

Generalized log-linear models with random effects, with application to smoothing contingency tables

Brent A Coull¹ and Alan Agresti²

¹Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

²Department of Statistics, University of Florida, Gainesville, FL, USA

Abstract: We define a class of generalized log-linear models with random effects. For a vector of Poisson or multinomial means $\boldsymbol{\mu}$ and matrices of constants \mathbf{C} and \mathbf{A} , the model has the form $\mathbf{C} \log \mathbf{A}\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$, where $\boldsymbol{\beta}$ are fixed effects and \mathbf{u} are random effects. The model contains most standard models currently used for categorical data analysis. We suggest some new models that are special cases of this model and are useful for applications such as smoothing large contingency tables and modeling heterogeneity in odds ratios. We present examples of its use for such applications. In many cases, maximum likelihood model fitting can be handled with existing methods and software. We outline extensions of model fitting methods for other cases. We also summarize several challenges for future research, such as fitting the model in its most general form and deriving properties of estimates used in smoothing contingency tables.

Key words: association model; generalized linear mixed model; marginal model; mixture model; multinomial; logit model; ordinal data; overdispersion; Poisson

Data and program available from <http://stat.uibk.ac.at/SMIJ>

Received July 2002; revised April 2003; accepted August 2003

1 The generalized log-linear mixed model

Let $\mathbf{y} = (y_1, \dots, y_I)$ denote a vector of counts that are either (1) independent Poisson variates, with means $\boldsymbol{\mu} = (\mu_1, \dots, \mu_I)$, or (2) conditional on $n = \sum y_i$, multinomial with means $\boldsymbol{\mu} = (n\pi_1, \dots, n\pi_I)$ with $\sum \pi_i = 1$, or (3) sets of independent multinomial probabilities. The ordinary log-linear model has the form

$$\log \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

where \mathbf{X} is a model matrix and $\boldsymbol{\beta}$ is a vector of parameters.

For matrices of constants \mathbf{C} and \mathbf{A} , the *generalized log-linear model*, abbreviated GLLM,

$$\mathbf{C} \log \mathbf{A}\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \tag{1.1}$$

Address for correspondence: Brent A Coull, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Ave., Boston, MA 02115, USA. E-mail: bcoull@hsph.harvard.edu

has been considered by various authors, perhaps beginning with Grizzle *et al.* (1969). The class describes a much wider variety of models than the ordinary log-linear model, including logit models, ordinal logit models such as using cumulative logits (Williams and Grizzle, 1972), association models using ordinal odds ratios (Goodman, 1979; Dale, 1986), models for marginal distributions of a contingency table and odds ratios in marginal faces of multi-way contingency tables (Haber, 1985a,b), and models that simultaneously describe the association structure and marginal distributions (Lang and Agresti, 1994).

This article extends the generalized log-linear model to include random effects. Let \mathbf{u} denote a vector of random effects having multivariate normal $N(0, \Sigma)$ distribution. Let \mathbf{y} denote counts that, conditional on \mathbf{u} , are independent Poisson, multinomial, or independent multinomial with means $\boldsymbol{\mu}$. We define the *generalized log-linear mixed model*, denoted by GLLMM, to have the form

$$\mathbf{C} \log \mathbf{A}\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad (1.2)$$

where \mathbf{Z} is a model matrix for \mathbf{u} .

This class provides a unifying framework for a wide variety of models, including the GLLMs discussed above. We first mention some special cases. A commonly used GLLMM is the *logistic-normal model*, by which conditionally on certain random effects and indices, response variates are independent binomial variates (e.g., Breslow and Clayton, 1993). Such a model is a special case of (1.2) in which $\boldsymbol{\mu}$ contains the pairs of success and failure probabilities for each binomial, \mathbf{A} is an identity matrix, and each row of \mathbf{C} has a +1 in the column corresponding to a particular success probability and -1 in the column corresponding to the failure probability and 0 everywhere else. With appropriate specification, (1.2) can be a multinomial logit model with random effects. Special cases include random effects models using the cumulative logit (Hedeker and Gibbons, 1994; Tutz and Hennevogel, 1996), continuation-ratio logit (Coull and Agresti, 2000), and adjacent-categories logit and baseline-category logit models (Hartzel *et al.*, 2001a,b). Applications of such models are increasingly widespread.

