Joint Correspondence Analysis (JCA) by Maximum Likelihood

Jeroen K. Vermunt[1] and Carolyn J. Anderson[2]

[1]Department of Methodology and Statistics, Tilburg University, The Netherlands,
[2]Department of Educational Psychology, University of Illinois, Urbana, IL, USA

Abstract. Parameter estimation in joint correspondence analysis (JCA) is typically performed by weighted least squares using the Burt matrix as the data matrix. In this paper, we show how to estimate the JCA model by means of maximum likelihood. For that purpose, JCA is defined as a model for the full *K*-way distribution by generalizing the correspondence analysis model for three-way tables proposed by Choulakian (1988a, 1988b). The advantage of placing JCA in a more formal statistical framework is that standard chi-squared tests can be applied to assess the goodness-of-fit of unrestricted and restricted models.

Keywords: Categorical data analysis, optimal scaling, correlation models, bilinear models, likelihood-ratio chi-squared tests

Introduction

Correspondence analysis (CA) is a popular technique for the exploratory analysis of two-way frequency tables. Widely used statistical software packages such as SPSS, SAS, and BMDP contain a CA routine. Two types of related extensions have been developed for the analysis of *K*-way frequency tables: Multiple correspondence analysis (MCA) and joint correspondence analysis (JCA; Greenacre, 1988, 1993). MCA is a form of principal component analysis, while JCA is factor-analytic technique for categorical variables (Boik, 1996).

Similarly to standard factor analysis, the JCA model is defined in terms of the second-order moments. Since we are dealing with categorical variables, the second-order moments are the two-way marginal frequencies. The matrix with one-way margins on the diagonal and two-way margins on the off-diagonal blocks is referred to as the Burt matrix. Parameter estimation is typically performed by weighted least squares (WLS; Greenacre, 1988; Boik, 1996) using this Burt matrix as data matrix. Recently, Tateneni and Browne (2000) presented a slightly different noniterative estimation procedure that is also based on the Burt matrix. The main advantage of ignoring higher-order moments is that it is possible to deal with large numbers of variables, which is important in exploratory data analysis. An important disadvantage is, however, that there are no formal statistical tests to assess whether a particular model fits the data. This makes it impossible to use JCA in a more confirmatory manner as can be done with standard factor analysis.

In this paper, we show how to estimate the JCA model by means of maximum likelihood (ML). For standard CA of two-way tables, ML estimation methods have been developed, yielding what is known as a row-column correlation model (Goodman, 1985, 1987) or canonical analysis of two-way tables (Gilula & Haberman, 1986; De Leeuw & Van der Heijden, 1991). To be able to estimate the JCA model by ML, it has to be defined as a model for the full K-way distribution rather than as a model for the bivariate marginal distributions. The formulation we propose is a generalization to *K*-way tables of the CA model for three-way tables proposed by Choulakian (1988a, 1988b). An important feature of our new model is that the bivariate marginal distributions are exactly in agreement with the constraints implied by JCA.

An advantage of the proposed ML method compared to the standard limited information WLS approach is that JCA is placed in a more formal statistical framework. We are now able to apply standard goodness-of-fit chi-squared tests to assess the overall fit of a model, as well as to compare competing models with one another in order to check whether certain restrictions hold.

The next section describes standard JCA. In Section 3 we derive the formulation of JCA as a model for a *K*-way table. Section 4 presents two empirical examples and Section 5 concludes.

DOI 10.1027/1614-1881.1.1.18

Joint Correspondence Analysis

Let $\pi_{y_1 y_2 \ldots y_K}^{Y_1 Y_2 \ldots Y_K}$ denote an expected cell proportion in the K-way contingency table formed by the categorical variables Y_1, Y_2, ..., and Y_K. The number of levels of variable Y_k is denoted by J_k and a particular level by y_k (i.e., y_k = 1, ..., J_k).

JCA can be defined as a model for all bivariate marginal distributions $\pi_{y_k y_l}^{Y_k Y_l}$. For each variable pair Y_k and Y_l, where $k \neq l$, the R-dimensional JCA model states that

$$\pi_{y_k y_l}^{Y_k Y_l} = \pi_{y_k}^{Y_k} \pi_{y_l}^{Y_l} \left(1 + \sum_{r=1}^{R} \lambda_r \eta_{r y_k}^{Y_k} \eta_{r y_l}^{Y_l} \right). \quad (1)$$

Here, $\pi_{y_k}^{Y_k}$ denotes an entry in the univariate marginal distribution of variable Y_k, $\eta_{r y_k}^{Y_k}$ is the quantification or scale value of category y_k of variable Y_k for dimension r, and λ_r is the singular value or the "average" correlation between the variables in dimension r.

It should be noted that with $K = 2$, the model described in Equation (1) equals a standard CA model. In addition, a MCA is obtained by dropping the condition that $k \neq l$.

For identification purposes, several constraints have to be imposed on the model parameters. The typical centering, scaling, and orthogonalization constraints are:

$$\sum_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} \eta_{r y_k}^{Y_k} = 0, \quad \sum_{k=1}^{K} \sum_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} (\eta_{r y_k}^{Y_k})^2 = K, \sum_{k=1}^{K} \sum_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} \eta_{r y_k}^{Y_k} \eta_{s y_k}^{Y_k} = 0,$$

for all r and $s \neq r$. As can be seen, each set of category quantifications is assumed to be centered. To identify λ_r, one has to impose one scaling constraint per dimension. Furthermore, to uniquely determine the various dimensions, one orthogonalization constraint must be imposed per pair of dimensions. The scaling and orthogonalization constraints involve a sum over all variables, as is common practice in MCA. Similar to factor analysis, other types of constraint can be used to identify λ_r, such as, for instance, imposing the scaling constraint on a single variable, say Y_1. The same applies to the necessary constraints to uniquely determine the various dimensions. Alternative identification constraints are, for example, equating r–1 scale values to zero in dimension r or orthogonalizing the scale values of a single variable.

