

# Visualizing, Summarizing, and Comparing Odds Ratio Structures

Mark de Rooij<sup>1</sup> and Carolyn J. Anderson<sup>2</sup>

<sup>1</sup>Leiden University, Institute for Psychological Research, The Netherlands

<sup>2</sup>Department of Educational Psychology, University of Illinois, Urbana, IL, USA

**Abstract.** The odds ratio is one of the main measures of association in  $2 \times 2$  tables. For larger tables, summary measures of association have been proposed as well as modeling strategies, where the odds ratios can be deduced from the parameters of a model. We propose a computationally simple scaling methodology, which gives a summary measure of association and a visualization of the odds ratio structure. Different variants of the methodology are discussed and compared, both theoretically and empirically. The methodology is also generalized to multiple tables, where not only the structure of each table is important, but also comparisons among tables.

**Keywords:** singular value decomposition, odds ratio, graphical displays, contingency tables

## Introduction

Without doubt the odds ratio is one of the main measures of association in  $2 \times 2$  frequency tables, for example in the disease-exposure paradigm. Let the cell probabilities of a  $2 \times 2$  table of variables  $X$  and  $Y$  be defined by  $\pi_{ij} = P(X = x_i \cap Y = y_j)$ , where  $x_i$  is the  $i$ -th category of  $X$  and  $y_j$  is the  $j$ -th category of  $Y$ . The odds ratio  $\gamma$  is a measure of association for the two binary variables  $X$  and  $Y$  and is defined by

$$\gamma = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}. \quad (1)$$

In practical research applications, the number of row classifications and column classifications often exceed two. When  $X$  and  $Y$  have  $I^*$  and  $J^*$  categories respectively and  $i < i'$  denote two different categories of the row variable  $X$ , and  $j < j'$  denote two different categories of the variable  $Y$ , the association is described by the set of odds ratios:

$$\gamma_{ij i' j'} = \frac{\pi_{ij}\pi_{i' j'}}{\pi_{i' j}\pi_{ij}}, \quad 1 \leq i < i' \leq I^*, \quad 1 \leq j < j' \leq J^*. \quad (2)$$

This defines a set of  $(I^*(I^* - 1))/2 \times (J^*(J^* - 1))/2$  different odds ratios, which will be called the *complete set*. However, this complete set is redundant and there exist *basic sets* of  $(I^* - 1)(J^* - 1)$  odds ratios that capture all the information about the association between the variables. Two basic sets will be considered here (for others see Agresti, 1984, p. 21), the *local odds ratios* defined by

$$\gamma_{ij}^{(l)} = \frac{\pi_{ij}\pi_{i+1 j+1}}{\pi_{i+1 j}\pi_{ij}}, \quad 1 \leq i \leq (I^* - 1), \quad 1 \leq j \leq (J^* - 1), \quad (3)$$

and the *spanning cell odds ratios* (here with the cell  $i = 1$ ,  $j = 1$  as spanning cell)

$$\gamma_{ij}^{(s)} = \frac{\pi_{11}\pi_{ij}}{\pi_{1j}\pi_{i1}}, \quad 2 \leq i \leq I^*, \quad 2 \leq j \leq J^*. \quad (4)$$

The superscripts distinguish the two basic sets of odds ratios. The spanning cell odds ratios are useful when one of the categories defines a control or reference group. In that case, all other categories are described against this reference group. The local odds ratios are useful for ordinal variables when all local odds ratios are larger or equal to 1 (i.e., a positive relationship). We will call such basic sets *odds ratio structures*.

Given either set of odds ratios, all other odds ratios may be derived. For the local odds ratio by

$$\frac{\pi_{ij}\pi_{i' j'}}{\pi_{i' j}\pi_{ij}} = \prod_{i \leq k < i'} \prod_{j \leq l < j'} \gamma_{kl}^{(l)}, \quad (5)$$

and for the spanning cell odds ratio by

$$\frac{\pi_{ij}\pi_{i' j'}}{\pi_{i' j}\pi_{ij}} = \frac{\gamma_{ij}^{(s)}\gamma_{i' j'}^{(s)}}{\gamma_{i' j}^{(s)}\gamma_{ij}^{(s)}}. \quad (6)$$

These relationships turn to additive relationships for the log of the odds ratios. Let us denote the  $\log(\gamma) = \delta$ . Then, for the local odds ratio

$$\log\left(\frac{\pi_{ij}\pi_{i' j'}}{\pi_{i' j}\pi_{ij}}\right) = \sum_{i \leq k < i'} \sum_{j \leq l < j'} \delta_{kl}^{(l)}. \quad (7)$$

Similarly, for the spanning cell odds

$$\log \left( \frac{\pi_{ij}\pi_{i'j'}}{\pi_{ij'}\pi_{i'j}} \right) = \delta_{ij}^{(s)} + \delta_{i'j'}^{(s)} - (\delta_{ij'}^{(s)} + \delta_{i'j}^{(s)}). \tag{8}$$

A basic set of log odds ratios will be gathered in a matrix. Define  $I = I^* - 1$  and  $J = J^* - 1$  and denote the  $I \times J$  matrix with local odds ratios by  $\Gamma_l = \{\gamma_{ij}^{(l)}\}$  and their logarithms by  $\Delta_l = \{\delta_{ij}^{(l)}\}$ . Similarly for the spanning cell odds ratios,  $\Gamma_s = \{\gamma_{ij}^{(s)}\}$  and  $\Delta_s = \{\delta_{ij}^{(s)}\}$ .

Although the total number of odds ratios to check is reduced from  $(I^*(I^* - 1))/2 \times (J^*(J^* - 1))/2$  to  $(I^* - 1)(J^* - 1)$ , in applications where  $I^*$  or  $J^*$  is large, the number of odds ratios needed to capture the association structure may still be too large to gain insight into the nature of the relationship between the variables. Therefore alternative strategies have been developed to overcome interpretational problems. The first strategy is modeling the frequencies, and the second is through summary measures of association.

An approach to the analysis of contingency tables is to analyze the data with some model (for example, the log-linear model) and see how the odds ratios are obtained from the parameters of the model at hand. Examples of such approaches include log-linear models, the  $RC(M)$ -association model, and the R + C association model.

In log-linear models, the logarithm of the expected frequencies ( $F_{ij}$ ) is described by

$$\log(F_{ij}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}, \tag{9}$$

see Bishop, Fienberg, and Holland (1975) or Agresti (2002). Given the log-linear model, the odds ratio is

$$\gamma_{i'j'} = \exp(\lambda_{i'j'}^{AB} + \lambda_{i'j}^{AB} - \lambda_{ij'}^{AB} - \lambda_{ij}^{AB}). \tag{10}$$

The R + C association model is defined as (Goodman, 1979)

$$F_{ij} = \alpha_i \beta_j \mu_i^i \nu_j^j, \tag{11}$$

where the  $\mu_i$ 's and  $\nu_j$ 's are row and column scores, respectively, and the superscripts indicate powers. To obtain the odds ratio from expression (11), define

$$\gamma_i = \mu_{i+1} / \mu_i$$

$$\gamma_j = \nu_{j+1} / \nu_j.$$

The values of local odds ratios are

$$\gamma_{ij}^{(l)} = \gamma_i \gamma_j. \tag{12}$$

This model is not invariant under permutations of row and/or column order, and is usually applied when the variables are ordinal.