The general class of models (1.2) has sufficient scope to contain a variety of models that have not yet seen much, if any, use. Our main purposes in posing the GLLMM class are (1) to unify a wide variety of models, including most models currently used in categorical data analysis, (2) to make some tentative suggestions of GLLMMs that may be useful and merit future study, (3) to propose using certain of these models for the application of smoothing contingency tables and modeling heterogeneous odds ratios, and (4) to describe a variety of challenges the GLLMM class holds for future research, including determining properties of the models and developing methods of maximum likelihood (ML) model fitting that apply to the general form of the model. We hope that purposes (2)–(4) can lead to interesting future research.

Section 2 discusses a potential application of the GLLMM, that of smoothing contingency tables either by smoothing cell counts or smoothing summary odds ratios. It makes some tentative proposals of models that generalize well-known existing models and that may be useful for this purpose. Section 3 discusses model fitting by maximum likelihood, focusing mainly on cases in which one can apply existing methods for generalized linear mixed models but also suggesting ways to fit more

complex cases. Section 4 presents a simulation study that compares the two methods of smoothing presented in Section 2, and Section 5 presents two examples. Section 6 discusses other potential applications of GLLMMs as well as further generalizations of the model and future research challenges.

2 Smoothing associations and counts in contingency tables

One potentially useful application of the GLLMM is to the smoothing of contingency tables. The goal may be to smooth counts in a large, sparse table or to smooth odds ratios describing association or interaction structure. In some applications, a particular unsaturated model may provide too severe a smoothing. Nonetheless, it might still be beneficial to smooth by shrinking toward the fit of the model. This provides a way of generating nonzero estimates of cell probabilities in cells that are sampling zeroes but not structural zeroes. It also has the potential of reducing the mean squared error of estimators of cell probabilities and of odds ratio effects. One normally thinks of Bayesian methods for performing shrinkage, but a classical approach using random effects provides an alternative way of improving estimates by ‘borrowing from the whole.’

2.1 Association models with random effects

We first consider smoothing of association structure. We extend association models for contingency tables (Goodman, 1979) to include random effects. Goodman’s association models describe the structure of local odds ratios in a two-way table with ordered classifications. They extend to other types of odds ratios such as global odds ratios (Dale, 1986) and to describing conditional odds ratios in stratified tables (e.g., Clogg, 1982).

Consider first an $I \times J$ table cross-classifying ordinal variables X and Y . Let θ_{ij} denote an odds ratio with first cell indexed by level i of X and j of Y . For instance, this may be a local or global odds ratio with split for X after row i and split for Y after column j . Goodman’s model of uniform association is

$$\log \theta_{ij} = \beta, \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1.$$

Having only a single parameter, this is a strong structure that is often too simplistic.

The uniform association model extends to a GLLMM,

$$\log \theta_{ij} = \beta + u_{ij} \tag{2.1}$$

where $\{u_{ij}\}$ are $N(0, \sigma_u^2)$ random effects. Correlation structure is possible on $\{u_{ij}\}$, although for large tables it is computationally much simpler to assume independence. At first glance, using random effects seems unorthodox here, since the cells are not sampled from a population. Nonetheless, such a model can be useful for smoothing odds ratio estimates without requiring them to perfectly satisfy a given model. The magnitude

of σ_u^2 dictates the amount of smoothing of log odds ratios towards β , with $\sigma_u^2 = 0$ corresponding to the extreme smoothness of uniform association. Although one could specify σ_u , for most applications it is natural to treat this smoothing index as a parameter and estimate it together with the other parameters using maximum likelihood. After doing so, we use the usual approach of predicting the random effects by their posterior mean, given the data. This expectation depends on the model parameters, and we use the empirical Bayes approach of substituting the ML estimates of those parameters.

