ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE

ANTOINE DE FALGUEROLLES
PETER G.M. VAN DER HEIJDEN

Reduced rank quasi-symmetry and quasi-skew symmetry : a generalized bi-linear model approach

Annales de la faculté des sciences de Toulouse 6^e série, tome 11, n° 4 (2002), p. 507-524

<http://www.numdam.org/item?id=AFST_2002_6_11_4_507_0>

Annales de la Faculté des Sciences de Toulouse Vol. XI, n° 4, 2002
pp. 507–524

Reduced rank quasi-symmetry and quasi-skew symmetry: a generalized bi-linear model approach (*)

ANTOINE DE FALGUEROLLES (1), PETER G.M. VAN DER HEIJDEN (2)

RÉSUMÉ. — Ce qui suit est une revue des modèles linéaires généralisés classiques pour tableau carrés et une étude des modèles d'interaction de rang réduit pour la symétrie et l'écart à la symétrie. Ces modèles offrent une description parcimonieuse de questions fondamentales et se prêtent à des visualisations de type « biplot ». Nous discutons aussi les procédures destimation.

ABSTRACT. — What follows is a review of standard generalized linear models for square tables and an investigation into reduced rank models for symmetry and departure from symmetry. These models offer parsimonious representations of substantive questions and lend themselves to biplot visualization. We also discuss estimation procedures.

1. Introduction

The terms of *quasi-independence* and *quasi-symmetry* entered the statistical literature in the context of truncated tables (see the seminal paper on quasi-symmetry by Caussinus [5] and an history of truncated tables in Stigler [20], chapter 19). In the context of square tables, these models have been motivated either by the absence of diagonal entries or by the desire to obtain analyses invariant with respect to these values. Square tables are two-way tables cross-classified by homologous factors (McCullagh [18]): the

(*) Reçu le 18 septembre 2001, accepté le 18 septembre 2002
(1) Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France. E-mail: falguero@cict.fr
(2) Department of Methodology and Statistics, Utrecht University, The Netherlands. E-mail: p.vanderheijden@fss.uu.nl

levels for the row and column factors are in a meaningful one-to-one relationship. In matrix form, the levels of the homologous factors are listed in the same order so that the main diagonal, often missing or uninteresting, corresponds to entries referring to the 'same' level.

This paper will discuss three topics. First, quasi-independence and quasi-symmetry models were originally developed in the context of multinomial, product multinomial and Poisson distributions. These are log-linear models. However, it is also possible and quite natural to develop them in the broader context of generalized linear models (Aitkin *et al.* [1] and McCullagh and Nelder [19]; compare also Caussinus and Falguerolles [6]).

Second, the hierarchy of quasi-independence, quasi-symmetry and saturated models is somewhat restrictive and it is quite natural to refine this grid. Not surprisingly, several useful intermediate models have already been introduced, but only for the context of contingency tables. Some are sub-models of quasi-symmetry: the reduced rank quasi-symmetry model (Becker [2]) and, for ordered categories, the uniform association model (Goodman [14]). Some have quasi-symmetry as a sub-model and introduce further asymmetry than that ascribed to the margins (van der Heijden and Mooijaart [22]). The general idea in these papers is to restrict the row $\times$ column interaction term in a parsimonious way. It turns out that this can be done in the framework of bi-linear models which preserve most of the graphical outputs of exploratory methods based on singular value decomposition (see Falguerolles [9]). What follows is an investigation into these models and their biplot visualization.

Third, in this paper we show both exploratory approaches as well as modelling approaches. We will illustrate that the possibilities offered by the former are also offered by the latter. However, the modelling approach has the advantage that it is possible to test (aspects of) the model such as model fit. These models are more flexible than exploratory approaches in the sense that they can be adjusted easily. The reason that this is possible is that in the modelling framework we have moved away from the matrix formulation of the data to an array formulation.

We start by introducing two examples, one where the data are assumed to be generated by a multinomial distribution, and one where the data are generated by a normal distribution. Then we describe three strategies for the analysis of square tables. First, we review purely exploratory approaches. Second, we discuss approaches where both modelling as well as exploratory approaches are used. And third, we discuss the generalized linear model (GLM) approach, with reduced rank interactions. In the presence of such an interaction we speak of a 'generalized bi-linear model' (GBM).

2. Examples

In this section we present two examples of square tables where quasi-symmetry has been applied under different distributional assumptions. Both the examples as well as their distributional assumptions have been considered by Henri Caussinus ([5] and [6]).

2.1. A social mobility table

The first example is one of the earliest applications of the quasi-symmetry model for count data (Example 3 in Caussinus [5], page 166). The data pertain to the social mobility of a sample of 1384 workers between two time points (1954 and 1962). The number of levels of social categories is $N = 6$, but their labels are not reported.

Table 1. — Social mobility between 1954 and 1962
(marginal totals in parentheses do not include diagonal counts).

	1964						total	
1957	1	2	3	4	5	6		
1	187	13	17	11	3	1	232	(45)
2	4	191	4	9	22	1	231	(40)
3	22	8	182	20	14	3	249	(67)
4	6	6	10	323	7	4	356	(33)
5	1	3	4	2	126	17	153	(27)
6	0	2	2	5	1	153	163	(10)
total	220	223	219	370	173	179	1384	
	(33)	(32)	(37)	(47)	(47)	(26)		(222)

In this example, it is quite natural to assume that the data come from a multinomial distribution with known parameter $n = 1384$ and unknown probabilities π_{ij}^{IJ}. In most practical circumstances, the modelling of the multinomial probabilities can be performed assuming that the entries y_{ij}^{IJ} in the table are independently Poisson distributed with unknown parameters $\lambda_{ij}^{IJ} = n\pi_{ij}^{IJ}$.