Goodman (1979, 1985) proposed the  $RC(M)$ -association model, defined as

$$F_{ij} = \alpha_i \beta_j \exp \left( \sum_{m=1}^M \varphi_m \mu_{im} \nu_{jm} \right), \tag{13}$$

where  $\mu_{im}$  and  $\nu_{jm}$  are row and column scores on dimension  $m$ , and  $\varphi_m$  is a measure of the strength of the association. Graphical displays of the  $RC(M)$ -association model represent two sets of vectors, one for the rows with coordinates  $\varphi_m^{1/2} \mu_{im}$  and one for the columns with coordinates  $\varphi_m^{1/2} \nu_{jm}$ . The interpretation is through inner products, i.e., the association equals the length of a row vector times the length of a column vector times the cosine of the angle between the two vectors. Given this model the odds ratio  $\gamma_{i'j'}$  is given by

$$\gamma_{i'j'} = \exp \left( \sum_m \varphi_m (\mu_{im} - \mu_{i'm}) (\nu_{jm} - \nu_{j'm}) \right). \tag{14}$$

Define the vectors  $\tilde{\mu}_{i'} = \mu_i - \mu_{i'}$ , where  $\mu_i = [\mu_{i1}, \dots, \mu_{iM}]^T$  and  $\tilde{\nu}_{j'} = \nu_j - \nu_{j'}$  with  $\nu_j = [\nu_{j1}, \dots, \nu_{jM}]^T$ , then (14) can be rewritten as

$$\gamma_{i'j'} = \exp \left( \tilde{\mu}_{i'}^T \mathbf{D}_\varphi \tilde{\nu}_{j'} \right) = \exp \left( \sum_m \varphi_m \tilde{\mu}_{i'm} \tilde{\nu}_{j'm} \right), \tag{15}$$

where  $\mathbf{D}_\varphi$  is a diagonal matrix with  $\varphi_1 \dots \varphi_M$  on the diagonal. This is again a usual inner product relationship, as is the interpretation of the vectors of the graphical display of the  $RC(M)$ -association model.

De Rooij and Heiser (2005) show that the graphical display resulting from the  $RC(M)$  association model can also be interpreted using a distance rule instead of an inner product rule. Define  $\mu_i^* = \mathbf{D}_\varphi^\tau \mu_i$  and  $\nu_j^* = \mathbf{D}_\varphi^{1-\tau} \nu_j$  for some power  $\tau$ , then the odds ratios can be expressed in terms of distances as

$$\gamma_{i'j'} = \exp \left( \frac{1}{2} d^2(\mu_i^*, \nu_j^*) + \frac{1}{2} d^2(\mu_{i'}^*, \nu_{j'}^*) - \frac{1}{2} d^2(\mu_i^*, \nu_{j'}^*) - \frac{1}{2} d^2(\mu_{i'}^*, \nu_j^*) \right), \tag{16}$$

where  $d^2(\mu_i^*, \nu_j^*)$  is the squared Euclidean distance between the points with coordinates  $\mu_i^*$  and  $\nu_j^*$ .

To summarize, we have discussed several modeling strategies for representing association between variables where association is defined in terms of odds ratios; however, in all of these models, odds ratios are functions of the model parameters, which obscures the relationship between the variables. In other words, the representation of the odds ratios is *indirect*. On the other hand, all odds ratios (not just a basic set) can be obtained from such procedures and the models can provide some insight into structure in the data.

The second approach is to summarize the association using one quantity. There is an extensive literature on such summary measures (see, for example, Altham, 1970; Agresti, 1980; Dale, 1986; Edwardes & Baltzan, 2000). Some

measures are typically designed for ordinal variables where a direction of association is sought, others provide only information about the amount of association, but no direction.

Altham (1970) proposed a measure based on spanning cell odds ratios, which is defined<sup>1</sup> as

$$\theta_{A_i}(p) = \left( \frac{1}{I \times J} \sum_i \sum_j |(\delta_{ij}^{(p)})|^p \right)^{1/p}, \quad p \geq 1. \quad (17)$$

When there are zero cells in the cross-classification of the variables .5 is added to each cell. This measure can also be defined on the local odds ratio which will be denoted by  $\theta_{A_i}(p)$ . In both cases the measure does not give a direction of association, and is therefore suitable for nominal variables, i.e., variables with unordered categories. For ordered variables  $\theta_{A_i}(p)$  defined on local odds ratios could be redefined without taking the absolute value to obtain a directed association measure ( $\theta_{A_i}(p)$ ). When  $p = 2$ ,  $\theta_{A_i}(p) = \theta_{A_i}(p)$ .

For a two-by-two table the odds ratio can be written in terms of Goodman and Kruskal's gamma,  $\gamma_{GK}$  (Goodman & Kruskal, 1979). Because  $\gamma_{GK}$  is also defined for larger tables with ordered variables, Agresti (1980) used this relationship to propose a generalized odds ratio ( $\theta_{GOR}$ ) measure for ordinal tables

$$\theta_{GOR} = \frac{1 + \gamma_{GK}}{1 - \gamma_{GK}}. \quad (18)$$

Edwardes and Baltzan (2000) provide an interpretation and discuss properties of this measure.

Another basic set of odds ratios is given by the global cross ratios (see Agresti, 1984), which are the odds ratios given by collapsing the row and column variables into dichotomies. A summary measure is obtained by taking the mean of the values in this basic set, i.e.,

$$\theta_{GCR} = \frac{1}{I \times J} \sum_i \sum_j \frac{(\sum_{a \leq i} \sum_{b \leq j} \pi_{ab}) (\sum_{a > i} \sum_{b > j} \pi_{ab})}{(\sum_{a \leq i} \sum_{b > j} \pi_{ab}) (\sum_{a > i} \sum_{b \leq j} \pi_{ab})}. \quad (19)$$

Edwardes and Baltzan (2000) use this summary measure as a measure of comparison, whereas Dale (1986) proposed modeling the basic set. There may be problems when some cells contain zero observations. The remedy is to add a .5. The measures by Agresti (1980) and Dale (1986) take into account the direction of association, i.e., they are suitable for ordinal variables. These summary measures of association do not provide detailed insight into which categories are (most) responsible for the association and which not.

In this paper, we provide a scaling methodology based on the singular value decomposition where the representation is *direct* (as opposed to indirect as in the models for

frequency tables discussed above) and from which we get a summary measure of association. The procedure can be applied to a table with a basic set of odds ratios or log odds ratios. It is well known that from these structures all other odds ratios can be obtained (equations 5, 6, 7, and 8) and it will be discussed whether this translates into our graphical representation, and if so, how it translates.

