In: Proceedings of Fifth International Workshop on Statistical Modelling. July 9-13, 1990 Toulouse: Université Paul Sabatier p. 99-10

1990

On the relation between latent class analysis and correspondence analysis[1]

Peter G.M. van der Heijden*, Jan de Leeuw** and Ab Mooijaart***

Latent class analysis (LCA) and correspondence analysis (CA) are closely related methods for the analysis of contingency tables. For two–way tables earlier work has shown that both models give reduced rank approximations. In this paper we show that a graphical representation of rescaled LCA parameters can be made that is very similar to the usual simple CA representation. Furthermore the reduced rank interpretation of LCA and simple CA is extended to higher–way tables, and it is shown that in this context LCA and multiple CA (a generalization of CA for the analysis of higher–way tables) are also very similar.
Key words: latent class analysis, correspondence analysis, canonical analysis, data analysis.

*University of Utrecht, Department V.O.S., Section Methods, The Netherlands.
**University of California Los Angeles, Departments of Psychology and Statistics, California, U.S.A.
***University of Leiden, Department of Psychometrics and Research Methods, The Netherlands.

1. Introduction

Latent class analysis (LCA) and correspondence analysis (CA) are methods for the analysis of contingency tables. In this paper we study some aspects of their relation. We discuss two subjects. First, we show for two–way contingency tables how graphical representations can be made using LCA parameters. These representations are very similar to CA representations. Second, we discuss the relation between LCA and CA for the analysis of higher way tables.

2. Two–way tables

2.1 Simple CA and LCA

For two–way contingency tables simple CA provides the decomposition

$$\boldsymbol{\Pi} = \mathbf{D}_r(\mathbf{uu}' + \mathbf{R\Lambda C}')\mathbf{D}_c, \tag{1}$$

where $\boldsymbol{\Pi}$ is a matrix of probabilities π_{ij} (i=1,...,I; j=1,...,J) of rank M; $\mathbf{D}_r$ is a diagonal matrix with marginal probabilities π_{i+}, $\mathbf{D}_c$ is a diagonal matrix with marginal probabilities π_{+j}; $\mathbf{u}$ is a unit column vector whose length depends on the context; $\mathbf{R}$ is a matrix with row scores r_{im} (m=1,...,M), $\mathbf{C}$ is a matrix with column scores c_{jm} (m=1,...,M), and $\boldsymbol{\Lambda}$ is a diagonal matrix with singular values λ_m in decreasing order. The matrices $\mathbf{R}$ and $\mathbf{C}$ have restrictions $\mathbf{u}'\mathbf{D}_r\mathbf{R} = 0 = \mathbf{u}'\mathbf{D}_c\mathbf{C}$, and $\mathbf{R}'\mathbf{D}_r\mathbf{R} = \mathbf{I} = \mathbf{R}'\mathbf{D}_c\mathbf{C}$. If M is chosen to be M=min (I–1,J–1), then every contingency table can be decomposed perfectly with (1).

CA provides a rank decomposition of a probability matrix (Gilula, 1979). This rank interpretation can be seen by rewriting (1a) as $\boldsymbol{\Pi} = \mathbf{R}^*\boldsymbol{\Lambda}^*\mathbf{C}^{*\prime}$, where $\mathbf{R}^*$ is a Ix(M+1) matrix with scores π_{i+} in column 1 and $\pi_{i+}r_{im}$ in column m+1, $\mathbf{C}^*$ is a Jx(M+1) matrix with scores π_{+j} in column 1 and $\pi_{+j}c_{jm}$ in column m+1, and $\boldsymbol{\Lambda}^*$ is a (M+1)x(M+1) matrix

[1] An extended version of this paper appeared in 1990 as internal report # 2 in the Series Methods published by ISOR, University of Utrecht.

with elements 1, λ_1, $\lambda_2,\ldots,\lambda_M$ in decreasing order.