As seen by inspection, the entries on the main diagonal of the table (the 'stayers') are much larger than the non-diagonal entries (the 'movers'): there are 1162 'stayers' and 222 'movers'. The analysis can be restricted to the non-diagonal entries and we can still assume that the non-diagonal entries are independently Poisson distributed.

The quasi-symmetry model is an obvious candidate for such data: the λ_{ij}^{IJ}, $i \neq j$, are decomposed on a log scale as the sum of a row effect (I), a column effect (J) and a symmetric interaction effect ($I \times J$):

$$\eta_{ij}^{IJ} = log(\lambda_{ij}^{IJ}) = \beta^{\emptyset} + \beta_i^I + \beta_j^J + \beta_{ij}^{IJ} \text{ where } \beta_{ij}^{IJ} = \beta_{ji}^{IJ}. \quad (1)$$

The model has $(N-1)(N-2)/2$ degrees of freedom. While the second and third terms are intended to reflect the marginal sociological changes between 1957 and 1962, the symmetric interaction term is introduced to investigate the symmetry of transitions from i to j and from j to i which expresses that comparable skills are needed for both transitions. In passing, we note that the parameters are not identified and that *ad hoc* constraints are needed. Moreover it turns out that the fit does not depend on the values of the diagonal entries. In this example, the deviance goodness-of-fit chi-square statistic for quasi-symmetry yields $p = 0.0027$ and the corresponding Pearson statistic yields $p = 0.0024$, suggesting the need for modelling further asymmetry beyond that induced by the margins.

2.2. A socio-matrix

Thomas' socio-matrix reproduces the grades given by each of the 24 pupils of a class to all other schoolmates [4, page 295]. The evaluation was unsupervised: each pupil was asked to grade on an integer scale ranging from 0 to 20 his/her 'affinity' with all other mates but no specific criteria were suggested. In some experiments the diagonal entries give self-assessment grades. But in this example the diagonal terms are missing and the analysis must take that fact into account. In other words it should not depend on imputed values for the diagonal values like the two principal components analyses which are reported in Caillez and Pagès [4]: one on the matrix of grade given and one on the matrix of grade received with self assessments (diagonal entries) set equal to 20.

It is quite natural to assume that the grades are normally and independently distributed with unknown constant variance and unknown expected values μ_{ij}^{IJ}. Again the quasi-symmetry model is a sensible candidate for modelling their expected values. In other words, the μ_{ij}^{IJ} are decomposed as the sum of a row effect (average rating attributed), a column effect (average rating received) and a symmetry effect (symmetric preferences).

Here quasi-symmetry and departure from quasi-symmetry are to be investigated. In this normal setting, the variance is unknown and the problem of model choice is different from that in the Poisson situation.

3. Exploratory strategies for the analysis of square tables

In this section we review some of the well established approaches for the exploratory analysis of the type of data which we just presented. Broadly speaking, they may be classified into two categories: 'purely' exploratory methods, and methods combining modelling and exploratory analysis of residuals.

3.1. Exploratory reduced rank approximations

Square tables are efficiently analyzed by methods based on singular value decomposition. We adopt here a simplified approach. A more sophisticated approach where correspondence analysis, a generalized singular value decomposition, is used can be found in Greenacre [17].

Exploratory reduced rank approximations have a few prerequisites:

- The data are in matrix form. In other words no other explanatory variables than the homologous 'row' and 'column' effects can be taken into account.
- The possibly missing values are imputed.
- The data are pre-processed with special attention to filtering the main 'row' and 'column' effects, and sometimes the diagonal.

Let M be the matrix under study. Our goal is to decompose it into symmetric and asymmetric parts. The symmetric part is intuitively estimated by

$$C = (\frac{1}{2}M + \frac{1}{2}M')$$

and departure from symmetry by

$$D = (\frac{1}{2}M - \frac{1}{2}M')$$

so that $M = C + D$. This is the so called Gower's decomposition of a square matrix into its symmetric part C and a skew symmetric part D ($D' = -D$). Each can then be submitted to a singular value decomposition (Constantine and Gower [7]). The pattern of singular elements are more heavily constrained by skew-symmetry than by symmetry. For the symmetric matrix C, it may occur that a left and a right singular vector associated with the same singular value have opposite signs. This is called an inversion (Greenacre [16, chapter 8]). One or more inversions occur when a symmetric matrix is not positive definite, since singular values are constrained to be

Table 2. — Thomas' socio-matrix: rows are grades given, columns are grades received (Gender: b for boy, g for girl).