It is also possible to drop the scaling constraints and absorb the λ_r parameters in the category quantifications. In that case, we obtain

$$\pi_{y_k y_l}^{Y_k Y_l} = \pi_{y_k}^{Y_k} \pi_{y_l}^{Y_l} \left(1 + \sum_{r=1}^{R} \eta_{r y_k}^{Y_k *} \eta_{r y_l}^{Y_l *} \right) \quad (2)$$

for all pairs Y_k and Y_l, with

$$\sum_{y_k=1}^{j_k} \pi_{y_k}^{Y_k} \eta_{r y_k}^{Y_k *} = 0, \sum_{k=1}^{K} \sum_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} \eta_{r y_k}^{Y_k *} \eta_{s y_k}^{Y_k *} = 0$$

for all r and $r \neq s$, respectively. It can easily be verified that $\eta_{r y_k}^{Y_k *} = \sqrt{\lambda_r}\, \eta_{r y_k}^{Y_k}$.

Estimation of the parameters is typically done by means of weighted least squares (Greenacre, 1988; Boik 1996). The appendix provides more detail on parameter estimation.

A Multivariate Correlation Model

The ML-variant of CA is called the row-column correlation model (RCCM, Goodman, 1985, 1987) or canonical analysis of two-way tables (Gilula & Haberman, 1986). Several extensions have been proposed for tables with more than two dimensions. Gilula and Haberman (1988) suggested dividing the cross-classified variables into two sets, each of which can be treated as a single polytomous variable. A restricted canonical correlation model is specified for this "two-way" table, where the category quantifications are linear functions of the original variables.

Choulakian (1988a) proposed the following extension of the RCCM for the trivariate case:

$$\pi_{y_1 y_2 y_3}^{Y_1 Y_2 Y_3} = \pi_{y_1}^{Y_1} \pi_{y_2}^{Y_2} \pi_{y_3}^{Y_3} \left[1 + \sum_{r=1}^{R} (\sigma_r \nu_{r y_1}^{Y_1} \nu_{r y_2}^{Y_2} + \sigma_r \nu_{r y_1}^{Y_1} \nu_{r y_3}^{Y_3} + \sigma_r \nu_{r y_2}^{Y_2} \nu_{r y_3}^{Y_3}) \right]. \quad (3)$$

This trivariate correlation model (TCM) can also be formulated in a slightly more compact manner as

$$\pi_{y_1 y_2 y_3}^{Y_1 Y_2 Y_3} = \left(\prod_{k=1}^{3} \pi_{y_k}^{Y_k} \right) \left(1 + \sum_{r=1}^{R} \sum_{k=1}^{3} \sum_{l=k+1}^{3} \sigma_r \nu_{r y_k}^{Y_k} \nu_{r y_l}^{Y_l} \right).$$

The meaning of the parameters is similar to the ones in JCA: $\pi_{y_k}^{Y_k}$ denotes the marginal probability that $Y_k = y_k$, R is the number of dimensions, ρ_r denotes the canonical correlation in dimension r, and $\nu_{r y_k}^{Y_k}$ is the quantification of category y_k of variable Y_k for dimension r.

The constraints on the $\nu_{r y_k}^{Y_k}$ parameters are the same as the ones on the $\eta_{r y_k}^{Y_k}$ parameters in the JCA model described in the previous section; that is,

$$\sum_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} \nu_{r y_k}^{Y_k} = 0, \sum_{k=1}^{3} \sum_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} (\nu_{r y_k}^{Y_k})^2 = 3, \sum_{k=1}^{3} \sum_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} \nu_{r y_k}^{Y_k} \nu_{s y_k}^{Y_k} = 0,$$

As can be seen, the quantifications are assumed to be centered for each variable and each dimension. It is important to note that the centering constraints are not arbitrary constraints needed for identification but real model restrictions. In fact, the centering restrictions are necessary to guarantee that the univariate marginal distributions are reproduced by the model. In addition to the centering restrictions, one scaling constraint has to be imposed per dimension in order to identify ρ_r. Furthermore, in order to uniquely define the various dimensions or solve the rotation problem, one orthogonalization constraint has to be imposed per pair of dimensions.

In the original paper, Choulakian (1988a) proposed imposing the orthogonalizing constraints per variable, $\Sigma_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} \nu_{ry_k}^{Y_k} \nu_{sy_k}^{Y_k} = 0$, but this was corrected in an Errata (Choulakian, 1988b). Another minor difference is that he imposed scaling constraints on each variable separately, $\Sigma_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} (\nu_{ry_k}^{Y_k})^2 = 1$, and was therefore able to identify different ρ_r per variable pair. Our representation and his are, however, equivalent.

