R} can be interchanged by redefining $\mathbf{X} = \mathbf{K}^{-1}\mathbf{R}$ and $\mathbf{R} = \mathbf{K}\mathbf{X}$ systematically (Takane and Hunter, 2001; Section 5.1). By redefining $\mathbf{R} = \mathbf{K}\mathbf{X} = \mathbf{K}\mathbf{Q}_{1/K}\mathbf{T}_1$ by the latter, for instance, \mathbf{E} in (8) becomes \mathbf{B} in (6). (Note that the new definitions of \mathbf{X} and \mathbf{R} also satisfy (22) and (23), respectively.)

4 Decompositions into finer components

When \mathbf{T}_1 and/or \mathbf{T}_2 consist of more than one column, their effects can further be split into finer components. For explanatory purposes, suppose that we have already split the entire predictive relationship (\mathbf{A}) into \mathbf{B} (\mathbf{T}_1 ignoring \mathbf{T}_2), and \mathbf{E} (\mathbf{T}_2 eliminating \mathbf{T}_1). Let \mathbf{T}_1 be partitioned into \mathbf{T}_{11} and \mathbf{T}_{12} , i.e., $\mathbf{T}_1 = [\mathbf{T}_{11}, \mathbf{T}_{12}]$, and let $\mathbf{T}_2 = [\mathbf{T}_{21}, \mathbf{T}_{22}]$. Let \mathbf{G}_1 and \mathbf{H}_1 denote the selection matrices to select \mathbf{T}_{11} and \mathbf{T}_{12} from \mathbf{T}_1 , respectively. That is, $\mathbf{T}_{11} = \mathbf{T}_1\mathbf{G}_1$, and $\mathbf{T}_{12} = \mathbf{T}_1\mathbf{H}_1$. Let \mathbf{G}_2 and \mathbf{H}_2 do the same for \mathbf{T}_2 . Note that $\mathbf{G}'_1\mathbf{H}_1 = \mathbf{0}$, and $\text{Sp}(\mathbf{G}_1) \oplus \text{Sp}(\mathbf{H}_1) = \text{Sp}(\mathbf{T}'_1)$, and that analogous properties hold for \mathbf{G}_2 and \mathbf{H}_2 . We generally have

$$\mathbf{P}_{Y/K} = \mathbf{P}_{YC/K} + \mathbf{P}_{Y(Y'KY)^{-1}S/K} \quad (36)$$

(Takane and Yanai 1999), where \mathbf{C} and \mathbf{S} are such that $\mathbf{C}'\mathbf{S} = \mathbf{0}$, and $\text{Sp}(\mathbf{C}) \oplus \text{Sp}(\mathbf{S}) = \text{Sp}(\mathbf{Y}')$, and $\mathbf{P}_{Y/K} = \mathbf{Y}(\mathbf{Y}'\mathbf{K}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{K}$ is the orthogonal projector onto $\text{Sp}(\mathbf{Y})$ in metric \mathbf{K} . Using this decomposition, we obtain

$$\mathbf{B} = \mathbf{P}_{Q_{1/K}T_1G_1/K}\mathbf{K}^{-1}\mathbf{F} + \mathbf{P}_{Q_{1/K}T_1(T'_1Q'_{1/K}KT_1)^{-1}H_1/K}\mathbf{K}^{-1}\mathbf{F}, \quad (37)$$

and

$$\mathbf{E} = \mathbf{P}_{K^{-1}T_2G_2/K}\mathbf{K}^{-1}\mathbf{F} + \mathbf{P}_{K^{-1}T_2(T'_2K^{-1}T_2)^{-1}H_2/K}\mathbf{K}^{-1}\mathbf{F}. \quad (38)$$

The two terms on the right hand side of (37) represent the effect of \mathbf{T}_{11} ignoring both \mathbf{T}_{12} and \mathbf{T}_2 , and the effect of \mathbf{T}_{12} ignoring \mathbf{T}_2 but eliminating \mathbf{T}_{11} , respectively. Similarly, the two terms on the right hand side of (38) represent the effect of \mathbf{T}_{21} ignoring \mathbf{T}_{22} but eliminating \mathbf{T}_1 , and the effect of \mathbf{T}_{22} eliminating both \mathbf{T}_{21} and \mathbf{T}_1 . These effects (the effect of a particular set of variables eliminating some of other effects but ignoring the rest) can also be obtained by differences in fit between two models, as will be illustrated in the discussion section. Analogous decompositions hold for \mathbf{J} and \mathbf{L} as well.

One important message so far is that whenever we fit particular effects (say \mathbf{T}_1), there are two ways to do so. One pertains to the particular effects (\mathbf{T}_1) ignoring, and the other eliminating, other effects (\mathbf{T}_2). This is true despite

the fact that \mathbf{T}_1 and \mathbf{T}_2 , as have been postulated earlier, are orthogonal in metric \mathbf{I} . They are usually not orthogonal in metric \mathbf{K} , which gives rise to the distinction between the effects of ignoring and eliminating.

There are situations in which all contrast vectors in \mathbf{T}_1 and \mathbf{T}_2 are also orthogonal under the \mathbf{K} metric. One obvious situation is in which \mathbf{K} is constant diagonal. This is a sufficient condition, but not necessary. There are two exceptions in which the effects of ignoring and eliminating coincide no matter what \mathbf{K} is. One is in which \mathbf{T}_1 is null and $\text{Sp}(\mathbf{T}_2) = \text{Sp}(\mathbf{Q}_{1/K})$, and the other case is in which \mathbf{T}_2 is null and $\text{Sp}(\mathbf{T}_1) = \text{Sp}(\mathbf{Q}_{1/K})$. These two cases correspond with the unconstrained case of NSCA. There are also situations in which only \mathbf{T}_1 and \mathbf{T}_2 are orthogonal. When $k_1 = k_3$ and $k_2 = k_4$, where k_j is the j^{th} diagonal element of \mathbf{K} , \mathbf{T}_1 and \mathbf{T}_2 are orthogonal, but not the two contrast vectors in \mathbf{T}_2 . In the example data set given in Table 1, the interaction between treatment and gender is orthogonal to both the treatment and gender main effects, and vice versa, under the \mathbf{K} metric, but the two main effects are not mutually orthogonal. This implies that the interaction effect remains the same whether the joint effects of the two main effects are ignored or eliminated. The latter also remains the same whether the interaction effect is ignored or eliminated.

5 Relations to Other Methods

In this section, we discuss the relationships between the method proposed in this paper and other existing techniques. More rigorous proofs of the relationships between CNSCA and nonsymmetric versions of CCA, CALC, and PCCA will be given in the Appendix.

Böckenholt and Böckenholt (1990) developed CALC, which finds a row representation \mathbf{U} of a two-way contingency table under the restriction that $\mathbf{R}'\mathbf{U} = \mathbf{0}$, where they further chose $\mathbf{R} = \mathbf{T}_1$ (respectively, $\mathbf{R} = \mathbf{T}_2$). As has been seen already, the nonsymmetric version of their CALC (NSCALC) is equivalent to CNSCA with \mathbf{T}_2 ignoring \mathbf{T}_1 (respectively, \mathbf{T}_1 ignoring \mathbf{T}_2). In their original formulation, \mathbf{R} also contained $\mathbf{K}\mathbf{1}$, so that $\mathbf{1}'\mathbf{K}\mathbf{U} = \mathbf{0}$ also holds. Because in our formulations the effect of $\mathbf{K}\mathbf{1}$ was *a priori* eliminated, $\mathbf{1}'\mathbf{K}\mathbf{U} = \mathbf{0}$ is automatically satisfied.