Gender:	g	b	b	b	b	b	g	b	g	b	g	g	b	g	g	g	g	g	b	b	g	g	g	b	total
g		5	12	14	15	3	18	10	10	10	14	14	5	20	17	15	19	17	10	16	12	16	15	12	299
b	10		17	16	14	20	17	19	13	17	11	13	8	10	6	1	17	17	16	17	13	17	17	19	325
b	15	9		11	19	17	13	7	16	8	9	2	3	15	14	10	13	14	14	10	10	16	16	17	278
b	15	12	11		10	17	12	16	13	15	12	10	8	8	11	11	13	13	17	16	14	14	15	18	301
b	15	4	19	11		17	15	8	16	9	3	1	2	15	10	10	10	3	13	9	13	15	15	13	246
b	9	19	18	13	18		10	5	12	2	2	1	5	2	12	3	17	15	15	11	11	12	16	14	242
g	16	15	14	10	15	15		10	16	10	8	2	1	16	6	13	20	2	8	10	8	18	16	14	263
b	17	18	15	13	16	18	12		15	16	17	13	11	16	10	14	16	15	12	19	18	18	17	18	354
g	14	14	12	10	8	10	15	10		9	14	8	5	14	13	11	14	9	10	11	13	18	19	15	276
b	19	10	15	12	10	15	13	20	15		12	8	12	17	12	3	12	12	10	20	14	15	15	15	306
g	19	12	15	12	15	8	14	14	17	9		15	5	19	16	19	17	16	10	10	19	18	18	15	332
g	20	15	14	17	16	13	13	10	17	13	14		13	19	10	12	11	19	8	14	14	18	18	9	327
b	13	2	15	11	3	19	10	12	12	18	8	19		12	16	12	12	20	4	18	10	15	16	17	294
g	20	5	13	14	15	5	18	16	10	10	14	13	5		17	16	17	13	10	16	11	15	15	13	301
g	14	10	8	6	14	10	5	9	15	7	11	6	3	13		13	19	14	1	10	7	18	17	2	232
g	10	10	5	3	6	6	8	7	9	6	15	12	1	11	10		8	10	4	5	11	13	8	3	181
g	18	17	10	10	14	17	20	17	16	10	16	1	1	18	17	15		1	10	9	10	19	17	13	296
g	15	12	10	10	3	16	10	13	10	10	11	20	17	15	18	10	10		10	13	19	14	14	10	290
b	12	14	14	13	17	18	14	12	13	11	10	5	12	12	12	11	11	10		11	11	11	12	14	280
b	18	9	11	16	6	15	13	20	17	19	12	8	13	18	19	10	11	11	14		14	19	17	15	325
g	20	8	14	12	16	8	7	10	10	6	19	11	4	19	7	16	11	16	2	16		14	13	12	271
g	14	13	13	12	13	10	16	12	18	12	10	12	12	14	18	9	15	10	9	13	10		19	18	302
g	14	13	13	9	10	11	15	9	19	10	12	12	6	14	13	12	13	13	13	12	13	18		16	290
b	12	15	16	19	15	15	11	12	13	13	4	12	13	12	12	7	12	11	14	13	11	17	17		296
total	349	261	304	274	288	303	299	278	322	250	258	218	165	329	296	253	318	281	234	299	286	368	362	312	

positive. For the skew-symmetric matrix D, the singular vectors occur in pairs corresponding to pairs of equal singular values (see Constantine and Gower [7]). It is appropriate to speak in this situation of 'bi-dimension'.

We illustrate the standard exploratory strategy on the socio-matrix. Firstly, unobserved diagonal entries are set equal to 20 (self-assessments equal to maximal grade). Secondly, the full matrix is doubly-centered (so that the margins of the resulting matrix are zero) and decomposed in its symmetric part C and its skew-symmetric part D. Table 3 gives the values of the four largest squared singular values of C and D. It also gives their associated so-called percentage of inertia: the ratios (in %) to their respective sums, Trace($C'C$) and Trace($D'D$), as well as their ratios to their grand sum Trace($M'M$) = Trace($C'C$) + Trace($D'D$). As seen in Table 3, singular value decomposition singles out symmetry as the dominant feature of the interaction in these data (79% versus 21%) and indicates that a rank two approximation of symmetry may well suffice to depict it (56% of symmetry or 45% of all).

For completeness we mention another strategy which consists in performing the complex singular value decomposition of $C + iD$. Comparable bi-dimensions arise in this context. This interesting strategy was suggested by Escoufier and Grorud [8] but is not considered in this article.

Table 3. — Exploratory analysis of Example 2.
Singular value k ($k = 1, ..., 4$) is denoted by σ_k.

	σ_1^2	σ_2^2	σ_3^2	σ_4^2	Total
Symmetric part: C	2122.7	1672.1	618.1	530.4	6797.3
% of Trace($C'C$)	31	25	9	8	100
% of Trace($M'M$)	25	20	7	6	79
Skew-symmetric part: D	241.2	241.2	202.0	202.0	1756.1
% of Trace($D'D$)	14	14	12	12	100
% of Trace($M'M$)	3	3	2	2	21

3.2. Biplot representations

The reduced rank approximations obtained in the analysis of C and D allow to consider biplot representations (for a seminal paper, see Gabriel [12, 13] and for a review of the topic, Gower and Hand [15]).

The biplot representations of the symmetric matrix C are quite standard. Notice that if there is no inversion in the singular elements retained, then there is only one marker by row and column level. When there is an

inversion, the biplot has different markers for the same row and column level, with opposite coordinates on the corresponding axis.

The biplot representations of the skew-symmetric matrix D are more intricate. Each bi-dimension give rise to a biplot which has a specific reading in terms of oriented area. The interpretation of the relation between two points in the biplot is not in terms of their inner product of the corresponding vectors, but in terms of the area of their triangle with the origin (compare Gower and Hand [15] and Greenacre [17] for more details).

