Log-Linear Models for Association and Agreement in Stratified Square Contingency Tables [1]

Gerhard Osius

Institut für Statistik, FB3, Universität Bremen,
Postfach 33 04 40, D–28334 Bremen, Germany.

Summary

In square contingency tables the agreement between row and column classification is often of primary interest. Using a decomposition of the total association into a symmetrical component describing relative agreement and an asymmetrical part, and then providing separate models for each part yields a variety of models for the expected table. Simple and easily interpretable models are introduced for classifications on nominal and ordinal scales which only restrict the symmetrical association, thus leaving the asymmetrical part and the margins completely arbitrary. Furthermore corresponding models for the asymmetric part of the association are discussed, and it is shown, that for nominal scales only two extreme models with zero or saturated asymmetry are suitable. All models considered here are log-linear models which are applicable under the usual sampling schemes (Poisson or multinomial) and can be fitted using standard software. The models are extended to a set of square tables arising from stratification according to additional observed covariables. The methods are applied to an original data set on the classification of uterine cancer and to two social mobility tables from the literature.

Keywords: Agreement, association, asymmetry, contingency table log-linear models, odds ratio, social mobility, stratification, symmetry, uterine cancer.

[1] Nov. 7, 1996. A shorter version appeared in *Computational Statistics* 12 (1997)

1 Introduction

The question of agreement in square contingency tables arises in a variety of observational studies. One typical situation occurs if each item under study is classified into J categories in two different ways, giving the classification factors A and B with values $1,...,J$. For example (e.g. application 1) the development of a particular tumor may be classified first using a low cost screening method A and later by a more expensive clinical diagnosis B. In another typical situation A and B represent the *joint* classification of a *pair*, using J categories for each member of the pair. In social mobility studies for example (e.g. applications 2 and 3), A and B may represent the occupational status of father and son, the item under study being the pair of father and son.

In the general setting, A and B can be arbitrary discrete random variables with values $1,...,J$ and positive probabilities $\pi_{jk} = P\{A{=}j, B{=}k\} > 0$ for all j and k. The data to be analyzed here arise from the classification of a number of items according to A and B and may be summarized by the total numbers of observed counts Y_{jk} for the classification $\{A=j, B=k\}$, thus giving rise to a $J{\times}J$ square contincency table $\mathbf{Y} = (Y_{jk})$. The sampling distribution of the table $\mathbf{Y}$ will only partly be specified here in order to cover several important sampling schemes in collecting the items under study. More precisely, we allow for *Poisson* or *multinomial sampling*, the latter fixing either the total Y_{++}, the row totals Y_{j+} or the column totals Y_{+k} in advance, where "$+$" indicates summation over the replaced subscript. The expected counts are denoted by $\mu_{jk} = E(Y_{jk})$.

Our aim is to provide models for *agreement* of A and B without restricting the marginal distributions of A and B by separating agreement from the *total association* of A and B. For a start let us look at the simple case with $J=2$ categories. In a $2{\times}2$ table the factors A and B agree exactly on the two main diagonal cells and disagree on the other two off-diagonal cells. Hence the odds ratio (or cross-product ratio) $\theta = \mu_{11}\,\mu_{22} / \mu_{12}\,\mu_{21}$ resp. its logarithm $\psi = \log\theta$, which characterizes the *association* independently of the margins, also describes the *agreement*. If the off-diagonal cells μ_{12} and μ_{21} both tend to zero (i.e. the table approaches complete agreement) then θ tends to infinity. Since the converse does not hold (take for example $\mu_{11} = \mu_{12} = \mu_{22} = \frac{1}{3}np$ and $\mu_{21} = n(1{-}p)$ where $\theta = \frac{1}{3}p/(1{-}p) \to \infty$ as $p \to 1$) the odds ratio appears as a measure of *relative agreement* rather than *absolute agreement*, and the term *agreement* hereafter is understood to refer to *relative agreement* even if not explicitely stated. The sign of ψ gives the direction of agreement (negative values indicating disagreement) and the absolute value measures its strength, which is 0, of course, if and only if A and B are independent.

For a general $J{\times}J$ table, the total agreement between A and B is determined by the agreement on *all* subsets $\{j,k\}$ for any two categories $1 \le j < k \le J$, which in turn are characterized by the odds ratios

$$\theta_{jk} = \mu_{jj}\,\mu_{kk} \,/\, \mu_{jk}\,\mu_{kj} \tag{1}$$

of all 2×2 subtables with upper left and lower right cells on the main diagonal. An important feature is that the odds ratio can also be written as the corresponding ratio of the probabilities π_{jk} or the conditional probabilities $\pi^B_{jk} = P\{B=k \mid A=j\}$ resp. $\pi^A_{jk} = P\{A=j \mid B=k\}$.

$$\theta_{jk} = \pi_{jj}\,\pi_{kk} \,/\, \pi_{jk}\,\pi_{kj} = \pi^B_{jj}\,\pi^B_{kk} \,/\, \pi^B_{jk}\,\pi^B_{kj} \;=\; \pi^A_{jj}\,\pi^A_{kk} \,/\, \pi^A_{jk}\,\pi^A_{kj}\,.$$

The following discussion is based on the conviction, that the odds ratio family (θ_{jk}) formally summarizes (and thus defines) a concept of *(relative) agreement* between A and B, which is independent of the marginal distributions. The *total association* beetween A and B however, is determined by the family of *all* odds ratios

$$\theta_{jklm} = \mu_{jk}\,\mu_{lm} / \mu_{jm}\,\mu_{lk} \tag{2}$$

for any $j, k, l, m \in \{1,...,J\}$, which contains all $\theta_{jk} = \theta_{jjkk}$. The association family (θ_{jklm}) can be decomposed into a symmetrical and an asymmetrical component (Sobel et al. 1985). The symmetrical part turns out to be uniquely determined by the subfamily (θ_{jk}) and thus represents *agreement*, whereas the asymmetrical part is a family (Δ_{jk}) containing the remaining association beetween A and B. The family (θ_{jk}) of agreement has four obvious, attractive properties:

- Each member θ_{jk} has a clear interpretation as an odds ratio.
- The family does not depend on the marginal distributions of the factors A and B,
- The value of a single member θ_{jk} depends only on the 2×2 subtable with rows and colums in $\{j,k\}$ and hence remains unchanged if one or more of the remaining categories are modified, e.g by combining adjacent categories.
- Suppose the factors are restricted to a subset $I \subset \{1,...,J\}$ by omitting one or more levels which may not be observed for *both* A and B. Then the family for the restricted table $\{Y_{jk} \mid j, k \in I\}$, i.e. for the *conditional* distribution of $\mathbf{Y}$ given $\{A, B \in I\}$, coincides with the *subfamily* $\{\theta_{jk} \mid j, k \in I\}$ of the unrestricted table $\mathbf{Y}$.

Our first goal is to present models for association by specifying *only* the family (θ_{jk}) of agreement, thus leaving the remaining structure of the expected table $\boldsymbol{\mu}$ completely arbitrary. Allowing for additional (asymmetrical) association besides agreement (i.e. symmetrical association) distinguishes the proposed models from popular *quasi-symmetry* models (cf. Agresti 1990, chapter 10) and leads to a broader class of models. Starting with models for *nominal* classifications (which do not depend on the possibly artificial ordering of the categories $1, ..., J$) we proceed to models for *ordinal* classifications which exploit the natural ordering or make use of a distance defined between the response categories. Although not of primary interest for agreement, corresponding models for the asymmetrical association family (Δ_{jk}) are also given.

As a second goal we wish to extend the above models to situations where an additional vector $\mathbf{X}$ of covariables is observed for each item under study and the *conditional agreement* of A and B *given the value of* $\mathbf{X}$ is of primary interest. We restrict our attention to *discrete* covariables taking a *finite* set of values (also applying for continuous variables after appropriate grouping), which allows to rely on standard asymptotic results. Since the vector $\mathbf{X}$ can only take a finite number S of values $\mathrm{x}_1, ..., \mathrm{x}_S$ and it may be coded into a single *stratum* factor C with values $1, ..., S$, i.e. $C = s$ is equivalent to $\mathbf{X} = \mathrm{x}_s$. Typical examples for the components of $\mathbf{X}$ are population characteristics like gender and age group or an observational period, e.g. calendar year.