Although Choulakian's TCM differs from JCA in that it is a model for a trivariate distribution rather than for the three bivariate distributions, it is strongly related to JCA. The exact relationship becomes visible if we derive the implications of Choulakian's model for the bivariate marginal tables. Let us take $\pi_{y_1y_2}^{Y_1Y_2}$ as an example:

$$\pi_{y_1y_2}^{Y_1Y_2} = \sum_{y_3=1}^{J_3} \pi_{y_1}^{Y_1}\pi_{y_2}^{Y_2}\pi_{y_3}^{Y_3} \left[1 + \sum_{r=1}^{R} (\sigma_r \nu_{ry_1}^{Y_1}\nu_{ry_2}^{Y_2} + \sigma_r \nu_{ry_1}^{Y_1}\nu_{ry_3}^{Y_3} + \sigma_r \nu_{ry_2}^{Y_2}\nu_{ry_3}^{Y_3} \right]$$

$$= \pi_{y_1}^{Y_1}\pi_{y_2}^{Y_2}\left[\sum_{y_3=1}^{J_3} \pi_{y_3}^{Y_3} + \sum_{r=1}^{R} \left(\sigma_r \nu_{ry_1}^{Y_1}\nu_{ry_2}^{Y_2} \sum_{y_3=1}^{J_3} \pi_{y_3}^{Y_3} + \sigma_r \nu_{ry_1}^{Y_1} \sum_{y_3=1}^{J_3} \pi_{y_3}^{Y_3}\nu_{ry_3}^{Y_3} + \sigma_r \nu_{ry_2}^{Y_2} \sum_{y_3=1}^{J_3} \pi_{y_3}^{Y_3}\nu_{y_3}^{Y_3} \right) \right]$$

$$= \pi_{y_1}^{Y_1}\pi_{y_2}^{Y_2} \left(1 + \sum_{r=1}^{R} \sigma_r \nu_{ry_1}^{Y_1}\nu_{ry_2}^{Y_2} \right).$$

The last simplification is based on the fact that $\Sigma_{y_3=1}^{J_3} \pi_{y_3}^{Y_3} = 1$ and $\Sigma_{y_3=1}^{J_3} \pi_{y_3}^{Y_3} \nu_{ry_3}^{Y_3} = 0$.

The above derivation shows that as far as the bivariate marginals are concerned, the model proposed by Choulakian is equivalent to JCA. In other words, the TCM can be seen as the underlying model for the three-way table when the JCA model holds for the two-way tables. Choulakian proposed estimating his model by means of ML yielding what could be called a ML variant of JCA for the three-variable case.

Using the results on the relationship between Choulakian's extended row-column correlation model and JCA, we propose the following extension to K-way tables:

$$\pi_{y_1y_2\ldots y_K}^{Y_1Y_2\ldots Y_K} = \left(\prod_{k=1}^{K} \pi_{y_k}^{Y_k} \right) \left(1 + \sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{l=k+1}^{K} \sigma_r \nu_{ry_k}^{Y_k} \nu_{ry_l}^{Y_l} \right). \quad (4)$$

We label this model a multivariate correlation model (MCM). The meaning of the parameters and the identifying constraints are the same as in the trivariate case, of course, with 3 replaced by K.

The proposed multivariate extension of the RCCM is similar to the class of row-column association models proposed by Anderson and Vermunt (2000). One of their row-column association models has exactly the same set of bi-linear terms as the model described in Equation (4).

As in JCA, it is possible to drop the scaling constraint and absorb the λ_r parameter in the category quantifications. In that case, we obtain

$$\pi_{y_1y_2\ldots y_K}^{Y_1Y_2\ldots Y_K} = \left(\prod_{k=1}^{K} \pi_{y_k}^{Y_k} \right) \left(1 + \sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{l=k+1}^{K} \nu_{ry_k}^{Y_k*} \nu_{ry_l}^{Y_l*} \right), \quad (5)$$

where $\nu_{ry_k}^{Y_k*} = \sqrt{\sigma_r}\, \nu_{ry_k}^{Y_k}$.

Because the MCM is a model for the joint distribution, its parameters can be estimated by means of maximum likelihood (ML) assuming a Poisson sampling scheme. In the computation of the ML estimates it is important to take into account the centering constraints, which are not arbitrary constraints needed for identification but are real model restrictions. Another important issue is that the algorithm should guarantee that all estimated cell entries are at least zero. The Appendix describes two algorithms for obtaining the ML estimation, a simple unidimensional Newton method and a Fisher scoring method.

Comparison of JCA and the MCM

It can be verified that the JCA model and the MCM defined in Equations (1) and (4) have the same number of free parameters. With R dimensions, the number of free parameters, T_R, equals

$$T_R = T_0 + R \cdot (T_0 + 1) - R \cdot \left(1 + \frac{(R-1)}{2} \right)$$
$$= T_0 + R \cdot \left(T_0 + 1 - \frac{(R+1)}{2} \right).$$

Here, T_0 denotes the number of parameters of the independence model, the model with 0 dimensions: $T_0 = \Sigma_{k=1}^{K} (J_k - 1)$.

Despite the fact that the bilinear structures appearing in the JCA model and the MCM are similar, the important difference is, of course, that the former is defined as a model for all two-way tables while the latter is a model for the K-way table. As in the trivariate case, collapsing $\pi_{y_1y_2\ldots y_K}^{Y_1Y_2\ldots Y_K}$, as defined in Equation (4), over all variables except for Y_k and Y_l yields

$$\pi_{y_ky_l}^{Y_kY_l} = \pi_{y_k}^{Y_k}\pi_{y_l}^{Y_l} \left(1 + \sum_{r=1}^{R} \sigma_r \nu_{ry_k}^{Y_k}\nu_{ry_l}^{Y_l} \right). \quad (6)$$

The bilinear terms involving variables other than Y_k

or Y_l, say Y_m, cancel because $\Sigma_{y_m=1}^{J_m} \pi_{y_m}^{Y_m} = 1$ and $\Sigma_{y_m=1}^{J_m} \pi_{y_m}^{Y_m} \nu_{ry_m}^{Y_m} = 0$.