Van der Heijden and Meijerink (1989) (see also Takane, Yanai, and Mayekawa (1991), Section 3.2) discussed the so-called zero average restriction, where the row representation is required to satisfy $\mathbf{R}'\mathbf{U} = \mathbf{0}$ as in CALC, but with $\mathbf{R} = \mathbf{K}\mathbf{T}_1$ (respectively, $\mathbf{R} = \mathbf{K}\mathbf{T}_2$). NSCALC with this restriction (denoted as NSCALC(\mathbf{R}) with $\mathbf{R} = \mathbf{K}\mathbf{T}_1$) turns out to be equivalent to CNSCA with \mathbf{T}_2 eliminating \mathbf{T}_1 (respectively, \mathbf{T}_1 eliminating \mathbf{T}_2). This amounts to GSVD($\mathbf{K}^{-1}(\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}^{-1})\mathbf{F}$) $_{K, I}$. Takane, Yanai, and Mayekawa (1991) discussed the zero average restriction as a special kind of CALC. After all, the only difference is whether we take $\mathbf{R} = \mathbf{T}_1$ (CALC) or $\mathbf{R} = \mathbf{K}\mathbf{T}_1$ (the zero average restriction). There was no discussion on the nature of the difference between the two in Takane, Yanai, and Mayekawa (1991). However, the

discussion in this paper makes clear that CALC pertains to ignoring, whereas the zero average restriction to eliminating.

CCA (ter Braak 1986) was developed as a method of finding the representation of row categories subject to linear constraints (i.e., $\mathbf{U} = \mathbf{X}\mathbf{U}^*$ for some \mathbf{U}^*), which is based on symmetric CA. Since CNSCA may be viewed as a nonsymmetric version of CCA (NSCCA; Takane and Jung 2009), equivalence can be readily established between CNSCA and NSCCA. From (26), NSCA(\mathbf{X}) with $\mathbf{X} = \mathbf{Q}_{1/K}\mathbf{T}_1$ (respectively, $\mathbf{X} = \mathbf{Q}_{1/K}\mathbf{T}_2$) is equivalent to CNSCA with \mathbf{T}_1 ignoring \mathbf{T}_2 (respectively, \mathbf{T}_2 ignoring \mathbf{T}_1), and NSCCA(\mathbf{X}) with $\mathbf{X} = \mathbf{K}^{-1}\mathbf{T}_1$ (respectively, $\mathbf{X} = \mathbf{K}^{-1}\mathbf{T}_2$) is equivalent to CNSCA with \mathbf{T}_1 eliminating \mathbf{T}_2 (respectively, \mathbf{T}_2 eliminating \mathbf{T}_1). Ter Braak (1986) only discusses the former (ignoring).

Ter Braak (1988) also developed partial CCA (PCCA) to eliminate the effect of extraneous variables (\mathbf{Z}) from the predictor variables (\mathbf{X}). In PCCA, residuals from \mathbf{X} left unaccounted for by \mathbf{Z} are taken as newly derived constraints in CCA. In typical applications of CCA, \mathbf{X} and \mathbf{Z} do not exhaust the entire space of $\mathbf{Q}_{1/K}$, so the situation is more like splitting \mathbf{T}_1 into $\mathbf{X} = \mathbf{T}_{11}$ and $\mathbf{Z} = [\mathbf{1}, \mathbf{T}_{12}]$. (Note that we need to include $\mathbf{1}$ as part of \mathbf{Z} to eliminate the effect of $\mathbf{K}\mathbf{1}$ from \mathbf{X} .) In this case, we obtain \mathbf{T}_{11} eliminating $[\mathbf{1}, \mathbf{T}_{12}]$ but ignoring \mathbf{T}_2 . (See Section 4 for the decomposition of a projector corresponding to this case.) That is, the complementary part to \mathbf{T}_1 is typically ignored. The nonsymmetric version of PCCA, denoted as NSPCCA(\mathbf{X}, \mathbf{Z}), can, however, be made equivalent to CNSCA with \mathbf{T}_1 eliminating \mathbf{T}_2 (respectively, \mathbf{T}_2 eliminating \mathbf{T}_1) by setting $\mathbf{X} = \mathbf{T}_1$ and $\mathbf{Z} = [\mathbf{1}, \mathbf{T}_2]$ (respectively, $\mathbf{X} = \mathbf{T}_2$ and $\mathbf{Z} = [\mathbf{1}, \mathbf{T}_1]$). This amounts to obtaining $\text{GSVD}(\mathbf{X}^*(\mathbf{X}^*\mathbf{K}\mathbf{X}^*)^{-1}\mathbf{X}^*\mathbf{F})_{K, I}$ where $\mathbf{X}^* = \mathbf{Q}_{Z/K}\mathbf{X}$. However, this has rarely been done, if at all.

D'Ambra and Lauro (1989) proposed nonsymmetric partial correspondence analysis (NSPCA) in the context of three-way contingency tables. Let us look at Table 1 again for illustration. NSPCA fits the gender main effect, and then fits the treatment main effect, and the interaction eliminating the gender main effect by conditioning on levels of gender. However, the conditioning strategy does not always work. For example, there is no way to perform CNSCA of \mathbf{F} with the treatment main effect eliminating both the gender main effect and the interaction effect by conditioning. This corresponds with the fact that log-linear analysis allows no closed-form solution when both the gender main effect and the interaction effect have to be fitted (van der Heijden and de Leeuw 1985; see the discussion section for examples).

6 Permutation Tests and the Bootstrap Method

One important aspect of NSCA and CNSCA is low dimensional displays of predictive relationships between the rows and columns of a contingency table via GSVD of relevant matrices. How many dimensions we should retain in these displays is extremely important in practical applications of (C)NSCA. Unfortunately, the CATANOVA test statistic cannot be used for this purpose.

So there remains a question of determining how many singular values are statistically significant. We use permutation tests for dimensionality selection, which have proven useful in similar contexts (e.g., Takane and Hwang 2006; Takane and Jung 2009). They are easy to apply, computationally simple, and do not require any specific distributional assumptions. Note that this is just for getting a rough idea as to how many components are reliable enough and so worth interpreting. (These tests are not designed to identify the “correct” dimensionality or the number of empirically meaningful components.)

The permutation tests work as follows for CNSCA: First, singular values (sv) are computed from the original data set. Then, the rows of predictor variables are randomly permuted, and sv’s are calculated from the permuted data set. The largest sv from the permuted data set is compared with that from the original data set. To test the statistical significance of the sv from the original data set, we repeat the same procedure many times (say, 1000), and count how many times the former is larger than the latter. If this count is smaller than 1000α (where α is the prescribed significance level), the largest sv being tested is significantly different from 0. Each subsequent sv can be tested in the same way after eliminating the effects of the preceding sv’s. Note that the square of a singular value indicates the sum of squares in the table that is explained by the corresponding component. The best dimensionality is the number of significant sv’s. (See Legendre and Legendre (1998) for more general discussions on permutation tests in similar contexts.)