In the case of multiple tables, it is often of interest whether the association differs among the tables or has a similar form. Association models have been generalized by Clogg (1982), but like association models for a single table, the representation of association is indirect. Summary measures can be obtained for each of the tables, and using confidence intervals, it can be verified whether the association differs. We will generalize our scaling method to such multiple tables, where again summary measures are obtained plus a direct graphical display.

The remainder of this paper is organized as follows. The next section will discuss the singular value decomposition on which our methodology is based. We will also generalize the singular value decomposition to the case of multiple tables where odds ratio structures have to be compared. Section 3 discusses the application of the singular value decomposition to tables with (log) odds ratio structures. Properties of these applications will be shown. In Section 4 the methods will be applied to empirical data sets and we will conclude (Section 5) with some discussion.

## The Singular Value Decomposition

### One Table

Let  $\mathbf{A}$  be a general  $I \times J$  matrix with rank  $M$ . There exists an  $I \times M$  matrix  $\mathbf{X}$  such that  $\mathbf{X}^T \mathbf{X} = \mathbf{I}_M$ , an  $J \times M$  matrix  $\mathbf{Y}$  such that  $\mathbf{Y}^T \mathbf{Y} = \mathbf{I}_M$  and an  $M \times M$  diagonal matrix  $\mathbf{D}$  with positive elements, such that

$$\mathbf{A} = \mathbf{X} \mathbf{D} \mathbf{Y}^T = \sum_{m=1}^M \mathbf{x}_m d_m \mathbf{y}_m^T, \quad (20)$$

where  $\mathbf{x}_m$  is the  $m$ -th column of  $\mathbf{X}$ ,  $\mathbf{y}_m$  is the  $m$ -th column of  $\mathbf{Y}$ , and the singular values are ordered,  $d_1 > d_2 > \dots > d_M$ . For a proof we refer to Magnus and Neudecker (1988). This decomposition is called the singular value decomposition (SVD).

The Eckart-Young (1936) theorem states that the best least squares rank  $M^*$  approximation to a matrix  $\mathbf{A}$  is given by the first  $M^*$  singular values plus corresponding singular vectors; that is,

1 Altham does not use the mean but the sum, i.e., without standardizing for table size.

$$\mathbf{A}^{M^*} = \sum_{m=1}^{M^*} \mathbf{x}_m d_m \mathbf{y}_m^T, \tag{21}$$

with loss equal to  $\sum_{m=M^*+1}^M d_m^2$ .

Decomposition (20) where  $\mathbf{A}$  equals  $\mathbf{\Gamma}_l$  or  $\mathbf{\Delta}_l$  was proposed in Goodman (1979) as Model I and Model II, respectively. These models were then further developed into the R + C and RC association models, discussed in the previous section. As shown earlier the representation of the odds ratio in either of these models is not direct. Here we will follow a route that leads to a direct representation.

A joint plot of a matrix  $\mathbf{A}$  with SVD  $\mathbf{A} = \mathbf{X}\mathbf{D}\mathbf{Y}^T$  is obtained by plotting the rows of  $\mathbf{A}$  as vectors with coordinates  $\mathbf{U} = \mathbf{X}\mathbf{D}^{\tau}$  and the columns of  $\mathbf{A}$  as vectors with coordinates  $\mathbf{V} = \mathbf{Y}\mathbf{D}^{1-\tau}$ , where  $\tau$  usually equals zero, one or a half. The inner product of the vectors represents the values in the cells of matrix  $\mathbf{A}$ . In other words,  $\hat{a}_{ij} = \sum_{m=1}^{M^*} u_{im} v_{jm}$  and for the graphical display the following relationship holds

$$\sum_{m=1}^{M^*} u_{im} v_{jm} = |\mathbf{u}_i| |\mathbf{v}_j| \cos(\mathbf{u}_i, \mathbf{v}_j), \tag{22}$$

where  $|\mathbf{u}_i|$  the length of vector  $\mathbf{u}_i$  and similarly for  $|\mathbf{v}_j|$ , and  $\cos(\mathbf{u}_i, \mathbf{v}_j)$  is the cosine of the angle of the two vectors at the origin. Alternatively, (22) can be interpreted in terms of the length of  $\mathbf{u}_i$  and the length of the projection of  $\mathbf{v}_j$  onto  $\mathbf{u}_i$ , or the other way around.

### Multiple Tables

In the case where there are  $K$  tables, the SVD can be generalized as follows

$$\mathbf{A}_k = \mathbf{X}_k \mathbf{D}_k \mathbf{Y}_k^T \tag{23}$$

for  $k = 1, \dots, K$ , and where  $\mathbf{X}_k^T \mathbf{X}_k = \mathbf{Y}_k^T \mathbf{Y}_k = \mathbf{I}$ , and  $\mathbf{D}_k$  are diagonal matrices with ordered values like before.

Dimensionality restrictions can be imposed as well as, restrictions on, for example, the row vectors such that  $\mathbf{X}_k = \mathbf{X}$  for  $k = 1, \dots, K$ . Similar restrictions can be imposed on the column vectors or the singular values. The loss can be measured using the least squares criterion

$$\sum_{k=1}^K \|\mathbf{A}_k - \hat{\mathbf{A}}_k\|^2. \tag{24}$$

The decomposition (23) applied to tables of conditional local (log) odds ratios, including adding restrictions on the parameters over tables, was proposed by Clogg (1982) for analyzing groups of two-way tables by the RC(M)-association model. When equality restrictions are placed on the

row and column vectors (i.e.,  $\mathbf{X}_k = \mathbf{X}$  and  $\mathbf{Y}_k = \mathbf{Y}$  for all  $k$ ) but not the singular values (i.e.,  $\mathbf{D}_k$ ), the decomposition has the same form as the CANDECOMP/PARAFAC decomposition (Carroll & Chang, 1970), which is a special case of Tucker's 3-mode decomposition (Tucker, 1964, 1966; Kroonenberg, 1983). The orthogonality restrictions  $\mathbf{X}^T \mathbf{X} = \mathbf{Y}^T \mathbf{Y} = \mathbf{I}$  are not necessary in the CANDECOMP decomposition; however, they will continue to be imposed in our generalization to multiple tables.

An algorithm to fit models to multiple tables with various combinations of equality restrictions on  $\mathbf{X}_k$ ,  $\mathbf{Y}_k$ , and  $\mathbf{D}_k$  is presented in the Appendix. The algorithm is implemented in MATLAB and SAS/IML, which can be obtained from the authors' websites<sup>2</sup>.

### Properties of Applied SVDs to (Log) Odds Ratio Structures

In this section we will apply the SVD to tables with basic sets of odds ratios or log odds ratios. Although the main interest is in the basic set chosen for a specific application, we will also check whether the relationships between a basic set and all odds ratios, as defined in (5), (6), (7), and (8) transfers to the graphical display obtained through the SVD.

Summary measures will be defined based on the singular values in each of the cases. We first discuss decompositions of  $\mathbf{\Delta}_s$  and  $\mathbf{\Delta}_l$  (basic sets of log odds ratios); thereafter, decompositions of  $\mathbf{\Gamma}_s$  and  $\mathbf{\Gamma}_l$  (basic sets of odds ratios).

