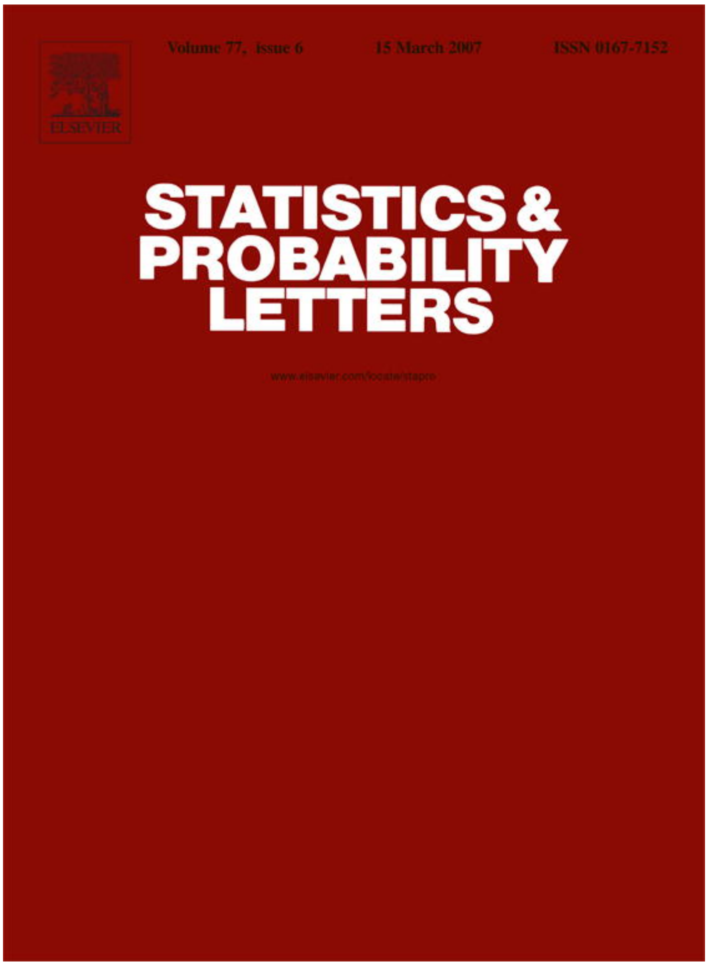
Volume 77, issue 6
15 March 2007
ISSN 0167-7152
ELSEVIER
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Statistics & Probability Letters 77 (2007) 598–603
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

A class of ordinal quasi-symmetry models for square contingency tables

Maria Kateri[a], Alan Agresti[b,*]

[a]Department of Statistics and Insurance Science, University of Piraeus, Greece
[b]Department of Statistics, University of Florida, USA

Received 11 July 2006; accepted 20 September 2006
Available online 18 October 2006

Abstract

Kateri and Papaioannou [1997. Asymmetry models for contingency tables. J. Amer. Statist. Assoc. 92, 1124–1131] proved that, under certain conditions, quasi-symmetry is the closest model to symmetry. A simpler ordinal quasi-symmetry model is the closest to symmetry, under a weaker condition of unequal marginal mean scores. It is a special case of a class of ordinal models based on *f*-divergence.

Keywords: Diagonals-parameter symmetry; Square contingency table; Symmetry; Marginal homogeneity; *f*-Divergence; Ordered categorical data

1. Introduction

Consider a square contingency table with row and column classification variables X and Y. Such tables are common with repeated measurement of a categorical variable, such as in panel studies and in the study of social mobility. The quasi-symmetry (QS) model is a key model for such tables that often fits well and has connections with other standard models for repeated categorical data, such as the Bradley–Terry model and the Rasch model (cf. Agresti, 2002). Kateri and Papaioannou (1997) proved that, under certain conditions pertaining to allowing marginal inhomogeneity, the QS model is the closest to complete symmetry in terms of the Kullback–Leibler distance. They introduced a generalized QS model, that under the same conditions is the closest model to complete symmetry in terms of the *f*-divergence (Csiszàr, 1963).

For ordinal classifications, Agresti (1983) introduced a special case of the QS model, which we refer to as the ordinal quasi-symmetry (OQS) model. It has only one more parameter than the complete symmetry model and simpler interpretation than the QS model. In this paper, we prove that OQS is the closest model to complete symmetry under certain conditions pertaining to allowing unequal marginal mean scores. We also introduce a generalized class of OQS models based on the *f*-divergence and discuss useful special cases (Section 2). Section 3 discusses the interpretation of these models. An alternative parameterization for the OQS model simplifies its

*Corresponding author.
E-mail address: aa@stat.ufl.edu (A. Agresti).