However there are obvious limitations. First, the analysis of the symmetric part depends on the values of the diagonal terms (although this can be alleviated as it will be shown in the next section). Since the diagonal values are often arbitrary or not the focus of interest, this is a drawback in this approach. Second, the simultaneous analyses of the symmetrical and the skew symmetric part by reduced rank approximations do not follow the accepted principle which states that departure from symmetry is to be analyzed if quasi-symmetry fails to represent the data. Researchers who prefer statistical models may therefore feel uncomfortable with this approach. Third, a natural question to ask about biplots is how many dimensions or bi-dimensions are needed to adequately represent the data and their is no clear answer to this.

3.3. Combining modelling and exploratory analysis of residuals

Combined analyses were devised in order to answer to the limitations just mentioned. They combine traditional modelling and exploratory methods based on singular value decomposition. For an example where correspondence analysis is used in a combined approach see van der Heijden *et al.* [21].

The natural framework for modelling is that of generalized linear models (see for example McCullagh and Nelder [19], Aitkin *et al.* [1]). This framework can be summarized as follows. A unidimensional response variable is observed at fixed values of explanatory variables (mostly fixed levels of factors in our context). These explanatory variables influence in turns the distribution of the response through a linear predictor. The mean of the response is smooth invertible function of the linear predictor whose inverse function is called the link function. The distribution of the response is assumed to follow the form of the exponential family of distribution. It involves a scale parameter (possibly known) and known prior weights (which can be used to weight out structurally valued or missing data). Some of the above assumptions can be somehow relaxed but we will not discuss these aspects except for the linearity of the predictor.

This framework can accommodate our examples. Their main differences are

- their distributions (Poisson versus normal);
- their link functions (identity versus logarithmic).

But they have many common features:

- the diagonal entries are weighted out;
- the same simplified hierarchy of linear models is of interest for the linear predictors η_{ij}^{IJ}.

The latter can be summarized as follows:

- Quasi-independence model $\eta_{ij}^{IJ} = \beta^{\emptyset} + \beta_i^I + \beta_j^J$ (marginal effects of I and J only).
- Quasi-symmetry model $\eta_{ij}^{IJ} = \beta^{\emptyset} + \beta_i^I + \beta_j^J + \beta_{ij}^{IJ}$ where $\beta_{ij}^{IJ} = \beta_{ji}^{IJ}$ (marginal effects of I and J, and a symmetrical $I \times J$ interaction).
- Saturated model $\eta_{ij}^{IJ} = \beta^{\emptyset} + \beta_i^I + \beta_j^J + \beta_{ij}^{IJ}$ (β_{ij}^{IJ} unconstrained).

Note that the flexibility of generalized linear model allows one to take into account other explanatory variables, as well as more complicated designs than the simple two-way table without diagonal which is considered in the examples.

Table 4. — Residual deviance from quasi-independence et quasi-symmetry (degrees of freedom given in parentheses).

	Quasi-independence	Quasi-symmetry
Example 1	101.1 (19)	27.2 (10)
Example 2	7119.9 (505)	1756.1 (253)

Now a hierarchical line can be held. If the *symmetric* interaction is of main interest, then the residuals from *quasi-independence* should be symmetricized and analyzed by singular value decomposition. If *departure* from symmetric interaction is of main interest, then the residuals from *quasi-symmetry* should be submitted to singular value decomposition.

However there are many definitions for the residuals: raw, Pearson, deviance ... Interestingly, the raw residuals are skew symmetric for canonical settings: quasi-symmetry for a two-way table of data and canonical link.

But, except for the normal distribution, their singular value decomposition has no direct connection with the deviance. By contrast, for the normal distribution with identity link, the situation is simple: the skew-symmetric part of Gower's decomposition of the matrix of residuals from independence or quasi-independence is nothing else than the matrix of residuals from quasi-symmetry (see Caussinus and Falguerolles [6]).

Table 5. — Raw and Pearson residuals from quasi-symmetry in Example 1.

	Raw residuals					
	1	2	3	4	5	6
1	0.00	2.67	-2.00	-0.67	-0.14	0.13
2	-2.67	0.00	-0.56	0.21	4.45	-1.43
3	2.00	0.56	0.00	-0.92	-0.28	-1.37
4	0.67	-0.21	0.92	0.00	1.38	-2.75
5	0.14	-4.45	0.28	-1.38	0.00	5.42
6	-0.13	1.43	1.37	2.75	-5.42	0.00

	Pearson residuals					
	1	2	3	4	5	6
1	0.00	0.83	-0.46	-0.20	-0.08	0.14
2	-1.04	0.00	-0.26	0.07	1.06	-0.92
3	0.45	0.21	0.00	-0.20	-0.07	-0.65
4	0.29	-0.09	0.31	0.00	0.58	-1.06
5	0.15	-1.63	0.14	-0.75	0.00	1.59
6	-0.36	1.89	1.72	1.84	-2.14	0.00

Table 6. — Combined analysis of Example 1.
Singular value analysis of the raw residuals from quasi-symmetry.

squared singular values:	61.8	61.8	13.7	13.7	151.0
% of their sum:	41	41	9	9	

Table 4 reports the deviance of the two baseline models applied in the two examples. For Example 1, a chi-square test of the quasi-symmetry model against the saturated model rejects quasi-symmetry. Table 5 gives the raw and Pearson residuals from quasi symmetry in Example 1. These are dominated by transitions between levels 5 and 6, and levels 2 and 5. Table 6 reports the squared singular values obtained in the analysis of the skew-symmetric matrix of raw residuals from the quasi-symmetry model. It turns out that a single bi-dimension dominates this skew symmetric matrix.