The classification of items according to A, B and C is now summarized by the observed counts Y_{jks} for the events $\{A{=}j,\ B{=}k,\ C{=}s\}$ giving rise to a $J{\times}J{\times}S$ contingency table $\mathbf{Y} = (Y_{jks})$. As above we allow the distribution of $\mathbf{Y}$ to arise from *Poisson* or *multinomial sampling*, the latter fixing in advance *any* set of marginal totals *except* (Y_{jk+}), which restricts the *joint* distribution of (A, B), and drawing items conditional on the corresponding values of A resp. B and/or C.

Denoting the expected value by $\mu_{jks} = E(Y_{jks})$, the (conditional) agreement of A and B given stratum s is as above characterized by the family of odds ratios

$$\theta_{jks} = \mu_{jjs}\, \mu_{kks} \,/\, \mu_{jks}\, \mu_{kjs} \tag{3}$$

for any two categories j and k. Again our main objective is to present models for association by specifying *only* the family (θ_{jks}) of agreement, thus leaving the remaining structure of the expected table $\boldsymbol{\mu}$ completely arbitrary. All models given here correspond to log-linear models for the expected table $\boldsymbol{\mu}$ thus allowing an analysis of the observed table $\mathbf{Y}$ by means of standard statistical software.

Application 1: Cytological and pathological classification

The application, which motivated this paper, is concerned with the diagnosis of uterine cancer (carcinoma of the cervix). In Germany a preventive check-up to detect various forms of cancer is offered free of charge by the health insurances. As part of the investigations for females, a vaginal smear is analyzed histologically and categorized according to a cytological classification given in table 1-1, which goes back to Papanicolaou (1943). For the categories *Pap III* , *IV* and *V*, a suspicion of a carcinoma cannot be ruled out, and a pretherapeutic cone-biopsy is suggested, in order to obtain a reliable diagnosis based on a pathological classification of the tissue using the same categories of table 1-1. Obviously, the accuracy of the cytological classification *CYT* as measured by its agreement with the final pathological classification *PAT* is of major importance.

Pap I	normal cells
Pap II	light inflamed or regenerative alterations
Pap III D L	light displasias
Pap III D M	medium displasias
Pap IV A S	severe displasias
Pap IV A Cis	carcinoma in situ
Pap IV B	carcinoma in situ with beginning infiltration
Pap V	carcinoma in epithelial tissue or adenoid carcinoma

Table 1-1: Categories for cytological and pathological classifications in application 1.

The data to be analyzed here are taken from the records of the *Cytological Lab* (headed by Prof. Dr. D. Langnickel) in the gynaecological clinic of the central hospital *Zentralkrankenhaus St. Jürgen Strasse* in Bremen, Germany. From the large body of data only those individual records from 12 years (1972-1983) are selected, for which both classifications *CYT (A)* and *PAT (B)* were given. Since a cone-biopsy was routinely performed only if *CYT* was *Pap III* or higher, our analysis is restricted to these cases. As mentioned earlier, the exclusion of the first two categories, i.e. conditioning upon the event $\{CYT, PAT \geq Pap\, III\}$, does not affect the agreement between the remaining categories. This leaves us with $J=6$ response categories from *Pap III DL* to *Pap V*, exhibiting a natural order corresponding to the severity of suspicion. In order to investigate whether agreement of both classifications has changed over time (which is the only question to be pursued here), the data is divided into $S=6$ strata according to two-year time intervalls (table 1-2). The (one-dimensional) covariate *time* in stratum s is taken as $x_s = 2\,(s-1)$ to represent the number of years elapsed since the beginning of the study. An explorative data analysis ignoring the stratification is given in Osius (1993).

Cytol. class.	Pathological classification III DL	III DM	IV AS	IV AC	IV B	V	Total
1972 – 1973							
III DL	2	0	1	1	0	0	4
III DM	0	0	0	1	1	0	2
IV AS	1	2	2	5	1	3	14
IV AC	2	4	6	51	9	13	85
IV B	0	0	1	5	7	31	44
V	0	1	1	28	10	139	179
1974 – 1975							
III DL	4	4	0	0	0	0	8
III DM	2	18	7	16	0	2	45
IV AS	0	3	4	7	1	1	16
IV AC	2	1	7	55	1	9	75
IV B	0	0	1	5	1	9	16
V	0	0	2	23	0	103	128
1976 – 1977							
III DL	8	15	6	3	0	0	32
III DM	7	17	17	26	0	1	68
IV AS	1	7	16	10	0	0	34
IV AC	1	2	6	55	3	4	71
IV B	0	0	1	10	2	7	20
V	0	0	0	14	1	90	105
1978 – 1979							
III DL	18	14	6	2	0	0	40
III DM	11	19	32	18	0	1	81
IV AS	1	2	20	19	0	2	44
IV AC	2	4	18	57	1	12	94
IV B	0	0	2	8	2	1	13
V	0	1	2	8	1	60	72
1980 – 1981							
III DL	9	4	1	2	0	1	17
III DM	9	26	16	11	0	0	62
IV AS	1	5	23	17	0	1	47
IV AC	1	3	10	39	0	4	57
IV B	0	0	0	3	2	0	5
V	0	0	2	4	1	73	80
1982 – 1983							
III DL	12	5	0	1	0	1	19
III DM	6	32	13	7	0	2	60
IV AS	0	8	43	18	1	1	71
IV AC	0	0	9	36	2	1	48
IV B	0	0	1	6	0	1	8
V	0	1	1	7	1	64	74

Table 1-2: Cytological and pathological classifications in Bremen during 1972-1983 for six two-year time strata (application 1).

Application 2: Social mobility in Britain

The cross-classification of british males according to the occupational status of father (A) and son (B) using $J = 8$ status categories given in table 1-3 is based on data by Glass (1954). Goodman (1979) has fitted several models to the observed table with the main diagonal *deleted* and we will fit models presented here to the *complete* table.

Father's Status	Son's Status 1	2	3	4	5	6	7	8	Total
1	50	19	26	8	7	11	6	2	129
2	16	40	34	18	11	20	8	3	150
3	12	35	65	66	35	88	23	21	345
4	11	20	58	110	40	183	64	32	518
5	2	8	12	23	25	46	28	12	156
6	12	28	102	162	90	554	230	177	1355
7	0	6	19	40	21	158	143	71	458
8	0	3	14	32	15	126	91	106	387
Total	103	159	330	459	244	1186	593	424	3498

Table 1-3: Classification of british males by occupational status of father and son (taken from Goodman 1979, based on Glass 1954).

Application 3: Social mobility in the United States

Another social mobility table taken from the 1973 Occupational Changes in a Generation Survey in the United States (whose primary analyses were reported by Featherman and Hauser, 1978) with $J = 5$ occupation categories is given in table 1-4. The data has been analyzed by Yamaguchi (1990) who focused on the *asymmetric* structure of association and our aim is to provide models for its *symmetric* structure.

Father's Occupation	Son's Occupation 1	2	3	4	5	Total
1: Upper nonmanual	1275	364	274	272	17	2202
2: Lower nonmanual	1055	597	394	443	31	2520
3: Upper manual	1043	587	1045	951	47	3673
4: Lower manual	1159	791	1323	2046	52	5371
5: Farmer	666	496	1031	1632	646	4471
Total	5198	2835	4067	5344	793	18237

Table 1-4: Social mobility in the United States from the 1973 Occupational Changes in a Generation Survey (from Yamaguchi 1990).

2 Symmetrical and asymmetrical components of association

For simplicity we begin the discussion of models for an *unstratified* $J{\times}J$-contingency table $\mathbf{Y} = (Y_{jk})$ with $J \geq 3$ and positive expectations $\mu_{jk} = E(Y_{jk}) > 0$ for each cell. For any pair of cells (j,k) and (l,m) the odds ratio θ_{jklm} for the $2{\times}2$ subtable with columns $\{j,l\}$ and rows $\{k,m\}$ is given by (2). It is well known (Plackett 1974, sec. 3.4) that the expected table $\boldsymbol{\mu}$ is uniquely determined by: the row margins (μ_{j+}), the column margins (μ_{+k}), and the odds ratio family (θ_{jklm}).