It will be clear that, apart from the notation, Equation (6) is equivalent to Equation (1). This shows that a MCM implies that the two-way tables are in agreement with a JCA model. Consequently, a R-dimensional JCA will exactly reproduce the Burt matrix obtained from the estimated frequencies of a R-dimensional MCM.

The relationship between MCM and JCA shown in Equation (6) also suggests how to obtain estimated cell entries in the K-way table using the results from a JCA; that is, how to derive the reversed relationship between the two models: We may fill in the JCA parameters in the MCM model; that is,

$$\pi_{y_1 y_2 \ldots y_K}^{Y_1 Y_2 \ldots Y_K} = \left(\prod_{k=1}^{K} \pi_{y_k}^{Y_k} \right) \left(1 + \sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{l=k+1}^{K} \lambda_r \eta_{ry_k}^{Y_k} \eta_{ry_l}^{Y_l} \right). \quad (7)$$

In the WLS estimation method of JCA, there is no guarantee that all estimated cell probabilities will be at least zero. Note that even some entries in the two-way tables, $\pi_{y_k y_l}^{Y_k Y_l}$, may be smaller than zero. When all estimated $\pi_{y_1 y_2 \ldots y_K}^{Y_1 Y_2 \ldots Y_K}$ are in the permissible range, estimation of a MCM using such a constructed K-way table as data matrix will give a perfect fit, and with the same identifying constraints also the same parameter estimates.

On the basis of the above comparison, it can be concluded is that the proposed MCM can be regarded as a full information ML-variant of JCA. The implication is that the MCM formulation can be used to place JCA within a more formal statistical framework. This makes it possible to assess the goodness-of-fit of the specified model using asymptotic chi-squared tests, as well as to perform more confirmatory analyses. On the other hand, JCA can seen as a limited information WLS-variant of the MCM. Actually, we are dealing with two equivalent models that are estimated in different manners.

Examples

Tables 1 and 2 present two small data sets that we will use to illustrate the new MCM, as well as to compare it with other factor-analytic techniques for categorical variables. Table 1 cross-tabulates 5 dichotomous political attitude variables from the Political Action Survey (Hagenaars, 1993). Table 2 is a four-way cross-tabulation taken from McCutcheon (1987). The items from the General Social Survey 1982 measure respondents' (Y_1 and Y_2) and interviewers' (Y_3 and Y_4) evaluation of the survey.

Tables 3 and 4 present the testing results for the various models we estimated using these two data sets. Besides the independence model and MCM models with different number of dimensions, we used Bock's nominal response model (NMR), the latent class cluster model (LCCM), the latent class factor model (LCFM), and the multivariate association model (MAM). Bock's (1972) NRM is an IRT model that could be used for these types of variables. The LCCM model was used because of the similarity between JCA and latent class analysis pointed out by Van der Heijden, Gilula, and Van der Ark (1999). The LCFM (Magidson & Vermunt, 2001) is similar to Bock's NRM, except for the fact that the latent variables are assumed to be dichotomous instead of continuous. The MAM is a factor-analytic model that has the same

Table 1. Cross-tabulation of five variables from the Political Action Survey.

System Responsive-ness (Y_1)	Ideological Level (Y_2)	Repression Potential (Y_3)	Protest Approval (Y_4)	Conventional Participation (Y_5) 1. Low	2. High
1. Low	1. Low	1. High	1. Low	109	8
			2. High	59	44
		2. Low	1. Low	28	18
			2. High	48	54
	2. High	1. High	1. Low	4	19
			2. High	7	32
		2. Low	1. Low	3	3
			2. High	10	26
2. High	1. Low	1. High	1.Low	49	92
			2. High	46	96
		2. Low	1. Low	16	16
			2. High	33	80
	2. High	1. High	1. Low	7	38
			2. High	10	63
		2. Low	1. Low	3	12
			2. High	8	55

Table 2. Cross-tabulation of four variables from the 1982 General Social Survey.

Purpose (Y_1)	Accuracy (Y_2)	Understanding (Y_3)	Cooperation (Y_4) 1. Interested	2. Cooperative	3.Hostile/ Impatient
1. Good	1. Mostly true	1. Good	419	35	2
		2. Fair/Poor	71	25	5
	2. Not true	1. Good	270	25	4
		2. Fair/Poor	42	16	5
2. Depends	1. Mostly true	1. Good	23	4	1
		2. Fair/Poor	6	2	0
	2. Not true	1. Good	43	9	2
		2. Fair/Poor	9	3	2
3. Waste	1. Mostly True	1. Good	26	3	0
		2. Fair/Poor	1	2	0
	2. Not true	1. Good	85	23	6
		2. Fair/Poor	13	12	8

Table 3. Testing results of the models estimated with the Political Action Survey data.

Model	G^2	*df*	*p*	*BIC*	ΔG^2
Independence	296.56	26	.00	113.19	.00
MCM(1)	105.09	21	.00	–43.02	.65
MCM(2)	29.39	17	.03	–90.51	.90
LCCM(2)	95.79	20	.00	–45.26	.68
LCCM(3)	24.28	14	.04	–74.46	.92
LCFM(1)	95.79	20	.00	–45.26	.68
LCFM(2)	11.73	14	.63	–87.00	.96
NRM(1)	98.46	21	.00	–49.64	.67
NRM(2)	15.92	16	.46	–96.92	.95
MAM(1)	98.49	21	.00	–49.61	.67
MAM(2)	16.21	17	.51	–103.69	.95

1. MCM = Multivariate Correlation Model; LCCM = Latent Class Cluster Model; LCFM = Latent Class Factor Model; NRM = Nominal Response Model; MAM = Multivariate Association Model. 2. ΔG^2 is the proportional reduction of G^2 compared to the independence model.