Table 7. — Combined analysis of Example 2.
Singular value analysis of the residuals from quasi-independence.

	σ_1^2	σ_2^2	σ_3^2	σ_4^2	Total
Symmetric part: C	1428.5	1045.6	609.0	492.6	5363.8
% of Trace($C'C$)	27	19	11	9	100
% of Trace($M'M$)	20	15	9	7	75
Skew-symmetric part: D	241.2	241.2	202.0	202.0	1756.1
% of Trace($D'D$)	14	14	12	12	100
% of Trace($M'M$)	3	3	3	3	25

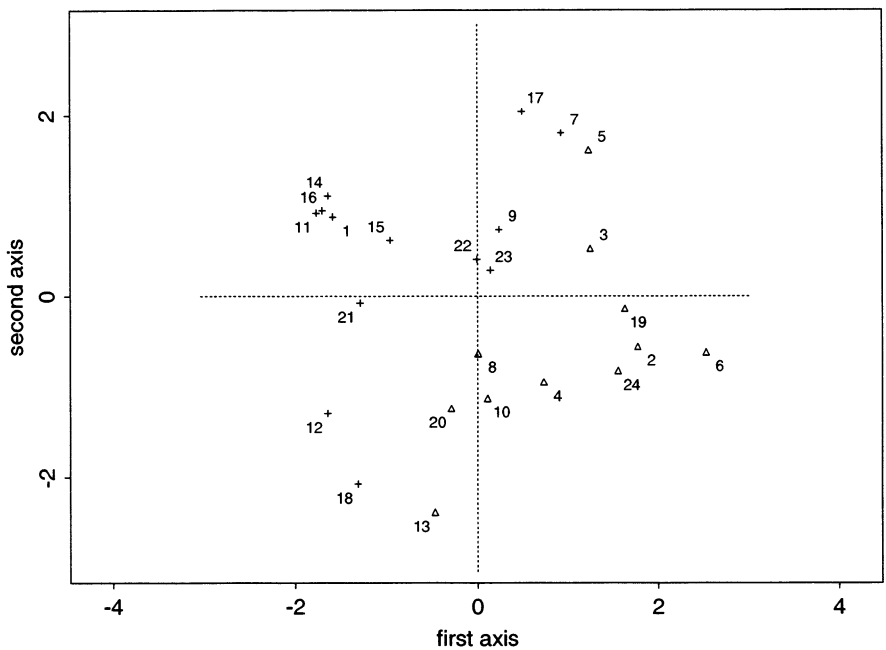

Figure 1. — Example 2. Biplot of the rank two approximation of the symmetric part C, boys are denoted by a +, girls by a $\triangle$.

For Example 2 where the variance is unknown, a F-test of quasi-independence against quasi-symmetry rejects quasi-independence (a F-statistics equal to 3.07 for 252 and 253 degrees of freedom). In this situation the particular properties of the normal distribution allow to analyze simultaneously the symmetric and the skew symmetric part. The results in Table 7 do not differ much from those of Table 3. Different values for the first two singular values for the symmetric part are due to the fact that the diago-

nal terms are not considered in the analysis described in Table 7. The fact that the singular values in Table 7 are lower shows that the diagonal values arbitrarily set to 20 set in the previous analysis were too high. Notice that Trace $(D'D)$ is equal in Tables 3 and 7 because the skew-symmetric parts are constructed in the same way. The associated biplot is given in Figure 1. It discriminates clearly the boys from the girls but not along the axes revealed by singular value decomposition.

One drawback of the method discussed in this section is that it combines approaches of different natures: modelling and exploratory methods for the residuals from modelling. These in turn use a different criterion: maximum likelihood and least squares which do not coincide except in the normal case. Thus it may look strange to pile up these two approaches. This point was raised in the discussion of van der Heijden *et al.* [21] by John Gower who suggested the introduction of reduced rank formulas in the linear predictor of generalized linear models (see also McCullagh and Nelder [19, subsection 6.5.3]). Actually more recent articles have introduced related models in the log-linear analysis of square tables. The quasi-symmetric model of Becker ([2]) and the log-bi-linear model for skew-symmetric of van der Heijden and Mooijaart ([22]) are notable examples. These can be further revisited in the unifying context of generalized bi-linear models.

Table 8. — Reduced quasi-symmetry based on gender in Example 2 (degrees of freedom given in parentheses).

quasi-independence	reduced quasi-symmetry	quasi-symmetry
7119.9 (505)	6434.0 (504)	1756.1 (253)

4. Generalized bi-linear models

In this section we describe classes of models that were intended by Gower (see the conclusion of the previous section), i.e., we are going to propose models that share the same objectives as the exploratory approach and the combined approach in section 3. An overall modelling approach has the advantage that the likelihood is maximized. It is also possible to propose parsimonious models, as these models turn out to be much more flexible than the matrix decompositions performed in the exploratory approaches. This is due to the fact that in the modelling approach the data are approached as an array. Thus models can be formulated in a flexible way.

As an example of this flexibility, we discuss a GLM for the second example of this paper. Here, gender information on the homologous factors

I and J can be exploited. A restricted quasi-symmetry model can be obtained by incorporating a symmetric interaction based on gender in the quasi-independence model: a score of $-\xi$ for boys and ξ for girls generates an explanatory variable with values ξ^2 for intra-gender ratings and $-\xi^2$ for inter-gender ratings. The deviance of this model is reported in Table 8. However the significant deviance reduction is smaller than the one associated with the first two axes obtained in singular value decomposition. This explains why this discrimination is observed in the biplot but not on the axes obtained by singular value decomposition. Interestingly, the positive sign of the corresponding parameter shows that in inter-gender grading the pupils are lower.