### Application to Log Odds Ratio Structures

A singular value decomposition on a basic set of log odds ratios is performed. Here the interest is in the relationship between a basic set and the complete set, and how the complete set can be obtained from a graphical representation of a basic set. The additivity relationships discussed in (7) and (8) transfer to an additivity relationship of singular vectors as shown in the following two theorems.

**Theorem 1:** Application of the SVD to  $\mathbf{\Delta}_l$  turns relationship (7) into additivity of singular vectors.

*Proof:* Each element of the matrix  $\mathbf{\Delta}_l$  is described by  $\delta_{ij}^{(l)} = \sum_m d_m x_{im} y_{jm} = \mathbf{u}_i^T \mathbf{v}_j$ . Then

$$\log \left( \frac{\pi_{ij} \pi_{i'j'}}{\pi_{ij'} \pi_{i'j}} \right) = \sum_{i < k < i'} \sum_{j < l < j'} \delta_{kl}^{(l)}$$

2 www.leidenuniv.nl/fsw/mderooij or www.psych.uiuc.edu/~canderso

$$\begin{aligned}
 &= \sum_{i \leq k < i'} \sum_{j \leq l < j'} \mathbf{u}_k^T \mathbf{v}_l \\
 &= \left( \sum_{i \leq k < i'} \mathbf{u}_k \right)^T \left( \sum_{j \leq l < j'} \mathbf{v}_l \right)
 \end{aligned} \tag{25}$$

□

**Theorem 2:** Application of the SVD to  $\Delta_s$  turns relationship (8) into subtraction of singular vectors.

*Proof:* Each element of the matrix  $\Delta_s$  is described by  $\delta_{ij}^{(s)} = \mathbf{u}_i^T \mathbf{v}_j$ . Then

$$\begin{aligned}
 \log \left( \frac{\pi_{ij} \pi_{i'j'}}{\pi_{ij'} \pi_{i'j}} \right) &= \delta_{ij}^{(s)} + \delta_{i'j'}^{(s)} - (\delta_{ij'}^{(s)} + \delta_{i'j}^{(s)}) \\
 &= \mathbf{u}_i^T \mathbf{v}_j + \mathbf{u}_{i'}^T \mathbf{v}_{j'} - \mathbf{u}_{i'}^T \mathbf{v}_j - \mathbf{u}_i^T \mathbf{v}_{j'} \\
 &= (\mathbf{u}_i - \mathbf{u}_{i'})^T (\mathbf{v}_j - \mathbf{v}_{j'}).
 \end{aligned} \tag{26}$$

□

The singular values can be used to construct a summary measure of association

$$\theta_{II}(p) = \left( \frac{1}{I \times J} \sum_{m=1}^M d_m^p \right)^{1/p}, \quad p \geq 1,$$

which can be used for spanning cell as well as local log odds ratios. When  $p = 2$ , this measure equals Altham's measure with  $p = 2$ .

### Application to Odds Ratio Structures

When the SVD is applied to a basic set of odds ratios (as compared to log odds ratios in the previous section) the multiplicative rules of obtaining the complete set (5) and (6) do not transfer to simple rules of the graphical display. For the spanning cell approach, however, we have a rather remarkable result.

**Theorem 3:** If the SVD is applied to  $\Gamma_s$  and only one dimension is retained, then relationship (6) results in odds ratios equaling unity.

*Proof:* Each element of  $\Gamma_s$  is described by  $\gamma_{ij}^{(s)} = d_1 x_{i1} y_{j1}$ . Then

$$\begin{aligned}
 \frac{\pi_{ij} \pi_{i'j'}}{\pi_{ij'} \pi_{i'j}} &= \frac{\gamma_{ij}^{(s)} \gamma_{i'j'}^{(s)}}{\gamma_{ij'}^{(s)} \gamma_{i'j}^{(s)}} \\
 &= \frac{(d_1 x_{i1} y_{j1})(d_1 x_{i'1} y_{j'1})}{(d_1 x_{i'1} y_{j1})(d_1 x_{i1} y_{j'1})} \\
 &= 1.
 \end{aligned} \tag{28}$$

□

Although the decomposition can be used to provide insight into a basic set, it should not be used to obtain values of the complete set.

A measure of overall association for an  $I^* \times J^*$  contingency table in case of the decomposition of a table with odds ratios is

$$\theta_{I*} = \sum_{m=1}^M \left| \log \left( \frac{d_m}{\sqrt{I \times J}} \right) \right|, \tag{29}$$

where the logarithm is taken to ensure the same effect of positive and negative association. When there is no association all odds ratios equal one. Applying the SVD to an  $I \times J$  table of ones gives constant singular vectors and a singular value equal to  $\sqrt{I \times J}$ . This latter value is thus the neutral value in a multiplicative sense and is used as a standardizing factor. For  $2 \times 2$  tables, this measure equals the absolute value of the log odds ratio. Taking the exponent of  $\theta_{I*}$  would give a measure with the same domain as the odds ratio. This measure does not take into account a direction of association. A modification for ordered categories is given by

$$\theta_{I*o} = \sum_{m=1}^M \log \left( \frac{d_m}{\sqrt{I \times J}} \right),$$

where an indication of direction is given. In such a case one should define the measure on the local odds ratio structure.

### Application to Empirical Data

In this section we will apply our methods to empirical data sets. The first is discussed in Edwardes and Baltzan (2000), the second in Williams and Grizzle (1972) and Dale (1986).

To obtain stability estimates of parameters, data reuse methods can be used, like the bootstrap and the jackknife. The jackknife procedure will be exploited here since for contingency tables it is computationally efficient (Clogg & Shidadeh, 1994, pp. 34–38; Dayton, 1998, pp. 22–23). Let  $\xi$  be any parameter of interest and  $\hat{\xi}$  be its estimator. For a data set with  $n$  observations, an estimate of  $\xi$  is obtained for  $n$  samples where the  $i$ -th observation is deleted,  $i = 1, \dots, n$ . Let  $\hat{\xi}_{-i}$  be the estimate with the  $i$ -th observation deleted. The estimated standard error using the jackknife is defined as (see Dayton, 1998, pp. 22–23)

$$SE_J(\hat{\xi}) = \sqrt{\frac{n}{n-1} \sum_{i=1}^n (\hat{\xi}_{-i} - \hat{\xi})^2}. \tag{31}$$

There is some evidence that the standard errors obtained from the jackknife procedure are somewhat too large (see Dayton, 1998, p. 37), although our personal experiences are that in the context of maximum likelihood the standard

Table 1. Summary measures of association for apnoea versus sleepiness data. Confidence intervals are obtained using a jackknife procedure; \* denotes that the exponent is taken of this measure since it is a measure on a log odds ratio scale

Measure	$\theta_{A_i}(1)^*$	$\theta_{A_i}(2)^*$	$\theta_{A_o}(1)^*$	$\theta_{GOR}$	$\theta_{GCR}$	$\theta_{II}(1)^*$	$\theta_{II}(2)^*$	$\theta_{I_o}^*$	$\theta_{I_e}^*$
Value	2.65	3.58	1.02	1.31	1.59	1.39	3.58	10.26	6.04
95% ci	(1.73, 4.05)	(2.18, 5.87)	(.95, 1.11)	(1.06, 1.56)	(0.76, 2.43)	(1.22, 1.59)	(2.18, 5.87)	(1.62, 65.07)	(.09, 399.66)

errors obtained through the jackknife are almost identical to those obtained with Newton-Raphson (if available). In general the number of computations to be done is equal to the sample size; however, for contingency tables each observation in a given cell produces the same calculation as other observations in this cell. Therefore, the number of calculations reduces from the sample size ( $n$ ) to the size of the contingency table ( $I^* \times J^*$ ). Confidence intervals are obtained from these estimated standard errors by  $\hat{\xi} \pm 1.96 \times SE_J(\hat{\xi})$ .