For a matrix **P** with observed proportions the full rank decomposition in (1) can be obtained from a generalized singular value decomposition (GSVD; Greenacre, 1984). By dropping the last columns of **R** and **C**, a lower rank approximation of **Π** is found. For M*<M the rank (M*+1) approximation is optimal w.r.t. $D_r^{-1/2}\Pi D_c^{-1/2}$ in a LS sense (see Greenacre, 1984). Recently CA is also estimated by ML (see Goodman, 1985; Gilula and Haberman, 1986; Francis and Saint-Pierre, 1986). Only for the full rank decomposition (i.e. rank M) the ML estimates are equal to the LS estimates.

LCA for a two-way contingency table also gives a rank decomposition of a probability matrix (see, for example, Gilula, 1979). For rank N LCA is

$$\Pi = \Pi_r \Pi_n \Pi_c' \tag{2}$$

where Π_r is an IxN matrix with probabilities π_{in} restricted as $\Sigma_i \pi_{in} = 1$, Π_c is a JxN matrix with probabilities π_{jn} restricted as $\Sigma_j \pi_{jn} = 1$, and Π_n is a diagonal NxN matrix with latent class sizes π_n ($\Sigma_n \pi_n = 1$) on the diagonal. The model has (I–N)(J–N) degrees of freedom. LCA gives a nonnegative rank decomposition of a probability matrix.

Gilula (1979) shows that a probability matrix can always be decomposed by LCA with N=min (I,J) latent classes. Hence for rank N=M*+1=min (I,J) LCA is equivalent to CA: for both models $\pi_{ij}=p_{ij}$. For rank N=M*+1=1 both models reduce to independence. De Leeuw and van der Heijden (1989) show that for rank N=M*+1=2 LCA and CA are equivalent. It follows that Gilula's statement (Gilula, 1979, 1984; Gilula and Haberman, 1986) that for rank 2 LCA does not necessarily imply CA, is incorrect. For 2 < N=M*+1 < min(I,J) CA always implies LCA but the reverse does not hold (Gilula, 1979).

2.2 Graphical representations of CA

CA is often used to make graphical representations. The full representation of a matrix **P** requires M=min(I–1,J–1) dimensions. Usually only 2 or 3 dimensions are studied. Thus the matrix **P** is approximated by a matrix **Π**. This approximation is optimal in a LS sense or an ML sense, depending on the estimation procedure chosen. We describe here the graphical representations based on **Π**.

Two graphical representations of **Π** are made: one for the rows, using rows of **RΛ** as coordinates, and one for the columns, using rows of **CΛ** as coordinates. In each representation separately distances between points are equal to so-called chi-squared distances. The chi-squared distance $\delta^2(i,i')$ between rows i and i' of the matrix **Π** is defined as

$$\delta^2(i,i') = \sum_{j=1}^{J} \frac{\left(\frac{\pi_{ij}}{\pi_{i+}} - \frac{\pi_{i'j}}{\pi_{i'+}}\right)^2}{\pi_{+j}} \tag{3}$$

(a similar equation can be given for $\delta^2(j,j')$). Distances (4) are defined between two vectors of conditional proportions π_{ij}/π_{i+}, which are called the 'profiles' of row i and i'. The two representations are related by 'transition formulas' $R\Lambda = D_r^{-1}\Pi C$, and $C\Lambda = D_c^{-1}\Pi' R$.

As an illustration we provide an example presented by Caussinus (1986). The matrix is a cross-classification of five age groups and four types of cancer. The fit of the rank 3 model, approximated by ML, is good: G^2=.31, df=2. Graphical displays of the parameters