Our primary interest focuses on the subfamily (θ_{jk}) given in (1) containing all odds ratios $\theta_{jk} = \theta_{jjkk}$ of "quadratic" $2{\times}2$ subtables along the main diagonal. We want to specify parametric models for the subfamily (θ_{jk}), leaving the remaining structure of the expected table $\boldsymbol{\mu}$ completely arbitrary. Before proposing specific models it is useful to give a convenient parametrization of the *saturated* model which imposes no structure on the expected table whatsoever. In the saturated model, the log-expectations $\eta_{jk} = \log \mu_{jk}$ can take arbitrary values, and a common parametrization is

$$\eta_{jk} = \lambda_0 + \lambda_j^A + \lambda_k^B + \lambda_{jk} \tag{4}$$

To identify the parameters, suitable linear constraints have to be imposed. We adopt the constraints $\lambda_1^A = 0$, $\lambda_1^B = 0$, $\lambda_{j1} = 0$ and $\lambda_{1k} = 0$ for all j and k, which allow the following interpretation of the parameters

$$\lambda_0 = \log \mu_{11}\ ,\ \ \lambda_j^A = \log(\mu_{j1}/\mu_{11})\ ,\ \ \lambda_k^B = \log(\mu_{1k}/\mu_{11})\ ,\ \ \lambda_{jk} = \log \theta_{11jk}\ .$$

The association parameters λ_{jk} thus completely determine the family (θ_{jklm}) of odds ratios resp. the family of logarithms

$$\psi_{jklm} = \log \theta_{jklm} = \eta_{jk} + \eta_{lm} - \eta_{jm} - \eta_{lk}\ . \tag{5}$$

Since we are dealing with a square table, we get a decomposition

$$\lambda_{jk} = \overline{\lambda}_{jk} + \Delta_{jk} \tag{6}$$

into the following parts

$$\overline{\lambda}_{jk} = \tfrac{1}{2}\,(\lambda_{jk} + \lambda_{kj})\ , \qquad \Delta_{jk} = \tfrac{1}{2}\,(\lambda_{jk} - \lambda_{kj}) = \tfrac{1}{2}\,(\psi_{11jk} - \psi_{11kj})\ . \tag{7}$$

In view of $\overline{\lambda}_{jk} = \overline{\lambda}_{kj}$ and $\Delta_{jk} = -\,\Delta_{kj}$ the family $(\overline{\lambda}_{jk})$ represents the *symmetrical* and (Δ_{jk}) the *asymmetrical* part of the association $\boldsymbol{\lambda}$. The main advantage of this decomposition is that the symmetrical part $\overline{\boldsymbol{\lambda}}$ completely characterizes the family (θ_{jk}) of agreement resp. the family of logarithms

$$\psi_{jk} = \log \theta_{jk} = \eta_{jj} + \eta_{kk} - \eta_{jk} - \eta_{kj} \,. \tag{8}$$

In fact, inserting (4) herein yields $\psi_{jk} = \lambda_{jj} + \lambda_{kk} - 2\,\overline{\lambda}_{jk}$. Using $\lambda_{jj} = \overline{\lambda}_{jj}$, gives $\boldsymbol{\psi}$ as a function of $\overline{\boldsymbol{\lambda}}$. And conversely, $\lambda_{jj} = \psi_{1j}$ yields $\overline{\boldsymbol{\lambda}}$ as function of $\boldsymbol{\psi}$:

$$\overline{\lambda}_{jk} = \tfrac{1}{2}\,(\psi_{1j} + \psi_{1k} - \psi_{jk}). \tag{9}$$

Inserting (6) and (9) into (4) now gives the desired parametrization of the *saturated model*

$$\text{(SM)} \qquad \eta_{jk} = \alpha_0 + \alpha_j^A + \alpha_k^B + \Delta_{jk} - \tfrac{1}{2}\,\psi_{jk} \,,$$

with new parameters $\alpha_0 = \lambda_0$, $\alpha_j^A = \lambda_j^A + \frac{1}{2}\psi_{1j}$ and $\alpha_k^B = \lambda_k^B + \frac{1}{2}\psi_{1k}$ satisfying for all j, k

$$\begin{array}{ll} \alpha_1^A = 0\,, & \alpha_1^B = 0\,, \\ \Delta_{1k} = 0\,, & \Delta_{jk} = -\Delta_{kj}\,, \\ \psi_{jj} = 0\,, & \psi_{jk} = \psi_{kj}\,. \end{array} \tag{10}$$

The symmetrical part $\boldsymbol{\psi}$ is determined by the subfamily ψ_{jk} with $1 \le j < k \le J$ through

$$\psi_{jk} = I\{j \neq k\} \cdot \psi_{(j\wedge k)\,(j\vee k)} \,. \tag{11}$$

Here $I\{E\}$ denotes the indicator for an event E (which equals 1 if E occurs and 0 otherwise), and $j \wedge k$ resp. $j \vee k$ is the minimum resp. maximum of j and k. And the asymmetrical part $\boldsymbol{\Delta}$ is determined by all Δ_{jk} with $1 < j < k \le J$ through

$$\Delta_{jk} = I\{1 < j \wedge k\} \cdot \operatorname{sgn}(k-j) \cdot \Delta_{(j\wedge k)\,(j\vee k)} \,, \tag{12}$$

where *sgn(m)* denotes the sign of m, taking values –1, 0, +1 for $m<0$, $=0$, >0. The dimensions of these parameter families are summarized in table 2-1.

Using (11) and (12) the saturated model may finally be written as

$$\begin{aligned} \text{(SM)}' \qquad \eta_{jk} = \alpha_0 + \alpha_j^A + \alpha_k^B + {} & I\{1 < j \wedge k\} \cdot \operatorname{sgn}(k-j) \cdot \Delta_{(j\wedge k)\,(j\vee k)} \\ & - \tfrac{1}{2}\, I\{j \neq k\} \cdot \psi_{(j\wedge k)\,(j\vee k)} \end{aligned}$$

with no additional restrictions for the parameters α_0, $\{\alpha_j^A \,|\, 1 < j \le J\}$, $\{\alpha_k^B \,|\, 1 < k \le J\}$, $\{\Delta_{jk} \,|\, 1 < j < k \le J\}$ and $\{\psi_{jk} \,|\, 1 \le j < k \le J\}$.

Source	Parameter	Dimension
constant	α_0	1
row factor A	$\boldsymbol{\alpha}^A$	J–1
column factor B	$\boldsymbol{\alpha}^B$	J–1
asymmetric association	$\boldsymbol{\Delta}$	$(J–1)(J–2)/2$
symmetric association	$\boldsymbol{\psi}$	$J\,(J–1)/2$
total		$J{\times}J$

Table 2-1: Dimensions of the parameters in the saturated model.

Our definition of the asymmetrical part $\boldsymbol{\Delta}$ is based on the subfamily (ψ_{11jk}) of log odds ratios, which completely determines the whole family (ψ_{jklm}). However, other minimal subfamilies may be used, e.g. the *local* log odds ratios

$$\psi^*_{jk} = \psi_{jk(j+1)(k+1)} = \eta_{jk} + \eta_{(j+1)(k+1)} - \eta_{j(k+1)} - \eta_{(j+1)k} \tag{13}$$

for all $j, k = 1, ..., J–1$. The corresponding *local* asymmetrical family

$$\Delta^*_{jk} = \tfrac{1}{2}\,(\psi^*_{jk} - \psi^*_{kj}) \tag{14}$$

is a linear function of $\boldsymbol{\Delta}$:

$$\Delta^*_{jk} = \Delta_{jk} + \Delta_{(j+1)(k+1)} - \Delta_{j(k+1)} - \Delta_{(j+1)k}\,. \tag{15}$$

Conversely, $\boldsymbol{\Delta}$ is uniquely determined by the local family $\boldsymbol{\Delta}^*$ through

$$\Delta_{jk} = \sum\nolimits_{1\le l<j}\sum\nolimits_{1\le m<k}\Delta^*_{lm} = \mathrm{sgn}(k-j)\cdot\sum\nolimits_{1\le l<j}\sum\nolimits_{j\le m<k}\Delta^*_{lm} \tag{16}$$

and this relation can be used to replace $\boldsymbol{\Delta}$ in (SM)$'$ by $\boldsymbol{\Delta}^*$, if preferred.