The bootstrap method (Efron and Tibshirani 1993) provides a way of assessing the stability of parameter estimates. A form of nonparametric bootstrapping is used to derive empirical distributions of parameter estimates. To illustrate, we repeatedly draw random bootstrap samples from the original data set with replacement. We analyze each of the bootstrap samples by CNSCA to obtain parameter estimates. We then calculate means and variances of the estimates, from which we estimate biases and standard errors. The bootstrap method may be used to test whether the estimated parameters are significantly positive or negative, and to construct 95% confidence regions (Ramsay 1978). Suppose that an estimate from the original data set is positive. We count the number of times that the estimate of the same parameter is negative in bootstrap samples. If the relative frequency of the bootstrap estimates crossing over zero is less than a prescribed significance level (e.g., .05 or .01), we conclude that it is significantly positive.

7 Examples of Application

In this section, we present two detailed applications of the proposed method. The first application concerns an example of Simpson’s paradox, and shows how the two parts of the CNSCA analysis help understand this paradox. The second application involves a three-way contingency table that was previously analyzed by CALC by Böckenholt and Böckenholt (1990).

7.1 Health Recovery Data

Let us begin by looking at Table 2 obtained by collapsing Table 1 across genders. The table shows that there is not much difference in the recovery rate between the medication group and the control group (in fact, the recovery rate is slightly higher for the medication group). Now let us look at Table 1 again. This table indicates that the recovery rate is higher for the control group for both male and female groups, which is the opposite from what we have just observed in Table 2. This kind of discrepancy is called Simpson's paradox (Blyth 1972; Shapiro 1982).

The difference is presumed to arise from the fact that Table 1 takes into account the effect of gender, whereas Table 2 ignores it. The treatment effect (T) in Table 1 represents the main effect of T eliminating the gender main effect (G). The treatment effect we saw in Table 2, on the other hand, corresponds to the T main effect ignoring the G main effect. What should we do about the interaction between the two (the GT interaction)? Should we ignore it or eliminate it in evaluating the kinds of T main effect? Ideally, we should look at both. In this particular data set, however, the interaction effect happens to be orthogonal to both the T and G main effects, as has been discussed at the end of Section 3. Thus, whether we ignore or eliminate the interaction makes no difference. This considerably simplifies the analyses to follow.

Let \mathbf{T}_1 and \mathbf{T}_2 be as given in (1) and (2), respectively. Let us be reminded that \mathbf{T}_1 represents the T main effect, and \mathbf{T}_2 the G main effect and the GT interaction. If we use this \mathbf{T}_1 in (6), we obtain the CNSCA of \mathbf{F} with \mathbf{T}_1 ignoring \mathbf{T}_2 (this analysis is equivalent to unconstrained NSCA of Table 2), and if we use this \mathbf{T}_2 in (8), we obtain the analysis of its residuals, that is, the CNSCA of \mathbf{F} with \mathbf{T}_2 eliminating \mathbf{T}_1 . If, on the other hand, we use the \mathbf{T}_1 in (10), we obtain the CNSCA of \mathbf{F} with \mathbf{T}_1 eliminating \mathbf{T}_2 , and if we use the \mathbf{T}_2 in (9), we obtain the CNSCA of \mathbf{F} with \mathbf{T}_2 ignoring \mathbf{T}_1 . For Simpson's paradox, we are interested in comparing the test of \mathbf{T}_1 ignoring \mathbf{T}_2 and that of \mathbf{T}_1 eliminating \mathbf{T}_2 .

There are several other effects of interest as preliminary to examining Simpson's paradox. One relates to the conditions related to "collapsibility" (e.g., Agresti 2004, Chapter 9), and another pertains to the GT interaction. The collapsibility conditions (see below for more details) are sufficient for nonexistence of Simpson's paradox. It is important to see them fail as prerequisites to examining Simpson's paradox. The test of the interaction effect is important because it concerns whether the effect of treatment remains the same across genders. That the interaction effect is insignificant is another prerequisite for Simpson's paradox.

The collapsibility holds, in the context of CNSCA, if there is no joint effect of the G main effect and the GT interaction after eliminating the T main effect, or if the G and T main effects are statistically independent and there is no GT interaction (Wermuth 1987). The first condition is tested by the test of \mathbf{E} (\mathbf{T}_2 eliminating \mathbf{T}_1) in Table 3. The condition clearly fails because the CATANOVA statistic for testing this effect is highly significant ($C_E = 74.95$

Table 3 Summary of the CATANOVA tests for the data in Table 1

Matrix	Constraints	Goodman-Kruskal	CATANOVA	
		τ index	C -statistic	(d.f.)
A	Unconstrained	.0944	75.45	(3)
B	\mathbf{T}_1 ignoring \mathbf{T}_2	.0006	.50	(1)
E	\mathbf{T}_2 eliminating \mathbf{T}_1	.0938	74.95	(2)
J	\mathbf{T}_2 ignoring \mathbf{T}_1	.0775	61.96	(2)
L	\mathbf{T}_1 eliminating \mathbf{T}_2	.0169	13.49	(1)
B with new \mathbf{T}_1	GT interaction	.0018	1.50	(1)

with 1 df, $p \ll .01$). The second condition also fails because the two main effects are not marginally independent, since $300 \times 300 - 100 \times 100 \neq 0$. The GT interaction is insignificant ($C_{Int.} = 1.50$ with 1 df, $p > .05$), as may be observed in the last row of Table 3. (When testing the interaction effect, whether we should ignore or eliminate the effect of the two main effects is again an issue. Fortunately, however, in this particular data set, the interaction effect remains the same whether we ignore or eliminate the two main effects for the same reason that the two main effects remain identical whether we ignore or eliminate the interaction. The unique interaction effect in this case can be analyzed in a number of ways, e.g., by defining \mathbf{T}_1 by the second column of \mathbf{T}_2 in (2), and use it in (6).)

Now we may look at other results reported in Table 3. We observe that the Goodman-Kruskal τ index and the CATANOVA C -statistic (along with its degrees of freedom) for \mathbf{T}_1 ignoring \mathbf{T}_2 (Matrix **B**) and \mathbf{T}_2 eliminating \mathbf{T}_1 (matrix **E**) add up to the same quantities for unconstrained NSCA (Matrix **A**), and similarly for \mathbf{T}_2 ignoring \mathbf{T}_1 (Matrix **J**) and \mathbf{T}_1 eliminating \mathbf{T}_2 (Matrix **L**). These observations are consistent with our theory presented in Section 2. Given the prerequisites for Simpson's paradox given above, we may compare the test of \mathbf{T}_1 ignoring \mathbf{T}_2 and that of \mathbf{T}_1 eliminating \mathbf{T}_2 . The CATANOVA test of the former is not significant ($C_B = .50$ with 1 df, $p > .05$), whereas that of the latter is significant ($C_L = 13.49$ with 1 df, $p < .01$). This observation is in agreement with the fact that in Table 2 the difference between the control and medication groups is rather small (although the latter has a slightly better recovery rate), while in Table 1 the control group has a better recovery rate for both males and females, indicating Simpson's paradox.

