Statistics & Probability Letters 1 (1983) 313–316
North-Holland October 1983

A Simple Diagonals-Parameter Symmetry and Quasi-Symmetry Model

Alan Agresti
Department of Statistics, University of Florida, Gainesville, FL, U.S.A.

Received March 1983

Abstract. Goodman (1979) proposed a class of diagonals-parameter symmetry models for square contingency tables with ordered categories, A simple version of that model is considered in which the log odds parameters have a linear pattern. The model is also a simple quasi-symmetry model. It fits well when there is an underlying bivariate normal distribution.

Keywords. Conditional symmetry, ordered categorical data, quasi-symmetry, square contingency table, underlying normality, Bradley–Terry paired comparisons.

1. Introduction

Suppose that an $r \times r$ contingency table has the same ordered categories in the row classification as in the column classification. Let π_{ij} denote the probability that an observation falls in the cell in row i and column j. For this setting, Goodman (1972, 1979) discussed the diagonal-parameter symmetry (DPS) model,

$$\pi_{ij} = \pi_{ji}\delta_{i-j}, \quad i > j. \tag{1.1}$$

The parameter δ_k represents the odds that an observation falls in a cell (i, j), satisfying $i - j = k$, instead of in the cell (j, i), $k = 1, \ldots, r-1$. In this model it is assumed that these odds depend only on the distance between the diagonal containing the cell and the main diagonal.

In the DPS model no assumption is made about the form of the odds parameters. In certain situations a more structured form for the $\{\delta_k\}$, giving a more parsimonious model, may better describe predicted departures from the symmetry case $\{\delta_k = 1\}$. The conditional symmetry model

$$\pi_{ij} = \pi_{ji}\delta, \quad i > j,$$

is one simplification of the DPS model that has received some attention (Bishop et al., 1975, pp. 285–286; McCullagh, 1978). This model implies that for $i > j$,

$$P(X = i, Y = j | X > Y) = P(X = j, Y = i | X < Y),$$

where (X, Y) is selected at random according to the $\{\pi_{ij}\}$ distribution. In many applications it may be more realistic to expect monotonicity in the $\{\delta_k\}$. In this article we consider a simple version of model (1.1) in which the $\{\log \delta_k\}$ have a linear pattern.

2. Linear diagonals-parameter model

Consider the model

$$\pi_{ij} = \pi_{ji}\delta^{i-j}, \quad i \geqslant j, \tag{2.1}$$

which is the special case of the DPS model in which $\delta_k = \delta^k$, $k = 1, \ldots, r-1$. Here, the log odds that an observation is a certain distance above the main diagonal, instead of the same distance below it, is assumed to depend linearly on the distance. Like the conditional symmetry model, this model has only one more parameter than the symmetry model, which is the special case $\delta = 1$. Both models imply a stochastic ordering of the marginal distributions. Though both models are so simple that they are inadequate for describing most square tables, there is an important setting in which model (2.1) has theoretical justification. For ordinal matched-pairs data, it is often reasonable to assume an underlying continuous distribution that is approximately bivariate normal with equal marginal standard deviations. The joint normal density satisfies

$$\frac{f(x,y)}{f(y,x)} = \delta^{x-y} \quad \text{for } x \geqslant y,$$

the form given by model (2.1) for the discrete case.

Model (2.1) is a loglinear model, and it can be expressed in terms of expected frequencies $\{m_{ij}\}$ as

$$\log m_{ij} = \mu + \lambda_i + \lambda_j + \beta(i-j) + \lambda_{ij}, \tag{2.2}$$

where, for all i and j,

$$\lambda_{ij} = \lambda_{ji}, \quad \sum_i \lambda_i = \sum_i \lambda_{ij} = 0, \quad 2\beta = \log \delta.$$