All models discussed here will only restrict the association but not the marginal distributions of A and B, i.e. the parameters α_0, $\boldsymbol{\alpha}^A$ and $\boldsymbol{\alpha}^B$ in the saturated model remain arbitrary. Our first goal is to construct submodels by imposing additional structure on the *symmetrical* part $\boldsymbol{\psi}$ of the association, thus leaving the asymmetrical part $\boldsymbol{\Delta}$ completely arbitrary. All models given here restrict $\boldsymbol{\psi}$ to a *linear* subspace $\mathscr{Q}$ of $\mathbb{R}^{J\times J}$. And in a second step we discuss structures for the *asymmetrical* family by restricting $\boldsymbol{\Delta}$ to a *linear* subspace $\mathscr{D}$ of $\mathbb{R}^{J\times J}$. Combining any model $\mathscr{Q}$ for $\boldsymbol{\psi}$ with any model $\mathscr{D}$ for $\boldsymbol{\Delta}$ provides a variety of *log-linear models* for the expected table $\boldsymbol{\mu}$ in which $\boldsymbol{\psi}\in\mathscr{Q}$ can vary *independently* of $\boldsymbol{\Delta}\in\mathscr{D}$, i.e the corresponding model formulae for $\boldsymbol{\psi}$ and $\boldsymbol{\Delta}$ have no common parameters. We will distinguish between models suitable for *nominal* responses and those who depend on the given ordering of the response categories. A model $\mathscr{Q}$ for the symmetrical part is called *suitable* for nominal responses, if it is *invariant* under *any* permutation p of the response categories $1, ..., J$, i.e. $\boldsymbol{\psi} = (\psi_{jk}) \in \mathscr{Q}$ implies $\boldsymbol{\psi}_p = (\psi_{p(j)\,p(k)}) \in \mathscr{Q}$. Suitability of a model $\mathscr{D}$ for $\boldsymbol{\Delta}$ is defined correspondingly.

3 Models for symmetry

Beginning with models suitable for nominal reponses (e.g. blood group), the most elementary models are *zero symmetry*

(ZS) $$\psi_{jk} = 0 \quad , \quad j \neq k$$

and *constant symmetry* with an additional parameter ψ_0:

(CS) $$\psi_{jk} = \psi_0 \quad , \quad j \neq k.$$

A natural extension allowing additive row and column effects is $\psi_{jk} = \psi_0 + \sigma_j + \nu_k$ for all $j \neq k$ with parameter vectors $\boldsymbol{\sigma}, \boldsymbol{\nu} \in \mathbb{R}^J$ satisfying $\sigma_1 = \nu_1 = 0$. However the constraints upon $\boldsymbol{\psi}$ require $\boldsymbol{\sigma} = \boldsymbol{\nu}$ thus reducing the above model to *additive symmetry*

(AS) $$\psi_{jk} = \psi_0 + \sigma_j + \sigma_k \quad , \quad j \neq k.$$

An equivalent formulation without using particular parameters is

(AS)′ For any fixed pair $j \neq l$ the differences $\psi_{jk} - \psi_{lk}$ are constant for all $k \neq j, l$.

A reparametrization $\sigma_j^\circ = \sigma_j + \frac{1}{2}\psi_0$ leads to

(AS)° $$\psi_{jk} = \sigma_j^\circ + \sigma_k^\circ \quad , \quad j \neq k,$$

where σ_j° represents the contribution of category j to each symmetric association involving j. For a satisfactory overall agreement we expect all parameters σ_j° to be *positive*.

Suppose now we are dealing with an *ordinal* response variable with category labels $1, ..., J$ reflecting the order of the response. The additive symmetry model may then be extended to an *order additive symmetry* model by allowing different contributions according to the order of the responses for the row and column factor

(OAS) $$\psi_{jk} = \psi_0 + \tau_{j \wedge k} + \nu_{j \vee k} \quad , \quad j \neq k$$

with parameter vectors $\boldsymbol{\tau}, \boldsymbol{\nu} \in \mathbb{R}^J$ satisfying suitable constraints, e.g. $\tau_1 = \tau_J = 0$, $\nu_1 = \nu_J = 0$. An equivalent description reveals the model to contain the additive symmetry model:

(OAS)′ For any fixed $j = 1, ..., J$ the differences $\psi_{jk} - \psi_{1k}$ are constant for all $k > j$ and $\psi_{jk} - \psi_{Jk}$ are constant for all $k < j$.

A reparametrization $\tau_j^\circ = \tau_{j-1} - \tau_j$ and $\nu_j^\circ = \nu_{j+1} - \nu_j$ for $1 < j < J$ yields

(OAS)° $$\psi_{jk} = \psi_0 - \sum_{1<l\leq j} \tau_l^\circ - \sum_{k\leq l<J} \nu_l^\circ \quad , \quad j < k$$

with the following interpretation of the parameters. The constant $\psi_0 = \psi_{1J}$ represents the agreement for the extreme responses 1 and J, and for $j < k$ the parameter $\tau_j^\circ = \psi_{(j-1)\,k} - \psi_{jk}$ resp. $\nu_k^\circ = \psi_{j\,(k+1)} - \psi_{jk}$ measures the *change* in agreement due to decreasing the lower category j resp. increasing the upper category k. For a satisfactory overall agreement, the parameter ψ_0 as well as all changes $\tau_2^\circ, ..., \tau_{J-1}^\circ$, $\nu_2^\circ, ..., \nu_{J-1}^\circ$ should be *positive.*

The model (OAS) has $2J - 3$ independent parameters whereas (AS) has only J parameters. However, for $J = 3$ both models coincide with the *saturated symmetry model* (which leaves $\boldsymbol{\psi}$ completely arbitrary), but for $J > 3$ the three models are strictly nested.

If in addition to the given ordering a *distance* $d_{jk} > 0$ between any pair $j \neq k$ of responses can be specified, further models for $\boldsymbol{\psi}$ based on the distance are straightforward, for instance the *quadratic distance symmetry model*

(QDS) $$\psi_{jk} = \psi_0 + \varrho\, d_{jk} + \varrho_2\, d_{jk}^2 \quad , \quad j \neq k\,.$$

A natural choice is $d_{jk} = |k-j|$, and more generally $d_{jk} = |z_k - z_j|$, provided the ordering of the response categories is induced by associated *scores* $z_1 < z_2 \ldots < z_J$. The most general model in which ψ_{jk} depends on j and k only through the distance $|j-k|$ is obtained by viewing the distance as a factor with J–1 levels, thus giving the *general distance symmetry model* with parameters $\varrho_1, ..., \varrho_{J-1}$:

(GDS) $$\psi_{jk} = \varrho_{\,|k-j\,|} \quad , \quad j \neq k.$$

4 Models for asymmetry

Even if prime interest focuses on agreement and hence on models for $\boldsymbol{\psi}$, it may be desirable to specifiy the asymmetrical association $\boldsymbol{\Delta}$ as well, in order to achieve a more parsimonious log-linear model for the expected counts $\boldsymbol{\mu}$. For instance, if the observed table $\mathbf{y} = (y_{jk})$ contains several zero entries, the maximum likelihood estimate $\hat{\boldsymbol{\mu}}$ may not exist for some models given by $\boldsymbol{\psi} \in \mathscr{Q}$ and *saturated* asymmetry, but only under a restriction $\boldsymbol{\Delta} \in \mathscr{D}$ (for the existence of maximum likelihood estimates, see Haberman 1974, Chapter 2). Clearly, the most restrictive model for $\boldsymbol{\Delta}$ is the *zero asymmetry* model

(ZA) $$\Delta_{jk} = 0 \ .$$

This model states that the association for the expected table $\boldsymbol{\mu}$ and its transpose $\boldsymbol{\mu}^T$ are equal, i.e. interchanging the factors A and B does not change the association. Note that the model may also be formulated in terms of the local family as $\boldsymbol{\Delta}^* = \mathbf{0}$.

A natural extension of zero asymmetry is *constant asymmetry*

(CA) $$\Delta_{jk} = \Delta_0 \quad , \quad 1 < j < k \leq J.$$

However, in contrast to constant *symmetry*, the model (CA) makes implicit use of the given ordering for the response categories (via the definition of Δ_{jk}) and hence is *not* suitable for *nominal* responses. In fact, the following consideration shows, that *any* log-linear model, specified by $\boldsymbol{\psi} \in \mathscr{Q}$ and $\boldsymbol{\Delta} \in \mathscr{D}$, which is suitable for *nominal* responses and contains *constant* asymmetry already contains the model with *saturated* asymmetry:

$$\overline{\mathscr{D}} = \{\, \boldsymbol{\Delta} \in \mathbb{R}^{J \times J} \mid \Delta_{jj} = 0 \text{ and } \Delta_{jk} = -\Delta_{kj} \text{ for all } j,\, k \,\}.$$

This is mainly due to $\Delta_{jk} = -\Delta_{kj}$, and a formal argument runs as follows. Suppose there exists a constant non-zero element $\boldsymbol{\Delta}^\circ \in \mathscr{D}$, i.e. $\Delta^\circ_{jk} = \Delta_0 \neq 0$ for all $1 < j < k \leq J$. Choosing the permutation p which swaps only 2 and 3, we conclude $\boldsymbol{\Delta}^\circ_p = (\boldsymbol{\Delta}^\circ_{p(j)\,p(k)}) \in \mathscr{D}$ and hence $\boldsymbol{\Delta} - \boldsymbol{\Delta}^\circ_p \in \mathscr{D}$. Let $\mathbf{e}_{lm} = (I\{j = l, k = m\})$ denote the $J \times J$ "unit table" with entry 1 in cell (l, m) and zero entries elsewhere. Then the representation $\boldsymbol{\Delta} - \boldsymbol{\Delta}^\circ_p = 2\Delta_0\, \mathbf{d}_{23}$ with $\mathbf{d}_{23} = \mathbf{e}_{23} - \mathbf{e}_{32}$ implies $\mathbf{d}_{23} \in \mathscr{D}$. Applying the permutation with exchanges 2 with j and 3 with k yields $\mathbf{d}_{jk} = \mathbf{e}_{jk} - \mathbf{e}_{kj}$ $\in \mathscr{D}$ for any $1 < j < k \leq J$ and hence $\mathscr{D}$ contains the linear space spanned by *all* such tables $\mathbf{d}_{jk}$, which is precisely $\overline{\mathscr{D}}$ above.

