DECOMPOSITIONS AND BIPLOTS IN THREE-WAY CORRESPONDENCE ANALYSIS

ANDRÉ CARLIER

LABORATOIRE DE STATISTIQUE ET PROBABILITÉS, UNIVERSITÉ PAUL SABATIER, TOULOUSE, FRANCE

PIETER M. KROONENBERG

DEPARTMENT OF EDUCATION, LEIDEN UNIVERSITY, LEIDEN, THE NETHERLANDS

In this paper correspondence analysis for three-way contingency tables is presented using three-way generalisations of the singular value decomposition. It is shown that in combination with Lancaster's (1951) additive decomposition of interactions in three-way tables, a detailed analysis is possible of the deviations from independence. Finally, biplots are shown to produce powerful graphical representations of the results from three-way correspondence analyses. An example from child development is used to illustrate the theoretical developments.

Key words: Additive decomposition of interactions, inertia, three-mode principal component analysis, PARAFAC, mother-child interaction.

Introduction

Correspondence analysis can be presented in a number of different ways (see e.g., Benzécri, 1970; Greenacre, 1984; Lebart, Morineau, & Warwick, 1984; and their references). Here correspondence analysis will be viewed as a technique in which a contingency table of counts is first processed in such a way that the resulting table only contains the dependence between the row and column variables. Then the singular value decomposition (SVD) is applied to the processed table to find a low-rank approximation, and the resulting approximation is displayed in a graph.

The major aim of the present paper is to extend correspondence analysis to three-way tables over and above earlier generalisations by Dequier (1973), Choulakian (1988b), and Kroonenberg (1989) using three-way generalizations of the singular value decomposition. Moreover, it will be shown that the combination of these generalizations and Lancaster's (1951) additive decomposition of X^2 provides a powerful vehicle for the decomposition of dependence in a three-way table.

In the example, biplot displays (Gabriel, 1971; Gabriel & Odoroff, 1990; Greenacre, 1993) will be used, rather than the until recently more common simultaneous displays (Benzécri, 1970). In correspondence analysis, the biplot can be used to provide a graphical representation of the low-rank approximation to the observed dependence in a contingency table (Gabriel & Odoroff, pp. 479ff; Greenacre), and these biplots can be enhanced by linearly marking the biplot axes (Gabriel & Odoroff), a procedure called *calibration* (Greenacre). To evaluate the interactions in three-way contingency tables, as analysed with three-way correspondence analysis, biplots can be fruitfully used, be it that it is not very well feasible to display the information of all three modes simul-

taneously, as biplots are essentially based on two sets of markers. Proposals how to deal with this situation will be presented.

The basics of (two-way) correspondence analysis will not be presented here but can be found in such standard reference works as Benzécri (1970) and Greenacre (1984). The core of the paper consists of (a) the presentation of three-way correspondence analysis, (b) the use of Lancaster's additive decomposition of X^2 to evaluate marginal dependence, and (c) the use of biplots to graph the dependence in three-way tables. The paper ends with an example from child development which will be used to discuss ways of interpreting the outcomes of an analysis.

Three-way Correspondence Analysis

With three-way correspondence analysis it is possible to produce good measures and graphical displays of the dependence in three-way tables, and it shares and extends many properties of ordinary (two-way) correspondence analysis. Previous work on extending (two-way) correspondence analysis to the three-way case, mostly reduced such tables to two-way tables using so-called interactive coding; see van der Heijden (1987), and van der Heijden, De Falguerolles, and de Leeuw (1989) for overviews of this approach. Papers in which three-way tables were analysed without reducing them to two-way tables are Choulakian (1988b), and Kroonenberg (1989), and especially Dequier (1973).

Measures of Global Dependence

The basic data structures in this paper are three-way contingency tables of order I, J and K with relative frequencies, p_{ijk}. They will be analysed with three-way correspondence analysis, and the starting point for the discussion of this technique is Pearson's *mean-square contingency coefficient*, Φ^2, also referred to as the *Inertia*. It is defined as

$$\Phi^2 = \frac{X^2}{n} = \sum_{i,j,k} \frac{(p_{ijk} - p_{i..}p_{.j.}p_{..k})^2}{p_{i..}p_{.j.}p_{..k}}, \tag{1}$$

where Φ^2 is based on the deviations from the three-way independence model, and as such contains all two-way interactions and the three-way interaction. In order to be able to define the orthogonality of two vectors (or three-way arrays) $\mathbf{Y} = (y_{ijk})$ and $\mathbf{Z} = (z_{ijk})$ in $\Re^{I \times J \times K}$, we define its inner product as

$$(\mathbf{Y} = (y_{ijk}), \ \mathbf{Z} = (z_{ijk})) = \sum_{i,j,k} p_{i..}p_{.j.}p_{..k}y_{ijk}z_{ijk}. \tag{2}$$

The distance between two vectors, which follows from this inner product, is

$$d(\mathbf{Y}, \mathbf{Z}) = \|\mathbf{Y} - \mathbf{Z}\| = \left(\sum_{i,j,k} p_{i..}p_{.j.}p_{..k}(y_{ijk} - z_{ijk})^2 \right)^{1/2}, \tag{3}$$

so that Φ^2 can be written as

$$\Phi^2 = \sum_{i,j,k} p_{i..}p_{.j.}p_{..k} \left[\frac{p_{ijk}}{p_{i..}p_{.j.}p_{..k}} - 1 \right]^2 = \sum_{i,j,k} p_{i..}p_{.j.}p_{..k}(\Pi_{ijk})^2 = \|\mathbf{\Pi}\|^2 \tag{4}$$

with $\mathbf{\Pi} = (\Pi_{ijk})$, and

$$\Pi_{ijk} = \frac{p_{ijk}}{p_{i..}p_{.j.}p_{..k}} - 1 = \frac{\Pr[i|j,\,k]}{\Pr[i]} \cdot \frac{\Pr[j|k]}{\Pr[j]} - 1. \tag{5}$$

Thus Pearson's Φ^2 is the weighted sum of the deviations of the observed relative frequencies from the expected values under the model of three-way independence. From equation (5) we see that $1 + \Pi_{ijk}$ can be interpreted as the product of two ratios. First, $\Pr[j|k]/\Pr[j]$ indicates the conditional probability of category j given that k has occurred, and $\Pr[i|j,\,k]/\Pr[i]$ measures the conditional probability of category i given that j and k have occurred. The symmetric statements after permutation of the indices hold as well. Thus Π_{ijk} can be considered as a measure of global dependence of the cell (i, j, k). Note that the weighted marginal totals summed over two indices are zero, for instance,

$$\Pi_{..k} = \sum_i \sum_j p_{i..}p_{.j.}\Pi_{ijk} = 0. \tag{6}$$

Measures of Marginal Dependence

So far no distinction has been made between possible two-way and three-way interactions, but all of them have been lumped together. Based on a different perspective, Kroonenberg (1989) argued that this seems a reasonable thing to do because it leads to a consistent definition of profiles in a three-way context (see also de Leeuw, 1983, p. 128ff.). However, for a proper evaluation of the global dependence, the contributions of the two-way interactions and the three-way interaction have to be ascertained. To this end, the concept of *marginal dependence* will be defined.

An orthogonal decomposition of three-way arrays. To define marginal dependence, it is instructive to identify the contributions of the main effects and interactions as in the classical ANOVA context of balanced experimental designs (for technical details see Carlier & Kroonenberg, 1995).

The elements of an arbitrary three-way array $\mathbf{X} = (x_{ijk})$ in $\Re^{I \times J \times K}$ can be written as

$$x_{ijk} = (x_{...}) + (x_{i..} - x_{...}) + (x_{.j.} - x_{...}) + (x_{..k} - x_{...}) + (x_{ij.} - x_{i..} - x_{.j.} + x_{...})$$

$$+ (x_{i.k} - x_{i..} - x_{..k} + x_{...}) + (x_{.jk} - x_{.j.} - x_{..k} + x_{...})$$

$$+ (x_{ijk} - x_{ij.} - x_{i.k} - x_{.jk} + x_{i..} + x_{.j.} + x_{..k} - x_{...}), \tag{7}$$

where a dot indicates that a weighted mean has been taken over the relevant index with respects to the associated weights $(p_{i..})$, $(p_{.j.})$, or $(p_{..k})$. Thus, for instance, $x_{ij.} = \sum_k p_{..k}x_{ijk}$; $x_{i..} = \sum_j \sum_k p_{.j.}p_{..k}x_{ijk}$, et cetera. The eight bracketed terms of (7) are the elements (i, j, k) of eight three-way arrays, denoted by $\mathbf{X}_{...}$, $\mathbf{X}_{I..}$, $\mathbf{X}_{.J.}$, $\mathbf{X}_{..K}$, $\mathbf{X}_{IJ.}$, $\mathbf{X}_{I.K}$, $\mathbf{X}_{.JK}$, and $\mathbf{X}_{IJK}$. For example, the element (i, j, k) of $\mathbf{X}_{IJ.}$ is $(x_{ij.} - x_{i..} - x_{.j.} + x_{...})$.