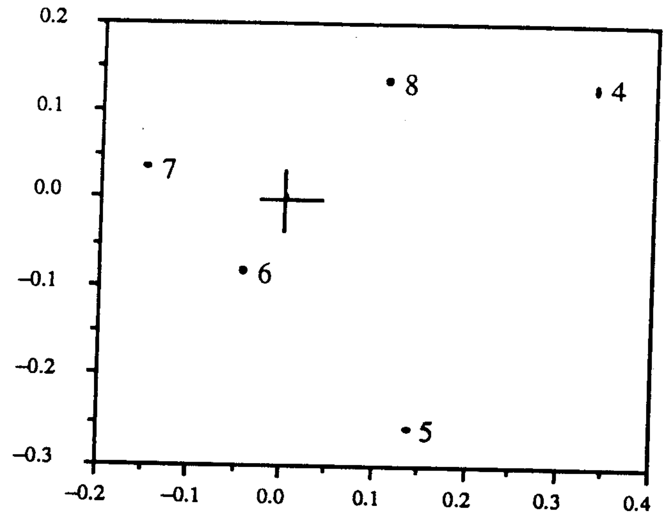

Figure 1a: CA solution, row parameters

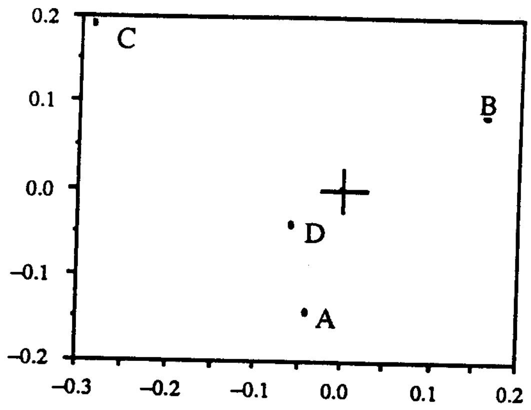

Figure 1b: CA solution, column parameters

are provided in figure 1a and 1b. These represent Π perfectly. Figure 1a shows that the profiles for age groups 4, 5 and 7 are most extreme, the profile of 6 is in between 5 and 7, whereas the profile of 8 is in between 4 and 7. The average profile with values π_{+j} is located in the origin. In figure 1b the (column) profiles of cancer types A, B and C are most extreme, D being in between A and the average profile with values π_{i+}. The transition formulas show that in age group 7 the conditional probability π_{ij}/π_{i+} for cancer C is higher than average (π_{+j}), whereas those for cancers A and B are lower than average. Similarly, the conditional probability π_{ij}/π_{i+} for cancer type A is highest for age group 5, and the conditional probability π_{ij}/π_{i+} for cancer type B is highest for age group 4.

2.3 Graphical displays of LCA

Usually LCA parameter estimates are not presented graphically. Here we propose a graphical presentation. Graphical displays can be made by rescaling the conditional probabilities in Π_r and Π_c (with $\Sigma_i\pi_{in}=1$ and $\Sigma_j\pi_{jn}=1$), into conditional probabilities in Π_r* and Π_c* restricted as $\Sigma_n\pi_{in}*=1$ and $\Sigma_n\pi_{jn}*=1$. The latter parameters are the probabilities to fall into class n given category i or j. We derive Π_r* (and, similarly, Π_c*) as

$$\Pi_r^* = D_r^{-1}\Pi_r\Pi_n \quad (4)$$

It is possible to represent a row in N dimensional space by using the row of Π_r* as coordinates. Each point lies in the N−1 dimensional simplex because for each point its coordinates add up to one. For example, if there are N=3 latent classes, the points lie in the two dimensional triangle spanned by (1,0,0), (0,1,0) and (0,0,1). Such graphical representations are common in compositional data analysis (Aitchison, 1986).

As an example we study again the data in table 1. The fit of the rank 3 model is G^2=.31 (df=2), identical to the fit of CA. For this example the models are equivalent. In figure 2a and 2b we find the representation of the rescaled parameters. The point 'mean' is the point with parameters π_n. The corner points specify the latent classes. The location of the row points shows their relation to each of the latent classes. In figure 2a the persons having age group 7 fall almost all into latent class 1, the persons having age group 5 fall all into class 3, and the persons having age group 4 fall for the largest part in latent class 2,