The special case $\log \theta_{ij} = 0$ for all i and j corresponds to independence of X and Y . Fitting model (2.1) with $\beta = 0$ shrinks the sample log odds ratios toward its fit. For a fixed value of σ_u^2 , the implied shrinkage is analogous to that obtained by smoothing cell counts via maximizing the penalized likelihood

$$L(\pi) = \log \text{likelihood} - \lambda \sum_i \sum_j (\log \theta_{ij})^2 \quad (2.2)$$

where λ is a smoothing parameter that dictates the weight of this penalty relative to the log likelihood. Green (1987) noted that these two models are equivalent only for the linear model for normal responses, in which case $\lambda = \sigma_\varepsilon^2 / \sigma_u^2$, where σ_ε^2 is the residual variance. Simonoff (1983, 1987) considered (2.2) when θ_{ij} are local odds ratios, and discussed selection strategies for the smoothing parameter λ based on mean squared error criteria.

Special cases of (2.1) that provide structure for the random effects may be of interest in some applications. One such model is

$$\log \theta_{ij} = \beta + u_i + v_j$$

with $u_i \sim N(0, \sigma_u^2)$ and $v_j \sim N(0, \sigma_v^2)$. This mimics the Goodman (1979) $R + C$ association model, but treats row and column effects as random rather than fixed. For large tables, comparing estimates of σ_u^2 and σ_v^2 would indicate whether departures from uniform association are greater across rows or across columns.

The model-based approach to smoothing in contingency tables extends naturally to multi-way tables. Consider a three-way table cross-classifying ordinal variables X and Y at several levels of Z . Let $\theta_{ij(k)}$ denote an odds ratio indexed by level i of X and j of Y in stratum k . The extension of model (2.1) to

$$\log \theta_{ij(k)} = \beta_k + u_{ijk} \quad (2.3)$$

shrinks toward uniform association within each stratum, where the level of association may vary across strata. Simpler models with $\beta_k = \beta$ and/or $u_{ijk} = u_{ij}$ allow commonality across strata in the level of association or the pattern of shrinkage.

Model (2.3) is directed primarily toward within-stratum smoothing, as the expected log odds ratio varies across strata but not within strata. In many applications, especially when the strata are a sample of the possible strata (such as hospitals in multi-center clinical trials or schools or geographical regions), it may be relevant to smooth among

strata or to describe variability among strata in the association. This suggests a model such as

$$\log \theta_{ij(k)} = b_k + u_{ijk} \quad (2.4)$$

with $b_k \sim N(\beta, \sigma_b^2)$ and $u_{ijk} \sim N(0, \sigma_u^2)$. The expected log odds ratio is then uniform across and within strata. The special case $\sigma_u^2 = 0$ gives a model where the log odds ratio is uniform within strata and varies among strata according to a $N(\beta, \sigma_b^2)$ distribution. The model estimates variation in association among strata while borrowing strength across strata to aid in the estimation for each one. This is useful when sample sizes within strata are small and so sample estimates are unappealing. An example is infinite sample odds ratios when all true cell probabilities are positive.

2.2 Cell-wise log-linear smoothing in contingency tables

The models in the previous subsection focused on smoothing associations by shrinking log odds ratios. An alternative approach focuses directly on the cell probabilities. One can perform such shrinkage with log-linear models containing cell-specific random effects. This makes sense when the main goal is to smooth the cell counts or when there is no obvious baseline for an association pattern. This is usually the case with nominal variables in two-way tables, for instance, or with very large, sparse tables.

Consider a two-way table with expected frequencies $\{\mu_{ij}\}$. The independence model usually provides too drastic a shrinkage, yet when variables are nominal no obvious model is more complex and yet unsaturated. A random effects generalization of the independence model is

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + u_{ij} \quad (2.5)$$

where $\{u_{ij}\}$ are $N(0, \sigma_u^2)$. Like model (2.1) with $\beta = 0$, fitting this model smooths toward independence. With ordinal variables, one could add to (2.5) a linear-by-linear association term,

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + x_i y_j \beta + u_{ij} \quad (2.6)$$

for fixed, monotone scores $\{x_i\}$ and $\{y_j\}$. The model then smooths toward an ordinal trend, namely uniform association for local odds ratios when $\{x_i\}$ and $\{y_j\}$ are equally spaced.