Our scaling procedure results in multidimensional solutions, and the interest is not only in the standard error per estimate but also in correlations between the estimates over dimensions. Therefore confidence regions should be obtained. The jackknife procedure results in  $IJ$  configurations each occurring  $n_{ij}$  times. Not all these configurations will be shown but clouds are represented that cover 95% of the configurations (see Meulman & Heiser, 1983). These clouds represent confidence intervals for the vectors. The SVD is unique up to a reflection of the dimensions, which can artificially enlarge the computed confidence regions. This is easily remedied by reflection of the jackknifed solution into the direction of the solution on the original table.

### Analysis of a Single Table

Table 1 of Edwardes and Baltzan (2000) classifies apnoea quartiles (1st; 2nd; 3rd; 4th) against scores of sleepiness (0–11). For such ordered variables it seems most natural to look at the local odds ratios. Since there are a large number of zeros in the data, the frequencies are smoothed first by adding a .5 to all. The values of all summary measures discussed in this paper are given in Table 1. From this table we see that the  $\theta_{II}(1)$  measure is similar to the *generalized odds ratio* ( $\theta_{GOR}$ ) and *global cross ratio* ( $\theta_{GCR}$ ), while  $\theta_{II}(2)$  is more in the range of Altham’s measures. The  $\theta_{I_o}$  and  $\theta_{I_e}$  values seem to have another range, but the confidence intervals show that these measures can vary widely. Both measures for ordinal association ( $\theta_{A_o}(1)$  and  $\theta_{I_o}$ ) are much smaller than their variants for unordered categories implying that the association is not in the same direction everywhere. Moreover, their confidence intervals include one, showing that there is no uniform direction of association,

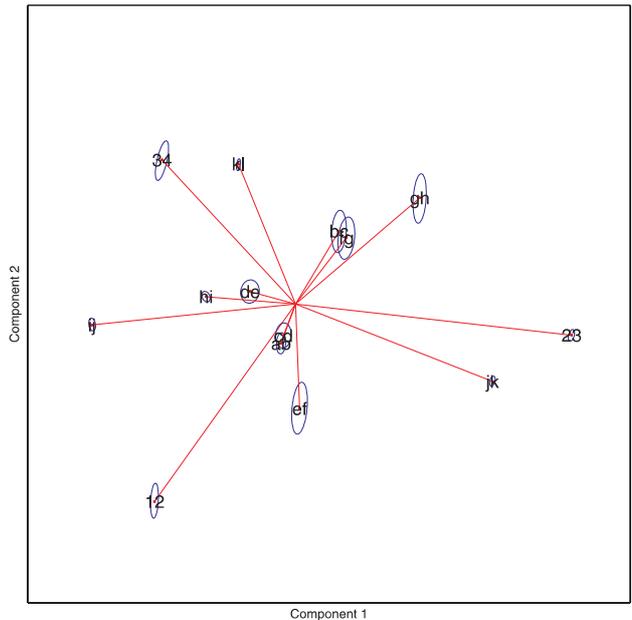


Figure 1. Two dimensional display of the local odds ratio structure for apnoea-sleepiness data. We used labels “12,” “23,” “34” for the rows (i.e., first quartile versus second; second versus third, third versus fourth) and “ab” to “kl” for the columns. 95%-confidence regions are obtained using a jackknife procedure.

which can also be concluded from the confidence interval of  $\theta_{GCR}$ . The value of  $\theta_{GOR}$  and  $\theta_{GCR}$  differ slightly from the values in Edwardes and Baltzan (2000), this is probably due to the manner of dealing with zeros<sup>3</sup>. Edwardes and Baltzan (2000) report  $\theta_{GOR} = 1.34$  with a 95% confidence interval (1.09, 1.65) which is similar to our jackknife results.

For the analysis of the log of the local odds ratios ( $\Delta_i$ ) the singular values are 6.36, 3.45, and 1.17 for dimensions 1 to 3, respectively, so the loss for the two-dimensional solution is only 2.5%. To see what the association pattern looks like, we give the two dimensional solution in Figure 1. In the figure, we use the labels “12,” “23” and “34” for the rows (i.e., first quartile *versus* the second, second *versus* the third, and third *versus* the fourth) and the labels “ab” to “kl” for the columns. The row categories are plotted using coordinates  $\mathbf{XD}^{1/2}$  the column categories as  $\mathbf{YD}^{1/2}$ . The odds

3 It is unclear to us how Edwardes and Baltzan (2000) do this.

ratios are represented in Figure 1 by the projection of the column points (e.g., “ab”) onto the row vectors. Alternatively, we can also look at the projections of the row points (e.g., “12”) onto the column vectors. The longer the projection, the higher the odds ratio.

From Figure 1 we see that the assumption of ordinal association is violated. If the variables were ordinal then all local log odds ratios should have been larger or equal to zero, i.e., have an angle less than 90°, which is clearly false. A closer look at the plot reveals that the local odds ratios for the first two rows are positive with “ab,” “cd,” “de,” “ef,” “hi,” and “ij,” while they are negative (obtuse angle) for “bc,” “fg,” “gh,” and “jk.” The ordering of local odds ratios for the first two rows “12” is

$$ij > ef > hi > ab > cd > de > 0 > jk > kl > fg > bc > gh.$$

For row 2 *versus* 3 (“23”) the order from low to high is

$$jk > gh > fg > bc > ef > 0 > cd > ab > de > kl > hi > ij$$

and finally for rows 3 *versus* 4 (“34”) the order is

$$kl > ij > hi > de > bc > fg > 0 > gh > cd > ab > ef > jk.$$

We can also work the other way around and project the row points onto the vector “ij” and we see that “12” projects higher, so the local odds ratio for rows 1 *versus* 2 and columns i *versus* j is higher than the local odds ratio in rows 1 *versus* 2 and these columns.

Since the vector for rows 2 *versus* 3 (compared to 12 and 34) is the longest, the association is strongest in this comparison.