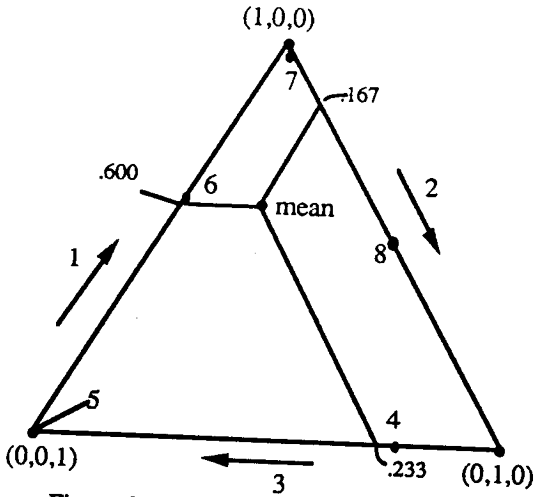

Figure 2a: solution for row parameters

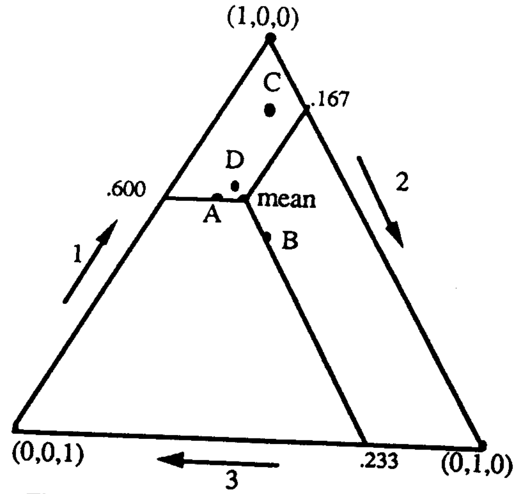

Figure 2b: solution for column parameters

but also in 3. The solution for the column parameters is found in figure 2b. Persons having cancer type C are more than average in class 1, and that the observations of B are more than average in class 2. The two figures supplement each other: for example, persons in age group 7 have more than average cancer type C.

This type of graphical representations were first made in van der Heijden et al. (1989) and de Leeuw et al. (1990) in the context of latent budget analysis (LBA), a reparametrization of LCA proposed by Clogg (1981). LBA is $\mathbf{D_r}^{-1}\Pi = \Pi_r{*}\Pi_c{'}$, which is equivalent to LCA and to $\Pi\mathbf{D_c}^{-1} = \Pi_r\Pi_c{*}'$: they all produce the same estimates of expected probabilities π_{ij}. In LBA $\Pi_r{*}$ and $\Pi_c{*}$ are found directly.

2.4 Comparing CA and LCA

For our example both graphical representations are very similar: if we rotate the LCA triangles and subsequently stretch or shrink their edges in an appropriate way, we find CA. This result holds always if the two models are equivalent. In such cases there exists a matrix $\mathbf{T}$ of order Nx(N–1) such that $\Pi_r{*}\mathbf{T} = \mathbf{R}$, since both $\Pi_r{*}$ and $\mathbf{R}$ give scores for the rows in a reduced rank approximation of $\mathbf{D_r}^{-1}\Pi$. A similar equation holds for $\Pi_c{*}$.

3. Higher–way tables

3.1 MCA and LCA

MCA is an extension of simple CA to the situation of more than two variables. MCA decomposes the so–called Burt matrix by a set of restricted CA's. The Burt matrix is a symmetric matrix that is a concatenation of diagonal univariate margins and the bivariate margins of the higher–way contingency table. For three variables 1, 2 and 3 let the probabilities be denoted as π_{ijk}, where i is the index for variable 1, j for 2 and k (k=1,...,K) for 3. This higher–way table is coded into the Burt matrix having order (I+J+K)x(I+J+K). The sub–matrices on the diagonal of the Burt matrix are diagonal with elements π_{i++}, π_{+j+} and π_{++k} respectively. The off–diagonal sub–matrices are the three bivariate margins that we denote by Π_{12}, Π_{13} and Π_{23}. The Burt matrix is decomposed by the following restricted CA's:

$$\Pi_{12} = \mathbf{D}_1(\mathbf{uu}' + \mathbf{R}_1\Lambda\mathbf{R}_2')\mathbf{D}_2 = \mathbf{R}_1^*\Lambda^*\mathbf{R}_2^{*\prime}, \tag{5a}$$

$$\Pi_{13} = D_1(uu' + R_1\Lambda R_3')D_3 = R_1^*\Lambda^* R_3^{*\prime}, \tag{5b}$$