Below we will work out models that have resemblance with what happens in a singular value decomposition. In the absence of additional information that can be incorporated in the model, the formulas suggested by singular value decomposition are good candidates mostly on the ground that they are parsimonious and that they lend themselves to biplot visualizations. We will discuss two types of reduced rank models: reduced rank models for symmetric interaction and reduced rank models for skew-symmetric interaction. These are particular instances of GBM's and relevant earlier work can be found in Falguerolles and Francis [10], van Eeuwijk [23] and Falguerolles [9].

4.1. Reduced rank quasi-symmetry

In the *reduced rank quasi-symmetry* model a reduced rank symmetrical interaction term is introduced in the predictor:

$$\eta_{ij}^{IJ} = \beta^{\emptyset} + \beta_i^I + \beta_j^J + \sum_{k=1}^{M} \sigma_k (-1)^{\ell_k} \xi_{k,i} \xi_{k,j} \qquad (2)$$

where ℓ_k is either equal to 1 or to 2 and M being 'small' for a parsimonious modelling. Clearly the parameters are not identified and it is quite natural to use in the bi-linear term constraints similar to those used in singular value decomposition. In particular, the ℓ_k is introduced in order to allow for inverse factors (compare Section 2).

Thus, taking strictly positive weights $w_1, \ldots, w_N$ such that $\sum_{i=1}^N w_i = 1$ (for example, for counts these could be marginal proportions), sensible identification constraints for the scores are:

$$\begin{aligned} \textstyle\sum_{i=1}^N w_i \xi_{k,i} &= 0 \quad , \\ \textstyle\sum_{i=1}^N w_i \xi_{k',i} \xi_{k'',i} &= \delta_{k'}^{k''} \quad , \\ \sigma_1 \geqslant \ldots \geqslant \sigma_k \geqslant \ldots \geqslant \sigma_M &> 0 \quad . \end{aligned}$$

Note that the choice of values for identification weights is often a matter of taste although it may have a great impact on visual aspect of biplots.

An order M reduced rank quasi-symmetry model adds $M(1+N-1)-\frac{M(M+1)}{2}$ independent parameters to quasi-independence. Hence, in the standard setting which is considered, it has $N^2-3N+1-MN+\frac{M(M+1)}{2}$ degrees of freedom and there might not be a value of M which exactly reconstructs quasi-symmetry ($\frac{N(N-3)}{2}$ extra parameters). In Example 2, M equal to 16 is maximal and defines a quasi-symmetric model of 257 degrees of freedom (compared to the 253 degrees of freedom for quasi-symmetry).

The parameters of the bi-linear term define the usual order-two contrast:

$$\eta_{ij}^{IJ} - \eta_{ij'}^{IJ} - \eta_{i'j}^{IJ} + \eta_{i'j'}^{IJ} = \sum_{k=1}^{M} \sigma_k (-1)^{\ell_k} (\xi_{k,i} - \xi_{k,i'})(\xi_{k,j} - \xi_{k,j'}).$$

This shows that this contrast is equal to the inner product of the vector connecting the row markers i to i' and of the vector connecting the column markers j to j'. Notice that for frequency data with a log link this equals to usual log odds ratio (compare Becker [2]).

An order two reduced rank quasi-symmetry model in Example 2 has a deviance equal to 4266.3 with 460 degrees of freedom. In Table 8 we found a deviance of 7119.9 for quasi-independence and a deviance of 1756.1 for a deviance of quasi-symmetry. This shows that a rank two model takes $(7119.9-4266.3)/(7119.9-1756.1) = .53$ of the quasi-symmetric association. This is more than found for the exploratory approach described in Table 7, where a similar calculation leads to a proportion of .46. This is due to the fact that the singular value decomposition of the symmetricized residuals from quasi-symmetry overestimates the number of dimensions allowed by the degrees of freedom.

We do not show a biplot of the parameter estimates because it is very similar to the biplot displayed in Figure 1.

4.2. Reduced rank skew-symmetry

In the *reduced rank skew-symmetry* model, a reduced rank skew-symmetric interaction term is introduced in the predictor:

$$\eta_{ij}^{IJ} = \beta^{\emptyset} + \beta_i^I + \beta_j^J + \beta_{ij}^{IJ} + \sum_{k=1}^{M} \sigma_k (\xi_{2k,i}\xi_{2k-1,j} - \xi_{2k-1,i}\xi_{2k,j}) \quad (3)$$