The CATANOVA tests, however, do not provide information regarding the direction of the effects. For this we may look at the dimension reduction results by GSVD, although in the present case the dimension reduction aspect of GSVD per se is not of much interest because there are only two criterion categories, for which the maximum number of extractable dimensions is one. Permutation tests in this case should give similar results to those obtained by the CATANOVA tests, as indeed was the case ($s_1^2 = .25, p > .418$ for \mathbf{T}_1 ignoring \mathbf{T}_2 , where s_1^2 indicates the sum of squares (the variation) in the contingency table that can be explained by the first component, which is equal to the squared largest singular value, and $s_1^2 = 6.75, p < .001$ for \mathbf{T}_1 eliminating

Table 4 The principal coordinates (Prn) of predictor categories, and the standard coordinates (Std) of criterion categories by CNSCA for the health recovery data. (Bootstrap standard error (SE) estimates are given in parentheses. “*” indicates a significance at the 5% level, and “***” at the 1% level.)

Crit	\mathbf{T}_1 ignoring \mathbf{T}_2		\mathbf{T}_1 eliminating \mathbf{T}_2	
	Std	(SE)	Std	(SE)
R	** .707	(.000)	** .707	(.000)
N	** -.707	(.000)	** -.707	(.000)
Pred	Prn	(SE)	Prn	(SE)
M-M	.018	(.025)	** -.053	(.013)
C-M	-.018	(.025)	** .159	(.040)
M-F	.018	(.025)	** -.159	(.040)
C-F	-.018	(.025)	** .053	(.013)

\mathbf{T}_2). Nonetheless, it is interesting to see how predictor categories are related to criterion categories through the coordinate values obtained by the GSVD. Table 4 gives estimates of principal coordinates of predictor categories, and standard coordinates of criterion categories with \mathbf{T}_1 ignoring \mathbf{T}_2 , and with \mathbf{T}_1 eliminating \mathbf{T}_2 along with their standard error estimates (in parentheses) obtained by the bootstrap method. In both cases, the recovery group takes a positive coordinate value. For \mathbf{T}_1 ignoring \mathbf{T}_2 , the medication groups take a slightly positive coordinate value, indicating a slightly closer relationship with recovery. For \mathbf{T}_1 eliminating \mathbf{T}_2 , the control groups are closer to the recovery group for both males and females. That is, the predictive relationship between treatment and prognosis is reversed in the two. These results are consistent with our earlier observation that Simpson’s paradox exists in this table. (That the coordinates of the criterion categories for \mathbf{T}_1 ignoring \mathbf{T}_2 are significant at the 1% significance level with 0 standard error is an artifact because the standard coordinates are always $.707 = \sqrt{2}/2$ or $-.707$ with two criterion categories, and with the reflection of singular vectors that is permitted they can always be made to coincide.) This data set will again be taken up in the discussion section.

7.2 Abortion Data

The second example pertains to the analysis of a $3 \times 3 \times 3$ contingency table obtained by cross-classifying subjects by religion, education, and attitude toward abortion. This data set was previously analyzed by Böckenholt and Böckenholt (1990; see also Takane, Yanai, and Mayekawa (1991)). The first two variables are interactively coded, constituting predictor categories, while the third variable is taken as the criterion variable. Religion has three categories: (np) Northern Protestants, (sp) Southern Protestants, and (ct) Catholic. The symbols in parentheses are then combined with numbers that indicate years of education of the subjects coded as 1 = less than 8 years, 2 = between 9 and 12, and 3 = more than 13. The resultant nine predictor categories are labeled as

Table 5 Attitudes toward non-therapeutic abortion: Cross classification by religion and education of respondents ($n = 3181$)

Religion	Abbr.	Education (Years)	Attitudes		
			Positive (pt)	Neutral (nt)	Negative (ng)
Northern Protestant	np1	≤ 8	49	46	115
	np2	9 – 12	293	140	277
	np3	≥ 13	244	66	100
Southern Protestant	sp1	≤ 8	27	34	117
	sp2	9 – 12	134	98	167
	sp3	≥ 13	138	38	73
Catholic	ct1	≤ 8	25	40	88
	ct2	9 – 12	172	103	312
	ct3	≥ 13	93	57	135

np1, np2, np3, sp1, sp2, sp3, ct1, ct2, and ct3. The criterion variable consists of three attitudes toward abortion: positive (pt), neutral (nt), and negative (ng). (These abbreviations are used to label categories in subsequent reports.) We denote the resultant 9 by 3 table by \mathbf{F} , which is given in Table 5.

Define

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ -2 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & -2 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{matrix} \text{np1} \\ \text{np2} \\ \text{np3} \\ \text{sp1} \\ \text{sp2} \\ \text{sp3} \\ \text{ct1} \\ \text{ct2} \\ \text{ct3} \end{matrix}, \text{ and } \mathbf{T}_2 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ -1 & 0 & 1 & 1 \\ 1 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & -1 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & -2 \\ 0 & -1 & 0 & -2 \end{bmatrix} \begin{matrix} \text{np1} \\ \text{np2} \\ \text{np3} \\ \text{sp1} \\ \text{sp2} \\ \text{sp3} \\ \text{ct1} \\ \text{ct2} \\ \text{ct3} \end{matrix}$$

in line with Takane, Yanai, and Mayekawa (1991). The first three columns of \mathbf{T}_1 indicate quadratic trends over the three education levels for each of the three religious groups, while the last column represents the interaction between the first two levels of religion (np and sp) and the first two levels of education (the low and medium education levels). The first column of \mathbf{T}_2 , on the other hand, represents the same linear trend over the three education levels for the first two levels of religion, while the second column represents a separate linear trend over the level of education for the catholic. The last two columns of \mathbf{T}_2 together represent the main effects of religion.

These two constraint matrices are motivated as follows: An unconstrained correspondence analysis was first applied to the data in Table 5 (Gilula and Haberman, 1988). The analysis revealed that the solution was essentially unidimensional, and that given religion, the scores on this dimension looked approximately linear in the three education levels. Furthermore, the linear spacing seemed similar for the two protestant groups, while different for the catholic. The first two columns of \mathbf{T}_2 capture these tendencies. There were also differences among the three religious groups in the degree of favorableness toward

Table 6 Summary of the CATANOVA tests for the abortion data

Matrix	Constraints	Goodman-Kruskal	CATANOVA	
		τ index	C-statistic	(d.f.)
A	Unconstrained	.0488	310.59	(16)
B	\mathbf{T}_1 ignoring \mathbf{T}_2	.0039	24.87	(8)
E	\mathbf{T}_2 eliminating \mathbf{T}_1	.0449	285.72	(8)
J	\mathbf{T}_2 ignoring \mathbf{T}_1	.0471	299.81	(8)
L	\mathbf{T}_1 eliminating \mathbf{T}_2	.0017	10.77	(8)

abortion. The last two columns in \mathbf{T}_2 capture these differences (the main effects of religion). On the other hand, \mathbf{T}_1 was derived as spanning the ortho-complement subspace to \mathbf{T}_2 . Because of the way these two constraint matrices were derived, we expect that \mathbf{T}_2 captures a majority of the total predictability, while \mathbf{T}_1 only a minor portion.