0167-7152/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2006.09.015

interpretation and the estimation of the parameter pertaining to divergence from symmetry. The final section uses a classic example to illustrate the ordinary OQS model and the OQS model that minimizes a Pearsonian distance.

2. A generalized ordinal quasi-symmetry model

Let $\boldsymbol{\pi} = (\pi_{ij})$ denote probabilities for a $r \times r$ contingency table, where π_{ij} is the probability that an observation falls in its (i,j) cell. Agresti (1983) proposed the model

$$\pi_{ij} = \pi_{ji}\delta^{i-j}, \quad i \geqslant j. \tag{1}$$

Under this model, the log odds that an observation falls a certain distance above the main diagonal instead of the same distance below it depends linearly on the distance. More generally, the distances in the above model can be expressed in terms of a set of known scores $u_1 \leqslant u_2 \ldots \leqslant u_r$ (with $u_1 < u_r$) instead of the category indices. Such models are special cases of Goodman's (1979) *diagonals-parameter symmetry* models, which replace δ^{i-j} in (1) by δ_{i-j}.

Agresti (1983) showed that model (1) fits well when the two categorical variables are discretized variables having an underlying bivariate normal distribution. Agresti (1993) showed that the maximum likelihood (ML) estimate of $\log(\delta)$ for this model is identical to the conditional ML estimate of the item parameter for an ordinal extension of a Rasch-type model using adjacent-category logits. Model (1) can be expressed in loglinear form as

$$\log(\pi_{ij}) = \lambda + \lambda_i^X + \lambda_j^X + \beta u_j + \lambda_{ij}^{XY}, \quad i,j = 1, \ldots, r, \tag{2}$$

with $\{\lambda_{ij}^{XY} = \lambda_{ji}^{XY}\}$. When the $\{u_i\}$ scores are equally spaced, model (2) is equivalent to (1), with $\beta = -\log(\delta)$ when $\{u_i = i\}$. Model (2) is the OQS model. The special case $\beta = 0$ is the *complete symmetry model*, denoted S, for which all $\pi_{ij} = \pi_{ji}$.

Model (2) is the special case of the ordinary QS model in which the column main effect parameters satisfy $\lambda_j^Y = \lambda_j^X + \beta u_j$. Since the OQS model is invariant under linear transformations of the scores, without loss of generality we take

$$\sum_i u_i = 0 \quad \text{and} \quad \sum_i u_i^2 = 1. \tag{3}$$

These constraints are useful for the interpretation of β, as shown in the next section, while the first of them is compatible with the sum-to-zero constraints often taken for the main effect parameters of the loglinear model.

Kateri and Papaioannou (1997) proved that, under the constraint that $\{\pi_{ij} + \pi_{ji},\ i>j\}$ and $\{\pi_{i\cdot}\}$ (and hence $\{\pi_{\cdot i}\}$) are given, QS is the closest model to S in terms of the Kullback–Leibler distance. The sample values of these measures are the sufficient statistics for the QS model. For the OQS model, the sufficient statistics are the sample values of $\{\pi_{ij} + \pi_{ji},\ i>j\}$ and the marginal mean $\sum_{i=1}^r u_i\pi_{i\cdot}$ (or $\sum_{i=1}^r u_i\pi_{\cdot i}$). Analogous to the result in Kateri and Papaioannou (1997), it follows from Theorem 2.1 below that the OQS model is the closest model to S in terms of the Kullback–Leibler distance, given $\{\pi_{ij} + \pi_{ji},\ i>j\}$ and $\sum_{i=1}^r u_i\pi_{i\cdot}$ (and hence $\sum_{i=1}^r u_i\pi_{\cdot i}$).

Replacing the Kullback–Leibler distance by the more general f-divergence, we next introduce a class of generalized OQS models. If $\boldsymbol{p} = (p_{ij})$ and $\boldsymbol{q} = (q_{ij})$ are two discrete finite bivariate probability distributions, then the f-divergence between $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined by

$$I^C(\boldsymbol{p} : \boldsymbol{q}) = \sum_{i,j} q_{ij} f\left(\frac{p_{ij}}{q_{ij}}\right),$$

where f is a real-valued convex function on $[0, +\infty)$ with $f(1) = 0$, $f(x) \to 0$ as $x \to 0$, $0f(0/0) = 0$ and $0f(x/0) = xf_\infty$ with $f_\infty = \lim_{x\to\infty}[f(x)/x]$.