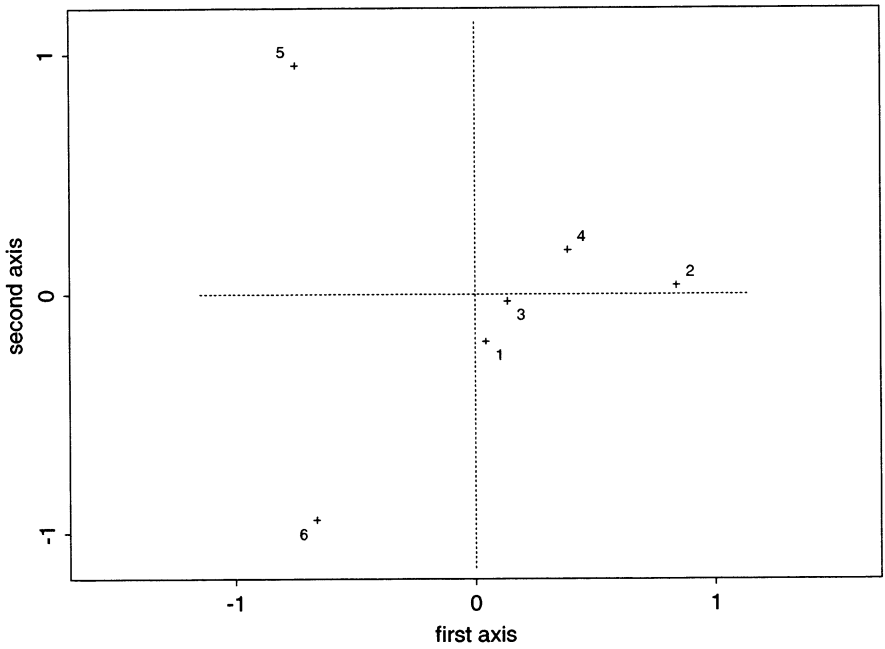

Figure 2. — Biplot of the first bi-dimension in Example 1.

where $\beta_{ij}^{IJ} = \beta_{ji}^{IJ}$ and $2M \leqslant |I| = |J| = N$, M being 'small' for a parsimonious modelling. Again, identification constraints need to be set for the score vectors. The number of degrees of freedom of this model is $2MN - 3M - 2M^2$. If N is even, then the maximal value of M is $N/2 - 1$; if N is odd, then the maximal value of M is $(N-1)/2$.

The parameters of the bi-linear term induce a special form for the usual order-two contrast:

$$\eta_{ij}^{IJ} - \eta_{ij'}^{IJ} - \eta_{i'j}^{IJ} + \eta_{i'j'}^{IJ}.$$

The latter is the sum of two terms namely, one for quasi-symmetry:

$$\beta_{ij}^{IJ} - \beta_{ij'}^{IJ} - \beta_{i'j}^{IJ} + \beta_{i'j'}^{IJ}$$

and one for reduced rank skew-symmetry:

$$\sum_{k=1}^{M} \sigma_k \left[(\xi_{2k,i} - \xi_{2k,i'})(\xi_{2k-1,j} - \xi_{2k-1,j'}) - (\xi_{2k-1,i} - \xi_{2k-1,i'})(\xi_{2k,j} - \xi_{2k,j}) \right].$$

Notice that the term given by reduced rank skew-symmetry, is a determinant, and therefore has also an interpretation in terms of an oriented area in the biplot.

Figure 2 shows the biplot of first skew-symmetric bi-dimension added to quasi-symmetry in the linear predictor of Example 1. The deviance is equal to 2.1 for 7 degrees of freedom. The visual aspect of Figure 2 is dominated by the triangle with vertices 5, 6 and the origin. Its area, the largest among those of similarly constructed triangles, represents the large value of the skew-symmetric transition parameter between 5 and 6: positive for the transition from 5 to 6 and negative for the transition from 6 to 5. Smaller values correspond to the transition between 2 and 5, and 6 and 2. Clearly, the model has picked up the skew-symmetry seen in Table 5.

4.3. Algorithmic approach

The models above are special cases of bi-linear models for reduced rank two-way interaction for non-homologous factors:

$$\eta_{ij}^{IJ} = \beta^{\emptyset} + \beta_i^I + \beta_j^J + \beta_{ij}^{IJ} + \sum_{k=1}^{M} \sigma_k \beta_{k,i}^I \beta_{k,j}^J, \tag{4}$$

and even of a particular tri-linear model for matched tables:

$$\eta_{sij}^{SIJ} = \beta^{\emptyset} + \beta_s^S + \beta_i^I + \beta_j^J + \beta_{ij}^{IJ} + \beta_{si}^{SI} + \beta_{sj}^{SJ} + \sum_{k=1}^{M} \sigma_k (-1)^{s+1} \beta_{k,i}^I \beta_{k,j}^J \tag{5}$$

where S has only two levels $s = 1, 2$ (see Falguerolles [9] for a review of these models).

All the models above can be fitted by alternative generalized linear regressions. Details of this strategy can be found in Falguerolles and Francis [10, 11]. Note that one way to alleviate the problem of local optima is to use several random starting values in the fitting process. It turns out that the generalized bi-linear models for symmetry and departure from symmetry can be fitted by using a trick analogous to the three-dimensional representation of a square table used to fit a quasi-symmetry model (see Bishop *et al.* [3, page 289]).

4.3.1. Reduced rank quasi-symmetry

The data are duplicated thus creating a third factor S with values $s = 1, 2$. The diagonal data are weighted out. Then the following generalized bi-linear model is considered:

$$\begin{aligned} \eta_{1ij}^{SIJ} &= \beta^{\emptyset} + \beta_i^I + \beta_j^J + \sum_{k=1}^{M} \sigma_k \beta_{k,i}^I \beta_{k,j}^J \\ \eta_{2ij}^{SIJ} &= \beta^{\emptyset} + \beta_i^I + \beta_j^J + \sum_{k=1}^{M} \sigma_k \beta_{k,j}^I \beta_{k,i}^J \end{aligned}$$

The common linear part $(\beta^{\emptyset}+\beta_i^I+\beta_j^J)$ corresponds to quasi-independence. The permutation of indices (i,j) in the bi-linear part constrains symmetry for the scores while its structure preserve the rank of the approximation in the alternating process.