For full multinomial or independent Poisson sampling, maximum likelihood estimates for model (2.2) can be easily obtained using the Newton–Raphson method. The sample frequencies $\{n_{ij}\}$ and expected frequency estimates satisfy, for all i and j,

$$\hat{m}_{ii} = n_{ii}, \quad \hat{m}_{ij} + \hat{m}_{ji} = n_{ij} + n_{ji},$$

$$\sum i\hat{m}_{i+} = \sum in_{i+}, \quad \sum j\hat{m}_{+j} = \sum jn_{+j}.$$

Notice that equal-interval scoring of categories implies that the observed and fitted marginal distributions have the same means. More generally, the integer scores in this model can be replaced by fixed by monotonic scores. In either case, the Pearson χ^2 or likelihood-ratio G^2 statistics have $\mathrm{d}f = \frac{1}{2}(r+1)(r-2)$ for testing goodness of fit, compared to $\mathrm{d}f = \frac{1}{2}(r-1)(r-2)$ for the general DPS model. The estimate of β in model (2.2) can also be obtained by fitting the no-intercept logit model $\log(m_{ij}/m_{ji}) = 2\beta(i-j)$, regarding the $\{n_{ij}, i>j\}$ as $\frac{1}{2}r(r-1)$ independent binomial random variables with sample sizes $\{n_{ij} + n_{ji}\}$.

Model (2.2) is a special case of the quasi-symmetry model for square tables,

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij},$$

where $\sum\lambda_i^X = \sum\lambda_j^Y = \sum_i\lambda_{ij} = 0$ and $\lambda_{ij} = \lambda_{ji}$. Like the general DPS model, this model has $\mathrm{d}f = \frac{1}{2}(r-1)(r-2)$. Breslow (1982) and McCullagh (1982) expressed the quasisymmetry model in Bradley–Terry paired comparison form with the logit model

$$\log(\pi_{ij}/\pi_{ji}) = \beta_i - \beta_j, \tag{2.3}$$

where $\{\beta_i = \lambda_i^X - \lambda_i^Y + c\}$ for any constant c. McCullagh suggested that linear or quadratic patterns should be considered for the $\{\beta_j\}$ in this model when the table has ordered categories. In fact, model (2.3) corresponds to model (2.1) when $\beta_{j+1} - \beta_j$ takes constant value for $j = 1,\ldots,r-1$, that value being $\log\delta$ in model (2.1). The linear DPS model is palindromic invariant, unlike the general quasi-symmetry model which is fully permutation invariant, ignoring the ordering of categories.

3. Examples

The linear DPS model generally gives a good fit when there is an underlying bivariate normal distribution. For example, suppose that the parameters of that distribution satisfy

$$\mu_Y = \mu_X + 0.2, \quad \sigma_X = \sigma_Y = \sigma, \quad \rho = 0.2,$$

and suppose that a 6×6 table is formed using cutpoints for each variable at μ_X, $\mu_X + 0.6\sigma$ and $\mu_X \pm 1.2\sigma$. Cell counts in the pattern of the resulting probabilities would yield $\hat{\delta} = 1.37$ and a Pearson chi-squared statistic, having $\mathrm{d}f = 14$, of only 14 per 100000 observations. By comparison, the corresponding chi-squared value for the conditional symmetry model is 484 per 100000 observations.

Breslow used the quasi-symmetry model (2.3) to analyze Table 1, in which 80 esophageal cancer patients are compared with controls on the number of beverages reported drunk at 'burning hot' temperatures. Taking $\hat{\beta}_1 = 0$, he obtained the estimates $\hat{\beta}_2 = 0.737$, $\hat{\beta}_3 = 1.299$ and $\hat{\beta}_4 = 2.573$, and $\chi^2 = 2.45$ with $\mathrm{d}f = 3$. Essentially as good a fit is given by the simpler linear DPS model, for

Table 1
Distribution of esophageal cancer case-control pairs by number of beverages drunk 'burning hot'. Estimated expected frequencies in parentheses for: (a) linear DPS, (b) quasi-symmetry models