Another disappointing feature of the constant asymmetry model is its dependence upon the chosen representation of the asymmetric family. The equivalent formulation in terms of the *local* asymmetric family $\boldsymbol{\Delta}^*$ is

(CA)$'$ $$\Delta^*_{j(j+1)} = \Delta_0 \quad , \quad \Delta^*_{jk} = 0 \quad , \quad 1 \leq j < k < J,\ \ k \neq j+1 \,,$$

and does *not* coincide with the corresponding model for *constant local asymmetry*

(CLA) $$\Delta^*_{jk} = \Delta^*_0 \quad , \quad 1 \leq j < k < J.$$

The latter model was introduced by Yamaguchi (1990) under the name of *uniform skew-symmetric association model* and can be rewritten in terms of $\boldsymbol{\Delta}$ as

(CLA)′ $$\Delta_{jk} = (j-1)(k-j)\,\Delta_0^* \quad , \quad 1 \leq j \leq k \leq J.$$

The model (CLA) suffers from the same drawbacks as (CA). It is not appropriate for *nominal* responses (the above argument still holds if $\boldsymbol{\Delta}$ is replaced by $\boldsymbol{\Delta}^*$) and depends upon the chosen asymmetric family $\boldsymbol{\Delta}^*$. For $J = 3$ however, both models (CA) and (CLA) already coincide with *saturated* asymmetry, and to avoid trivialities, we assume $J \geq 4$ in the remaining discussion for asmmetric models.

Further models for the asymmetric family $\boldsymbol{\Delta}$ or the local family $\boldsymbol{\Delta}^*$ can be obtained from any symmetry model if ψ_{jk} is replaced by Δ_{jk} for all $1 < j < k \leq J$ or by Δ_{jk}^* for all $1 < j < k \leq J$. For example, the corresponding models for (AS) are the *additive asymmetry* resp. *additive local asymmetry model*

(AA) $$\Delta_{jk} = \sigma_j + \sigma_k \quad , \quad 1 < j < k \leq J,$$

(ALA) $$\Delta_{jk}^* = \sigma_j^* + \sigma_k^* \quad , \quad 1 \leq j < k < J,$$

which may be rewritten using (15) resp. (16) as

(AA)′ $$\Delta_{j(j+1)}^* = \sigma_{j+1} + \sigma_k \quad , \quad \Delta_{jk}^* = 0 \quad , \quad 1 \leq j < j+1 < k < J,$$

(ALA)′ $$\Delta_{jk} = (k-2j+1)\,\tilde{\sigma}_j^* + (j-1)\,\tilde{\sigma}_k^* \quad , \quad 1 < j < k \leq J,$$

where $\tilde{\sigma}_j^* = \sigma_1^* + ... + \sigma_{j-1}^*$. Again, both models are not suitable for nominal responses by the above argument. For completeness, the *order additive asymmetry* resp. *local asymmetry model* corresponding to (OAS) are given below in terms of both $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^*$:

(OAA) $$\Delta_{jk} = \Delta_0 + \tau_j + \nu_k \quad , \quad 1 < j < k \leq J$$

(OAA)′ $$\Delta_{j(j+1)}^* = \Delta_0 + \tau_j + \nu_k \quad , \quad \Delta_{jk}^* = 0 \quad , \quad 1 \leq j < j+1 < k < J$$

with constraints $\tau_2 = \tau_J = \nu_2 = \nu_J = 0$, resp.

(OALA) $$\Delta_{jk}^* = \Delta_0^* + \tau_j^* + \nu_k^* \quad , \quad 1 \leq j < k < J$$

(OALA)′ $$\Delta_{jk} = (j-1)(k-j)\Delta_0^* + (k-j)\,\tilde{\tau}_j^* + (j-1)(\tilde{\nu}_k^* - \tilde{\tau}_j^*)$$

$$\tilde{\tau}_j^* = \tau_1^* + ... + \tau_{j-1}^* \quad , \quad \tilde{\nu}_k^* = \nu_1^* + ... + \nu_{k-1}^* \quad , \quad 1 < j < k \leq J$$

with constraints $\tau_1^* = \tau_{J-1}^* = \nu_1^* = \nu_{J-1}^* = 0$.

The additive models (AA) and (ALA) have $J-1$ independent parameters while the order additive models (OAA) and (OALA) have $2J-5$ parameters. Hence, for $J = 4$ the additive model and the corresponding order additive model already coincide with *saturated* asymmetry, but for $J > 4$ the three models are strictly nested.

Finally, the class of *skew-symmetric level models* introduced by Yamaguchi (1990) are based on the family

$$\Phi_{ijk} = (\eta_{ij} - \eta_{ji}) + (\eta_{jk} - \eta_{kj}) + (\eta_{ki} - \eta_{ik}) = 2\,(\Delta_{ij} + \Delta_{jk} + \Delta_{ki})$$

for $i < j < k$, and can be reformulated in terms of $\Delta_{jk} = \frac{1}{2}\,\Phi_{1jk}$. In particular, the *triangles-parameter skew-symmetry* model, given by $\Phi_{ijk} = 2\Phi$ for $i < j < k$, coincides with the model (CA) for $\Delta_0 = \Phi$. And the *middle-value-effect skew-symmetry* model, given by $\Phi_{ijk} = 2\Phi_j$ for $i < j < k$, is a *submodel* of (OAA), given by $\boldsymbol{\nu} = \mathbf{0}$ with $\Phi_j = \Delta_0 + \tau_j$.

Resuming the above discussion, we conclude that among log-linear models only zero or saturated asymmetry are suitable for nominal as well as ordinal responses. Since zero asymmetry is clearly too restrictive for many applications, we generally advocate models with saturated asymmetry, unless the asymmetric structure is of primary interest.

5 Joint models for symmetry and asymmetry

Combining any of the above models for the symmetrical part $\boldsymbol{\psi}$ (including saturated symmetry) with one for the asymmetrical part $\boldsymbol{\Delta}$ yields a variety of log-linear models for the expected counts, some of which have been introduced before by other authors under different names (cf. Agresti 1990, Chapter 10). In particular, the model with zero asymmetry and saturated symmetry, i.e. omitting $\boldsymbol{\Delta}$ in (SM), was introduced by Caussinus (1965) as the *quasi-symmetry model.* And submodels thereof, given by additive symmetry (AS) resp. (AS+QDS) with ϱ=0 are known as *quasi-independence* resp. *quasi-(uniform) association models* (an epidemiological application is given by Velema et al. 1991 which is re-analyzed in Osius 1993). Tanner and Young (1985) already used models with constant symmetry (CS) and constant asymmetry (CA) as well as a combination of general distance symmetry (GDS) with the corresponding model for asymmetry. However, models for ordinal responses containing *order additive* symmetry (OAS) or asymmetry (OAA) seem to be new, as well as models using *saturated* asymmetry.