Because a given log-linear model implies an association structure, this cell-wise shrinkage also provides another way of smoothing association structure. Model (2.5) with uncorrelated random effects implies that the local log odds ratios are normal with mean 0 and standard deviation 2σ . Unlike model (2.1) with $\beta = 0$, which applied directly to the odds ratios, any two local log odds ratios that share common cells are correlated. Thus, the fits of the two models are not equivalent. Generally, cell-specific models such as (2.5) imply that models such as (2.1) have a certain correlation pattern for the random effects.

One can also use cell-wise random effects to smooth cell counts in multi-way tables or to summarize heterogeneity in conditional associations. For instance, consider a three-way table in which X and Y have ordered categories. For expected counts $\{\mu_{ijk}\}$, the heterogeneous XY linear-by-linear association model generalizes to

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \beta_k x_i y_j + u_{ijk}$$

With uncorrelated random effects, the local log odds ratios in stratum k are normal with mean β_k and standard deviation 2σ . Like model (2.3), this provides within-stratum shrinkage while permitting heterogeneity across strata. If the focus is instead on summarizing variability in an overall summary of association across a large number of strata, one could replace $\{\beta_k\}$ in this model by random effects $\{b_k\}$. Like model (2.4), it then has a variance component for within-stratum heterogeneity and a separate variance component for between-strata heterogeneity. Setting $\sigma_u^2 = 0$ focuses on summarizing $\{b_k\}$ by their mean and standard deviation.

The log-linear models in this subsection treat both X and Y as responses. If only Y is a response, one could instead add the random effects to logit models to perform this shrinkage, as Section 5.2 illustrates.

When smoothing by adding random effects to association models or to loglinear models, it may not be obvious how to choose the base model. For instance, should one add random effects to an independence model or to a uniform association model? Simulations discussed later suggest that often the choice of base model may not be crucial to the quality of estimation of cell probabilities. However, lacking particular theoretical reasons to use a particular form of model, one could compare the base models using standard criteria. For instance, in deciding whether to use unspecified β or take $\beta = 0$ in models (2.1) and (2.6), one could use a likelihood-ratio test of $\beta = 0$ or compare the models with a criterion such as AIC.

3 Fitting GLLMMs

Maximum likelihood fitting of GLLMMs can be difficult except in certain special cases. In fact, this is also true for the generalized log-linear model (1.1), which is simpler in that it does not have random effects. For many models in that family, one cannot recover $\boldsymbol{\mu}$ from $\mathbf{C} \log \mathbf{A}\boldsymbol{\mu}$. Then, one cannot express the likelihood directly in terms of $\boldsymbol{\beta}$. A simple example of this is the model of marginal homogeneity in square contingency tables. See, for instance, Haber (1985a,b), Fitzmaurice *et al.* (1993), Lang and Agresti (1994), and Glonek and McCullagh (1995). In simple cases in which one can express $\boldsymbol{\mu}$ in terms of $\boldsymbol{\beta}$, model fitting can often use standard methods for GLMMs, as discussed next.

3.1 Adaptive Gauss–Hermite quadrature

For some cases, the link function is a one-to-one function of $\boldsymbol{\mu}$ and the components of $\boldsymbol{\beta}$ and \mathbf{u} are functionally independent. Then, one can express the conditional distribution

of \mathbf{y} given the fixed effects and the random effects, $f(\mathbf{y}|\mathbf{u}; \boldsymbol{\beta})$, in term of those effects. An example is the multi-center clinical trial example of Section 5.2.