We first applied the CATANOVA tests for unconstrained NSCA of \mathbf{F} as well as for various CNSCA induced by the above \mathbf{T}_1 and \mathbf{T}_2 . The results are summarized in Table 6, which is similar to Table 3 seen previously. As before, we see similar additivity in Goodman-Kruskal's τ index and the CATANOVA C -statistic for \mathbf{T}_1 ignoring \mathbf{T}_2 and \mathbf{T}_2 eliminating \mathbf{T}_1 , and also for \mathbf{T}_2 ignoring \mathbf{T}_1 and \mathbf{T}_1 eliminating \mathbf{T}_2 . Four τ 's are significantly different from 0, while the one associated with \mathbf{T}_1 eliminating \mathbf{T}_2 is insignificant. As expected, the effect of \mathbf{T}_2 (regardless of ignoring or eliminating) is much more dominant than that of \mathbf{T}_1 .

Böckenholt et al. (1990) applied $\text{CALC}(\mathbf{T}_1)$ to this data set. The nonsymmetric version of CALC is equivalent to CNSCA of \mathbf{F} with \mathbf{T}_2 ignoring \mathbf{T}_1 . We look at this analysis a little more closely. Permutation tests indicated that the first component was highly significant ($s_1^2 = 94.89, p < .001$), while the second was not ($s_2^2 = 0.49, p > .57$). We also look at CNSCA of \mathbf{F} with \mathbf{T}_2 eliminating \mathbf{T}_1 . Permutation tests for this analysis indicated, similarly to the above analysis, that the first component was highly significant ($s_1^2 = 90.60, p < .001$), while the second was not ($s_2^2 = 0.29, p > .744$).

Table 7 gives estimates of principal coordinates for predictor categories, and standard coordinates for criterion categories pertaining to the first component derived from CNSCA with \mathbf{T}_2 ignoring \mathbf{T}_1 , and with \mathbf{T}_2 eliminating \mathbf{T}_1 as well as their standard error estimates (in parentheses) obtained by the bootstrap method. In both cases, the positive side of this component represents a more favorable attitude toward abortion (pt is positive, and ng is negative). We see that more highly educated people tend to be more positive toward abortion than less educated people, as do protestants than catholic. Overall, there is not much difference in the effects of \mathbf{T}_2 between ignoring and eliminating \mathbf{T}_1 .

One final analysis concerns the CNSCA of \mathbf{F} with \mathbf{T}_1 ignoring \mathbf{T}_2 . This analysis is for the sake of completeness only. It is complementary to the CNSCA of \mathbf{T}_2 eliminating \mathbf{T}_1 . (The effect of \mathbf{T}_1 eliminating \mathbf{T}_2 was not significant by the CATANOVA test, and consequently no further analysis is empirically warranted.) In this analysis, permutation tests indicated that the first compo-

Table 7 The principal coordinates (Prn) of predictor categories, and the standard coordinates (Std) of criterion categories by CNSCA for the abortion data. (Bootstrap standard error (SE) estimates are given in parentheses. “**” indicates a significance at the 5% level, and “***” at the 1% level.)

Crit	\mathbf{T}_2 ignoring \mathbf{T}_1		\mathbf{T}_2 eliminating \mathbf{T}_1	
	Std	(SE)	Std	(SE)
pt	**-.749	(.016)	**-.752	(.016)
nt	*-.092	(.041)	*-.099	(.041)
ng	**-.657	(.024)	**-.652	(.025)
Pred	Prn	(SE)	Prn	(SE)
np1	**-.192	(.026)	**-.138	(.031)
np2	**-.059	(.013)	**-.051	(.008)
np3	**-.310	(.020)	**-.249	(.020)
sp1	**-.277	(.026)	**-.371	(.035)
sp2	-.026	(.018)	-.001	(.012)
sp3	**-.225	(.026)	**-.261	(.028)
ct1	**-.224	(.034)	**-.291	(.039)
ct2	**-.133	(.016)	**-.061	(.008)
ct3	-.043	(.031)	**-.096	(.031)

ment was significant ($s_1^2 = 5.17, p < .03$), as was the second ($s_2^2 = 2.75, p < .05$). Thus, the best dimensionality was found to be two, despite the fact that the size of the overall effect of \mathbf{T}_1 ignoring \mathbf{T}_2 is much smaller than that of \mathbf{T}_2 eliminating \mathbf{T}_1 . We present the two-dimensional configuration as the best solution in Figure 1. The predictive power of a particular predictor category on a particular criterion category can be evaluated by the magnitude of the inner product between the two vectors representing the two categories. For example, sp1 is closest to pt (the positive attitude toward abortion). This means that southern protestants with less than 8 years of education are most favorable to abortion. This is fundamentally different from what we have observed in Table 7, where more highly educated protestants tend to favor abortion most. This is because the two analyses reflect two independent portions of the relationships between rows and columns of the same contingency table. (Figure 1 represents \mathbf{T}_1 ignoring \mathbf{T}_2 , whereas Table 7 \mathbf{T}_2 eliminating \mathbf{T}_1 .)

8 Summary and Discussion

This paper presented a method of analyzing two complementary parts of the predictive relationships between the rows and columns of a contingency table, one part that can be explained by the constraints on predictive categories, and the other that cannot be explained by the constraints. The former pertains to the effects of the constraints ignoring other effects, while the latter to the other effects eliminating the effects of the constraints. The method partitions the Goodman-Kruskal (1954) τ index as well as the CATANOVA C -statistic, which allows asymptotic tests of the two complementary parts, the test of ignoring and the test of eliminating. These tests help understand Simpson’s

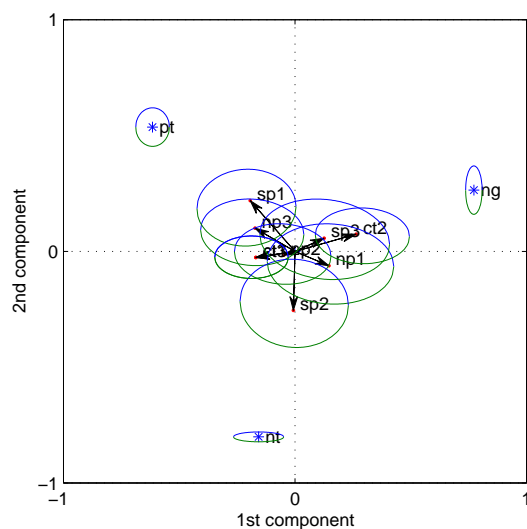


Fig. 1 Two-dimensional configuration of the abortion data obtained by the CNSCA with T_1 ignoring T_2 . Asterisks indicate criterion categories (attitudes towards abortion), and arrows indicate predictor categories. To make the graph look nicer, the estimated principal coordinates of predictor categories are multiplied by 3, and the size of confidence regions for these coordinates adjusted accordingly.

paradox, in which the effect of one variable on another depends on whether or not the effect of a third variable is taken into account.