Thus the fitting process gives scores such as $\beta_{k,i}^I = (-1)^{\ell_k}\beta_{k,i}^J$ with $\ell_k = 1$ or 2 for $i = 1,\ldots,N$ and $k = 1,\ldots,M$.

4.3.2. Reduced rank skew-symmetry

Again the data are duplicated but the replication is flipped over. A tri-linear model, with an even number of dimensions, is then fitted:

$$\eta_{sij}^{SIJ} = \beta^{\emptyset} + \beta_s^S + \beta_i^I + \beta_j^J + \beta_{ij}^{IJ} + \beta_{si}^{SI} + \beta_{sj}^{SJ} + \sum_{k=1}^{2M} \sigma_k (-1)^{s+1} \beta_{k,i}^I \beta_{k,j}^J.$$

In this model the linear part $\beta^{\emptyset}+\beta_s^S+\beta_i^I+\beta_j^J+\beta_{ij}^{IJ}+\beta_{si}^{SI}+\beta_{sj}^{SJ}$ imposes the baseline quasi-symmetry while the duplication $(s = 1,2)$ coerces the even number of dimensions into M bi-dimensions modelling skew-symmetry.

Bibliography

[1] Aitkin (M.), Anderson (D.), Francis (B.) and Hinde (J.). — *Statistical Modelling in GLIM*, Oxford University Press, Oxford, 1989.

[2] Becker (M. P.). — Quasisymmetric models for the analysis of square contingency tables, *Journal of the Royal Statistical Society*, 52 (1990), 369–378.

[3] Bishop (Y. M. M.), Fienberg (S. E.) and Holland (P. W.). — *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Mass, 1975.

[4] Caillez (F.), Pagès (J.-P.). — *Introduction à l'analyse des données*, SMASH, Paris, 1976.

[5] Caussinus (H.). — Contribution à l'analyse statistique des tableaux de corrélation, *Annales de la Faculté des Sciences de Toulouse*, 29 (année 1965), (1966), 77–183.

[6] Caussinus (H.), de Falguerolles (A.). — Tableaux carrés : modélisation et méthodes factorielles, *Revue de Statistique Appliquée*, 35 (1987), 35–52.

[7] Constantine (A. G.), Gower (J. C.). — Graphical representation of asymmetric matrices, *Applied Statistics*, 27 (1978), 297–304.

[8] Escoufier (Y.), Grorud (A.). — L'analyse factorielle des matrices carrées non symmétriques, in E. Diday, L. Lebart, J.P. Pagès and R. Tomassone, editors, *Data Analysis and Informatics 2*, pages 263–273, North–Holland, 1980.

[9] de Falguerolles (A.). — GBMs: GLMs with bilinear terms, in J. Bethlehem & P.G.M. van der Heijden, editors, *COMPSTAT 2000, proceedings in computational statistics*, Physica-Verlag, 2000.

[10] de Falguerolles (A.), Francis (B.). — Algorithmic approaches for fitting bilinear models, in Y. Dodge and J. Whittaker, editors, *COMPSTAT'92 Proceedings in Computational Statistics*, volume 1, Physica-Verlag, 1992.

[11] DE FALGUEROLLES (A.), FRANCIS (B.). — Fitting bilinear models in GLIM, *GLIM Newsletter*, 25 (1995), 9–20.

[12] GABRIEL (K. R.). — The biplot graphic display of matrices with application to principal component analysis, *Biometrika*, 58 (1971), 453–467.

[13] GABRIEL (K. R.). — Le biplot – outil d'exploration de données multidimensionnelles, *to appear in Journal de la Société Française de Statistique*, 2002.

[14] GOODMAN (L.). — Simple models for the analysis of association in cross-classifications having ordered categories, *Journal of the American Association*, 74 (1979), 537–552.

[15] GOWER (J. C.), HAND (D. J.). — *Biplots*, Chapman and Hall, London, 1996.

[16] GREENACRE (M.). — *Theory and Applications of Correspondence Analysis*, Academic Press, New York, 1984.

[17] GREENACRE (M.). — Correspondence analysis of square asymmetric matrices, *Applied Statistics*, 49 (2000), 297–310.

[18] MCCULLAGH (P.). — Invariance and factorial models (with discussion), *Journal of the Royal Statistical Society (series B)*, 62 (2000), 209–256.

[19] MCCULLAGH (P.), NELDER (J. A.). — *Generalized Linear Models*, Chapman and Hall, London, 2nd edition, 1989.

[20] STIGLER (S. M.). — *Statistics on the Table: the History of Statistical Concepts and Methods*, Harvard University Press, Cambridge, Massachusetts, 1999.

[21] VAN DER HEIJDEN (P. G. M.), DE FALGUEROLLES (A.), DE LEEUW (J.). — A combined approach to contingency table analysis using correspondence analysis (with Discussion), *Applied Statistics*, 38 (1989), 249–292.

[22] VAN DER HEIJDEN (P. G. M.), MOOIJAART (A.). — Some new-bilinear models for the analysis of asymmetry in a square contingency table, *Sociological Methods and Research*, 24 (1995), 7–29.

[23] VAN EEUWIJK (F. A.). — Multiplicative interaction in generalized linear models, *Biometrics*, 51 (1995), 1017-1032.