It can be shown (see Carlier & Kroonenberg, 1995) that these arrays are pairwise orthogonal with respect to the inner product defined in (2).

Decomposition of the array $\mathbf{\Pi}$. If the partitioning (7) is applied to $\mathbf{\Pi} = (\Pi_{ijk})$, it follows from the vanishing of the doubly weighted sum of the Π_{ijk} (see (6) as an example) that the first four terms of the decomposition of $\mathbf{\Pi}$ are equal to zero. Moreover, the bracketed terms can be greatly simplified.

$$\Pi_{ijk} = (\Pi_{ij.}) + (\Pi_{i.k}) + (\Pi_{.jk}) + (\Pi_{ijk} - \Pi_{ij.} - \Pi_{i.k} - \Pi_{.jk}), \qquad (8)$$

where the last term will be designated as $_\alpha\Pi_{ijk}$, following notation of Darroch (1974). The terms with two indices have the form

$$\Pi_{ij.} = \sum_k p_{..k}\Pi_{ijk} = \sum_k p_{..k} \frac{p_{ijk} - p_{i..}p_{.j.}p_{..k}}{p_{i..}p_{.j.}p_{..k}} = \frac{p_{ij.} - p_{i..}p_{.j.}}{p_{i..}p_{.j.}}, \qquad (9)$$

and they are the elements approximated by two-way correspondence analysis. As we may define $\Pi_{i.k}$ and $\Pi_{.jk}$ in a similar manner, the complete decomposition of Π can be written as

$$\Pi_{ijk} = \Pi_{ij.} + \Pi_{i.k} + \Pi_{.jk} + _\alpha\Pi_{ijk}$$

$$= \frac{p_{ij.} - p_{i..}p_{.j.}}{p_{i..}p_{.j.}} + \frac{p_{i.k} - p_{i..}p_{..k}}{p_{i..}p_{..k}} + \frac{p_{.jk} - p_{.j.}p_{..k}}{p_{.j.}p_{..k}} + \frac{p_{ijk} - _\alpha p_{ijk}}{p_{i..}p_{.j.}p_{..k}}, \qquad (10)$$

where $_\alpha p_{ijk}$ is equal to $p_{ij.}p_{..k} + p_{i.k}p_{.j.} + p_{.jk}p_{i..} - 2p_{i..}p_{.j.}p_{..k}$. The decomposition (10) was introduced by Lancaster (1951, 1960, 1980), and has also been used by Dequier (1973), Choulakian (1988a) and Yoshizawa (1975, 1988). The most important property of this decomposition is that it leads to an *additive* partitioning of the squared norm of Π.

Partitioning of the squared norm of Π. In terms of arrays, (10) may be written as

$$\Pi = \Pi_{IJ.} + \Pi_{I.K} + \Pi_{.JK} + \Pi_{IJK}, \qquad (11)$$

and the pairwise orthogonalities of the arrays lead to the additive partitioning of the squared norm of Π as

$$\|\Pi\|^2 = \|\Pi_{IJ.}\|^2 + \|\Pi_{I.K}\|^2 + \|\Pi_{.JK}\|^2 + \|\Pi_{IJK}\|^2. \qquad (12)$$

This equation may also be expressed in a more familiar form (see also Lancaster, 1980, p. 142) as

$$\Phi^2 = \sum_{ij} p_{i..}p_{.j.}\left(\frac{p_{ij.} - p_{i..}p_{.j.}}{p_{i..}p_{.j.}}\right)^2 + \sum_{ik} p_{i..}p_{..k}\left(\frac{p_{i.k} - p_{i..}p_{..k}}{p_{i..}p_{..k}}\right)^2$$

$$+ \sum_{jk} p_{.j.}p_{..k}\left(\frac{p_{.jk} - p_{.j.}p_{..k}}{p_{.j.}p_{..k}}\right)^2 + \sum_{ijk} p_{i..}p_{.j.}p_{..k}\left(\frac{p_{ijk} - _\alpha p_{ijk}}{p_{i..}p_{.j.}p_{..k}}\right)^2. \qquad (13)$$

This is clearly an additive definition of the interaction in a three-way array (see Darroch, 1974, for a comparative discussion of this *additive* definition of interaction and the *multiplicative* definition as used in loglinear analysis).

To summarize these results we can say that

1. Formula (10) shows that the deviations from three-way independence can be orthogonally decomposed into deviations from independence for the two-way margins of the three-way table, and a three-way interaction term.
2. Equation (13) shows that the global measure of dependence can be split into separate measures of dependence: there are three measures for the dependence due to each two-way margin which are identical to those used in two-way

correspondence analysis, and one measure for the three-way interaction. Such a partitioning can be the first step in the analysis of a three-way table.

A special case: The absence of one two-way interaction. A special case, discussed by Choulakian (1988a, p. 34ff.; see also Lancaster, 1960), occurs if one of the three terms $\|\mathbf{\Pi}_{IJ.}\|$, $\|\mathbf{\Pi}_{I.K}\|$ or $\|\mathbf{\Pi}_{.JK}\|$ turns out to be small (thus e.g., $\|\mathbf{\Pi}_{IJ.}\| \simeq 0$). The model of independence can then be assumed to hold for this two-way margin and it can be verified that Φ^2, obtained from the three-way table, is equal to the Φ^2 of the two-way table with K rows and $I \times J$ coded interactively. In such a case, the approximation of Π can be obtained by a single correspondence analysis on the table (p_{ijk}) with rows indexed by k and columns by (i, j). For details the reader is referred to Choulakian (1988a).

Modeling Global Dependence

Given a measure of global dependence and its partitioning into separate marginal measures and the three-way interaction, the next problem is how to find an appropriate model for these measures. For two-way tables, correspondence analysis is based on the generalized singular value decomposition (see e.g., Greenacre, 1984, p. 39), for three-way tables a three-way analogue of the GSVD is desired. Unfortunately, there is not one, but there are several possible generalizations. Here, only the two most common ones will be considered.

Definition and general properties. The two different three-way singular value decompositions or, for short, three-way models, that are used here, are those introduced in the context of the decomposition of general three-way arrays. The first of these is the PARAFAC model with no, one, two, or three orthonormal constraints on the component matrices (see Harshman, 1970; Harshman & Lundy, 1984; and Carroll & Chang, 1970; the latter authors used the name CANDECOMP model). With this model, the Π_{ijk}, are modeled as

$$\Pi_{ijk} = \sum_{s=1}^{S} g_{sss} a_{is} b_{js} c_{ks} + e_{ijk}. \tag{14}$$

The vectors $\{\mathbf{a}_s\}$, $\{\mathbf{b}_s\}$ and $\{\mathbf{c}_s\}$ are assumed to have unit lengths, and some (or none) of those sets are orthonormal in their respective spaces $\mathfrak{R}^I$, $\mathfrak{R}^J$, and $\mathfrak{R}^K$. There is a subtle difference in the present usage in that, analogous to two-way correspondence analysis, orthonormality is defined with respect to *weighted* metrics defined by $\{p_{i..}\}$, $\{p_{.j.}\}$, and $\{p_{..k}\}$, respectively (see (2)). Thus, it would be appropriate to refer to the model as a *generalized three-way singular value decomposition* or three-way GSVD. The g_{sss} are the three-way analogues of the singular values, and the e_{ijk} represent the errors of approximation.