One important implication of the above is that whenever we fit particular effects, there are two ways to do so, one ignoring, and the other eliminating, other effects. They almost always produce distinct effects having distinct meanings. The distinction arises whenever predictor variables are correlated as in multiple regression analysis, where the correlated predictor variables are rules than exceptions, and in nonorthogonal ANOVA (Maxwell and Delaney 2004), where the confounding between the effects arises from disproportionate numbers of observations per cell. Quite often, however, only one of them is highlighted, neglecting the other. For example, in CA we tend to focus more on the effects of ignoring, whereas in log-linear analysis we do just the opposite.

The two distinct effects of the same variables, however, necessarily complicate the situation. Table 8 shows the results of fitting all possible CNSCA models to the health recovery data (Example 1) discussed earlier. With three predictor variables (the treatment main effect (T), the gender main effect (G), and the interaction between them (GT)), there are six possible models (Models (2) to (7) in the table) excluding the independence model (1) and the saturated model (8). Two distinct effects are associated with each of these six models, the effects of ignoring in the second and third columns of the table and those eliminating in the last two columns. There are six other identifiable effects, all of which represent the effects of some variables eliminating certain

Table 8 The CATANOVA C statistics for all possible CNSCA models.

Model	Effects ignoring		Effects eliminating	
	other effects	χ^2 (df)	other effects	χ^2 (df)
(1)	independence	0 (0)	T,G,GT	75.45 (3)
(2)	G	60.46 (1)	T,GT	14.99 (2)
(3)	T	.50 (1)	G,GT	74.95 (2)
(4)	GT	1.50 (1)	G,T	73.95 (2)
(5)	G,T	73.95 (2)	GT	1.50 (1)
(6)	G,GT	61.96 (2)	T	13.49 (1)
(7)	T,GT	2.00 (2)	G	73.45 (1)
(8)	saturated (G,T,GT)	75.45 (3)	\emptyset	0 (0)

others but ignoring the rest. In the table, these effects can be calculated by the difference between two models. For example, G eliminating T but ignoring GT is obtained by the difference between Models (5) and (2). (See Section 4 for direct calculation of these quantities.) In any case, there are 20 identifiable effects altogether. This number grows very quickly as the number of predictor variables increases, and soon becomes unmanageable. One good strategy to avoid all these complications is, if at all possible, to design the study in such a way to make row marginal frequencies equal, so that the effects of ignoring and eliminating are identical.

The same phenomenon occurs regularly in log-linear analysis of contingency tables, although the problem has rarely been pointed out in the literature (see, however, Cheng, Liou, Liou, and Aston 2006). In log-linear analysis, performed tests are usually those of eliminating, but no consensus seems to exist that there are two distinct tests for the effects of the same variable. For comparison with CNSCA, Table 9 presents the results of fitting all possible predictive log-linear models (which are essentially the same as the logistic models with categorical predictor variables). In the table, P stands for Prognosis (the criterion variable). The independence model includes the row marginal effects (T, G, and TG), and P, which serves as the baseline model. All other models add one or more effects to this baseline model. For example, GP is added to the independence model to obtain Model (2), which is indicated by symbol “+GP”. This is analogous to adding the G main effect to the predictor set in CNSCA. In log-linear analysis, residual effects (corresponding to the effects of eliminating) reported in the last two columns of Table 9 are typically given. In the present case, the fitted effects (corresponding to the effects of ignoring) reported in the second and third columns of the table had to be recovered by subtracting the chi-square value corresponding to the residual effects from that of the saturated model. Overall, Table 9 looks very much like Table 8, and conclusions similar to those drawn from CNSCA can also be drawn from the log-linear analysis. Note, however, that certain log-linear models (those marked by # in the table) cannot be fitted in closed form, and some iterative estimation procedure is necessary. (Log-linear models that admit closed-form solutions (those without #) are in fact identical to the corresponding CN-

Table 9 The log-likelihood chi-square for all possible log-linear logistic models (logistic models with categorical predictor variables). The superscript # indicates that the model cannot be fitted in closed form.

Model	Fitted effects	χ^2 (df)	Residual effects	χ^2 (df)
	independence			
(1)	(G,T,GT,P)	0 (0)	GP,TP,GTP	78.67 (3)
(2)	+GP	61.33 (1)	TP,GTP	17.34 (2)
(3)	+TP	.50 (1)	GP,GTP	78.17 (2)
(4)	+GTP	1.51 (1)	GP,TP	77.16 (2)
(5)	+GP,TP#	76.54 (2)	GTP	2.13 (1)
(6)	+GP,GTP#	62.96 (2)	TP	15.71 (1)
(7)	+TP,GTP#	2.01 (2)	GP	76.66 (1)
(8)	saturated	78.67 (3)	\emptyset	0 (0)
	(+GP,TP,GTP)			

SCA models. However, the chi-square values differ due to the difference in statistics used, the likelihood ratio chi-square in log-linear analysis, and the CATANOVA C -statistic in CNSCA.)

In this paper, the two parts of the analysis are almost exclusively discussed in the context of NSCA. Essentially the same thing can be done for symmetric CA as well. In this case, we analyze the mutual relationship between the rows and columns, and instead of decomposing the Goodman-Kruskal τ index and the CATANOVA statistic, we decompose Cramér’s contingency coefficient and the Pearson’s chi-square statistic in a manner analogous to the former. Technically, the only difference in symmetric CA is to postmultiply matrices involved (\mathbf{A} , \mathbf{B} , \mathbf{E} , etc.) by \mathbf{D}^{-1} , and use \mathbf{D} as the column metric in GSVD rather than the identity column metric. One possible ramification of the procedure in this case may be to incorporate constraints similar to those imposed on rows in columns of contingency tables as well.

A comment is in order on a popular method in ecology (e.g., Bocard, Legendre, and Drapeau 1992; Anderson and Gribble 1998; Økland, 2003). This method attempts to “partition” the total variation in the criterion variable into non-overlapping portions. For example, in the simplest case of two sets of predictor variables, there is a portion that can only be explained by each of the two sets (the unique variation), and there is a portion that can be explained by both of them (the shared variation). The latter is defined as the joint variation (the variation that can be jointly explained by the two) minus the sum of the unique variations. One difficulty with this procedure is that this quantity can be negative (the sum of the unique variations exceeds the joint variation). Thus, the name “shared variation” is conceptually inadequate. In addition, the effects unaccounted for by the two sets of variables are simply ignored. The possibility of eliminating these effects seems totally out of sight.

The analysis presented in this paper is based on the asymptotic behavior of the CATANOVA statistic. Little has been done so far on the small sample behavior of the statistic (but see Margolin and Light (1974) for some attempt to compare this statistic against other competing statistics such as Pearson’s

chi-square statistic, deviance, etc. in small samples). Obviously more systematic studies are necessary on this point. The proposed method works strictly within the least squares (LS) framework. It is relatively straightforward to extend the method to allow the ridge type of regularized LS estimation (Hoerl and Kennard 1971). As has been demonstrated by Takane and Jung (2009), this type of estimation tends to provide estimates of model parameters which are on average closer to true population values in similar contexts.