$$\Pi_{23} = D_2(uu' + R_2\Lambda R_3')D_3 = R_2^*\Lambda^* R_3^{*\prime}, \tag{5c}$$

where $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ are diagonal with elements π_{i++}, π_{+j+} and π_{++k} respectively, and $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ give the scores for variables 1, 2 and 3, restricted as $\mathbf{u}'\mathbf{D}_1\mathbf{R}_1 = \mathbf{u}'\mathbf{D}_2\mathbf{R}_2 = \mathbf{u}'\mathbf{D}_3\mathbf{R}_3 = \mathbf{0}$ and $\mathbf{R}_1'\mathbf{D}_1\mathbf{R}_1 = \mathbf{R}_2'\mathbf{D}_2\mathbf{R}_2 = \mathbf{R}_3'\mathbf{D}_3\mathbf{R}_3 = \mathbf{I}$. Due to the restriction that the scores for a variable are identical in distinct decompositions, maximally M=(I−1)+(J−1)+(K−1) sets of orthogonal scores are required to give a joint decomposition of the univariate and bivariate marginal matrices (see, for example, Greenacre, 1984). The set of decompositions can be easily generalized if there are more than three variables by calculating a set of scores $\mathbf{R}_s$ for each additional variable s.

The scores of MCA can be obtained by a GSVD of the Burt matrix (see Greenacre, 1984). By ommitting the last columns of scores a scaled version of the Burt matrix is approximated in a LS sense. No ML approximations of the Burt matrix are proposed in the literature, although some models are inspired by MCA (see below).

LCA for higher–way tables is a straightforward extension of the model for two–way tables. For three variables the model for the probabilities π_{ijk} is

$$\pi_{ijk} = \sum_{n=1}^{N} \pi_n \pi_{in} \pi_{jn} \pi_{kn}, \tag{6}$$

with restrictions $\Sigma_n \pi_n = \Sigma_i \pi_{in} = \Sigma_j \pi_{jn} = \Sigma_k \pi_{kn}$. For more than three variables model (6) has an extra set of conditional probabilities for each extra manifest variable.

It is not straightforward to compare MCA and LCA, because MCA is defined in terms of the bivariate margins, whereas LCA is defined in terms of the higher–way table. A possible approach to solve this is by defining a model for MCA in terms of the higher–way table. This is done by de Leeuw (1983) and Green (1988, 1989), for example (compare also Choulakian, 1988, and Kohlmann, 1990). For a three–way table their proposal is

$$\log \pi_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + \sum_{p=1}^{P} \mu r_{ip} r_{jp} + \sum_{p=1}^{P} \psi r_{ip} r_{jp} + \sum_{p=1}^{P} \gamma r_{jp} r_{kp} \tag{7}$$

which is a model for log π_{ijk} with bilinear terms to describe two–factor interactions. This model is similar to MCA in the sense that sets of scores are estimated for each variablethat play the same role in each two–variable interaction. In going from MCA to (7) the following assumptions are made. First, the three– and more–variable interactions in the higher–way table can be neglected. This is motivated by the fact that in the Burt matrix only bivariate margins are studied. Second, in going from MCA to logbilinear decompositions, it is assumed that it is allowed to approximate log (1+x) by x, which is only allowed if x is small compared to zero. Third, it is assumed that the two–factor interaction in the full matrix has the same form as the two–factor interaction in each of the bivariate marginal tables.

The similarity of LCA to (7) can be discussed by using the loglinear formulation of LCA in terms of the unobserved contingency table (see Haberman, 1979)

$$\log \pi_{ijkn} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(n)} + u_{14(in)} + u_{24(jn)} + u_{34(kn)} \qquad (8)$$

where the u–parameters add up to zero over each index. In order to find a model for the manifest variables only, we have to add up (8) over the latent variable, indexed by n. The collapsibility theorem in Bishop et al. (1975) shows that this wil change the interaction terms $u_{12(ij)}$, $u_{13(ik)}$, $u_{23(jk)}$ and $u_{123(ijk)}$, that will generally not be equal to zero anymore if we add up over the latent variable. Notice that if the three–factor interaction $u_{123(ijk)}$ in the collapsed version of (8) would be zero, and the two–factor interactions $u_{12(ij)}$, $u_{13(ik)}$ and $u_{23(jk)}$ would be restricted as in (7), then we have the MCA–like model (7). Not much work is done on model (7) yet, and it is not known whether adding up (8) over the N latent classes will lead to the restricted two–factor interactions in (7).