We use the term ‘cluster’ to refer to a level of the random effect. We assume that $f(\mathbf{y}|\mathbf{u}; \boldsymbol{\beta}) = \prod_{s=1}^S f_s(\mathbf{y}_s|\mathbf{u}_s; \boldsymbol{\beta})$, where f_s refers to cluster s . When the dimension of the random effects vector \mathbf{u}_s is not large, one can integrate out the random effects with Gauss–Hermite quadrature methods and then use standard methods such as Newton–Raphson to find the ML estimates. Adaptive quadrature (e.g., Liu and Pierce, 1994) centers the Gauss–Hermite nodes with respect to the mode of the function being integrated and scales them according to the estimated curvature at the mode. From our experience, this dramatically reduces the number of quadrature points needed to approximate the integrals effectively. This method is available in PROC NLMIXED in SAS. The appendix illustrates the implementation of the global odds ratio model with PROC NLMIXED using (3.1).

The form of $f(\mathbf{y}|\mathbf{u}; \boldsymbol{\beta})$ depends on the type of odds ratios used in the analysis. Association models for local odds ratios can be written as log-linear models, for which existing methods for fitting generalized linear mixed models are applicable. More specifically, Lang *et al.* (1999) noted that any association model for a table of cell probabilities $\boldsymbol{\pi}$ having the form

$$\mathbf{L} \log \boldsymbol{\pi} = \mathbf{W} \mathbf{a}$$

is expressible as log-linear model $\log \boldsymbol{\pi} = \mathbf{W}^* \mathbf{a}^*$, where $\mathbf{W}^* = [\mathbf{P}; \mathbf{L}^\top (\mathbf{L} \mathbf{L}^\top)^{-1} \mathbf{W}]$ and the column space of \mathbf{P} equals the null space of \mathbf{W} (see also Goodman, 1979; Glonek, 1996).

GLLMMs for global odds ratios represent mixed extensions of models studied by Plackett (1965) and Dale (1986) in the bivariate case and extended to the multivariate case by Molenberghs and Lesaffre (1994). In these settings, mixed association models do not correspond to a generalized linear mixed model. One must explicitly calculate the multinomial probabilities implied by the model for the global odds ratios from formulas presented by Dale (1986), Molenbergh and Lesaffre (1994), and Ten Have and Morabia (1999).

For instance, consider the two-way table that cross-classifies ordinal variables X and Y with cell probabilities π_{ij} . Suppose we posit a model for the global odds ratios

$$\theta_{ij} = \frac{\Pr[X \leq i, Y \leq j] \Pr[X > i, Y > j]}{\Pr[X > i, Y \leq j] \Pr[X \leq i, Y > j]}$$

and marginal probabilities $\gamma_i^X = \Pr(X \leq i)$ and $\gamma_j^Y = \Pr(Y \leq j)$. If $F_{ij} = \Pr(X \leq i, Y \leq j)$, then

$$F_{ij} = \begin{cases} \frac{1}{2}(\theta_{ij} - 1)^{-1} [1 + (\gamma_i^X + \gamma_j^Y)(\theta_{ij} - 1) - S(\theta_{ij}, \gamma_i^X, \gamma_j^Y)], & \text{if } \theta_{ij} \neq 1 \\ \gamma_i^X \gamma_j^Y, & \text{if } \theta_{ij} = 1 \end{cases} \quad (3.1)$$

where

$$S(\theta_{ij}, \gamma_i^X, \gamma_j^Y) = \{[1 + (\gamma_i^X + \gamma_j^Y)(\theta_{ij} - 1)]^2 + 4\theta_{ij}(1 - \theta_{ij})\gamma_i^X\gamma_j^Y\}^{1/2}.$$

For given $\{\gamma_i^X\}$ and $\{\gamma_j^Y\}$, a linear transformation of $\{F_{ij}\}$ yields $\{\pi_{ij}\}$ and hence the log-likelihood $\sum_{ij} n_{ij} \log \pi_{ij}$. We use PROC NL MIXED to carry out this strategy in Section 5.2.

3.2 GLLM algorithms nested within Gauss–Hermite quadrature

The extension of these methods when the conditional distribution of \mathbf{y} is not expressible as a function of the model parameters is more complex. In the fixed effects setting, several authors have nested iterative algorithms for obtaining the response probabilities as a function of the parameters within iterative schemes for model fitting. For instance, Glonek and McCullagh (1995) and Glonek (1996) proposed using Newton–Raphson to calculate response probabilities given parameter values, and Fitzmaurice and Laird (1993) and Molenberghs and Lesaffre (1999) used iterative proportional fitting (IPF).