Theorem 2.1. *Let f be a twice differentiable and strictly convex function, and let $F(x) = f'(x)$ for all x. For probabilities $\boldsymbol{\pi}$ for a $r \times r$ contingency table, let $\{\pi_{ij}^S = (\pi_{ij} + \pi_{ji})/2\}$ be related values satisfying symmetry. Then, in the class of models with given sums $\{\pi_{ij} + \pi_{ji}\}$ and given marginal mean $\sum_{i=1}^r u_i\pi_{i\cdot}$ (or $\sum_{i=1}^r u_i\pi_{\cdot i}$), the model*

$$\pi_{ij} = \pi_{ij}^S F^{-1}(\alpha u_i + \gamma_{ij}), \quad i,j = 1, \ldots, r, \tag{4}$$

with $\gamma_{ij} = \gamma_{ji}$, is the closest to model S in terms of the f-divergence.

Appendix A gives a proof. We denote model (4) as OQS[f]. Its residual degrees of freedom (df) for testing lack of fit equal $r(r-1)/2-1$. This is just one less than df for the S model, which is the special case of OQS[f] with $\alpha=0$. Model OQS[f] is itself a special case of the QS[f] model of Kateri and Papaioannou (1997), which replaces αu_i in (4) with α_i, $i=1,\ldots,r$. For the probabilities in (4) to be positive, we need $F^{-1}(\alpha u_i+\gamma_{ij})>0$ for all i,j. Depending on the form of f, this can require a constraint on α.

From (4) and the constraints $\pi_{ij}^S=\pi_{ji}^S$ and $\gamma_{ij}=\gamma_{ji}$,

$$\pi_{ij}+\pi_{ji}=\pi_{ij}^S[F^{-1}(\alpha u_i+\gamma_{ij})+F^{-1}(\alpha u_j+\gamma_{ij})].$$

Thus,

$$F^{-1}(\alpha u_i+\gamma_{ij})+F^{-1}(\alpha u_j+\gamma_{ij})=2,\quad i,j=1,\ldots,r, \tag{5}$$

so γ_{ij} in model (4) is redundant, given α. This reflects OQS[f] having only a single extra parameter compared to S (namely, α) and one less residual df. The form of OQS[f] shows it provides a divergence from model S. The next section explains the role of the u-scores and α in describing this divergence.

We next consider some particular members of the OQS[f] class (4):

1. For $f(x)=x\log(x)$, $x>0$, the f-divergence I^C is the Kullback–Leibler distance. Then, $F^{-1}(y)=e^{y-1}$, and model (4) becomes

$$\pi_{ij}=\pi_{ij}^S\frac{2e^{\alpha u_i}}{e^{\alpha u_i}+e^{\alpha u_j}},\quad i,j=1,\ldots,r. \tag{6}$$

This is an equivalent expression for the OQS model (2), with $\alpha=-\beta$. Expression (6) reveals the "divergence from symmetry" nature of the OQS model in terms of α.

2. For $f(x)=(1-x)^2$, the f-divergence I^C is the Pearsonian distance. Model (4) then simplifies to

$$\pi_{ij}=\pi_{ij}^S(1+a(u_i-u_j)),\quad i,j=1,\ldots,r, \tag{7}$$

with $a=\alpha/4$, where the positivity of $\{\pi_{ij}\}$ requires $|a|<(u_r-u_1)^{-1}$. We refer to this model as the *Pearsonian OQS model*.

3. When I^C is the Cressie–Read power divergence (Read and Cressie, 1988), f depends on a real-valued parameter λ and equals $f_\lambda(x)=(1/\lambda(\lambda+1))(x^{\lambda+1}-x)$, $x>0$. Model (4) then becomes

$$\pi_{ij}=\pi_{ij}^S\{1+\lambda(\alpha u_i+\gamma_{ij})\}^{1/\lambda},\quad i,j=1,\ldots,r. \tag{8}$$

When $\lambda=0$ or -1, $f_0(x)=\lim_{\lambda\to 0}[f_\lambda(x)]$ and $f_{-1}(x)=\lim_{\lambda\to -1}[f_\lambda(x)]$. Model (8) reduces to model (6) when $\lambda=0$ and to (7) when $\lambda=1$. We denote models (8), (6) and (7) as OQS_λ, OQS_0, and OQS_1.

Analogous results occur in the literature for association models and correlation models (Goodman, 1985), in terms of distance from the independence model. Gilula et al. (1988) showed that association models are closest to independence in terms of the Kullback–Leibler distance, while correlation models are closest in terms of Pearsonian distance. In comparing association and correlation models, Goodman (1985, p. 32) pointed out that the parametric scores in correlation models must satisfy certain constraints to ensure the positivity of the cell probabilities, but this was not true for the corresponding scores in association models. Analogously, in our context, OQS_1 requires a constraint, whereas OQS_0 does not.