Case	Controls			
number	0	1	2	3
0	31	5	5	0
	–	(5.6[a], 5.5[b])	(3.6, 4.1)	(0.6, 0.4)
1	12	1	0	0
	(11.4, 11.5)	–	(0.3, 0.4)	(0.2, 0.1)
2	14	1	2	1
	(15.4, 14.9)	(0.7, 0.6)	–	(0.7, 0.4)
3	6	1	1	0
	(5.4, 5.6)	(0.8, 0.9)	(1.3, 1.6)	–

which $\chi^2 = 2.41$, based on $df = 5$. The assumed common value of $\{\beta_{j+1} - \beta_j\}$ in the linear DPS model is estimated by $\log \hat{\delta} = 0.723$. Hence, the probability that a case drank k more beverages burning hot than did the control is estimated to be $(2.06)^k$ times the probability that the control drank k more beverages burning hot than did the case. The $\log \hat{\delta}$ value divided by its asymptotic standard error of 0.215 gives strong evidence that the esophageal cancer patients tended to drink more beverages burning hot than did controls. Most cell counts are very small in Table 1, thus one should be cautious in applying any asymptotic approximations. However, it can be pointed out that the estimated expected frequencies given in that table show adequate and similar fits for both models with the linear DPS model having the advantages of parsimony and simple interpretation. The general DPS model fits only slightly better, with $\chi^2 = 1.16$ based on $df = 3$.

McCullagh (1978) and Goodman (1979) illustrated their models using the data in Table 2 on unaided distance vision of 7477 women. For those data, the quasi-symmetry model has $\chi^2 = 7.26$ with $df = 3$, the conditional symmetry model has $\chi^2 = 7.23$ with $df = 5$, and the linear DPS model has $\chi^2 = 7.27$ with $df = 5$. The maximum likelihood estimate of δ in (2.1) is 0.898. Hence, the probability that left eye vision is k grades higher than right eye vision is estimated to be $(0.898)^k$ times the probability than left eye vision is k grades lower than right eye vision. The standard error of $\hat{\beta} = (\frac{1}{2}) \log \hat{\delta} = 0.054$ in expression (2.2) is 0.016, so that the grade distribution is significantly lower for the left eye.

The estimated expected frequencies in Table 2 show that the fit is again essentially the same for the linear DPS and quasi-symmetry models, though the linear DPS model is more parsimonious. Both models give negative residuals in the cells (i, j) with $i - j = 1$ or 3, and positive residuals in the cells with $i - j = 2$. The corresponding residuals

Table 2
Unaided distance vision of women. Estimated expected frequencies in parentheses for: (a) linear DPS, (b) quasi-symmetry models

Right eye grade	Left eye grade			
	best	second	third	worst
best	1520	266	124	66
	–	(263.4[a], 263.4[b])	(133.4, 133.6)	(59.1, 59.0)
second	234	1512	432	78
	(236.6, 236.6)	–	(418.2, 419.0)	(88.5, 88.4)
third	117	362	1772	205
	(107.6, 107.4)	(375.8, 375.0)	–	(202.3, 201.6)
worst	36	82	179	492
	(42.9, 43.0)	(71.5, 71.6)	(181.7, 182.4)	–

above the main diagonal have opposite signs, because of the constraint $\hat{m}_{ij} + \hat{m}_{ji} = n_{ij} + n_{ji}$. Goodman (1979) noted that the pattern in Table 2 is explained by the estimates $\hat{\delta}_1 = 0.86$, $\hat{\delta}_2 = 0.99$, $\hat{\delta}_3 = 0.55$, obtained for the general DPS model. For these data the general model gives a better fit, with $\chi^2 = 0.50$ with $df = 3$. However, it can be argued that the simpler linear DPS and conditional symmetry models are also adequate, considering the large sample size.

Acknowledgment

I would like to thank an anonymous referee for some helpful comments.

References

Bishop, Y., S.E. Fienberg and P.W. Holland (1975), *Discrete Multivariate Analysis* (MIT Press, Cambridge).

Breslow, N. (1982), Covariance adjustment of relative-risk estimates in matched studies, *Biometrics* **38**, 661–672.

Goodman, L.A. (1972), Some multiplicative models for the analysis of cross classified data, *Proc. 6th Berkeley Symp.* **1**, 649–696.

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

McCullagh, P. (1978), A class of parametric models for the analysis of square contingency tables with ordered categories, *Biometrika* **65**, 413–418.

McCullagh, P. (1982), Some applications of quasisymmetry, *Biometrika* **69**, 303–308.