Another insight comes from the work of Whittaker (1989). He suggests that (7) can be derived as follows: let there be three categorical variables that are conditionally independent given N independent normally distributed continuous variables, then, if we add up over these continuous variables, we find (7). These results are relevant for our discussion above. First, as in LCA the manifest variables are conditionally independent given the latent variables. Second, by adding up over the latent variables – which is also done in LCA – Whittaker (1989) arrives at model (7) that was suggested for MCA. The difference with the situation for LCA is, first, that there are N latent continuous variables in Whittaker's formulation, whereas there is one latent variable with N classes in LCA, and, second, by assuming a normal distribution for each of the continuous variables, model (7) does not need three– and higher–factor terms.

A new way to relate MCA and LCA is presented in section 3.4.

3.2 Graphical displays of MCA

In MCA graphical representations are made for the category points. Usually only the first few dimensions of the full–dimensional solution are studied, for example, for the variable s only the first few columns of $\mathbf{R}_s\Lambda$ serve as coordinates. We give an example derived from data published in McCutcheon (1987).

A data set with four variables is analyzed, dealing with the attitude of respondents towards survey research. Respondents could judge the *Purpose* of survey research as 'good', 'depends' and 'waste of time', the *Accuracy* as 'mostly true' and 'not true'; the *Understanding* of respondents was judged by the interviewer as 'good' or 'fair, poor', and the *Cooperation* as 'interested', 'cooperative' or 'impatient/hostile'. The first two dimensions of the solution are given in figure 3. The first dimension shows the distinction between 'good' respondents, who have a positive attitude towards the purpose and accuracy, and seem to be interested and understand it well, versus 'bad' respondents on the right. On the second dimension we find at the top that being merely cooperative or even impatient/hostile goes together more often than average with a fair/poor understanding, whereas a judged purpose of the survey as depends or waste is found more often than average with an answer 'not true' on accuracy.

3.3 Graphical displays of LCA

We can make graphical representations of LCA parameters by rescaling the parameters in the same way as we did in section 2.2 for two variables: i.e., we collect the parameters π_{in}, π_{jn} and π_{kn} in matrices Π_1, Π_2 and Π_3, and rescale these as in (4) to Π_1*, Π_2* and Π_3*. As before, a rescaled parameter specifies the probability to fall into latent class n given that someone falls into a specific manifest category. The rescaled parameters are used to make a graphical representation for each of the manifest variables separately. For the example discussed earlier in section 3.2 we will overlay these representations. The fit for N=3 latent