There is a convenient description for models by means of model formulae (McCullagh and Nelder 1989, 3.4), which is used by computer programs like GLIM (Francis et al. 1993). Let *SYM* and *ASYM* denote the model formulae for the symmetrical and asymmetrical components $\boldsymbol{\psi}$ and $\boldsymbol{\Delta}$ of association. The corresponding model formulae for the log-expectation (linear predictor) $\boldsymbol{\eta} = \log \boldsymbol{\mu}$ is then given by

$$\boldsymbol{\eta} = 1 + A + B + S.SYM + AS.ASYM$$

where S and AS represent covariates given by

$$S(j,k) = -\tfrac{1}{2}\, I\{j \neq k\}\,, \qquad AS(j,k) = I\{1 < j \wedge k\} \cdot \mathrm{sgn}(k-j).$$

In the *saturated* model *SYM* is given by a factor *SatSym*, taking $\binom{J}{2}$ different values for all subsets $\{j,k\}$ of $\{1,\ldots,J\}$ with $j \neq k$, and an additional value on the diagonal $j=k$. A convenient version is

$$SatSym(j,k) = 1 + I\{j \neq k\} \cdot \left[\, (j \wedge k) + \tfrac{1}{2}\,(j \vee k - 1)\,(j \vee k - 2)\,\right],$$

with levels ranging from 1 to $1+\binom{J}{2}$. And *ASYM* is given by a factor *SatAsym* taking $\binom{J-1}{2}$ different values for all subsets $\{j,k\}$ of $\{2,\ldots,J\}$ with $j \neq k$, and an additional value in the remaining cases $j=k$, $j=1$, $k=1$, e.g.

$$SatAsym(j,k) = 1 + I\{j \neq k\,, 1 < j \wedge k\} \cdot \left[\, (j \wedge k - 1) + \tfrac{1}{2}\,(j \vee k - 2)\,(j \vee k - 3)\,\right]$$

with levels ranging from 1 to $1+\binom{J-1}{2}$. For a 6×6 table both factors *SatSym* and *SatAsym* are pictured in table 5-1.

SatSym

	1	2	3	4	5	6
1	1	2	3	5	8	12
2	2	1	4	6	9	13
3	3	4	1	7	10	14
4	5	6	7	1	11	15
5	8	9	10	11	1	16
6	12	13	14	15	16	1

SatAsym

	1	2	3	4	5	6
1	1	1	1	1	1	1
2	1	1	2	3	5	8
3	1	2	1	4	6	9
4	1	3	4	1	7	10
5	1	5	6	7	1	11
6	1	8	9	10	11	1

Table 5-1: The factors *SatSym* and *SatAsym* in a 6×6 table

The model of order additive symmetry, for example, can briefly be described by $SYM = Min + Max$ with factors *Min* and *Max* representing the minimum and maximum of A and B. Leaving the remaining structure of $\boldsymbol{\mu}$ unspecified produces the model

$$\boldsymbol{\eta} = 1 + A + B + S.(Min + Max) + AS.SatAsym\,.$$

Viewing $J{\times}J$ *tables* as *vectors* of length J^2 (and vice versa) this model may be written in matrix notation as $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\theta}$ with a parameter vector $\boldsymbol{\theta} = (\alpha_0, \boldsymbol{\alpha}^A, \boldsymbol{\alpha}^B, \boldsymbol{\Delta}, \psi_0, \boldsymbol{\tau}^{\circ}, \boldsymbol{\nu}^{\circ})$ of dimension, say R, and a corresponding $J^2{\times}R$ model matrix $\mathbf{X}$. The *columns* of $\mathbf{X}$ are pictured as $J{\times}J$ *tables* in table 5-2 for each component of $\boldsymbol{\theta}$, except for the first column which is the constant table $\mathbf{1}$ corresponding to α_0.

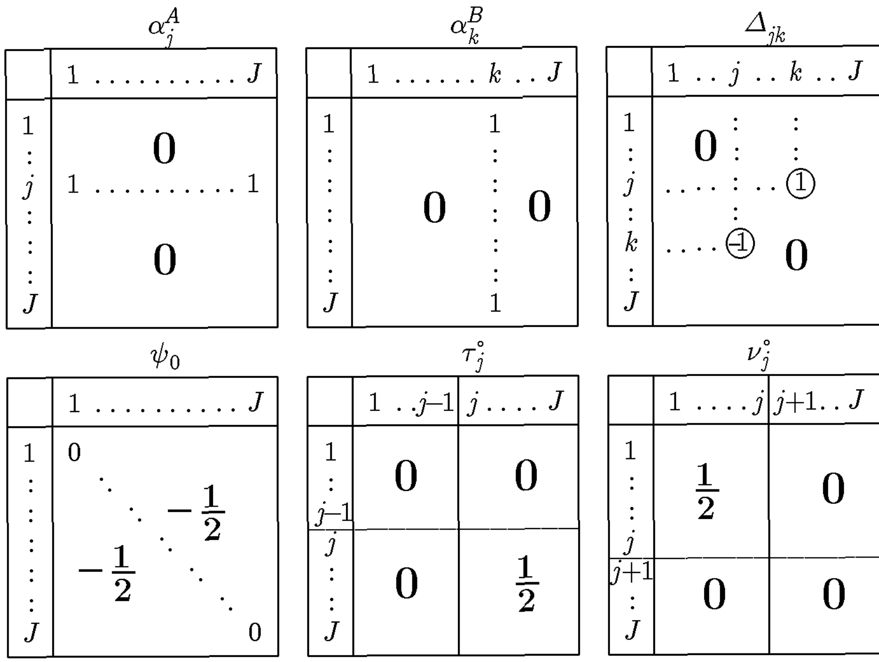

Table 5-2: Parameters and associated columns (viewed as $J{\times}J$ tables) of the matrix $\mathbf{X}$ for the model (OAS) with saturated asymmetry.

6 Models for stratified square tables

Let us now generalize the previous models to the case where an additional stratum factor C with S levels is observed. The previous notations will be extended by a subscript $s = 1, ..., S$ referring to stratum s. The saturated model for the log-expectation $\boldsymbol{\eta}$ may be parametrized using (SM) for each stratum s thus giving the *saturated stratum model*

$$\text{(SSM)} \qquad \eta_{jks} = \alpha_{0s} + \alpha^A_{js} + \alpha^B_{ks} + \Delta_{jks} - \tfrac{1}{2}\,\psi_{jks}$$

with the obvious constraints corresponding to (10) for each s. As before our main interest focuses on models for the symmetrical part (ψ_{jks}) and possibly on (Δ_{jks}) leaving the remaining parameters (α_{0s}), (α^A_{js}) and (α^B_{ks}) completely arbitrary to cover all major sampling models for the observed table $\mathbf{Y}$. Concerning the dependence of ψ_{jks} on stratum s, the most simple model is *constant stratum symmetry*

$$\text{(CSS)} \qquad \psi_{jks} = \beta_{0jk} \quad , \quad j \neq k$$

with a parameter matrix $\boldsymbol{\beta}_0 \in \mathbb{R}^{J \times J}$ satisfying $\beta_{0jk} = \beta_{0kj}$.

If stratum s is defined as the subset $\{\mathbf{X} = \mathbf{x}_s\}$ of the population under study on which a vector $\mathbf{X} = (X_1, ..., X_R)$ of R covariables takes a given value $\mathbf{x}_s = (x_{s1}, ..., x_{sR})$, then a *linear stratum symmetry* model for $\boldsymbol{\psi}$ is available

(LSS) $$\psi_{jks} = \beta_{0jk} + \sum_{1 \le r \le R} \beta_{rjk}\, x_{sr} \quad , \quad j \ne k.$$

The parameter matrix $\boldsymbol{\beta}_0 \in \mathrm{IR}^{J \times J}$ gives the *baseline symmetry* corresponding to $\mathbf{x} = \mathbf{0}$ and for each r the *slope* $\boldsymbol{\beta}_r \in \mathrm{IR}^{J \times J}$ measures the increase of symmetry per unit increase of x_r. Note that the constraints $\beta_{rjk} = \beta_{rkj}$ for all r (including $r = 0$), j and k are required to guarantee $\psi_{jks} = \psi_{kjs}$ for all possible values $\mathbf{x}_s \in \mathrm{IR}^R$.

The linear stratum model represents a *log-linear* model for the expected counts. A number of interesting submodels are obtained by imposing additional structure on the baseline $\boldsymbol{\beta}_0$ and the slopes $\boldsymbol{\beta}_r$ along the lines of the preceeding section, e.g. *order additive baseline* resp. *slope symmetry*

(OABS) $$\beta_{0jk} = \psi_{00} + \tau_{0(j \wedge k)} + \nu_{0(j \vee k)} \quad , \quad j \ne k,$$

(OASS) $$\beta_{rjk} = \psi_{r0} + \tau_{r(j \wedge k)} + \nu_{r(j \vee k)} \quad , \quad j \ne k.$$

It may also be desirable (in particular in the presence of many strata), to restrict the dimension of asymmetrical part $\boldsymbol{\Delta}$ by additional model assumptions, as*linear stratum asymmetry*

(LSA) $$\Delta_{jks} = \delta_{0jk} + \sum_{1 \le r \le R} \delta_{rij}\, x_{sr} \quad , \quad j \ne k.$$

The *baseline asymmetry* $\boldsymbol{\delta}_0 \in \mathrm{IR}^{J \times J}$ and *slopes of asymmetry* $\boldsymbol{\delta}_r \in \mathrm{IR}^{J \times J}$ (with the constraints $\delta_{rjk} = -\delta_{rkj}$ for all $r \ge 0$ and all j, k) may, if needed, be restricted further to one of the models for asymmetry in the previous section, e.g. order additive baseline and slopes.