Table 4. Testing results of the models estimated with the 1982 Social Survey data.

Model	G^2	*df*	*p*	*BIC*	ΔG^2
Independence	257.26	29	.00	51.60	.00
MCM(1)	98.59	23	.00	– 64.54	.62
MCM(2)	28.57	18	.05	– 99.08	.89
LCCM(2)	79.34	22	.00	– 76.68	.69
LCCM(3)	21.89	15	.11	– 84.48	.91
LCFM(1)	79.34	22	.00	– 76.68	.69
LCFM(2)	10.93	15	.76	– 95.45	.96
NRM(1)	81.43	23	.00	– 81.68	.68
NRM(2)	12.40	17	.78	–108.16	.95
MAM(1)	80.34	23	.00	– 82.80	.69
MAM(2)	13.13	18	.78	–114.53	.95

1. MCM = Multivariate Correlation Model; LCCM = Latent Class Cluster Model; LCFM = Latent Class Factor Model; NRM = Nominal Response Model; MAM = Multivariate Association Model. 2. ΔG^2 is the proportional reduction of G^2 compared to the independence model.

types of bilinear terms as the MCM described in this paper (Anderson & Vermunt, 2000).

The measures reported in Tables 3 and 4 are the likelihood-ratio chi-squared (G^2), its associated number of degrees of freedom and p value, the Bayesian information criterion *(BIC)*, and the proportional reduction in G^2 compared to the independence model (ΔG^2).

As far as the testing results are concerned, we see the

Table 5. Parameter estimates and standard errors for MCM(2) obtained with the Political Action data.

	Unrotated solution		Rotated solution	
	$r = 1$	$r = 2$	$r = 1$	$r = 2$
σ_r	0.145 (0.010)	0.089 (0.024)	0.129 (0.011)	0.105 (0.025)
$\nu^{Y_1}_{r1}$	–0.858 (0.167)	0.725 (0.216)	–1.090 (0.133)	0.028 (0.151)
$\nu^{Y_1}_{r2}$	0.731 (0.142)	–0.618 (0.184)	0.929 (0.113)	–0.024 (0.129)
$\nu^{Y_2}_{r1}$	–0.613 (0.070)	0.238 (0.178)	–0.655 (0.074)	–0.199 (0.102)
$\nu^{Y_2}_{r2}$	1.751 (0.200)	–0.678 (0.509)	1.870 (0.211)	0.567 (0.292)
$\nu^{Y_3}_{r1}$	–0.618 (0.359)	–1.251 (0.160)	0.000 (0.000)	–1.362 (0.275)
$\nu^{Y_3}_{r2}$	1.111 (0.646)	2.251 (0.288)	0.000 (0.000)	2.451 (0.495)
$\nu^{Y_4}_{r1}$	–1.245 (0.182)	–0.877 (0.777)	–0.728 (0.152)	–1.464 (0.602)
$\nu^{Y_4}_{r2}$	0.900 (0.132)	0.634 (0.561)	0.526 (0.110)	1.058 (0.435)
$\nu^{Y_5}_{r1}$	–1.559 (0.301)	1.287 (0.258)	–1.967 (0.170)	0.028 (0.185)
$\nu^{Y_5}_{r2}$	0.958 (0.185)	–0.791 (0.158)	1.209 (0.105)	–0.017 (0.114)

Table 6. Parameter estimates and standard errors for MCM(2) obtained with the 1982 General Social Survey data.

	Unrotated solution		Rotated solution	
	$r = 1$	$r = 2$	$r = 1$	$r = 2$
σ_r	0.182 (0.021)	0.119 (0.032)	0.170 (0.026)	0.131 (0.029)
$\nu^{Y_1}_{r1}$	–0.777 (0.164)	0.415 (0.132)	–0.874 (0.113)	–0.035 (0.102)
$\nu^{Y_1}_{r2}$	1.777 (0.465)	–0.461 (0.520)	1.824 (0.503)	0.501 (0.453)
$\nu^{Y_1}_{r3}$	2.954 (0.678)	–1.864 (0.569)	3.426 (0.441)	–0.114 (0.394)
$\nu^{Y_2}_{r1}$	–0.915 (0.178)	0.536 (0.345)	–1.046 (0.285)	0.000 (0.000)
$\nu^{Y_2}_{r2}$	0.991 (0.193)	–0.581 (0.374)	1.133 (0.309)	0.000 (0.000)
$\nu^{Y_3}_{r1}$	–0.268 (0.229)	–0.671 (0.377)	–0.010 (0.070)	–0.715 (0.422)
$\nu^{Y_3}_{r2}$	1.185 (1.010)	2.963 (1.662)	0.044 (0.311)	3.155 (1.863)
$\nu^{Y_4}_{r1}$	–0.364 (0.163)	–0.434 (0.429)	–0.184 (0.048)	–0.558 (0.430)
$\nu^{Y_4}_{r2}$	1.615 (0.771)	1.978 (1.964)	0.798 (0.268)	2.523 (1.969)
$\nu^{Y_4}_{r3}$	3.134 (1.284)	3.500 (3.618)	1.671 (0.437)	4.604 (3.614)

same kind of pattern in both data sets. Although the MCM with two dimensions does not fit perfectly, it describes around 90 percent of the association between the variables (see ΔG^2). This means that there is clear evidence that there are two underlying dimensions. This is confirmed by the results obtained with the other four methods. However, the two-dimensional NRM, LCFM, and MAM fit the data somewhat better, which illustrates that working with odds-ratios instead of correlations gives somewhat more flexibility when modeling relationships between categorical variables (Goodman, 1991).