9 Appendix

In this appendix, we show equivalences between CNSCA of \mathbf{T}_1 eliminating \mathbf{T}_2 and several existing methods, namely NSCCA(\mathbf{X}), NSCALC(\mathbf{R}), and NSPCCA(\mathbf{X}, \mathbf{Z}) by defining \mathbf{X} , \mathbf{R} , and \mathbf{Z} appropriately. These methods are nonsymmetric versions of CCA (ter Braak 1986), CALC (Böckenholt and Böckenholt 1990), and PCCA (ter Braak 1988), respectively.

Throughout this appendix, we set $\mathbf{T}_2^* = [\mathbf{1}, \mathbf{T}_2]$. The CNSCA of \mathbf{T}_1 eliminating \mathbf{T}_2 amounts to:

$$\text{GSVD}(\mathbf{K}^{-1}\mathbf{T}_1(\mathbf{T}_1'\mathbf{K}^{-1}\mathbf{T}_1)^{-1}\mathbf{T}_1'\mathbf{K}^{-1}\mathbf{F})_{K,I}, \quad (39)$$

(this is the same as (10)), which is equivalent to:

$$\text{GSVD}(\mathbf{K}^{-1}\mathbf{Q}'_{1/K}(\mathbf{I} - \mathbf{K}\mathbf{T}_2(\mathbf{T}_2'\mathbf{Q}'_{1/K}\mathbf{K}\mathbf{T}_2)^{-1}\mathbf{T}_2'\mathbf{Q}'_{1/K})\mathbf{F})_{K,I} \quad (40)$$

due to Khatri's lemma (Khatri, 1966; Takane, Yanai, and Mayekawa 1991, p. 675).

NSCCA(\mathbf{X}), on the other hand, obtains

$$\text{GSVD}(\mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{F})_{K,I}. \quad (41)$$

With $\mathbf{X} = \mathbf{K}^{-1}\mathbf{T}_1$, (41) reduces to (39).

NSCALC(\mathbf{R}) obtains

$$\text{GSVD}(\mathbf{K}^{-1}(\mathbf{I} - \mathbf{R}(\mathbf{R}'\mathbf{K}^{-1}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}^{-1})\mathbf{F})_{K,I}. \quad (42)$$

With $\mathbf{R} = \mathbf{K}\mathbf{T}_2^*$, this reduces to:

$$\text{GSVD}(\mathbf{K}^{-1}(\mathbf{I} - \mathbf{K}\mathbf{T}_2^*(\mathbf{T}_2^{*'}\mathbf{K}\mathbf{T}_2^*)^{-1}\mathbf{T}_2^{*'})\mathbf{F})_{K,I}, \quad (43)$$

where

$$\begin{aligned} \mathbf{K}\mathbf{T}_2^*(\mathbf{T}_2^{*'}\mathbf{K}\mathbf{T}_2^*)^{-1}\mathbf{T}_2^{*'} &= \mathbf{P}'_{\mathbf{T}_2^*/K} \\ &= \mathbf{P}_{\mathbf{K}\mathbf{T}_2^*/K^{-1}} \\ &= \mathbf{P}_{\mathbf{K}1/K^{-1}} + \mathbf{P}_{\mathbf{Q}_{\mathbf{K}1/K^{-1}}\mathbf{K}\mathbf{T}_2/K^{-1}} \end{aligned} \quad (44)$$

and $\mathbf{P}_{Y/K} = \mathbf{Y}(\mathbf{Y}'\mathbf{K}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{K}$. The last term in the above equation can be further rewritten as

$$\begin{aligned} & \mathbf{P}_{\mathbf{Q}_{K1/K^{-1}KT_2/K^{-1}}} \\ &= \mathbf{Q}'_{1/K}\mathbf{K}\mathbf{T}_2(\mathbf{T}'_2\mathbf{K}\mathbf{Q}_{1/K}\mathbf{K}^{-1}\mathbf{Q}'_{1/K}\mathbf{K}\mathbf{T}_2)^{-1}\mathbf{T}'_2\mathbf{K}\mathbf{Q}_{1/K}\mathbf{K}^{-1} \\ &= \mathbf{Q}'_{1/K}\mathbf{K}\mathbf{T}_2(\mathbf{T}'_2\mathbf{Q}'_{1/K}\mathbf{K}\mathbf{T}_2)^{-1}\mathbf{T}'_2\mathbf{Q}'_{1/K}. \end{aligned} \quad (45)$$

Hence,

$$\begin{aligned} & \mathbf{K}^{-1}(\mathbf{I} - \mathbf{K}\mathbf{T}_2^*(\mathbf{T}'_2\mathbf{K}\mathbf{T}_2^*)^{-1}\mathbf{T}'_2)\mathbf{F} = \\ & \mathbf{K}^{-1}\mathbf{Q}'_{1/K}(\mathbf{I} - \mathbf{K}\mathbf{T}_2(\mathbf{T}'_2\mathbf{Q}'_{1/K}\mathbf{K}\mathbf{T}_2)^{-1}\mathbf{T}'_2\mathbf{Q}'_{1/K})\mathbf{F}, \end{aligned} \quad (46)$$

whose GSVD is nothing but (40).

NSPCCA(\mathbf{X}, \mathbf{Z}) reduces to:

$$\text{GSVD}(\mathbf{X}^*(\mathbf{X}^*\mathbf{K}\mathbf{X}^*)^{-1}\mathbf{X}^*\mathbf{F})_{K,I}, \quad (47)$$

where

$$\mathbf{X}^* = \mathbf{Q}_{Z/K}\mathbf{X} = \mathbf{Q}_{T_2^*/K}\mathbf{T}_1, \quad (48)$$

when $\mathbf{X} = \mathbf{T}_1$ and $\mathbf{Z} = \mathbf{T}_2^*$. Note that

$$\mathbf{K}\mathbf{Q}_{T_2^*/K} = \mathbf{T}_1(\mathbf{T}'_1\mathbf{K}\mathbf{T}_1)^{-1}\mathbf{T}'_1 = \mathbf{Q}'_{T_2^*/K}\mathbf{K} \quad (49)$$

by Khatri's lemma. Then,

$$\begin{aligned} & \mathbf{X}^*(\mathbf{X}^*\mathbf{K}\mathbf{X}^*)^{-1}\mathbf{X}^*\mathbf{F} \\ &= \mathbf{Q}_{T_2^*/K}\mathbf{T}_1(\mathbf{T}'_1\mathbf{K}\mathbf{Q}_{T_2^*/K}\mathbf{T}_1)^{-1}\mathbf{T}'_1\mathbf{Q}'_{T_2^*/K}\mathbf{F} \\ &= \mathbf{K}^{-1}\mathbf{K}\mathbf{Q}_{T_2^*/K}\mathbf{T}_1(\mathbf{T}'_1\mathbf{K}\mathbf{Q}_{T_2^*/K}\mathbf{T}_1)^{-1}\mathbf{T}'_1\mathbf{Q}'_{T_2^*/K}\mathbf{K}\mathbf{K}^{-1}\mathbf{F} \\ &= \mathbf{K}^{-1}\mathbf{T}_1(\mathbf{T}'_1\mathbf{K}^{-1}\mathbf{T}_1)^{-1}\mathbf{T}'_1\mathbf{T}_1(\mathbf{T}'_1\mathbf{T}_1)^{-1} \times \\ & \quad \mathbf{T}'_1\mathbf{K}^{-1}\mathbf{T}_1(\mathbf{T}'_1\mathbf{T}_1)^{-1}\mathbf{T}'_1\mathbf{T}_1(\mathbf{T}'_1\mathbf{K}^{-1}\mathbf{T}_1)^{-1}\mathbf{T}'_1\mathbf{K}^{-1}\mathbf{F} \\ &= \mathbf{K}^{-1}\mathbf{T}_1(\mathbf{T}'_1\mathbf{K}^{-1}\mathbf{T}_1)^{-1}\mathbf{T}'_1\mathbf{K}^{-1}\mathbf{F}, \end{aligned} \quad (50)$$

whose GSVD is equivalent to (39). This is in fact the basis for calling $\mathbf{K}^{-1}\mathbf{T}_1(\mathbf{T}'_1\mathbf{K}^{-1}\mathbf{T}_1)^{-1}\mathbf{T}'_1\mathbf{K}^{-1}\mathbf{F}$ (the matrix whose GSVD is obtained in (39)) the effect of T1 eliminating T2.