The other model employed in this context is the Tucker3 model, also referred to as *three-mode factor analysis model* (Tucker, 1966),

$$\Pi_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr} + e_{ijk}, \tag{15}$$

where the three sets of vectors are generally taken to be orthonormal (without restriction of generality). The g_{pqr}, often referred to as core elements or elements of the core matrix, are another generalization of the singular values. For an extensive discussion of

the relative merits of these models, see for instance Harshman and Lundy (1984), and for a tensorial approach to the analysis of three-way arrays, see Franc (1992). In three-way correspondence analysis, the only modification that differentiates such decompositions from the usual ones is the use of a weighted least-squares criterion: the parameters g_{pqr}, a_{ip}, b_{jq} and c_{kr} are those which minimize

$$\sum_{i,j,k} p_{i..} p_{.j.} p_{..k} e_{ijk}^2.$$

As for fixed vectors $\{\mathbf{a}_p\}$, $\{\mathbf{b}_q\}$, and $\{\mathbf{c}_r\}$, the parameters g_{pqr} can be obtained by weighted regression of $\mathbf{\Pi}$ on the set of arrays $\mathbf{a}_p \otimes \mathbf{b}_q \otimes \mathbf{c}_r$ (where $\otimes$ indicates the Kronecker product), $\|\mathbf{\Pi}\|^2$ can be orthogonally decomposed as

$$\|\mathbf{\Pi}\|^2 = \|\hat{\mathbf{\Pi}}\|^2 + \|\mathbf{e}\|^2. \tag{16}$$

It follows that, using three-way GSVD as a model, Pearson's Φ^2 can be split into a fitted part and a residual part.

Additional properties of orthogonal models. For three-orthogonal models, that is the PARAFAC model with three orthogonality constraints, or the Tucker3 model, it can verified that the orthogonality of the three sets of vectors $\{\mathbf{a}_p\}$, $\{\mathbf{b}_q\}$ and $\{\mathbf{c}_r\}$, implies the orthogonality of the arrays $\mathbf{a}_p \otimes \mathbf{b}_q \otimes \mathbf{c}_r$ in the Euclidean space $\mathcal{R}^{I \times J \times K}$ using the definition of the metric defined by (2). As a consequence, we have the additional decomposition of Φ^2, assuming unit length components,

$$\|\hat{\mathbf{\Pi}}\|^2 = \sum_s g_{sss}^2 \text{ (three-orthonormal PARAFAC model)}$$

$$\|\hat{\mathbf{\Pi}}\|^2 = \sum_{pqr} g_{pqr}^2 \text{ (Tucker3 model)}, \tag{17}$$

which shows that the explained part of the Φ^2 can be further decomposed into parts referring to each element of the core matrix (see also ten Berge, de Leeuw, & Kroonenberg, 1987).

Modeling of Marginal Dependence

One of the attractive features of using Lancaster's approach over a loglinear modeling framework, is that one single decomposition of the global dependence is made, and that the marginal dependence can directly be modeled and assessed from the global decomposition. The contributions to the global dependence can be evaluated without having to construct special decompositions for lower order interactions.

Orthogonal decomposition of $\hat{\mathbf{\Pi}}$. In the previous section $\hat{\mathbf{\Pi}}$ was modeled by three-way GSVD. But in (11), the decomposition of the global dependence into marginal dependences and a three-way interaction was considered. The seemingly most straight-forward way to decompose $\hat{\mathbf{\Pi}}$ would be to claim that (12) would also hold for $\hat{\mathbf{\Pi}}$. This is unfortunately not true, because the first four terms of the decomposition vanish for $\mathbf{\Pi}$, but not for $\hat{\mathbf{\Pi}}$. However, $\mathbf{\Pi}$, $\hat{\mathbf{\Pi}}$ and $\mathbf{e}$ can each be orthogonally decomposed as follows

$$\mathbf{\Pi} = 0 \quad + 0 \quad + 0 \quad + 0 \quad + \mathbf{\Pi}_{IJ.} + \mathbf{\Pi}_{I.K} + \mathbf{\Pi}_{JK} + \mathbf{\Pi}_{IJK}$$

$$\hat{\mathbf{\Pi}} = \hat{\mathbf{\Pi}}_{...} + \hat{\mathbf{\Pi}}_{I..} + \hat{\mathbf{\Pi}}_{.J.} + \hat{\mathbf{\Pi}}_{..K} + \hat{\mathbf{\Pi}}_{IJ.} + \hat{\mathbf{\Pi}}_{I.K} + \hat{\mathbf{\Pi}}_{JK} + \hat{\mathbf{\Pi}}_{IJK}, \tag{18}$$

$$\mathbf{e} = \mathbf{e}_{...} + \mathbf{e}_{I..} + \mathbf{e}_{.J.} + \mathbf{e}_{..K} + \mathbf{e}_{IJ.} + \mathbf{e}_{I.K} + \mathbf{e}_{JK} + \mathbf{e}_{IJK}.$$

The uniqueness of the decomposition and the linearity of the mapping that associates an element of its decomposition (e.g., $\Pi \rightarrow \Pi_{IJ}$) with an array implies that we can equate each of the eight components of $\hat{\Pi} + e$ to each of the eight components of Π, respectively. (see Carlier & Kroonenberg, 1995). This leads to eight equations. The last four equations have the form $\Pi_{IJ.} = \hat{\Pi}_{IJ.} + e_{IJ.}$, $\Pi_{I.K} = \hat{\Pi}_{I.K} + e_{I.K}$, et cetera. The first four terms in (18) lead to the equations $e_{...} = -\hat{\Pi}_{...}$, $e_{I..} = -\hat{\Pi}_{I..}$, et cetera.

Modeling Two-Way and Three-Way Interactions

Given the appropriate expressions for the decomposition of $\hat{\Pi}$, and its constituent parts such as $\hat{\Pi}_{IJ.}$, expressions (14) or (15) can be used to derive submodels for such parts, that is for the marginal dependences.

Using PARAFAC to model the terms of (18), the following expressions can be derived

$$\hat{\Pi}_{...}(i, j, k) = \sum_s g_{sss} a_{.s} b_{.s} c_{.s}$$

$$\hat{\Pi}_{I..}(i, j, k) = \sum_s g_{sss}(a_{is} - a_{.s}) b_{.s} c_{.s}$$

$$\hat{\Pi}_{IJ.}(i, j, k) = \sum_s g_{sss}(a_{is} - a_{.s})(b_{js} - b_{.s}) c_{.s} \qquad (19)$$

$$\hat{\Pi}_{IJK}(i, j, k) = \sum_s g_{sss}(a_{is} - a_{.s})(b_{js} - b_{.s})(c_{ks} - c_{.s}).$$

The submodels for the other terms can be derived from the ones above by permutations of the indices. In this manner, the single model for the global dependence is used to model in a natural way the marginal dependence. In particular, the submodels for the two-way and the three-way interactions are given by the third and the fourth equations of (19), respectively, and they consist of centring and/or averaging the components of two or three of the modes. At the same time, expressions have been acquired for the so-called *partial residuals*, such as $e_{...} = -\hat{\Pi}_{...}$ and $e_{I..} = -\hat{\Pi}_{I..}$.

Removing marginal dependences. If it is desired to remove the dependence due to the margin $I \times J$ from the global dependence, for instance because it needs to be established what the third dimension K adds to dependence of I and J, one has to use the array $\hat{\Pi}_{..K} + \hat{\Pi}_{I.K} + \hat{\Pi}_{.JK} + \hat{\Pi}_{IJK}$.

Writing this expression for a single element, using the PARAFAC model (14), gives

$$\hat{\Pi}_{..K}(i, j, k) + \hat{\Pi}_{I.K}(i, j, k) + \hat{\Pi}_{.JK}(i, j, k) + \hat{\Pi}_{IJK}(i, j, k)$$

$$= \hat{\Pi}_{..K}(i, j, k) + \sum_s g_{sss}((a_{is} - a_{.s})(c_{ks} - c_{.s})b_{.s}$$

$$+ (b_{js} - b_{.s})(c_{ks} - c_{.s})a_{.s} + (a_{is} - a_{.s})(b_{js} - b_{.s})(c_{ks} - c_{.s}))$$

$$= \hat{\Pi}_{..K}(i, j, k) + \sum_s g_{sss} a_{is} b_{js}(c_{ks} - c_{.s}) - \sum_s g_{sss} a_{.s} b_{.s}(c_{ks} - c_{.s})$$

$$= \hat{\Pi}_{..K}(i, j, k) + \sum_s g_{sss} a_{is} b_{js}(c_{ks} - c_{.s}) - \Pi_{..K}(i, j, k)$$

$$= \sum_s g_{sss} a_{is} b_{js}(c_{ks} - c_{.s}). \qquad (20)$$

Removing the dependencies due to two two-way margins (e.g., the dependencies due to the margins $I \times J$ and $I \times K$), leads to

$$\hat{\Pi}_{.JK}(i, j, k) + \hat{\Pi}_{IJK}(i, j, k)$$

$$= \sum_s g_{sss}((b_{js} - b_{.s})(c_{ks} - c_{.s})a_{.s} + (a_{is} - a_{.s})(b_{js} - b_{.s})(c_{ks} - c_{.s}))$$

$$= \sum_s g_{sss} a_{is}(b_{js} - b_{.s})(c_{ks} - c_{.s}). \tag{21}$$

From the above the following rule may be inferred: in order to remove interactions due to one two-way margin (e.g., $I \times J$), the family of vectors associated with the indices that do not belong to the margins ($\{\mathbf{c}_s, s = 1, \ldots, S\}$) have to be centred. Removing more than one two-way margin effect can be done by repeating this operation for one or two of the remaining families of vectors. The above results will be used in biplots to display global dependence as well as marginal dependence.