7 Data analysis and applications

The analysis of an observed stratified table $\mathbf{Y} = (Y_{jks})$ using the above models for $\boldsymbol{\psi}$ and $\boldsymbol{\Delta}$ is straight forward, because these models belong to the family of log-linear models for the expected counts $\boldsymbol{\mu}$. Since log-linear models are treated in detail by many textbooks, e.g. Haberman (1974), Bishop, Fienberg and Holland (1975), McCullagh and Nelder (1989) and Agresti (1990) only a few remarks are required here.

The models for $\boldsymbol{\psi}$ and $\boldsymbol{\Delta}$ given here restrict the joint distribution of (A,B,C) *only* by specifying the *conditional* distribution $(A,B|C)$ of A and B given C, but not the (marginal) distributions of the pairs (A,C) and (B,C), because the corresponding family of nuisance parameters α_{0s}, α^A_{js} and α^B_{ks} are unrestricted. Results from Haberman (1974, Theorems 2.4, 4.1) imply, that an analysis based on the above models is *in-*

variant under the allowed sampling distributions for $\mathbf{Y}$ given earlier (namely Poisson and certain multinomial models) in the following sense:

Theorem (Invariance property): *The maximum likelihood estimate* $(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\Delta}})$ *and its estimated asymptotic covariance matrix coincide under all allowed sampling distributions for* $\mathbf{Y}$.

The asymptotics employed here requires the number J of reponse categories and the number S of strata to remain *fixed* as the total number Y_{+++} of items resp. its expectation μ_{+++} tends to infinity. And this is the only reason why the vector $\mathbf{X}$ of covariables defining the strata has been restricted to take only a finite number S of values.

The invariance property is inherited by the estimates of all parameters used to describe the symmetrical resp. asymmetrical component $\boldsymbol{\psi}$ resp. $\boldsymbol{\Delta}$ in a particular model. Hence estimation and asymptotic inference about $\boldsymbol{\psi}$ and $\boldsymbol{\Delta}$ in the above models can be based on *any* of the allowed distributions for the observed table $\mathbf{Y}$, instead of the actual sampling distribution (which must be of the allowed type). This allows a flexible analysis of the data according to the availability of suitable software. For example, GLIM may be applied on the basis of the Poisson model.

For an explorative data analysis it is instructive to compute for each stratum s the (corrected) observed odds ratios

$$\theta^c_{jklms} = (Y_{jks}+c)\,(Y_{lms}+c)\,/\,(Y_{jms}+c)\,(Y_{lks}+c)$$

and investigate the corresponding families (ψ^c_{jks}) and (Δ^c_{jks}) of agreement and asymmetry. A good choice for the correction is $c = -\frac{1}{2}$ (Cox and Snell 1989, 2.1.6), but if all Y_{jks} are positive, the ML estimate under the saturated model (given by $c = 0$) may be preferable.

Application 1: Cytological and pathological classification

We first assume a linear trend for the total association, i.e. linear stratum symmetry (LSS) in combination with linear stratum asymmetry (LSA), which leads to the following log-linear model of *linear stratum association*

$$\eta_{jks} = \alpha_{0s} + \alpha^A_{js} + \alpha^B_{ks} + \delta_{0jk} - \tfrac{1}{2}\,\beta_{0jk} + \left(\delta_{jk} - \tfrac{1}{2}\,\beta_{jk}\right) x_s \tag{17}$$

with obvious constraints for the parameters. In view of $x_1 = 0$, the baseline symmetry $\boldsymbol{\beta}_0$ describes agreement in the first stratum 1972-73, i.e. $\psi_{jk1} = \beta_{jk0}$, and the slope $\boldsymbol{\beta}$ of symmetry represents the increase of agreement per year, i.e. $\psi_{jk(s+1)} - \psi_{jks} = 2\,\beta_{jk}$. And similar interpretations apply to the baseline asymmetry $\boldsymbol{\delta}_0$ and slope $\boldsymbol{\delta}$. Restricting in (17) the parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$ by (OABS) and (OASS) respectively, but leaving the parameters $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}$ arbitrary, constitutes the *basic model* for our analysis.

The basic model provides a satisfactory fit (based on an examination of the residuals) and has a deviance $D = 107.51$ with 112 d.f. The estimates $\hat{\beta}_{0\,jk}$ resp. $\hat{\beta}_{jk}$ for the baseline resp. slope of agreement and their standard errors are given in table 7-1. Not surprisingly, all estimates are positive, and the majority even *significantly* positive, based on Wald's one sided test on the 5% level. In particular, the baseline $\hat{\beta}_{0jk}$ was not significant only for two adjacent pairs. And the slope was not significant for all pairs containing category *Pap IV B*, as well as for two "adjacent" pairs (i.e. with adjacent categories) indicating a non significant increase over time for these pairs. However, the estimates $\hat{\psi}_{jks}$ (not tabled here) obtained from (LSS) show a steady increase with s, and are significantly positive for $s > 2$ and all j, k. In summary, the basic model reveals positive agreement increasing over time for all pairs, being significantly positive for most pairs.

Although the basic model gives a satisfactory description of the data, it is instructive to investigate whether the above conclusions for agreement depend on the chosen model for the remaining association Δ_{jks}. Passing to the enlarged model which leaves Δ_{jks} completely arbitrary - by abandoning the condition (LSA) of linear stratum asymmetry - yields a decrease of the deviance by 27.50 with 40 d.f. which is not significant. Hence there is no need for a model extension concerning asymmetry. But even if the enlarged model is used, the conclusions concerning $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$ are similar to those for the basic model, except that (due to an increase of standard errors up to 13%) three further slopes $\hat{\beta}_{24}$, $\hat{\beta}_{26}$ and $\hat{\beta}_{36}$ are no longer significantly positive.

On the other hand, the submodel of the basic model given by $\boldsymbol{\delta} = \mathbf{0}$, i.e. assuming the (saturated) asymmetric part Δ_{jks} to be the same for all strata s, increases the deviance by 19.13 with 10 d.f., leading to a rejection of the submodel ($p = 3.9\%$). And passing even further to zero asymmetry, i.e. $\boldsymbol{\delta}_0 = \boldsymbol{\delta} = \mathbf{0}$, yields a totally inacceptable fit. Finally, the submodel of the basic model postulating an *additive* instead of an *ordered additive* symmetric structure corresponding to (AS) for both parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$ in (17) yields an extremely poor fit too.

Cytological classification (j)		Pathological classification (k) III DL (1)	III DM (2)	IV AS (3)	IV AC (4)	IV B (5)
III DM	(2)	0.46 ± 0.78				
IV AS	(3)	2.14 * ± 1.03	0.04 ± 0.65			
IV AC	(4)	4.35 * ± 0.96	2.26 * ± 0.60	1.29 * ± 0.50		
IV B	(5)	7.50 * ± 1.21	5.40 * ± 0.97	4.44 * ± 0.91	2.00 * ± 0.57	
V	(6)	8.52 * ± 1.09	6.42 * ± 0.82	5.46 * ± 0.75	3.02 * ± 0.30	1.31 * ± 0.50
III DM	(2)	0.30 ± 0.23				
IV AS	(3)	0.77 * ± 0.33	0.51 * ± 0.18			
IV AC	(4)	0.65 * ± 0.32	0.38 * ± 0.20	0.16 ± 0.14		
IV B	(5)	0.59 ± 0.43	0.32 ± 0.35	0.09 ± 0.32	0.04 ± 0.24	
V	(6)	0.94 * ± 0.38	0.67 * ± 0.28	0.44 * ± 0.25	0.39 * ± 0.13	0.93 * ± 0.28

Table 7-1 (application 1): Estimated baselines $\hat{\beta}_{0\,jk}$ (top) resp. slopes $\hat{\beta}_{jk}$ (bottom) of agreement with standard errors (±S.E.) for the basic model (* indicates a *significantly positive* entry at the 5% level).