Tables 5 and 6 report the parameter estimates and the estimated standard errors for the two-dimensional MCM. For both data sets, the unrotated solution shows a pattern that is well-known from unrotated factor solutions: All items are positively related to the first dimension, while some are positively and others are negatively related to the second dimension. The rotated solutions were obtained by setting one category quantification equal to zero rather than using the orthogonality constraint. In the model for the Political Action data, $\nu^{Y_3}_{11}$ was set to zero. It can now be seen, that the other four items are strongly related to the first dimension and that Y_2, Y_3 and Y_4 are related to the second dimension. After setting $\nu^{Y_2}_{21} = 0$ in the model for the General Social Survey data, we obtained a solution in which Y_1, Y_2 and Y_4 are related to the first dimension and Y_3 and Y_4 to the second dimension.

Because several parameter estimates reported in Tables 5 and 6 do not differ significantly from zero, it makes sense to impose additional constraints. Setting $\nu^{Y_1}_{21} = \nu^{Y_5}_{21} = 0$ in the first data set, for example, yields a very small increase in G^2 compared to the unrestricted MCM(2) model(i.e., $G^2 = 29.43$ versus $G^2 = 29.39$ on 2 degrees of freedom). A similar small increase in G^2 is found in the second data set by setting $\nu^{Y_3}_{11} = \nu^{Y_3}_{12} = \nu^{Y_1}_{21} = 0$: $G^2 = 30.82$ versus $G^2 = 28.57$ on 3 degrees of freedom.

Conclusions

In this paper we developed a ML variant of JCA called the multivariate correlation model. A nice feature of the proposed ML variant of JCA is that it provides formal tests to check whether a specified model fits the data. Another improvement over WLS estimation is that it seems to be more stable: For instance, Heywood cases as reported by Boik (1996) are less likely to occur. A disadvantage of ML estimation is that it may take somewhat longer, especially in large frequency tables.

In the empirical application, we compared the performance of the MCM with other models that could be used for the same types of data. Although the MCM yielded the same conclusion in terms of number of dimensions, models based on odds-ratios seem to be superior to the MCM in terms of model fit.

A consequence of defining JCA as a statistical model is that it can be applied in a more confirmatory manner. As in confirmatory factor analysis, an interesting type of constraint is to set the category scale values of variable Y_k on dimension r equal to zero, which is similar to setting a factor loading equal to zero. A related extension would be to allow for correlated dimensions. This was illustrated in the examples. Constraints that make sense with ordinal variables are fixed (equal-interval) or monotone category scale values. Another interesting extension is the inclusion of grouping variables or covariates influencing the correlation parameters or the category quantifications.

References

Anderson, C.J., & Vermunt, J.K. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology, 30,* 81–121.

Bock, D.R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37,* 29–51.

Boik, R.J. (1996). An efficient algorithm for joint correspondence analysis. *Psychometrika, 61,* 255–269.

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

Choulakian, V. (1988b). Errata. *Psychometrika, 53,* 593.

De Leeuw, J., & Van der Heijden, P.G.M. (1991). Reduced rank models for contingency tables. *Biometrika, 78,* 229–232.

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

Gilula, Z., & Haberman, S.J. (1988). The analysis of multivariate contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association, 83,* 760–771.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61,* 215–231.

Goodman, L.A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models etc. *The Annals of Statistics, 13,* 10–69.

Goodman, L.A. (1987). New methods for analyzing the intrinsic character of qualitative variables using cross-classified data. *American Journal of Sociology, 93,* 529–583.

Goodman, L.A. (1991). Measures, models, and graphical displays in the analysis of cross-classified data. *Journal of the American Statistical Association, 86,* 1085–1139.

Greenacre, M.J. (1988). Correspondence analysis of multivariate categorical data by weighted least-squares. *Biometrika, 75,* 457–467.

Greenacre, M.J. (1993). *Correspondence analysis in practice.* London: Academic Press.

Hagenaars, J.A. (1993). *Loglinear models with latent variables.* Newbury Park, CA: Sage.

Magidson, J., & Vermunt, J.K. (2001). Latent class factor and cluster models, biplots and related graphical displays, *Sociological Methodology, 31,* 223–264.

McCutcheon, A.L. (1987). *Latent class analysis.* Newbury Park: Sage Publications.

Tateneni, K., & Browne, M.W. (2000). A noniterative method of joint correspondence analysis. *Psychometrika, 65,* 157–165.

Van der Heijden, P.G.M., Gilula, Z., & Van der Ark, L.A. (1999). An extended study into the relationship between correspondence analysis and latent class analysis. *Sociological Methodology, 29,* 147–186.

Address for correspondence:

Jeroen K. Vermunt
Tilburg University
Faculty of Social and Behavioral Sciences
Department of Methodology and Statistics
Postbus 90153
NL-5000 LE Tilburg
The Netherlands
Tel. +31 13 466 2748
Fax +31 13 466 3002
E-mail J.K.Vermunt@uvt.nl

Appendix: Estimation Issues

WLS Estimation of the JCA Model

The parameters of the JCA model are usually estimated by weighted least squares (WLS); that is, by minimizing:

$$W = 1/2 \sum_{k=1}^{K} \sum_{l \neq k} \sum_{y_k=1}^{J_k} \sum_{y_l=1}^{J_l} \frac{(p_{y_k y_l}^{Y_k Y_l} - \pi_{y_k y_l}^{Y_k Y_l})^2}{p_{y_k}^{Y_k} p_{y_l}^{Y_l}}. \quad (8)$$

Here, $p_{y_k y_l}^{Y_k Y_l}$ and $p_{y_k}^{Y_k}$ denote observed sample proportions. It can easily be verified that the WLS estimates for the marginal probabilities $\pi_{y_k}^{Y_k}$ are equal to their sample equivalents $p_{y_k}^{Y_k}$.