The Choulakian and Dequier Models

As mentioned in the introduction both Choulakian (1988b) and Dequier (1973) have described models for three-way contingency tables along similar lines as those developed in this paper. In this section, the similarities and differences between our proposals and those of Choulakian and Dequier will be explained below.[1]

The elements of the decomposition of $\hat{\Pi}$ in (18) are of two different types. Firstly, $\hat{\Pi}_{IJ}$, $\hat{\Pi}_{IK}$, $\hat{\Pi}_{JK}$ and $\hat{\Pi}_{IJK}$ are respectively the approximations of Π_{IJ}, Π_{IK}, Π_{JK} and Π_{IJK}. Secondly, $\hat{\Pi}_{...}$, $\hat{\Pi}_{I..}$, $\hat{\Pi}_{.J.}$ and $\hat{\Pi}_{..K}$, the negations of the partial residuals, have no corresponding terms in the decomposition of Π, and ignoring them leads to a modified fitted matrix $\hat{\Pi}^*$, equal to $\hat{\Pi}_{IJ} + \hat{\Pi}_{IK} + \hat{\Pi}_{JK} + \hat{\Pi}_{IJK}$.

The new approximation $\hat{\Pi}^*$ is a better approximation of Π than $\hat{\Pi}$. Using (18), gives the following orthogonal decomposition of $\Pi - \hat{\Pi}$,

$$\Pi - \hat{\Pi} = (\mathbf{e}_{...}) + (\mathbf{e}_{I..}) + (\mathbf{e}_{.J.}) + (\mathbf{e}_{..K}) + (\Pi - \hat{\Pi}^*).$$

This expression leads to

$$\|\Pi - \hat{\Pi}^*\|^2 = \|\mathbf{e}_{...}\|^2 + \|\mathbf{e}_{I..}\|^2 + \|\mathbf{e}_{.J.}\|^2 + \|\mathbf{e}_{..K}\|^2 + \|\Pi - \hat{\Pi}\|^2.$$

Note, that $\hat{\Pi}^*$ is the sum of the two-way and three-way terms in (19), and this sum does not constitute a proper or complete Tucker3 or PARAFAC model. To see this in more detail, the PARAFAC model for the Π_{ijk} can be written as,

$$\Pi_{ijk} = \sum_s g_{sss}(a_{is} - a_{.s})(b_{js} - b_{.s})c_{.s} + \sum_s g_{sss}(a_{is} - a_{.s})(c_{ks} - c_{.s})b_{.s}$$

$$+ \sum_s g_{sss}(b_{js} - b_{.s})(c_{ks} - c_{.s})a_{.s} + \sum_s g_{sss}(a_{is} - a_{.s})(b_{js} - b_{.s})(c_{ks} - c_{.s}) + E_{ijk},$$

$$\tag{22}$$

where the residual E_{ijk} is equal to the four last partial residuals in (18). Using the Tucker3 model, another model can be obtained for Π_{ijk}, which can be written as follows

---

[1] Unfortunately, the proposal of Yoshizawa (1975; see also Yoshizawa, 1988) is written in Japanese and come too late to our notice to be included.

$$\Pi_{ijk} = \underbrace{\sum_{pq}}_{pq} \lambda_{pq}(a_{ip} - a_{.p})(b_{jq} - b_{.q}) + \sum_{pr} \mu_{pr}(a_{ip} - a_{.p})(c_{kr} - c_{.r})$$

$$+ \sum_{qr} \nu_{qr}(b_{jq} - b_{.q})(c_{kr} - c_{.r})$$

$$+ \sum_{pqr} g_{pqr}(a_{ip} - a_{.p})(b_{jq} - b_{.q})(c_{kr} - c_{.r}) + E_{ijk}, \tag{23}$$

where the parameters λ_{pq}, μ_{pr}, ν_{qr} are used to indicate the sum over r of $g_{pqr}c_{.r}$, the sum over q of $g_{pqr}b_{.q}$, and the sum over p of $g_{pqr}a_{.p}$, respectively, and where E_{ijk} is defined as above.

The former model (22) is similar to one proposed by Choulakian (1988b) and the latter (23) to one proposed in by Dequier (1973). The difference between these models and the present ones is that the above authors assume that the family of centred vectors (e.g., vectors of coordinates $(a_{is} - a_{.s})_{i=1,\ldots,I}$ in $\Re^I$) are orthogonal. On the other hand, in the present proposals the orthogonality is true for the uncentred, but not for the centred components. The effect of this is that a different estimation procedure is needed for Choulakian's and Dequier's models. The former sketches only implicitly an algorithm for his model, while the latter is not concerned with estimation.

Representating Dependence Graphically: Joint and Interactive Biplots

Greenacre (1993) contains an extensive discussion of the properties of biplots (which were introduced independently by Tucker, 1960, and Gabriel, 1971) in two-way correspondence analysis. Especially important in the present context is the concept of *metric-preserving* biplots. In row-isometric biplots, the distances between the row markers are faithfully represented, but those between the columns are not, with the reverse for column-isometric biplots.

So far the three ways of the contingency table have been treated in an entirely symmetric fashion. The symmetry can, however, not be maintained when graphical representations are considered, as (so far) no spatial representations or triplots (a term suggested by a reviewer) exist to portray all three ways simultaneously in one graph. A strict parallel with classical correspondence analysis cannot be maintained where the biplot can be viewed as a natural extension (see Greenacre, 1993).

To display the approximation of dependence obtained by a three-way correspondence analysis, two kinds of biplots are considered. They are presented here within the context of an approximation using the Tucker3 model, but a similar approach can be taken in the PARAFAC context.

The first kind, the so-called *joint (bi)plot* (see Kroonenberg, 1983, p. 164ff.) is based on the following decomposition

$$\hat{\Pi}_{ijk} = \sum_{r=1}^{R} c_{kr}\left[\sum_{p=1}^{P}\sum_{q=1}^{Q} g_{pqr}a_{ip}b_{jq}\right] \tag{24}$$

$$= \sum_{r=1}^{R} c_{kr}d_{(ij)r}, \tag{25}$$

where $d_{(ij)r}$ denotes the term in square brackets in equation (24). Each slice k of the array $\hat{\Pi}$ is a linear combination of the $\mathbf{D}_r = (d_{(ij)r})$ with as coefficients the c_{kr}. The object of joint plots is to construct graphical displays for each component r of the third

way. First, for each r, a singular value decomposition of the r-th slice of the core matrix G_r is performed, that is, $G_r = U_r \Lambda_r V_r'$ with $U_r' U_r = I_T$ and $V_r' V_r = I_T$, where $T = \min(P, Q)$. Thus $D_r = AU_r \Lambda_r V_r' B' = \tilde{A}_r \Lambda_r \tilde{B}_r'$ is the singular value decomposition with the same weighted metrics for $\tilde{A}_r$ and $\tilde{B}_r$ as for A and B, respectively. This choice of metrics will allow comparisons between the biplots of the $(I \times J)$ matrix D_r and the biplot obtained from the $I \times J$ margin of Π.