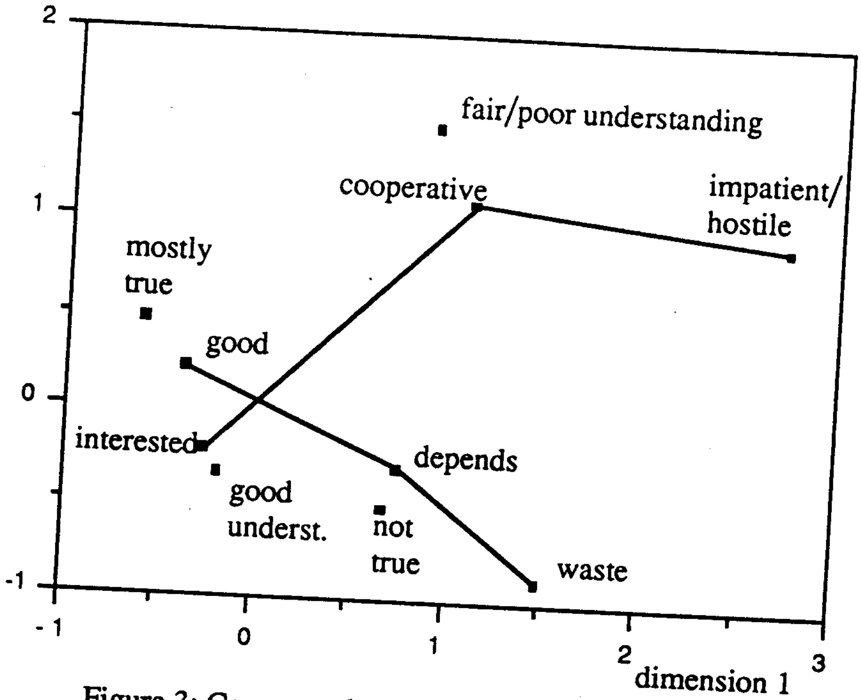

Figure 3: Correspondence analysis for example 1

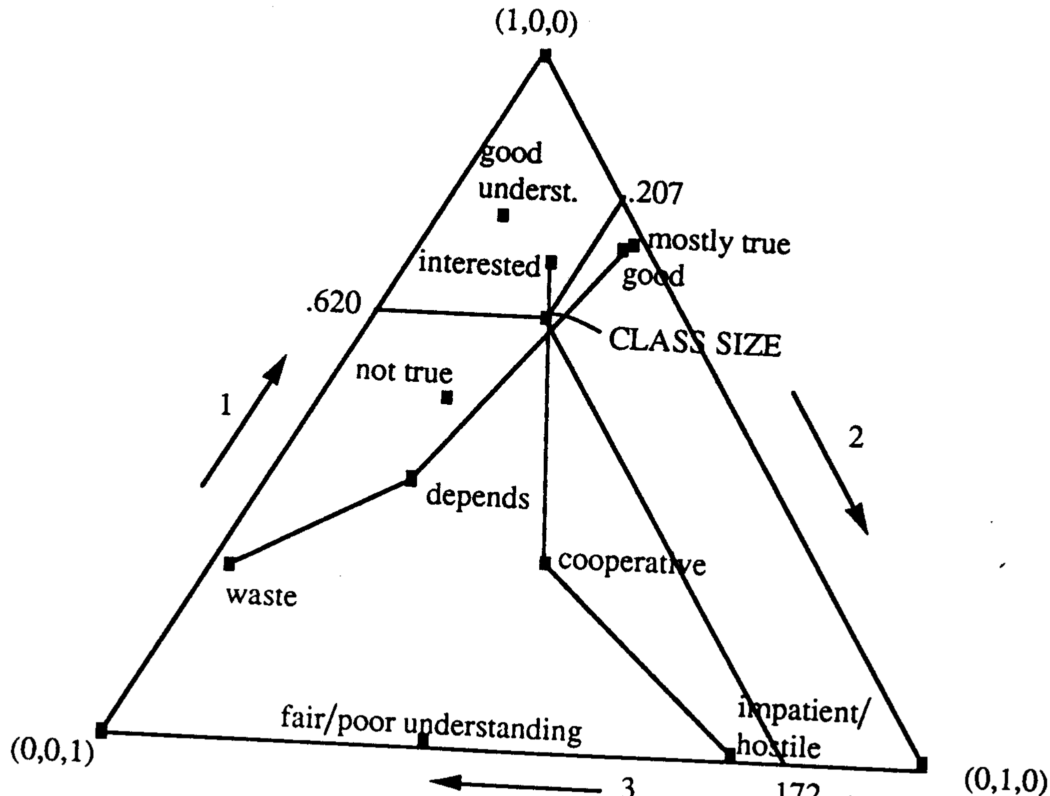

Figure 4: Latent class analysis, rescaled parameter estimates

classes is adequate: $G^2 = 21.89$ (df=16). A graphical representation is given in figure 4.

3.4 Comparing MCA and LCA

For the two examples the configurations of points for LCA are remarkably similar (but not identical) to those for MCA. In other words, there exists a matrix **T** of order (N−1)xN that makes $\Pi_1 T$ similar to R_1, $\Pi_2 T$ similar to R_2, and $\Pi_3 T$ similar to R_3 (compare 2.4).