Application 2: Social mobility in Britain

Using the *complete* table 1-3 (i.e. including the main diagonal) we find that the order additive symmetry model (AOS) with saturated asymmetry provides an acceptable fit with a deviance $D = 24.38$ having 15 d.f. ($p = 5.9\%$) and no extreme residuals. The estimates and their standard errors of the parameters in $(\mathrm{AOS})^\circ$ are given in table 7-2. Note that only four parameters are not significantly positive on the 5% level.

Passing to the submodel with zero aymmetry reduces the deviance by 21.25 with 21 d.f. which is not significant ($p = 44\%$) but yields two larger residuals. The estimates $\hat{\psi}_0$, $\hat{\tau}_j^\circ$ and $\hat{\nu}_j^\circ$ in the submodel do not differ much (at most by 50% of their standard errors) from the previous ones and their standard errors are only slightly (at most 5%) smaller than before. Hence, the choice of the model for the asymmetrical part $\boldsymbol{\Delta}$ is not crucial here concerning conclusions for the symmetrical part, but this is not generally the case (cf. application 3). Reducing on the other hand the ordered additive symmetry model to additive symmetry (AS) gives a very poor fit, even under saturated asymmetry.

ψ_0	$\overset{\circ}{\tau}_2$	$\overset{\circ}{\tau}_3$	$\overset{\circ}{\tau}_4$	$\overset{\circ}{\tau}_5$	$\overset{\circ}{\tau}_6$	$\overset{\circ}{\tau}_7$
6.904* ± 0.322	1.769* ± 0.338	2.042* ± 0.282	0.743* ± 0.210	− 0.297 ± 0.278	1.600* ± 0.276	0.194 ± 0.213
	$\overset{\circ}{\nu}_2$	$\overset{\circ}{\nu}_3$	$\overset{\circ}{\nu}_4$	$\overset{\circ}{\nu}_5$	$\overset{\circ}{\nu}_6$	$\overset{\circ}{\nu}_7$
	0.603 ± 0.479	1.887* ± 0.320	0.913* ± 0.334	0.017 ± 0.286	1.273* ± 0.171	0.326* ± 0.181

Table 7-2 (application 2): Estimates with standard errors for the order additive symmetry model (OAS)° and saturated asymmetry (* indicates a significant *positive* entry at the 5%-level).

Application 3: Social mobility in the United States

Yamaguchi (1990) focused on the asymmetric (or *skew-symmetric* in his terms) part of the association and fitted several models (including non-log linear models) which restrict only the asymmetric part of association, leaving the symmetrical part completely arbitrary. Being primarily concerned with agreement here, we are using models for the symmetric part $\boldsymbol{\psi}$ only, thus allowing for saturated asymmetry. Fitting the order additive symmetry model (AOS) with saturated asymmetry yields a deviance $D = 7.30$ with 3 d.f. ($p = 6.3\%$) and no extreme residuals. In view of the rather large sample size this indicates a satisfactory fit, which might possibly be improved by partitioning the population into more homogeneous strata. The estimates and their standard errors for the parameters in (AOS)° are given in table 7-3. The only negative estimate does not differ significantly from zero, all other estimates being significantly *positive*. In contrast to the British mobility table, the restriction of the asymmetric part $\boldsymbol{\Delta}$ to *any* of the discussed models extremely deteriorate the above fit ($p < 0.3\%$) thus clearly demonstrating the need for *saturated* asymmetry here. And passing on the other hand to submodels of (OAS) for $\boldsymbol{\psi}$ produces unacceptable fits too.

ψ_0	$\overset{\circ}{\tau}_2$	$\overset{\circ}{\tau}_3$	$\overset{\circ}{\tau}_4$	$\overset{\circ}{\nu}_2$	$\overset{\circ}{\nu}_3$	$\overset{\circ}{\nu}_4$
4.112 ± 0.124	0.730 ± 0.081	0.774 ± 0.083	− 0.138 ± 0.182	0.953 ± 0.099	0.399 ± 0.075	2.076 ± 0.118

Table 7-3 (application 3): Estimates with standard errors for the order additive symmetry model (OAS)° and saturated asymmetry.

8 Concluding remarks

Starting with a single quadratic contingency table $\mathbf{Y}$ arising from cross-classification according to factors A and B, we discussed several models for the association specified by separate models for the symmetrical resp. asymmetrical part of the association and *unrestricted* marginal distributions of A and B (to cover all major sampling schemes). The symmetrical part $\boldsymbol{\psi}$ provides a concept of *agreement* which is independent of the margins in the sense that $\boldsymbol{\psi}$ is given by the family of odds ratios for all 2×2 subtables along the main diagonal obtained from restricting (or conditioning) A and B to two categories. Hence, if agreement (i.e. symmetric association) is of primary interest, we advocate the use of semi-parametric models for the expected table $\boldsymbol{\mu}$, which only restrict the symmetric part $\boldsymbol{\psi}$, thus leaving the remaining structure completely arbitrary. Among the models for $\boldsymbol{\psi}$, we distinguished among those suitable for *nominal* or *ordinal* responses and those depending on a given *distance* between the response categories. Particulary for ordinal responses, a rather general *order additive symmetry model* has been introduced and sucessfully applied to social mobility tables a medical data set. Moreover, all models for $\boldsymbol{\psi}$ and their parameters are easy to interpret in terms of odds ratios.

Concerning models for the asymmetric part, an argument revealed, that only the two extreme models, namely *zero* and *saturated asymmetry*, are suitable for *nominal* respones. For ordinal responses however, any of the previous models for symmetry gives rise to a corresponding model for asymmetry. Unfortunately, these models depend on the chosen representation $\boldsymbol{\Delta}$ describing asymmetry and differ (except in trivial cases) from the corresponding models based on other representations, e.g. the *local* family $\boldsymbol{\Delta}^*$ derived from local odds ratios. The lack of non-trivial *canonical* models for asymmetry (even for ordinal scales) strongly suggests the use of the *saturated* asymmetry model, unless the asymmetric structure itself is on focus.

The models for symmetry and asymmetry for a *single* table extend in a natural way to a *finite set* of tables arising from stratification according to an additional observed vector of covariables by applying the above models separately to each stratum and allowing common parameters across strata. Again, with focus on agreement, the asymmetric association as well as the margins may be left completely arbitrary within each strata, even if this results in a large number of nuisance parameters as in our medical example.

We restricted our attention to log-linear models which provide a comfortable framework from a theoretical point of view and are easily fitted to observed data using standard statistical software. However, in situations where log-linear are not found flexible enough further models for symmetry and asymmetry may be designed in similar ways.

References

Agresti, A. (1990). *Categorical data analysis.* New York, Wiley.

Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* The MIT Press, Cambridge, Massachusettes, and London, England.

Cox, D.R. & Snell, E.J. (1989). *Analysis of Binary Data.* 2nd edn. Chapman and Hall, London, New York.

Featherman, D.L. and Hauser, R.M. (1978). *Opportunity and Change.* Academic Press, New York.

Francis, B., Green, M., Payne, C. (Eds.) (1993). *The GLIM system. Release 4 Manual.* Clarendon Press, Oxford.

Glass, D.V. (ed.) (1954). *Social Mobility in Britain.* Glencoe, Ill., Free Press.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Ass. 74*, 537-552.

Haberman, S.J. (1974). *The Analysis of Frequency Data.* The University of Chicago Press, Chicago and London.

McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*, 2nd edn. Chapman and Hall, London, New York.

Osius, G. (1993). *Separating agreement from association in log-linear models for square contingency tables with applications.* Mathematik-Arbeitspapiere, Preprint Nr. 43, Universität Bremen.

Papanicolaou, G.N and Traut, H.F. (1943). *Diagnosis of uterine cancer by the vaginal smear.* The Commonwealth Fund, New York..

Plackett, R.L. (1974). *The analysis of categorical data.* Griffin, London.

Sobel, M.E. (1988). Some Models for the Multiway Contingency Table with a One-to-One Correspondence among Categories. *Sociological Methodology 18*, 165-192.

Sobel, M.E., Hout, M. and Duncan, O.D. (1985). Exchange, Structure, and Symmetry in Occupational Mobility. *American Journal of Sociology 91*, 359-372.

Tanner, M.A. and Young, M.A. (1985). Modeling ordinal scale disagreement. *Psych. Bull. 98*, 408-415.

Velema, J.P., Blettner, M., Restrepo, M. and Muños, N. (1991). The evaluation of agreement by means of log-linear models: proxy interviews on reproductive history among floriculture workers in Columbia. *Epidemiology 2*, 107-115.

Yamaguchi, K. (1990). Some Models for the Analysis of Asymmetric Association in Square Contingency Tables with Ordered Categories. *Sociological Methodology 20*, 181-212.