When the dimension of \mathbf{u} in a GLLMM is not large, one can nest these approaches for inverting $\boldsymbol{\mu}$ within a Gauss–Hermite quadrature scheme. For density function g for the random effects, the marginal likelihood is

$$L(\boldsymbol{\beta}; \mathbf{y}) = \int f(\mathbf{y}|\mathbf{u}; \boldsymbol{\beta})g(\mathbf{u}) \, d\mathbf{u} \quad (3.2)$$

Here, the notation reflects the fact that $\boldsymbol{\mu}$ is a function of both the fixed effects $\boldsymbol{\beta}$ and the random effects \mathbf{u} . A quadrature approximation to (3.2) is

$$L(\boldsymbol{\beta}; \mathbf{y}) = \sum_{q_1=1}^Q \cdots \sum_{q_p=1}^Q f(\mathbf{y}|\mathbf{z}_{q_1, \dots, q_p}; \boldsymbol{\beta})w_{q_1, \dots, q_p} \quad (3.3)$$

where $\{\mathbf{z}_{q_1, \dots, q_p}\}$ and $\{w_{q_1, \dots, q_p}\}$ are pre-determined quadrature nodes and weights, respectively (Aitkin, 1996). As usual, a larger number of quadrature points are necessary for adequate accuracy when the variances of the random effects are large. Then, (3.3) is the objective function to be maximized with respect to $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\theta})$. Maximization requires inverting the set of equations $\boldsymbol{\eta} = \mathbf{C} \log \mathbf{A}\boldsymbol{\mu}(\boldsymbol{\beta}, \mathbf{z}_{q_1, \dots, q_p})$, $q_k = 1, \dots, Q$, $k = 1, \dots, p$, with p being the dimension of the random effects. Thus, we propose nesting existing algorithms for performing this inversion (IPF or Newton–Raphson) within a Newton–Raphson scheme for maximizing (3.3). We plan to discuss this approach in more detail in a separate paper. Our studies so far suggest that nesting the IPF algorithm of Molenberghs and Lesaffre (1999) within a Gauss–Hermite quadrature scheme works well.

3.3 Monte Carlo EM

When a closed-form expression exists for $f(y|\mathbf{u}; \boldsymbol{\beta})$ but the dimension of the random effects renders quadrature methods infeasible, one can use Monte Carlo algorithms for model fitting (Chan and Ledolter, 1995; McCulloch, 1997; Booth and Hobert, 1999). The EM algorithm iterates between calculating the expectation of the complete data log-likelihood given the observed data (E-step) and maximizing this expectation with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ (M-step). Denote the associated conditional density as $h(\mathbf{u}|\mathbf{y}, \boldsymbol{\psi}^{(r)})$. The calculation of the normalizing constant for $h(\mathbf{u}|\mathbf{y}, \boldsymbol{\psi}^{(r)})$ entails computing an intractable integral, and one must replace the required expectation with an approximation. Monte Carlo EM constructs Monte Carlo estimates of these expectations at each E-step. Existing implementations sample from the distribution $h(\mathbf{u}|\mathbf{y}, \boldsymbol{\psi}^{(r)})$ using either Markov chain Monte Carlo (McCulloch, 1997; Quintana *et al.*, 1999) or independent samples based on importance or rejection sampling (Booth and Hobert, 1999).

3.4 Penalized quasi-likelihood

Breslow and Clayton (1993) proposed a penalized quasi-likelihood (PQL) method to fitting GLMMs. This is computationally simpler than ML, but in some cases (e.g., binary data, large variance components) may provide poor approximations for ML estimates. Because ML is feasible for the smoothing applications presented in Section 5, we have not discussed this method. However, in the simulation study discussed in Section 4, we compared the quality of the ML and PQL estimates of the cell probabilities. We observed little difference between these estimates. This result is not surprising, as there is no reason to expect ML results to outperform approximate methods when the random effects are simply a tool for smoothing towards a given model. Hobert and Wand (2000) and Wager *et al.* (2003) reported similar simulation results in logistic and spatial smoothing applications. The PQL method may still be useful for complex versions of GLLMMs that are not readily amenable to ML model fitting.