Using the formulation of JCA in which the λ_r parameters are absorbed into the category quantifications (see Equation 2), the above loss function can also be written as follows:

$$W = 1/2 \sum_{k=1}^{K} \sum_{l \neq k} \sum_{y_k=1}^{J_k} \sum_{y_l=1}^{J_l} p_{y_k}^{Y_k} p_{y_l}^{Y_l} \left(\xi_{y_k y_l}^{Y_k Y_l} - \sum_{r=1}^{R} \eta_{r y_k}^{Y_k *} \eta_{r y_l}^{Y_l *} \right)^2, \quad (9)$$

where $\xi_{y_k y_l}^{Y_k Y_l} = \left(p_{y_k y_l}^{Y_k Y_l} - p_{y_k}^{Y_k} p_{y_l}^{Y_l} \right) / \left(p_{y_k}^{Y_k} p_{y_l}^{Y_l} \right)$. One of the algorithms that has been proposed to obtain the category score scale values $\eta_{r y_k}^{Y_k *}$ makes use of MCA; that is, it performs a singular value decomposition on the matrix collecting the elements $\sqrt{p_{y_k}^{Y_k} p_{y_l}^{Y_l}} \xi_{y_k y_l}^{Y_k Y_l}$ for all two-way tables (Greenacre, 1988). The difference with MCA is that the diagonal elements of this matrix are updated at each iteration cycle. More precisely, the diagonal elements are estimated with the provisional parameter values from the previous iteration:

$$\hat{\xi}_{y_k y_k}^{Y_k Y_k (t)} = \sum_{r=1}^{R} \hat{\eta}_{r y_k}^{Y_k * (t-1)} \hat{\eta}_{r y_k}^{Y_k * (t-1)}.$$

In a certain sense, this iterative algorithm is similar to an EM algorithm; that is, fill in the expected values for the missing data in one step and solve the "maximization" problem in a second step.

Other more efficient algorithms have been proposed that minimize W directly making use of its first and second derivatives with respect to $\eta_{r y_k}^{Y_k *}$ (see Boik, 1996). These derivatives equal

$$\nabla W(\eta_{r y_k}^{Y_k}) = \sum_{l \neq k} \sum_{y_l=1}^{J_l} p_{y_k}^{Y_k} p_{y_l}^{Y_l} \left(\sum_{r=1}^{R} \eta_{r y_k}^{Y_k *} \eta_{r y_l}^{Y_l *} - \xi_{y_k y_l}^{Y_k Y_l} \right) \eta_{r y_l}^{Y_l *},$$

$$\nabla^2 W(\eta_{r y_k}^{Y_k}) = \sum_{l \neq k} \sum_{y_l=1}^{J_l} p_{y_k}^{Y_k} p_{y_l}^{Y_l} (\eta_{r y_l}^{Y_l *})^2 .$$

A simple unidimensional Newton or alternating least squares updating scheme for the category quantifications $\eta_{r y_k}^{Y_k *}$ involves adjusting the parameters as follows:

$$\hat{\eta}_{r y_k}^{Y_k * (t)} = \hat{\eta}_{r y_k}^{Y_k * (t-1)} + \frac{\nabla W(\eta_{r y_k}^{Y_k})}{\nabla^2 W(\eta_{r y_k}^{Y_k})}.$$

After updating the rth set of scale values for variable Y_k, they should be centered. The orthogonalization of the scale values for the various dimensions can be done afterwards, for instance, by performing one cycle of the algorithm described above.

ML Estimation of the Multivariate Correlation Model

For the ML estimation of the MCM, we use the formulation of Equation (5) in which the σ_r parameters are absorbed in the category quantifications. In order to deal with the centering constraints, we simply write the category quantification for the last category J_k as a function of the quantifications for the other categories:

$$\nu_{r J_k}^{Y_k *} = - \sum_{y_k=1}^{J_k - 1} \frac{p_{y_k}^{Y_k}}{p_{J_k}^{Y_k}} \nu_{r y_k}^{Y_k *}.$$

Let I denote the number of cells in the contingency table and i a particular cell entry. The observed cell frequencies are denoted by n_i, the total sample size by N, and an expected cell proportion by $\pi_i(\beta)$, where β is the vector of unknown parameters. Furthermore, let $x_{i y_k}^{Y_k}$ equal 1 if $Y_k = y_k$, $-p_{y_k}^{Y_k} / p_{J_k}^{Y_k}$ if $Y_k = J_k$, and 0 otherwise. Let $z_{i y_k}^{Y_k}$ equal 1 if $Y_k = y_k$, –1 if $Y_k = J_k$, and 0 otherwise.