The second kind of biplot will be called an *interactive biplot*, and uses the same equation (24) as its base. Here each pair of indices (i, j) will be represented by a single marker, that is, the first and second modes are coded interactively, and hence the name of the biplot. As one reviewer remarked the name could be interpreted incorrectly, and *concatenated* biplot was suggested instead, but we prefer the former name because of its close relation to interactive coding. The number of biplots does not depend on P or Q, but only on R, and is equal to $R/2$ if R is even. The interactive biplot is especially useful when the number of elements in $I \times J$ is not too large, or when one of the two sets I or J is ordered (e.g., is associated with time). Bradu and Gabriel (1978) and Cox and Gabriel (1982) already used interactive biplots for tables with a continuous dependent variable and a three-way factorial design. The present use is different in that the scores of the interactively coded markers are structured by the three-way decomposition, while this structure is not explicitly modeled in their biplot. Bradu and Gabriel aptly remark that "[h]igher order tables can only be biplotted if they are collapsed into two-way tables" (p. 66). In comparison with their approach, we do not collapse tables resulting from the decomposition, but either combine two of the modes into a single one, or make the biplot "conditional" on the third mode.

Assuming j is an ordered mode, trajectories can be drawn in the biplot by connecting, for each i, the points (i, j) in their proper order. This will greatly facilitate interpretation, especially if j is a time mode (see also Bradu & Gabriel, 1978, Fig. 8A).[2] On the other hand, if there is no order in any of the modes and if the number of levels in the interactive modes is very large, there may be too many markers to produce an intelligible interactive biplot.

The nonsymmetry with respect to the three ways of the table leads to the choice of a "reference mode" (here the third one), the two other modes playing symmetric roles with respect to each other. The reference mode will often be the one leading to the smallest number of biplots: it will be the mode that is most easily summarized.

Implementation

The methods and graphical procedures described above have been programmed by the first author in S-Plus (Statistical Science; for a description of the language see Becker, Chambers, & Wilks, 1988). Most calculations are fairly straightforward except the three-way generalizations of the singular value decomposition. The technical basis for the algorithms can, for instance, be found in Harshman and Lundy (1984; PARA-FAC), and Kroonenberg (1983; Tucker3 model). We have not been able to find an explicit reference for the three-way orthogonal PARAFAC algorithm, but its development is straightforward given an algorithm for an one-way orthogonal PARAFAC. The former is incorporated in our S program; detailed algorithms for one-way and two-way orthogonal PARAFAC can, for instance, be found in Kiers and Krijnen (1991, pp. 150–151).

---

[2] For an example, see Carlier and Kroonenberg (in press).

Application: Mother-Child Interactions over Time

In this section data collected by van den Boom (1988; van den Boom & Hoeksma, 1994) will be analysed to illustrate some of the basic properties of three-way correspondence analysis, the Lancaster decomposition, and the associated biplots.

In her study of (Dutch) irritable infants, van den Boom and Hoeksma (1994) collected data of 30 infant-mother pairs during the first six months of life (for a discussion of irritability, see van den Boom, 1988, p. 70ff.). Each month, each mother-infant pair was observed at home in two sessions of forty minutes which were video-taped. The video tapes were coded by trained observers, and each six seconds the most salient behaviour of both the infant and the mother was coded, for instance, infant *cried* and mother *soothed*. The original 14 categories for infant behavior were reduced for this analysis to 7 categories and those of the mother to 6 categories. For each month and each mother-infant pair a 7 by 6 co-occurrence matrix was constructed from the categorical longitudinal sequences. Subsequently, the matrices were aggregated over mother-infant pairs, so that statements could be made about mother-infant interaction irrespective of the individual pairs.

The seven infant categories were crying, exploring, sucking, smile and similar positive social behaviour, inactivity, that is, the infant does not do anything in particular, looking at the mother, and vocalizing. The six mother behaviors were soothing, looking, stimulating, offering, contact seeking or maintaining with the infant, and other, that is behavior not directed at the infant.

Thus the data set under consideration form a 7 (infant behaviors) × 6 (mother behaviors) × 6 (months) three-way contingency table. The underlying structure for this table is that there are two response variables and one design variable (Time). In other words, the $p_{..k}$ are not really stochastic quantities, but proportions fixed by the design, and ideally they should (or could) have been equal, and their relative sizes are not something that needs to be explained.

Decomposition of X^2.   The decomposition of the X^2_{total} of the three-way table is given in Table 1. In absolute terms, the most important effects are the two-way interactions Infant × Mother, and the Infant × Time ones, followed by the three-way interaction, while the Mother × Time interaction is the smallest, be it that in terms of X^2/df ratios the last two change places. This indicates that in the first years of an infant's life, there is a distinct interaction pattern between mother and infant indepen-

Table 1

Van den Boom Data: Analysis of Fit

Source	X^2_{total}	df	X^2_{total}/df	% of Total X^2
Two-way interactions				
Infant x Mother	70687	30	2356	66%
Infant x Time	23472	30	782	22%
Mother x Time	4852	25	194	5%
Three-way Interaction	8505	65	131	8%
Total	107516	150	717	100%

dent of time (e.g., Crying generally goes together with Soothing). The larger interaction of the infant with time suggests that it is the infant rather than the mother who changes its behavior (e.g., over the six months the infant starts exploring), be it that a strict causal interpretation is, of course, not possible on the basis of the data alone. The smaller change in the mother's behavior over time suggests that her *overall* behavior patterns tend to be stable over time. However, the three-way interaction indicates that the changes in the interactions between mother and infant over time are not the same for all infant behaviors.

Joint biplots. It is known that, in contrast with the two-way singular value decomposition, the Tucker3 model is not an embedded (or nested) model (see Kroonenberg, 1983, p. 93ff.). This means that it is not possible to deduce a solution with a triple (P, Q, R) from a solution obtained with another triple (P', Q', R') with $P \leq P'$, $Q \leq Q'$ end $R \leq R'$ by simple removing some terms in the sum. On the other hand, the fit of the reduced model is always less than or equal to the model with more components. One way to assess how many components are adequate for any one analysis is to investigate how much each combination of components contributes to the overall fit. After inspecting several high-dimensional solutions, we decided that a $(P = 4, Q = 4, R = 2)$-solution, which had an overall fit of 90.3%, provided an adequate compromise between accuracy of approximation and simplicity of description. The core matrix of the solution is presented in Table 2. Because the time mode could be summarised with the least number of components, it was selected as the reference mode. Moreover, joint plots for the Mother and Child categories were deemed most informative, because their relationships were of prime interest in the analysis. There will be two joint plots, one for each of the time components.

With two components for the reference mode Time, (25) becomes

$$\hat{\Pi}_{ijk} = \sum_{r=1}^{2} c_{kr} d_{ijr} \text{ with } d_{ijr} = \sum_{p=1}^{P} \sum_{q=1}^{Q} g_{pqr} a_{ip} b_{jq}. \tag{26}$$

Table 2

Percentages Accounted for by Elements of the Core Matrix

(Tucker3 Model, 4*4*2-Solution)

	r_1				r_2			
	q_1	q_2	q_3	q_4	q_1	q_2	q_3	q_4
p_1	**37.4**	0.6	0.0	0.0	0.0	0.0	0.0	0.0
p_2	0.0	0.0	**4.2**	0.0	**1.2**	**13.4**	0.0	0.1
p_3	0.0	0.6	0.5	**4.3**	0.1	0.1	**1.5**	0.1
p_4	0.0	0.1	0.3	0.4	0.0	0.1	**2.0**	0.1

Note: p_i = i-th child component; q_j = j-th mother component, and r_k = k-th time component

Writing $\hat{\Pi}_{IJk}$ for the approximation of the k-th slice of Π (Π_{IJk} in $\mathfrak{R}^{I \times J}$), (26) can also be written as

$$\hat{\Pi}_{IJk} = c_{k1}\mathbf{D}_1 + c_{k2}\mathbf{D}_2. \tag{27}$$

In this case, one generally would not choose to make an interactive biplot with 7×6 interactive Child $\times$ Mother markers, because no trajectories can be drawn as Time is the reference mode. Therefore, it seems better to use joint biplots for visualizing the results of the decomposition.

Subdividing the error of approximation. To construct the joint biplot of $\mathbf{D}_r$, the singular value decomposition the core slice $\mathbf{G}_r$ is computed as explained above. Thus

$$\hat{\mathbf{D}}_r = \sum_{\ell=1}^{L_r} \lambda_\ell^r \tilde{\mathbf{a}}_\ell^r \otimes \tilde{\mathbf{b}}_\ell^r.$$

with $L_r \le \min(P, Q)$. Depending on the purpose of the biplot the λ_ℓ^r will be multiplied with the $\tilde{\mathbf{a}}_\ell^r$ or $\tilde{\mathbf{b}}_\ell^r$ to create a row-isometric or a column-isometric biplot, respectively.