There is another way of relating NSPCCA(\mathbf{X}, \mathbf{Z}) with CNSCA with \mathbf{T}_1 eliminating \mathbf{T}_2 . Set $\mathbf{X} = \mathbf{I}$ and $\mathbf{Z} = \mathbf{T}_2^* = \mathbf{K}^{-1}\mathbf{R}$. Then NSPCCA(\mathbf{X}, \mathbf{Z}) reduces to NSCALC(\mathbf{R}) (Takane, Yanai, and Mayekawa 1991, p. 679), which is equivalent to NSCA with \mathbf{T}_1 eliminating \mathbf{T}_2 , as shown above.

Acknowledgements The work reported here has been supported by a grant 10630 from the Natural Sciences and Engineering Research Council of Canada. Correspondence regarding this article should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada.

References

- Agresti A (2002) *Categorical data analysis* (2nd ed.). Wiley, New York
- Anderson M J, Gribble N A (1998) Partitioning the variation among spatial, temporal and environmental components in a multivariate data set. *Aust J of Ecol* 23:158-167
- Anderson R J, Landis J R (1980) CATANOVA for multidimensional contingency tables: Nominal-scale response. *Commun Stat A-Theor* 11:1191-1205
- Blyth C R (1972) On Simpson's paradox and the sure-thing principle. *J Am Statist Assoc* 67:364-366
- Bocard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology* 73:1045-1055.
- Böckenholt U, Böckenholt I (1990) Canonical analysis of contingency tables with linear constraints. *Psychometrika* 55:633-639
- Böckenholt U, Takane Y (1994) Linear constraints in correspondence analysis. In: Greenacre M J, Blasius J (eds.) *Correspondence analysis in social sciences*. Academic Press, London, pp 112-127
- Cheng P E, Liou J W, Liou M, Aston, J A D (2006) Data information in contingency tables: A fallacy of hierarchical loglinear models. *J Data Sci* 4:387-398
- D'Ambra L, Beh E J, Amenta P (2005) CATANOVA for two-way contingency tables with ordinal variables using orthogonal polynomials. *Commun Stat A-Theor* 34:1755-1769
- D'Ambra L, Lauro N C (1989) Nonsymmetric correspondence analysis for three-way contingency tables. In: Coppi R, Bolasco S (eds.) *Multiway data analysis*. Elsevier, Amsterdam, pp 301-315
- Efron B, Tibshirani R J (1993) *An introduction to the Bootstrap*. CRC Press, Boca Raton, Florida
- Gilula Z, Haberman S J (1988) The analysis of contingency tables by restricted canonical and restricted association models. *J Am Statist Assoc* 83:760-771
- Goodman L, Kruskal W (1954) Measures of association for cross-classifications. *J Am Statist Assoc* 49:732-764
- Gray L N, Williams J S (1981) Goodman and Kruskal's tau b: Multiple and partial analogs. *Sociol Methods Res* 10:50-62
- Greenacre M J (1984) *Theory and applications of correspondence analysis*. Academic Press, London
- Hoerl K E, Kennard R W (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12:55-67
- Hwang H, Takane Y (2002) Generalized constrained multiple correspondence analysis. *Psychometrika* 67:215-228
- Khatri C G (1966) A note on a MANOVA model applied to problems in growth curves. *Ann I Stat Math* 18:75-86
- Lauro N C, D'Ambra L (1984) Nonsymmetric correspondence analysis. In: Diday E et al. (eds.) *Data analysis and informatics*. Elsevier, Amsterdam, pp 433-446
- Legendre P, Legendre L (1998) *Numerical ecology* (2nd English ed.). Elsevier, Oxford
- Light R J, Margolin B H (1971) An analysis of variance for categorical data. *J Am Stat Assoc* 66:534-544
- Margolin B H, Light R J (1974) An analysis of variance of categorical data II. Small samples comparisons with chi-square and other competitors. *J Am Stat Assoc* 69:755-764
- Maxwell S E, Delaney H D (2004) *Designing experiments analyzing data* (2nd ed.). Earlbaum Associates, Mahwah, NJ
- Økland R H (2003) Partitioning the variation in a plot-by-species data matrix that is related to n sets of explanatory variables. *J Veg Sci* 14:693-700
- Ramsay J O (1978) Confidence regions for multidimensional scaling analysis. *Psychometrika* 43:145-160
- Shapiro S H (1982) Collapsing contingency tables – A geometric approach. *Am Stat* 36:43-46
- Simpson E H (1951) The interpretation of interaction in contingency tables. *Am Stat* 13:238-241

-
- Takane Y, Hunter M A (2001) Constrained principal component analysis. *Appl Algeb Eng Comm Comput* 12:391-419
- Takane Y, Hwang H (2006) Regularized multiple correspondence analysis. In: Blasius J, Greenacre M J (eds.) *Multiple correspondence analysis and related methods*. Chapman and Hall, London, pp 259-279
- Takane Y, Jung S (2009) Regularized nonsymmetric correspondence analysis. *Comput Statist Data An* 53:3159-3170
- Takane Y, Shibayama T (1991) Principal component analysis with external information on both subjects and variables. *Psychometrika* 56:97-120
- Takane Y, Yanai H (1999) On oblique projectors. *Linear Algebra Appl* 289:297-310
- Takane Y, Yanai H, Mayekawa S (1991) Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika* 56:667-684
- ter Braak C J F (1986) Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167-1179
- ter Braak C J F (1988) Partial canonical correspondence analysis. In: Bock H H (ed.) *Classification and related methods of data analysis*. North-Holland, Amsterdam, pp 551-558
- van der Heijden P G M, de Leeuw J (1985) Correspondence analysis used complementary to loglinear analysis. *Psychometrika* 50:429-447
- van der Heijden P G M, Meijerink F (1989) Generalized correspondence analysis of multi-way contingency tables and multi-way (super-) indicator matrices. In: Coppi R, Bolasco S (eds.) *Multiway data analysis*. Elsevier, Amsterdam, pp 185-202
- Wermuth N (1987) Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *J Roy Stat Soc B* 49:353-364
- Yule G U (1903) Notes on the theory of association of attributes in statistics. *Biometrika* 2:121-134