Assuming Poisson sampling, ML estimation involves maximizing the kernel of the log-likelihood function

$$L(\beta) = \sum_{i}^{I} [n_i \ln \pi_i(\beta) - N \pi_i(\beta)].$$

In order to simplify notation, we define $\pi_i(\beta) = \pi_i^0 \gamma_i$, where

$$\pi_i^0 = \prod_{k=1}^{K} \left(\sum_{y_k=1}^{J_k} \pi_{y_k}^{Y_k} z_{i y_k}^{Y_k} \right),$$

and

$$\gamma_i = \left[1 + \sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{l=k+1}^{K} \sum_{y_k=1}^{J_k} \sum_{y_k=1}^{J_l} (\nu_{r y_k}^{Y_k *} x_{i y_k}^{Y_k}) (\nu_{r y_l}^{Y_l *} x_{i y_l}^{Y_l}) \right].$$

The two algorithms described below use the first-order derivatives of $\pi_i(\beta)$ with respect to the unknown parameters; that is,

$$\nabla\pi_i(\nu_{ry_k}^{Y_k*}) = \pi_i^0 \sum_{l\neq k} \sum_{y_l=1}^{J_l} (\nu_{ry_l}^{Y_l*} x_{iy_l}^{Y_l})\, x_{iy_k}^{Y_k},$$

$$\nabla\pi_i(\pi_{y_k}^{Y_k}) = \left[\prod_{l\neq k}^{K}\left(\sum_{y_l=1}^{J_l} \pi_{y_l}^{Y_l} z_{iy_l}^{Y_l}\right)\right]\gamma_i z_{iy_k}^{Y_k}.$$

Goodman (1985) and Choulakian (1988a) proposed using a simple uni-dimensional Newton algorithm for ML estimation of correlation models for two-and three-way tables. This algorithm can easily be generalized to deal with the MCM. The ML estimates for the $\pi_{y_k}^{Y_k}$ terms are simply their sample equivalents $\pi_{y_k}^{Y_k}$. The unidimensional updating scheme for the $\nu_{ry_k}^{Y_k*}$ parameter is defined as follows:

$$\hat{\nu}_{ry_k}^{Y_k*(t)} = \hat{\nu}_{ry_k}^{Y_k*(t-1)} - \frac{\nabla L(\nu_{ry_k}^{Y_k*})}{\nabla^2 L(\nu_{ry_k}^{Y_k*})},$$

where the first and second derivatives of $L(\beta)$ with respect to $\nu_{ry_k}^{Y_k*}$ equal

$$\nabla L(\nu_{ry_k}^{Y_k*}) = \sum_i \left(\frac{n_i - n\pi_i(\beta)}{\pi_i(\beta)}\right)\nabla\pi_i(\nu_{ry_k}^{Y_k*}),$$

$$\nabla^2(\nu_{ry_k}^{Y_k*}) = -\sum_i \frac{n_i}{\pi_i(\beta)^2}\nabla\pi_i(\nu_{ry_k}^{Y_k*})^2.$$

One set of quantifications is updated at a time fixing all the other parameters at their current values. The identifying orthogonality constraints can be imposed afterwards, for instance, by a singular value decomposition. As was already mentioned, an alternative way to deal with the rotation problem is to equate certain scale values to zero.

Gilula and Haberman (1986) proposed obtaining ML estimates for the parameters of the bivariate correlation model by Fisher scoring. The same procedure can also be applied in the context of the MCM. An important difference with a standard Fisher scoring algorithm is that the identifying (orthogonality) constraints should be defined as side constraints in the maximization problem. We will denote these constraints by $\mathbf{h}(\beta) = \mathbf{0}$. The task to be performed is finding the parameter estimates $\hat{\beta}$ that fulfill the following two conditions:

$$\mathbf{g}(\hat{\beta}) = \nabla L(\hat{\beta}) + \lambda'\nabla \boldsymbol{h}(\hat{\beta}) = \mathbf{0},$$

$$\mathbf{h}(\hat{\beta}) = \mathbf{0},$$

where λ is a vector of Lagrange multipliers. The gradient vectors for β and λ, $\mathbf{g}(\beta)$ and $\mathbf{g}(\lambda)$, and the Fisher information matrix, $\mathbf{F}(\beta)$, are obtained by

$$\mathbf{g}(\beta) = \sum_i \pi_i(\beta)^{-1}\,[n_i - N\pi_i(\beta)]\,\nabla\pi_i(\beta) + \lambda'\nabla h(\beta),$$

$$\mathbf{g}(\lambda) = \mathbf{h}(\beta),$$

$$\mathbf{F}(\beta) = N\sum_i \pi_i(\beta)^{-1}\,\nabla\pi_i(\beta)\,\nabla\pi_i(\beta)'.$$

After collecting β and λ into a single vector θ, the Fisher-scoring updating scheme can be defined as follows

$$\hat{\theta}^{(t)} = \hat{\theta}^{(t-1)}\,\mathbf{H}(\hat{\theta})^{-1}\,\mathbf{g}(\hat{\theta}),$$

where

$$\mathbf{H}(\theta) = \begin{bmatrix} F(\beta) & -\nabla \boldsymbol{h}(\beta) \\ -\nabla \boldsymbol{h}(\beta) & \mathbf{0} \end{bmatrix}. \quad (10)$$

The upper left part of the inverse of $\mathbf{H}(\hat{\theta})$ contains the estimated variances and covariances of the unknown parameters.

A problem with the ML estimation of the MCM model is that there is no guarantee that all $\pi_i(\beta) \geq 0$. This problem may occur when some of the observed cell frequencies are equal to zero. The Fisher-scoring method described above can, however, easily be modified to include nonnegativity constraints on the expected cell proportions: A term $[\pi_i(\beta)\text{-}\varepsilon]$ is added to the vector $\mathbf{h}(\beta)$ for each i. Because we are dealing with inequality constraints, the corresponding Lagrange multipliers should be at least zero, which means that the equality constraint $\pi_i(\beta) = \varepsilon$ is only activated if the corresponding inequality $\pi_i(\beta) \geq \varepsilon$ would otherwise be violated. The value of ε can be set very near to zero, say 10^{-8}, but not exactly equal to zero.