3. Parameter interpretation and a basic property

The parameter α in the OQS[f] model (4) has straightforward interpretations in terms of departures from complete symmetry, which is the special case $\alpha=0$. For u-scores satisfying (3),

$$\alpha=\frac{1}{r}\sum_{i,j}u_i\left(F\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right)-F\left(\frac{\pi_{ji}}{\pi_{ij}^S}\right)\right). \tag{9}$$

For the standard OQS_0 model, (9) becomes

$$\alpha = \frac{1}{r}\sum_{i,j} u_i \log\left(\frac{\pi_{ij}}{\pi_{ji}}\right)$$

while for OQS_1 it reduces to

$$a = \frac{1}{r}\sum_{i,j} u_i \frac{(\pi_{ij} - \pi_{ji})}{2\pi_{ij}^S}.$$

These relations and the model expressions for QS[f] and its special cases show that $\alpha > 0$ indicates that the lower triangle of the table is more probable than the upper. This inequality depends linearly on the known scores in the F-scale, since

$$F\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right) - F\left(\frac{\pi_{ji}}{\pi_{ij}^S}\right) = \alpha(u_i - u_j).$$

The magnitude of the difference is determined by the absolute value of α. For the OQS_0 model, the F-scale is the log-scale and this relation becomes $\pi_{ij}/\pi_{ji} = e^{\alpha(u_i - u_j)}$. For the OQS_1 model, it is

$$\frac{\pi_{ij}}{\pi_{ji}} = \frac{1 + a(u_i - u_j)}{1 - a(u_i - u_j)}.$$

A standard property of the QS and OQS models relating complete symmetry and marginal homogeneity applies also for the OQS[f] model, namely

Complete symmetry = Marginal homogeneity + OQS[f].

Complete symmetry is the special case of OQS[f] in which $\alpha = 0$, so one can test marginal homogeneity (with df = 1) by comparing the fits of the S and OQS[f] models. Departures from S expressed by α also describe marginal inhomogeneity.

4. Example

Goodman (1979), Agresti (1983), and Kateri and Papaioannou (1997) analyzed Table 1, which has become a classic data set for illustrating methods for square ordinal tables. The table cross classifies 7477 women according to the unaided distance vision level of their right and left eyes. Agresti (1983) applied the ordinary OQS model. We shall additionally apply the OQS_1 model and compare results.

Table 1
Cross classification of 7477 women by unaided distance vision of right and left eyes

	Left eye grade			
Right eye grade	Best	Second	Third	Worst
Best	1520 –	266 (263.37[a]/263.35[b])	124 (133.35/133.37)	66 (59.12/59.17)
Second	234 (236.63/236.65)	1512 –	432 (418.23/418.20)	78 (88.53/88.54)
Third	117 (107.65/107.63)	362 (375.77/375.80)	1772 –	205 (202.27/202.25)
Worst	36 (42.88/42.83)	82 (71.47/71.46)	179 (181.73/181.74)	492 –

Parenthesized values are ML estimates of the expected frequencies under the models OQS_0 and OQS_1.
[a]OQS_0.
[b]OQS_1.

The likelihood equations for estimating parameters for these two models are relatively simple. Let p_{ij} denote the sample proportion for cell (i,j). For either model, $\hat{\pi}_{ij}^S = (p_{ij} + p_{ji})/2$, and so we need only consider α. For the OQS_1 model, the likelihood equation is

$$\sum_{i,j} \frac{p_{ij}}{\hat{\pi}_{ij}} \hat{\pi}_{ij}^S (u_i - u_j) = 0.$$

For the OQS_0 model, the likelihood equation is

$$\sum_{i,j} (p_{ij} - \hat{\pi}_{ij}) \frac{\hat{\pi}_{ji}}{\hat{\pi}_{ij}^S} (u_i - u_j) = 0.$$

With equidistant scores for successive categories that satisfy (3), the likelihood-ratio statistic for testing the goodness of fit of model OQS_0 applied to Table 1 equals $G_0^2 = 7.280$ with df $= 5$, while $\hat{\alpha} = -0.239$ (s.e. $= 0.069$), which equals $-\hat{\beta}$ in expression (2). (Agresti (1983) reported a different value, $\hat{\beta} = 0.054$, because he used a different set of equidistant u-scores, but the fits are identical.) For the OQS_1 model, $G_1^2 = 7.271$, and $\hat{a} = -0.119$ (s.e. $= 0.034$). The ML fits of both these models are shown in parentheses in Table 1. In practical terms, the fits are nearly equivalent, differing only in the second decimal place. The P-values for the G^2 goodness-of-fit statistics are both 0.201 to three decimal places.