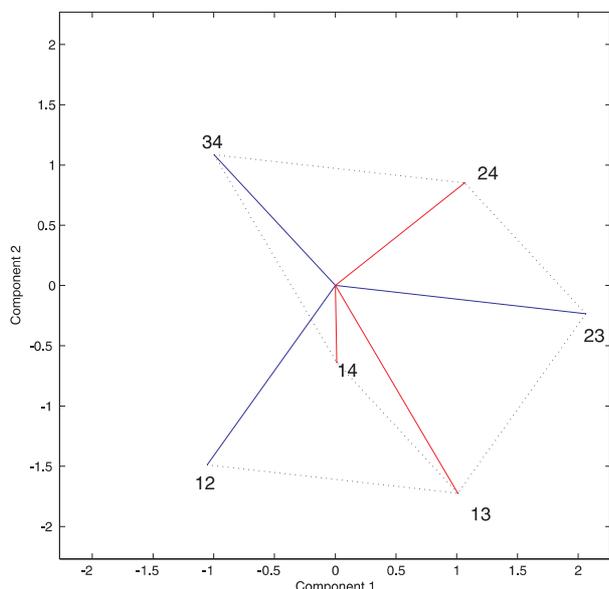


Figure 2. The additivity property discussed in Theorem 1 in practice for the rows. The vector “13” is obtained by summing “12” and “23.”

The additivity relation discussed in Theorem 1 is shown in Figure 2 for the row categories. For example, to get “13” we add vectors “12” and “23.” The same could be done for the column categories, and when represented in the same plot gives the complete set of odds ratios.

### Analysis of Multiple Tables

The second analysis is originally from Williams and Grizzle (1972, see also Dale, 1986, and Edwardes & Baltzan, 2000<sup>4</sup>) and classifies pain levels ( $I = 3$ ; none (1); slight (2); moderate (3)) against medication use ( $J = 4$ ; never (a); seldom (b); occasionally (c); regularly (d)) after four kinds of surgery for duodenal ulcer ( $K = 4$ ; A: VD; B: VA; C: RA; D: VH). The question is whether pain level and medication are associated, and whether the association between pain and medication is different for the four operations. Since the variables pain and medication appear to be ordinal, we analyzed  $\Delta_i$ , which are defined on each of the slices, i.e., after each of the four operations. Table 2 presents the loss values of the different models. Based on the loss values presented in Table 2, the choice of a final model is not clear cut in this case. If a plot is made of loss values against the number of parameters of all models, those with two dimensions are clearly superior to those with only one dimension. Of those with two dimensions, the models basically fall on a line (i.e., fewer parameter, more loss). Considering goodness-of-fit of the model to data, parsimony, and interpretation, two models stand out as good representations of the data: the model defined by  $A_k = X_k D Y^T$  having loss 7.34 with 9 estimated parameters and model  $A_k = X D_k Y^T$  with loss 5.97 and 12 estimated parameters. The interpretation of the first model is that there is an equal association between pain and medication after the four kinds of operations, but the pain levels have a different interpretation for the groups, i.e., “slight” pain has another meaning in group A compared to group B and similarly for the other categories and other groups. The interpretation of the second model is

Table 2. Loss values for different constrained singular value decompositions log odds ratios from data set 2 (in parenthesis are the number of parameters estimated)

Model	2 dim	1 dim
$X_k D_k Y_k^T$	0 (24)	15.01 (16)
$X_k D_k Y^T$	3.35 (15)	21.25 (10)
$X D_k Y_k^T$	0.38 (21)	21.28 (13)
$X_k D Y_k^T$	2.21 (18)	16.22 (13)
$X_k D Y^T$	7.34 (9)	24.30 (6)
$X D Y_k^T$	3.97 (15)	23.24 (10)
$X D_k Y^T$	5.97 (12)	23.79 (7)
$X D Y^T$	9.86 (6)	26.46 (4)

4 The data in Edwardes and Baltzan (2000) differs from that reported in Williams and Grizzle (1972) and Dale (1986), we used the data as reported in Williams and Grizzle.

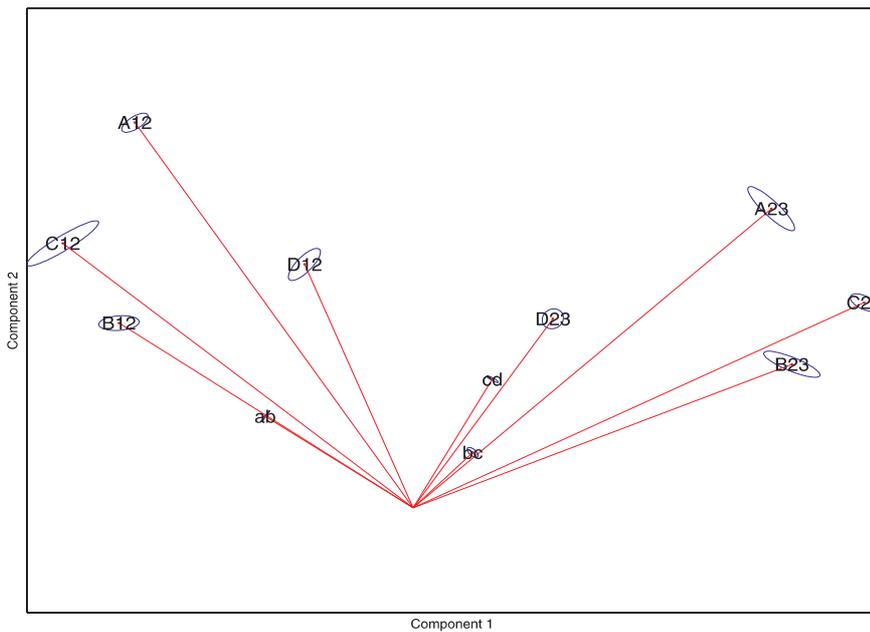


Figure 3. Two dimensional display of the local odds ratio structure for pain-medication data. We used labels “A12” and “A23” for the rows (i.e., first pain level versus second; second versus third) in slice A, and “ab” to denote columns ab, i.e., never versus seldom. 95%-confidence regions are obtained using a jackknife procedure

that the amount of association is different after the four operations, but that the pain levels and medication use have identical interpretations in each of the four groups. Our preference is for this second model, which we discuss further.

Although the amount of association is different, the pain and medication categories have the same meaning after the four operations. The summary measures of association are  $\theta_{II}(1) = 0.91, 0.68, 0.87,$  and  $0.47$  for operations (slices) A through D, respectively; the values  $\theta_{II}(2) = 1.58, 1.24, 1.54,$  and  $0.86$  for the operations A through D, respectively. It seems that after operation D the amount of association is much smaller than after the other three, a conclusion that is in agreement with that in Dale (1986) and Williams and Grizzle (1972). The graphical display with  $\mathbf{XD}_k$  versus  $\mathbf{Y}$  is shown in Figure 3.