For the two–variable case LCA and simple CA could be linked by showing that they are reduced rank decompositions for the two–way contingency table. For the situation with more than three variables earlier work has tried to link LCA and MCA by reformulation MCA as a model for a higher–way table (see section 3.1).

However, we can also try to relate LCA and MCA by studying what LCA implies in terms of the Burt matrix. For higher–way tables LCA can be considered as a reduced rank model in the sense that, for N latent classes, the matrix of probabilities π_{ijk} is the sum of N independent matrices (see (6)). So the higher–way matrix is approximated by a higher–way rank N model. By adding up (6) over the third variable, for example, we find (2). Thus the three bivariate margins are also approximated *implicitly* by the rank N matrix. It follows that LCA and MCA are related via the Burt matrix because they both provide reduced rank approximations of the bivariate margins in the Burt matrix.

References

Aitchison, J. (1986). *The statistical analysis of compositional data.* London: Chapman and Hall.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). *Discrete multivariate analysis.* Cambridge: M.I.T.Press

Caussinus, H. (1986). In discussion of: L.A. Goodman, Some useful extensions of the usual correspondence analysis approach and the usual loglinear models approach in the analysis of contingency tables. *International statistical review, 54,* 243–309.

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

Clogg, C.C. (1981) Latent structure models of mobility. *American Journal of Sociology, 86,* 836–868.

de Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. *Journal of Econometrics, 22,* 113–137.

de Leeuw, J., and van der Heijden, P.G.M. (1989). *Reduced rank models for contingency tables.* Leiden: Department of Psychometrics and Research Methods, internal report PRM 89–04.

de Leeuw, J., van der Heijden, P.G.M., and Verboon, P. (1990). A latent time budget model. *Statistica Neerlandica.*

Francis, B. and Saint Pierre, J. (1986). Correspondence analysis models using GLIM. *COMPSTAT 86 Short communications and posters.* Rome: Dipartimento di Statistica Probabilita e Statistiche Applicate, Universita 'La Sapienza'.

Gilula, Z. (1979). Singular value decomposition of probability matrices: probabilistic aspects of latent dichotomous variables. *Biometrics, 66,* 339–344.

Gilula, Z. (1984). On some similarities between canonical correlation models and latent class models for two–way contingency tables. *Biometrika,* 71, 523–529.

Gilula, Z. and Haberman, S.J. (1986). Canonical analysis of contingency tables by maximum likelihood. *Journal of the American Statistical Association,* 81, 780–788.

Goodman, L.A. (1985). The analysis of cross–classified data having ordered and/or unordered categories: association models, correlation models and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics, 13,* 10–69.

Green, M. (1988) A modelling approach to multiple correspondence analysis. In: *Proceedings COMPSTAT 1988.* Vienna: Physica Verlag.

Green, M. (1989). Generalizations of the Goodman association model for the analysis of multi–dimensional contingency tables. In: A. Decarli, B.J. Francis, R.Gilchrist and G.U.H.Seeber (Eds.). *Statistical Modelling. Proceedings, Trento, 1989.* Berlin, Springer Verlag.

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

Haberman, S.J. (1979). *Analysis of qualitative data (2 vols.).* New York: Academic Press.

Kohlmann, T. (1990). *Homogeneity analysis and analysis of partial association in the quantification of multiple categorical indicators.* Paper presented at the SMABS 90 conference, Marburg.

McCutcheon, A.L. (1987). *Latent class analysis. Series: Quantitative applications in the social sciences, 64.* London: Sage.

van der Heijden, P.G.M., Mooijaart, A. and de Leeuw, J. (1989). Latent budget analysis. In: A. Decarli, B.J. Francis, R.Gilchrist and G.U.H.Seeber (Eds.). *Statistical Modelling. Proceedings, Trento, 1989.* Berlin, Springer Verlag.

Whittaker, J. (1989). In discussion of: van der Heijden, P.G.M., de Falguerolles, A., and de Leeuw, J.. A combined approach to contingency table analysis using correspondence analysis and loglinear analysis. *Applied Statistics, 38,* 249–292.