4 A simulation study illustrating benefits of GLLMM smoothing

One would hope that when a simplistic model such as independence or uniform association does not hold, fitted proportion estimates based on GLLMM smoothing would tend to have smaller mean squared errors than the estimates based on the simpler model or the sample proportion estimates based on the saturated model. We do not expect a uniform domination, as when the simpler model nearly holds it has the benefit of model parsimony unless the sample size is very large. In addition, the GLLMM framework enables one to smooth either the associations directly or in a cellwise fashion.

We conducted a simulation study to investigate these issues. We compared estimates from different models for log odds ratios: (a) the independence model [i.e. (2.1) with $\boldsymbol{\beta} = 0$ and $\boldsymbol{\sigma} = 0$], (b) the uniform association model (i.e., (2.1) with $\boldsymbol{\sigma} = 0$), (c) the

sample proportions, (d) the local odds ratio association model (2.1) with $\beta = 0$, (e) the local odds ratio association model (2.1) with unrestricted β and σ , (f) the mixed log-linear model (2.5), and (g) the mixed log-linear model (2.6) having uniform association term. In addition to these special cases of GLLMMs, we also considered the local linear likelihood estimates as outlined in Simonoff (1996; Section 6.4, 1998). For this fully nonparametric approach, we chose the smoothing parameter for each simulated table by minimizing the corrected AIC (AICC) (Simonoff, 1998). We compared the estimators for 36 different cases, corresponding to

1. Two table sizes (3×3 and 6×6)
2. Three sample sizes ($n = 50, 200, 1000$)
3. Two models for the true probabilities (uniform association, with scores 1, 2, 3 for the 3×3 case and 1, 2, 3, 4, 5, 6 for the 6×6 case; nonuniform association, with scores 1.0, 2.5, and 3.0 for the 3×3 case and 1.0, 2.8, 4.2, 5.2, 5.8, 6.0 for the 6×6 case)
4. Three levels of association ($\beta = 0.1, 0.4$ and 1.0)

Table 1 shows the n -normalized average MSE for each of the estimators in 3×3 tables, based on 1000 simulated tables for each scenario. Monte Carlo standard errors for all estimates are less than or equal to 0.0025. Results for 6×6 tables are not shown here, since they were qualitatively similar. Table 1 shows that the relative performances of the different smoothing methods depend on the strength of association and sample size.

Table 1 Monte Carlo estimates of n -normalized MSE for cell probability estimates based on Gaussian quadrature-based ML estimation in GLLMMs, simpler models without random effects (GLM), local likelihood (LL) using AICC-chosen smoothing, and sample proportions (SP) for 3×3 tables

		Estimator								
Scores	β	n	SP	GLM		Cellwise GLLMM		Association GLLMM		LL
				Indep.	$L \times L$	Indep.	$L \times L$	Indep.	$L \times L$	
Unif.	0.1	50	0.099	0.053	0.064	0.053	0.064	0.059	0.066	0.038
		200	0.099	0.062	0.064	0.062	0.064	0.065	0.066	0.039
		1000	0.099	0.106	0.065	0.104	0.065	0.085	0.067	0.040
	0.4	50	0.098	0.092	0.064	0.089	0.065	0.076	0.067	0.042
		200	0.098	0.222	0.064	0.146	0.064	0.084	0.066	0.052
		1000	0.098	0.915	0.064	0.106	0.064	0.091	0.066	0.109
	1.0	50	0.093	0.291	0.067	0.120	0.067	0.084	0.068	0.067
		200	0.093	1.039	0.067	0.099	0.067	0.087	0.067	0.145
		1000	0.093	5.034	0.064	0.095	0.064	0.091	0.065	0.585
Non unif.	0.1	50	0.099	0.054	0.063	0.055	0.063	0.060	0.066	0.039
		200	0.099	0.062	0.063	0.062	0.063	0.064	0.065	0.039
		1000	0.099	0.114	0.071	0.110	0.071	0.080	0.072	0.049
	0.4	50	0.098	0.102	0.072	0.094	0.072	0.079	0.073	0.049
		200	0.098	0.265	0.096	0.138	0.094	0.086	0.090	0.083
		1000	0.098	1.138	0.233	0.104	0.147	0.092	0.103	0.268
	1.0	50	0.091	0.354	0.102	0.107	0.099	0.081	0.092	0.107
		200	0.091	1.306	0.230	0.092	0.130	0.084	0.096	0.322
		1000	0.091	6.368	0.902	0.092	0.100	0.090	0.091	1.469