The negative sign of $\hat{\alpha}$ (or $\hat{a}$) indicates that the lower triangle of Table 1 is less probable than the upper. That is, the grade distribution is lower for the left eye. According to the OQS models' structure, the odds of an observation falling a certain distance under the main diagonal of the table (instead of the same distance above it) are estimated as $\hat{\pi}_{ij}/\hat{\pi}_{ji} = \mathrm{e}^{-0.239(u_i - u_j)}$ for the OQS_0 model and $\hat{\pi}_{ij}/\hat{\pi}_{ji} = (1 - 0.119(u_i - u_j))(1 + 0.119(u_i - u_j))^{-1}$ for the OQS_1 model, $i > j$.

The OQS_0 and OQS_1 models do not always exhibit such similarity. In practice, often they do not when $\hat{a}$ in model OQS_1 takes value at the boundary in order for probability estimates to fall between 0 and 1 (that is, $|\hat{a}| = (u_r - u_1)^{-1}$). Such a case is Table 10.5 in Agresti (2002, p. 421), for which $G_0^2 = 2.1$ and $G_1^2 = 74.7$, with df $= 5$. Fitting these models to various data sets, we have observed that boundary solutions for model OQS_1 often, but not always, occurred if $p_{1r} = 0$ or $p_{r1} = 0$. We also observed that OQS_1 need not provide a poorer fit than OQS_0, and their G^2 values can also differ substantially even in non-boundary cases.

Appendix A. Proof of Theorem 2.1

This is a constraint minimization problem, solved by the method of Lagrange multipliers. The function to be minimized is $I^C(\boldsymbol{\pi} : \boldsymbol{\pi}^S)$, subject to the constraints $\pi_{ij} + \pi_{ji} = 2\pi_{ij}^S$ $(i,j = 1, \ldots, r)$ and $\sum_{i=1}^r u_i \pi_{i\cdot} = v$, for some constant v. The Lagrange function is

$$L(\{\pi_{ij}\}) = I^C(\boldsymbol{\pi} : \boldsymbol{\pi}^S) + b\left(\sum_{i=1}^r u_i \pi_{i\cdot} - v\right) + \sum_{i,j} c_{ij}(\pi_{ij} + \pi_{ji} - 2\pi_{ij}^S),$$

where b and $\{c_{ij}\}$ are the Lagrange multipliers. Setting $\partial L / \partial \pi_{ij} = 0$, we obtain

$$f'\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right) + b u_i + c_{ij} + c_{ji} = 0.$$

With $\alpha = -b$ and $\gamma_{ij} = -(c_{ij} + c_{ji})$, for which $\gamma_{ij} = \gamma_{ji}$, and using $F = f'$, we obtain

$$F\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right) = \alpha u_i + \gamma_{ij}. \tag{10}$$

To solve (10) with respect to π_{ij}, the existence of F^{-1} is required. This is ensured by the strict monotonicity of F, because $F'(x) = f''(x) > 0$ for all x, since f is strictly convex. Thus, (10) leads to (4). Finally, L has a minimum at the solution, since the Hessian matrix is positive definite ($f'' > 0$).

References

Agresti, A., 1983. A simple diagonals-parameters symmetry and quasi-symmetry model. Statist. Probab. Lett. 1, 313–316.

Agresti, A., 1993. Computing conditional maximum likelihood estimates for generalized Rasch models using simple loglinear models with diagonals parameters. Scand. J. Statist. 20, 63–71.

Agresti, A., 2002. Categorical Data Analysis, second ed. Wiley, New York.

Csiszàr, I., 1963. Eine informationstheoretische Ungleichung und ihre Anwendungen auf den Beweis der Ergozitat von Markoffschen Ketten. A Mayar Tudomanyos Academia Mathematikai Kutato Intezelent Kozlemezyri 8, 85–108.

Gilula, Z., Krieger, A.M., Ritov, Y., 1988. Ordinal association in contingency tables: some interpretive aspects. J. Amer. Statist. Assoc. 83, 540–545.

Goodman, L.A., 1979. Multiplicative models for square contingency tables with ordered categories. Biometrika 66, 413–418.

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

Kateri, M., Papaioannou, T., 1997. Asymmetry models for contingency tables. J. Amer. Statist. Assoc. 92, 1124–1131.

Read, T.R.C., Cressie, N.A.C., 1988. Goodness-of-fit Statistics for Discrete Multivariate Data. Springer, New York.