Due to the choice of metrics for $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ (see the discussion after (14)), the set of vectors $\tilde{\mathbf{a}}_\ell^r \otimes \tilde{\mathbf{b}}_\ell^r \otimes \mathbf{c}_r$ is orthogonal in $\mathfrak{R}^{I \times J \times K}$, which leads to the following additive decomposition of the squared norm of $\hat{\Pi}$

$$\|\hat{\Pi}\|^2 = \sum_{r=1}^{2} \sum_{\ell=1}^{L_r} (\lambda_\ell^r)^2.$$

Numerically, the explained variability or inertia $\|\hat{\Pi}\|^2$ of the first step, which represents 90.3% of the total dependence, can be decomposed into the inertia explained by the two slices $\mathbf{D}_1$ (64.6%) and $\mathbf{D}_2$ (25.7%). Table 3 provides more details about the quality of approximations of the different components of the dependence. It shows that all parts are well explained with the largest error for the three-way interaction which has a proportional error of 37%.

Interpretation. The structure of the time components, that is the coefficients c_{kr}, can be inspected via Table 4.

Table 3

Global and Marginal Quality Indexes

Source	X^2_{total}	% of Total X^2	X^2_{error}	% of Total X^2	% of Total X^2_{error}	X^2_{error}/X^2_{total}
Main effects	0.0	0.0	11.0	0.0	0.1	-
Two-way interactions						
I x J	70687.0	65.7	2742.9	2.6	26.3	3.9
I x K	23472.3	21.8	3795.9	3.5	36.4	16.2
J x K	4851.7	4.5	718.8	0.7	6.9	14.8
Three-way interaction	8504.8	7.9	3153.8	2.9	30.3	37.1
Total	107515.8	100.0	10422.5	9.7	100.0	9.7

Table 4

Time Components (Tucker3 Model, 4*4*2-Solution)

Month	r_1	r_2
1	1.16	-1.42
2	1.27	-0.63
3	0.93	-0.06
4	0.93	0.49
5	0.81	1.11
6	0.81	1.45

- The coefficients c_{k1} are approximately equal to one, but slightly decreasing (their mean $c_{.1}$ is equal to .98). As a first approximation, all these coefficients can be considered to be equal, and thus the slice D_1 (see Table 5) does not explain any time effect. If more precision is required, one may say that because these elements are slightly decreasing, this first component implies a slight decrease for all the interactions. But the variations of the interactions are primarily accounted for by the second component.

- The coefficients c_{k2} are regularly increasing from -1.42 to 1.45, and their mean is approximately equal to zero ($c_{.2} = .15$). In other words, the corresponding biplot portrays those category combinations which change most over time. The products

Table 5

The Slices D_1 and D_2.

Bold values are larger than 1.00, *italic* values exceed 0.75)

|	Mother Categories					
D_1	OTHER	LOOK	STIMULATE	OFFER	CONTACT	SOOTH
Inactive	-0.389	0.692	-0.524	-0.368	**1.307**	*-0.955*
Smile	-0.739	-0.056	**1.560**	0.466	-0.009	*-0.871*
Look	0.061	0.121	-0.186	-0.041	0.218	*-0.888*
Vocal	-0.303	-0.041	0.733	0.244	-0.032	-0.734
Explore	0.561	-0.256	-0.395	0.050	-0.510	*-0.866*
Cry	-0.064	-0.307	*-0.899*	-0.654	-0.593	**8.457**
Suck	0.501	-0.025	*-0.797*	-0.170	-0.093	-0.444
D_2	OTHER	LOOK	STIMULATE	OFFER	CONTACT	SOOTH
Inactive	-0.432	**-1.426**	-0.745	-0.553	**-1.828**	-0.279
Smile	-0.180	-0.309	0.046	-0.053	-0.391	-0.168
Look	0.081	-0.138	-0.083	-0.009	-0.224	-0.038
Vocal	0.125	-0.108	0.058	0.078	-0.217	-0.108
Explore	**1.592**	0.597	0.376	*0.832*	0.201	-0.105
Cry	0.088	-0.121	-0.092	-0.011	-0.201	-0.006
Suck	**1.002**	0.215	0.095	0.453	-0.092	-0.071
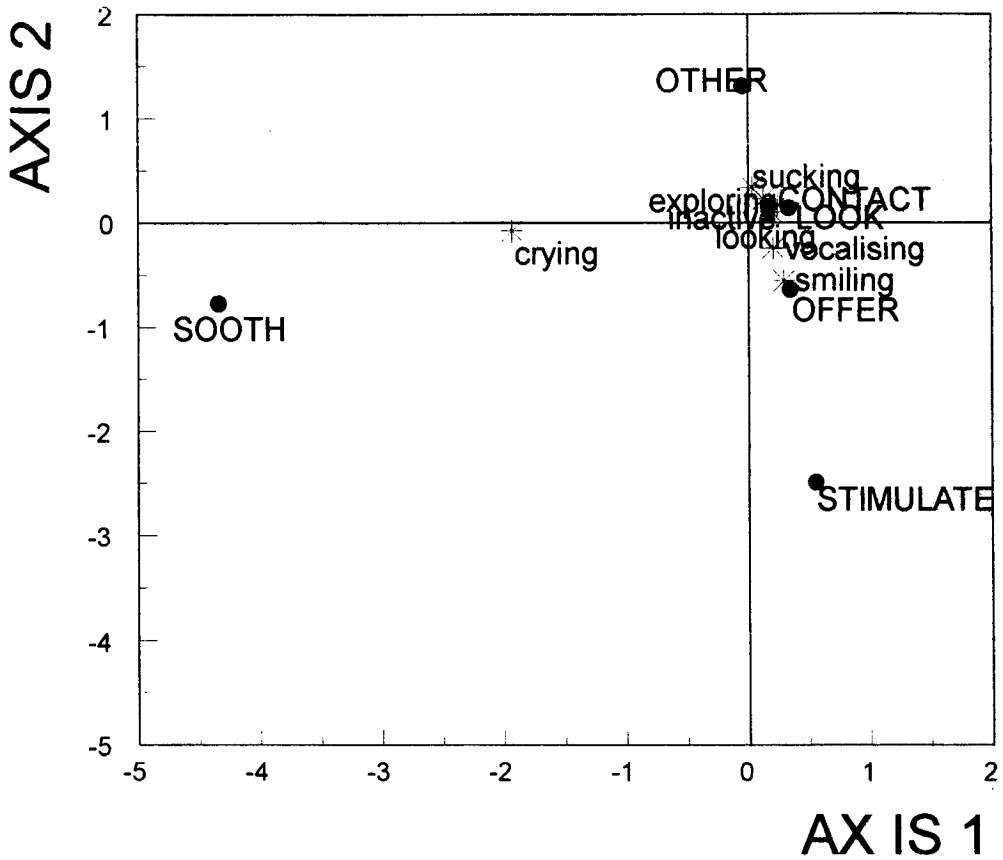
FIGURE 1.
Axis 1 versus Axis 2 of the Biplot of Slice D_1 associated with the First Time Component.

of the coefficients c_{k2} with the larger values in D_2 (see Table 5) represent large contributions to Π (see the left-hand side of (26)). If the inner products are positive, for instance Exploring with OTHER (= 1.59) then in combination with the large negative value of c_{12} and increasing towards large positive values of c_{k2} (k = 2 ... 6), this means that the child initially does *not* explore in combination with the mother doing non-child related things, while this combination occurs more and more frequently in the later months. The reverse is true for negative inner products, such as Inactive with CONTACT (= -1.83). In other words, while in the first month the mother often seeks contact with the inactive child, this occurs less and less over the next six months.

Because the coefficients c_{k1} are close to one, (27) becomes $\hat{\Pi}_{IJk} \simeq D_1 + c_{k2}D_2$. Furthermore, because $c_{.2} \simeq 0$ it can be deduced that D_1 is approximately equal to $\Pi_{IJ.}$, and that it contains the part of dependence that does not involve time. For the same reason and as a consequence of (11), the array which has as its k-th slice $c_{k2}D_2$, approximates $\Pi - \Pi_{IJ.} = \Pi_{I.K} + \Pi_{.JK} + \Pi_{IJK}$ which does depend on time.