The graphical display reveals that the association is strongest for operations (slices) A and C since the vectors representing these slices are relatively long. The association is not very strong for operation D, i.e., the vectors are relatively short. Most of the angles are smaller than  $90^\circ$ , indicating positive relationships. The exceptions in all operations (slices) are the local odds ratios of rows 2 versus 3 (slight and moderate pain) and columns a and b (never and seldom medication use), and in slices B and C the local odds ratios for rows 1 and 2 (none and slight pain) and columns b and c (seldom and occasionally use of medication). Most of the association is, however, positive. Because the vectors for operations A and D are close, the association patterns of slices A and D are similar. The pattern in A is stronger, which can be seen by the length of the vectors. Similarly the association patterns of operations (slices) B and C are similar.

For none versus slight pain (rows 1 and 2) after operations (slices) A and D the order of local odds ratios is

$$ab > cd > bc,$$

while for operations B and C we have the order

$$ab > cd > 0 > bc.$$

For slight versus moderate pain (rows 2 and 3) in all slices we have the order of local odds ratios

$$cd > bc > 0 > ab.$$

To verify whether the association none versus slight pain (in rows 1 and 2) with never versus seldom medication use (columns a and b) is larger after operation A or B (slices), we can project the points A12 and B12 onto the vector “ab,” and see that A12 project higher, i.e., has a larger value of the local odds ratio.

## Discussion

For two-way frequency tables, the joint distribution of  $X$  and  $Y$  is completely determined by the marginal distributions and a basic set of odds ratios. An analysis of odds ratios was proposed in this paper. Together with the marginal distributions the analysis provides us a complete description of the contingency table.

Our new proposed methods provide summary measures as well as direct graphical displays of the odds ratio structures. The summary measures were compared to existing measures in Table 1. The proposed analysis is based on ideas presented in Goodman (1979). From these ideas the R+C association model and the  $RC(M)$  association model were built in Goodman (1979, 1985). Our scaling method is computationally simpler compared to the maximum likelihood estimation of the  $RC(M)$  association model. Moreover, our method is invariant when all cells in a row or column are multiplied by a constant, since under such

transformations the (log) odds ratios remain the same. For the  $RC(M)$  association model this is not true, i.e., the row and column scores change when a row or column of the contingency table is multiplied by a constant.

The generalization to multiple tables was proposed by Clogg (1982), but like Goodman he parameterized the ideas to models for the frequencies in a contingency table. Our parametrization has the advantage that it provides a direct interpretation of the odds ratio structure in the data. Furthermore, our estimation method always results in the global optimum; whereas, maximum likelihood estimation of the  $RC(M)$  association models and their three-way generalizations sometimes yield a local optimum.

Comparing the decompositions on odds ratios or log odds ratios we can conclude the following. Methods for the log of the odds seem better than for the odds themselves. This is probably due to the fact that the distribution of the odds ratio is very skew, i.e., the range 0–1 equals the range 1– $\infty$ . It seems harder for the singular value decomposition to represent this skewness. Moreover, the translation of the results obtained by the SVD for the spanning odds ratios in a single dimension to the complete set, as discussed in Theorem 3, is a real drawback and complicates its use of spanning odds ratios. Moreover, the summary measures obtained by such a decomposition  $\theta_1$  seem very unstable (see confidence intervals in Table 1). Along with a particular decomposition comes a measure of association. As the decomposition of the log odds ratios has some advantages the association measure is  $\theta_{II}(p)$ , where only a choice of  $p$  is needed. The choice  $p = 2$  has the additional advantage that the measure is equal to that of Altham (1970), for which properties were deduced.

The choice between an analysis of local or spanning cell odds ratios should be made on data theoretical grounds, as was done in the examples. If the categories have a natural ordering an analysis of the local odds ratios is recommended; whereas, when there is a comparison or baseline category, the spanning cell approach seems to be the better choice.

To conclude we note two analysis procedures for contingency tables that employ the singular value decomposition. The first was proposed by Becker (1992) to obtain an exploratory analysis of association models. In his procedure first log-linear models are estimated and then the SVD is applied to the centered interaction parameters. It is shown that the estimates do approximate the maximum likelihood estimates of association models quite well. Graphical displays obtained in this procedure have to be interpreted in a similar fashion as in association models, i.e., odds ratios are represented indirectly. Secondly, correspondence analysis also uses the singular value decomposition. In correspondence analysis, the residuals from the independence model are decomposed using the SVD. Recently, Heiser (2004) showed that from the correspondence analysis solution the odds (and thus their ratio) can be obtained by a ratio of two distances that can be read from the display, but this yet again is an indirect representation.

## Acknowledgments

This research was sponsored by the Netherlands Organisation for Scientific Research (NWO), Innovational Grant, no. 452-06-002 to the first author, and by the National Science Foundation award (#SES-0351175) to the second author.

## References

- Agresti, A. (1980). Generalized odds ratios for ordinal data. *Biometrics*, 36, 59–67.
- Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.
- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.
- Altham, P.M.E. (1970). The measurement of association of rows and columns for an  $r \times s$  contingency table. *Journal of the Royal Statistical Society*, B32, 63–73.
- Becker, M.P. (1992). Exploratory analysis of association models using loglinear models and singular value decompositions. *Computational Statistics and Data Analysis*, 13, 253–267.
- Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.
- Carroll, J.D., & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an  $n$ -way generalization of Eckart-Young decomposition. *Psychometrika*, 35, 283–329.
- Clogg, C.C. (1982). Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association*, 77, 803–815.
- Clogg, C.C., & Shidadeh, E.S. (1994). *Statistical models for ordinal variables*. Thousand Oaks, CA: Sage.
- Dale, J.R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, 42, 909–917.
- Dayton, C.M. (1998). *Latent class scaling analysis*. Thousand Oaks, CA: Sage.
- De Rooij, M., & Heiser, W.J. (2005). Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data. *Psychometrika*, 70, 99–123.
- Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211–218.
- Edwardes, M.D. d., & Baltzan, M. (2000). The generalization of the odds ratio, risk ratio and risk difference to  $r \times k$  tables. *Statistics in Medicine*, 19, 1901–1914.
- Goodman, L.A. (1979). Simple models for the analysis of association in cross classifications having ordered categories. *Journal of the American Statistical Association*, 74, 537–552.
- Goodman, L.A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetric models for contingency tables with or without missing entries. *The Annals of Statistics*, 13, 10–69.
- Goodman, L.A., & Kruskal, W.H. (1979). *Measures of association for cross classifications*. New York: Springer-Verlag.
- Heiser, W.J. (2004). Geometric representation of association between categories. *Psychometrika*, 69, 513–545.
- Kiers, H.A.L. (1990). Majorization as a tool for optimizing a class of matrix functions. *Psychometrika*, 55, 417–428.

Kiers, H.A.L., & Ten Berge, J.M.F. (1991). Minimization of a class of matrix trace functions by means of refined majorization. *Psychometrika*, 57, 371–382.

Kroonenberg, P.M. (1983). *Three-mode principal component analysis*. Leiden, The Netherlands: DSWO.

Magnus, J.R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.