The biplot for the axes 1 and 2 of D_1 will be called the "1 × 2" biplot of D_1, and it is displayed in Figure 1. This biplot is very similar to that obtained from a correspondence analysis of $\Pi_{IJ.}$. The contribution of the interaction Crying × SOOTHING on the first axis is very important with Crying accounting for 90% of the variance of the first child axis and SOOTHING for 92% of the first mother axis. In addition to this
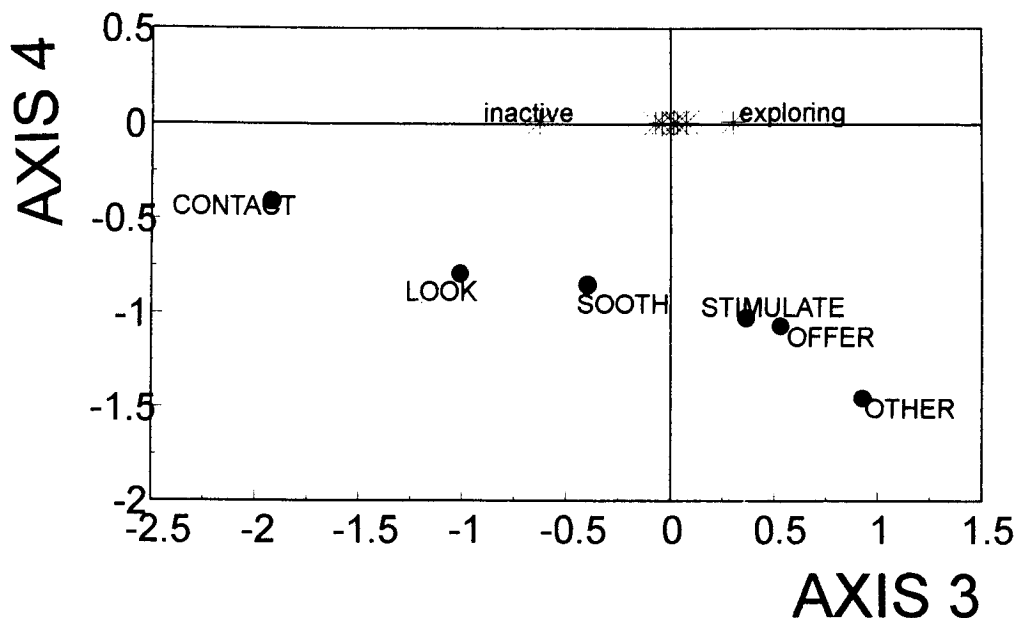
FIGURE 2.
Axis 3 versus Axis 4 of the Biplot of Slice D_1 associated with the First Time Component.

interaction the larger interactions are that the infant is Sucking and Exploring when the mother does OTHER nonchild-related things, and the infant is Smiling, and to a lesser degree, Vocalizing when the mother STIMULATES.

To supplement these larger scale patterns, one may study the remainder of the interaction. Instead of using a 4-dimensional biplot, which is not easy to visualize in practice, a second bidimensional biplot will be used to visualize the rank-two matrix $\lambda_3^1 \bar{a}_3^1 \bar{b}_3^1 + \lambda_4^1 \bar{a}_4^1 \bar{b}_4^1$, which is the difference between the 4th and 2nd order approximation of D_1. Such a figure contains "corrections" to the second-order approximation of the dependence and accounts for 7% of the inertia (compared to 93% of the 1 × 2 biplot). Figure 2 shows primarily interactions of Inactivity of the infant with several mother behaviors. On the one hand, the mother tends to seek CONTACT, LOOK at the infant when it is inactive. On the other hand when the mother is not STIMULating, OFFERing, nor engaged in OTHER activities the infant is also Inactive as follows from the negative projections on the respective axes. Further a clear interaction exists between Exploring of the infant and lack of CONTACT seeking of the mother.

As indicated above, the second slice D_2 contains the interactions which explicitly involve time. In other words, to study what is changing, only the component $r = 2$ of the decomposition needs to be used. The biplot associated with the decomposition of D_2 of order 2 (representing 99% of the variability of D_2) is given in Figure 3. The interactions involving Exploring are positive for all behaviors of the mother except SOOTHING, which means that they increase over time, when taking the values of the coefficients c_{k2} into account. At the opposite end, the interactions involving the Inactivity of the child are decreasing (all the corresponding elements of D_2 are negative). More precisely, the interactions which increase most are those linking Exploring of the child with OTHER (= 1.59) and OFFERing (= 0.83) behavior of the mother, and those linking Sucking with OTHER (= 1.00). Similarly the interactions that decrease most are those linking Inactivity of the child with CONTACT seeking (= −1.83) and LOOKING (= −1.43) of the mother.
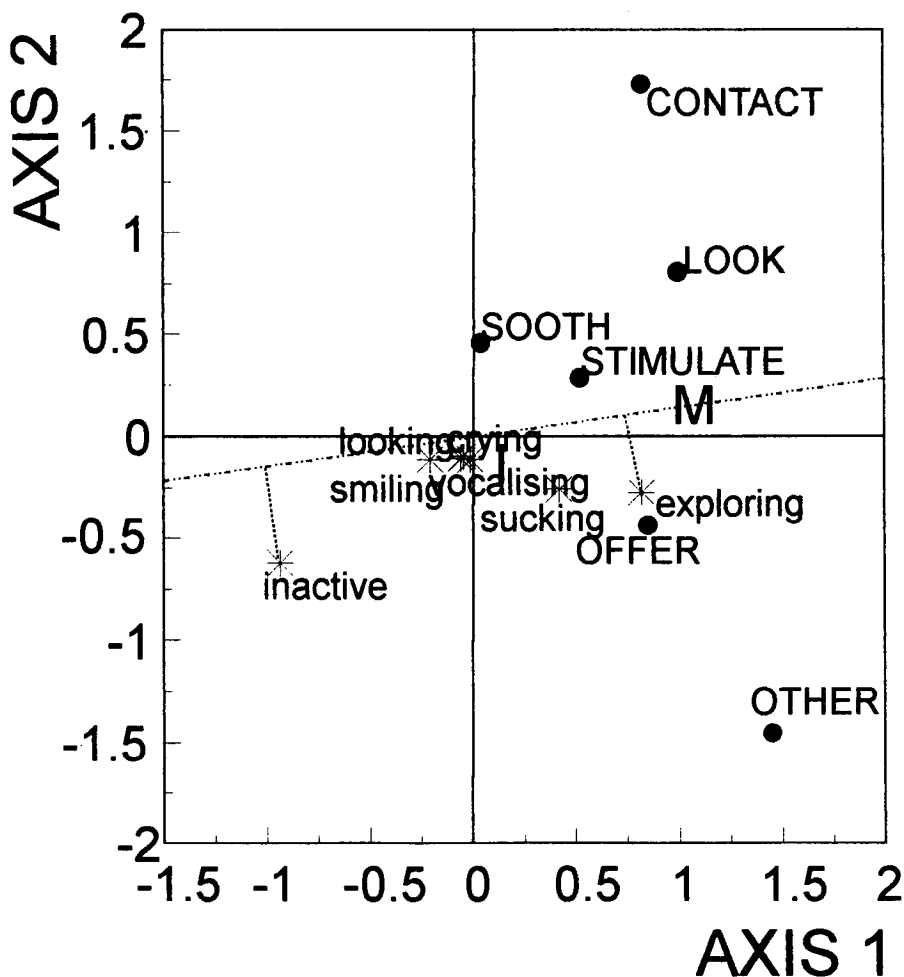
FIGURE 3.
Axis 1 versus Axis 2 of the Biplot of Slice D_2 associated with the Second Time Component. (I = centroid infant behaviours; M = centroid mother behaviors; the biplot axis through M can be used to evaluate the Infant × Time interaction.)

The marginal effects of Infant by Time ($I \times K$) and Mother by Time ($J \times K$) can be studied in the same biplot via combined usage of one set of markers and the axis determined by the centroid of the other set. Recall that the marginal effect of Infant by Time ($I \times K$) is given by $\Pi_{i.k} = \Sigma_j \, p_{.j} \Pi_{ijk}$ (see also (9)). Its approximation can be expressed in terms of the d_{ijr},

$$\hat{\Pi}_{i.k} = \sum_j p_{.j} \sum_r c_{kr} d_{ijr} = \sum_r c_{kr} d_{i.r} = \sum_r c_{kr} \sum_\ell \lambda_\ell^r \, \bar{a}_{i\ell}^r \, \bar{b}_{.\ell}^r$$

with $\bar{b}_{.\ell}^r = \Sigma_j \, p_{.j} \bar{b}_{j\ell}$. Thus by taking the centroid for the mother behaviors, and combining it with the markers of the infant behaviors, it is possible to assess from the biplot of D_r the value of $\hat{d}_{i.r}$, and to evaluate the Infant by Time interaction. The evaluation of the Mother by Time effects proceeds analogously.

The centroids of the markers of the mother behaviors and of the child behaviours are close to the origin in the biplot of D_1 and are not shown in the plot. This closeness to the origin shows that D_1 does not contribute to the change of the infant and mother

behaviors. On the biplot of D_2 the centroid of the infant behaviours (marked by an "I" in Figure 3) is again close to the origin, indicating that the Mother × Time interaction is small overall (see also Table 1). However, the centroid of the mother behaviors (marked by an "M" in Figure 3) is clearly away from the origin in accordance with the larger Infant × Time interaction indicated in Table 1. Projecting the infant behaviors on the axis through the centroid (see Figure 3) shows that Exploring is increasing over time and that Inactivity is decreasing over time, as one would expect from normal infants.

Conclusion

In this paper, it is explained how Lancaster's additive definition of interaction in contingency tables is well suited for an exploratory approach to analyzing (large) three-way contingency tables. It is interesting to note that his additive approach has fallen into disuse after the advent of loglinear modeling. Apart from the papers by Choulakian and Dequier cited above, only a few technical mathematical papers dealing with additive modeling have appeared after the comparative review by Darroch (1974), in particular Streitberg (1990) and the papers mentioned by him. The power of the additive decomposition lies in the possibility of fitting one model to the complete deviation from independence, and deriving the contributions of the separate lower-order terms from that one model.

So far only limited experience has been acquired with practical applications (see, however, for another example Carlier & Kroonenberg, in press), but as shown above three-way correspondence analysis in combination with various biplots can be fruitfully used to analyse large three-way contingency tables. In comparison with loglinear and/or logit modeling, three-way correspondence analysis as proposed here is not so much a procedure for fitting models to contingency tables, but primarily a technique to investigate and portray the main features of dependence in large three-way contingency tables with large and significant two-way and three-way interactions. In this sense, it is similar in spirit to the use of correspondence analysis complementary to loglinear analysis as suggested by van der Heijden et al. (1989). Given the complexity of higher-way data and their interpretation, it will not be easy to use some form of higher-way correspondence analysis in practical applications, even though mathematically it is should not pose too many problems (see Franc, 1992, p. 108).

References

Becker, R. A., Chambers, J. M., & Wilks, A. R. (1988). *The new S language. A programming environment for data analysis and graphics*. Pacific Grove, CA: Wadsworth.

Benzécri, J. P. (1970). *L'Analyse des données* [Data analysis]. Paris: Dunod.

Bradu, D., & Gabriel, K. R. (1978). The biplot as a diagnostic tool for models of two-way tables. *Technometrics, 20*, 47–68.

Carlier, A., & Kroonenberg, P. M. (1995). *Biplots and decompositions in two-way and three-way correspondence analyis* (Technical report No. 01-93, revised). Toulouse, France: Université Paul Sabatier, Laboratoire de Statistique et Probabilités.

Carlier, A., & Kroonenberg, P. M. (in press). Three-way correspondence analysis. The case of the French cantons. In J. Blasius and M. J. Greenacre (Eds.),Visualization of categorical data. London: Academic Press.

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 283–319.

Choulakian, V. (1988a). Analyse factorielle des correspondances de tableaux multiples [Correspondence analysis of multiway tables]. *Revue de Statistiques Appliquées, 36*(4), 33–42.

Choulakian, V. (1988b). Exploratory analysis of contingency tables by loglinear formulations and generalizations of correspondence analysis. *Psychometrika, 53*, 235–250.

Cox, C., & Gabriel, K. R. (1982). Some comparisons of biplot display and pencil-and-paper E.D.A. methods. In R. L. Launer & A. F. Siegel (Eds.), *Modern data analysis* (pp. 45–82). London: Academic Press.

Darroch, J. N. (1974). Multiplicative and additive interaction in contingency tables. *Biometrika, 61*, 207–214.

de Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. *Journal of Econometrics, 22*, 113–138.

Dequier, A. (1973). *Contribution à l'étude des tables de contingence entre trois caractères* [A contribution to the study of three-way contingency tables]. Unpublished doctoral thesis, Université de Paris VI, Paris.

Franc, A. (1992). *Étude algebrique des multitableaux: Apports de l'algèbre tensorielle* [Algebraic study of multiway tables: Contributions of tensor algebra]. Unpublished doctoral thesis, Université Montpellier II, France.

Gabriel, K. R. (1971). The biplot graphic display with application to principal component analysis. *Biometrika, 58*, 453–467.

Gabriel, K. G., & Odoroff, C. L. (1990). Biplots in biomedical research. *Statistics in Medicine, 9*, 469–485.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.

Greenacre, M. J. (1993). Biplots in correspondence analysis. *Journal of Applied Statistics, 20*, 251–269.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and contributions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics, 16*, 1–84. (Also available as University Microfilms, No. 10,0085)

Harshman, R. A., & Lundy, M. E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Synder, Jr., R. P. McDonald, & J. Hattie. (Eds.), *Research methods in multimode data analysis* (pp. 122–214). New York: Praeger.

Kiers, H. A. L., & Krijnen, W. P. (1991). An efficient algorithm for PARAFAC of three-way data with large numbers of observation units. *Psychometrika, 56*, 147–152.

Kroonenberg, P. M. (1983). *Three-mode principal component analysis: Theory and applications*. Leiden: DSWO Press.

Kroonenberg, P. M. (1989). Singular value decompositions of interactions in three-way contingency tables. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 169–184). Amsterdam: North Holland.

Lancaster, H. O. (1951). Complex contingency tables treated by the partition of the chi-square. *Journal of Royal Statistical Society, Series B, 13*, 242–249.

Lancaster, H. O. (1960). On tests of independence in several dimensions. *Journal of the Australian Mathematical Society, 1*, 241–254.

Lancaster, H. O. (1980). Orthogonal models for contingency tables. In P. R. Krishnaiah (Ed.), *Developments in statistics* (Vol. 3). New York: Academic Press.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis: Correspondence analysis and related techniques for large matrices*. New York: Wiley.

Streitberg, B. (1990). Lancaster interactions revisited. *Annals of Statistics, 18*, 1878–1885.

ten Berge, J. M. F., de Leeuw, J., & Kroonenberg, P. M. (1987). Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika, 52*, 183–191.

Tucker, L. G. (1966). Intra-individual and inter-individual multidemsionality. In H. Gulliksen & S. Messick (Eds.), *Psychometric Scaling: Theory and Applications*. New York: Wiley.

Tucker, L. R. (1960). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31*, 279–311.

van den Boom, D. C. (1988). *Neonatal irritability and the development of attachment: Observation and intervention*. Unpublished doctoral dissertation, Leiden University.

van den Boom, D. C., & Hoeksma, J. B. (1994). The effect on infant irritability on mother-infant interaction: A growth-curve analysis. *Developmental Psychology, 30*, 581–590.

van der Heijden, P. G. M. (1987). *Correspondence analysis of longitudinal categorical data*. Leiden: DSWO Press.

van der Heijden, P. G. M., De Falguerolles, A., & de Leeuw, J. (1989). A combined approach to contingency table analysis and log-linear analysis (with discussion). *Applied Statistics, 38*, 249–292.

Yoshizawa, T. (1975). Models for quantification techniques in multiple contingency tables—the theoretical approach. *Koudoukeiryougaku* [Japanese Journal of Behaviormetrics], *3*, 1–11. (in Japanese)

Yoshizawa, T. (1988). Singular value decomposition of multiarray data and its applications. In C. Hayashi, E. Diday, M. Jambu, & N. Ohsumi (Eds.), *Recent developments in clustering and data analysis* (pp. 24–257). New York: Academic Press.