Meulman, J.J., & Heiser, W.J. (1983). *The display of bootstrap solutions in multidimensional scaling* (tech. report). Leiden University, The Netherlands.

Tucker, L. (1964). The extension of factor analysis to three-dimensional matrices. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology* (pp. 110–127). New York: Holt, Rinehart, and Winston.

Tucker, L. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 279–311.

Williams, O.D., & Grizzle, J.E. (1972). Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association*, 67, 55–63.

Mark de Rooij

Leiden University  
 Institute for Psychological Research  
 Methodology and Statistics Unit  
 PO Box 9555  
 NL-2300 RB Leiden  
 The Netherlands  
 E-mail rooijm@fsw.leidenuniv.nl

### Appendix: Algorithm

An algorithm for finding the singular values and vectors for multiple matrices with and without equality restrictions over matrices on the vectors or values is presented. The algorithm is described for general matrices,  $\mathbf{A}_k$ , but it can also be applied to matrices of (log) odds ratios,  $\mathbf{A}_k$ . We seek the  $\mathbf{X}_k$ ,  $\mathbf{Y}_k$ , and  $\mathbf{D}_k$  that minimize the least squares loss function,

$$\sum_{k=1}^K \| \mathbf{A}_k - \mathbf{X}_k \mathbf{D}_k \mathbf{Y}_k^T \|^2,$$

where  $\mathbf{X}_k^T \mathbf{X}_k = \mathbf{Y}_k^T \mathbf{Y}_k = \mathbf{I}_{M^*}$ , and  $\mathbf{D}_k$  is an  $(M^* \times M^*)$  diagonal matrix.

- The algorithm is an application of an optimization method proposed by Kiers (Kiers, 1990; Kiers & Ten Berge, 1991). The basic idea is that the least squares loss function is iteratively replaced by a majorizing function that has an easily found minimum. The algorithm proceeds as follows:
1. Obtain starting values by finding the first  $M^*$  singular values and vectors of  $\mathbf{A}_k$  for  $k = 1, \dots, K$ .
  2. Update  $\mathbf{X}_k^{[l]}$ : For each  $k$ , use the current estimates of  $\mathbf{X}_k^{[l]}$ ,  $\mathbf{Y}_k^{[l]}$  and  $\mathbf{D}_k^{[l]}$  to compute the singular value decomposition of

$$\mathbf{X}_k^{[l]} + \frac{1}{d_{11k}^2} (\mathbf{A}_k \mathbf{Y}_k^{[l]} \mathbf{D}_k^{[l]} - \mathbf{X}_k^{[l]} \mathbf{D}_k^{[l]2}) = \mathbf{P}_k \mathbf{A}_k \mathbf{Q}_k^T \quad (32)$$

where  $d_{11k}^2$  is the largest diagonal element of  $\mathbf{D}^{[l]2}$ , and set

$$\mathbf{X}^{[l+1]} = \mathbf{P}_{M^*k} \mathbf{Q}_{M^*k}^T \quad (33)$$

where  $\mathbf{P}_{M^*k}$  and  $\mathbf{Q}_{M^*k}$  are the first  $M^*$  left and right singular vectors found in (32).

3. Update  $\mathbf{Y}_k^{[l]}$ : For each  $k$ , use the current estimates of  $\mathbf{X}_k^{[l]}$ ,  $\mathbf{Y}_k^{[l]}$ , and  $\mathbf{D}_k^{[l]}$  to compute the singular value decomposition of

$$\mathbf{Y}_k^{[l]} + \frac{1}{d_{11k}^2} (\mathbf{A}_k^T \mathbf{X}_k^{[l]} \mathbf{D}_k^{[l]} - \mathbf{Y}_k^{[l]} \mathbf{D}_k^{[l]2}) = \mathbf{P}_k \mathbf{A}_k \mathbf{Q}_k^T, \quad (34)$$

where  $d_{11k}^2$  is the largest diagonal element of  $\mathbf{D}_k^{[l]}$ , and set

$$\mathbf{Y}^{[l+1]} = \mathbf{P}_{M^*k} \mathbf{Q}_{M^*k}^T \quad (35)$$

where  $\mathbf{P}_{M^*k}$  and  $\mathbf{Q}_{M^*k}$  are the first  $M^*$  left and right singular vectors found in (34).

4. Update  $\mathbf{D}_k^{[l]}$ : For each  $k$ , use the current estimates of  $\mathbf{X}_k^{[l]}$ ,  $\mathbf{Y}_k^{[l]}$ , and  $\mathbf{D}_k^{[l]}$  to compute

$$\mathbf{D}_k^{[l+1]} = \text{diag} (\mathbf{X}_k^{[l]T} \mathbf{A}_k \mathbf{Y}_k^{[l]}). \quad (36)$$

5. Check for convergence. If the solution has not converged, go back to step 2.

Equality restrictions on  $\mathbf{X}_k$ ,  $\mathbf{Y}_k$ , and  $\mathbf{D}_k$  can be imposed over  $k$  by appropriate modifications of the up-dating equations. To impose the restriction that  $\mathbf{X}_1 = \dots = \mathbf{X}_K = \mathbf{X}$ , equations (32) and (33) would be replaced by

$$\mathbf{X}^{[l]} + \frac{1}{\sum_k d_{11,k}^2} \left( \sum_k (\mathbf{A}_k \mathbf{Y}_k^{[l]} \mathbf{D}_k^{[l]}) - \mathbf{X}^{[l]} \sum_k (\mathbf{D}_k^{[l]2}) \right) = \mathbf{P} \mathbf{A} \mathbf{Q}^T,$$

and

$$\mathbf{X}^{[l+1]} = \mathbf{P}_{M^*} \mathbf{Q}_{M^*}^T,$$

respectively, and  $\mathbf{X}_k^{[l]}$  would be replaced by  $\mathbf{X}^{[l]}$  in equations (34)–(36). Likewise, the restriction  $\mathbf{Y}_1 = \dots = \mathbf{Y}_K = \mathbf{Y}$  can be imposed by replacing equations (34) and (35) by

$$\mathbf{Y}^{[l]} + \frac{1}{\sum_k d_{11,k}^2} \left( \sum_k (\mathbf{A}_k^T \mathbf{X}_k^{[l]} \mathbf{D}_k^{[l]}) - \mathbf{Y}^{[l]} \sum_k (\mathbf{D}_k^{[l]2}) \right) = \mathbf{P} \mathbf{A} \mathbf{Q}^T,$$

and

$$\mathbf{Y}^{[l+1]} = \mathbf{P}_{M^*} \mathbf{Q}_{M^*}^T,$$

respectively, and replacing  $\mathbf{Y}_k^{[l]}$  by  $\mathbf{Y}^{[l]}$  in all other equations. The up-dating equation to impose the restriction  $\mathbf{D}_1 = \dots = \mathbf{D}_K = \mathbf{D}$  is

$$\mathbf{D}^{[l+1]} = \frac{1}{K} \sum_k \text{diag} (\mathbf{X}_k^{[l]T} \mathbf{A}_k \mathbf{Y}_k^{[l]}).$$