For small samples and/or small β , the local likelihood estimates outperform both their parametric and GLLMM counterparts. This is somewhat surprising, as one would expect that the maximum likelihood estimates from the true model would perform best. Thus, the asymptotic optimality of the ML estimates does not hold for the small to moderate samples of this setting. When the association is large ($\beta = 1.0$), the performance of the local likelihood estimates deteriorates rapidly with increasing sample size, whereas the GLLMM estimates behave well. Also, it seems as if shrinking in the GLLMMs with $\beta = 0$ (i.e. shrinking to independence) is essentially as good as shrinking to uniform association unless uniform association truly holds, there is a strong association, and the sample size is relatively large (e.g., $\beta = 0.4$ or $\beta = 1.0$, $n = 1000$).

Results from this study also provided direct comparison among the different GLLMM formulations. These comparisons showed no systematic differences between the performances of the cellwise and association models. Overall, these results suggest that an effective way to smooth large, sparse contingency tables is to fit mixed log-linear models that contain cell-specific random effects. In applications in which an alternative association structure is particularly relevant, however, the general GLLMM formulation provides the flexibility to smooth on the relevant association scale.

5 Examples of smoothing using GLLMMs

We now present examples of smoothing cell counts and associations using GLLMMs. Section 5.1 uses GLLMMs to smooth a large sparse contingency table. Section 5.2 uses GLLMMs to characterize heterogeneity in stratified ordinal contingency tables.

5.1 Smoothing counts in a large sparse contingency table

Table 2 presents data from Simonoff (1996; p. 229) on responses of 55 first-year MBA students at New York University's Stern School of Business to questions about the importance of statistics and economics in business education. Responses were coded on a seven-point scale from 1 = 'completely useless' to 7 = 'absolutely crucial.' Simonoff (1996) smoothed the sparse counts using local polynomial estimators based on kernel density estimation. He noted that this approach can yield negative probability estimates, but this can be avoided by using local polynomial likelihood estimates (Simonoff, 1996, Section 6.4; 1998).

Simulation results presented in Section 4 suggest that the association and cell-wise approaches to smoothing cell probabilities yield estimates having similar mean squared errors, with the cell-wise approach having the added advantage that maximum likelihood (ML) fitting is straightforward. Thus, for simplicity here we report only shrinkage estimates of the cell probabilities by fitting the cell-wise model (2.6), both with unspecified β and with $\beta = 0$ (shrinking towards independence). The model with unspecified β seems natural here, to reflect the ordinal classifications. Model fitting for these GLLMMs used the adaptive Gauss-Hermite quadrature approach to ML estimation discussed in Section 3.1. We also compared these estimates to those obtained using the local linear likelihood approach of Simonoff (1998).

Table 2 Opinions of MBA students about the importance of statistics and economics in business education, and fitted values from association and cell-wise smoothing

Statistics response	Economics response						
	1	2	3	4	5	6	7
2	0	1	0	0	0	1	0
	(0.04) ^a	(0.04)	(0.14)	(0.46)	(0.58)	(0.62)	(0.07)
	(0.50) ^b	(0.28)	(0.48)	(0.52)	(0.18)	(0.04)	(0.00)
3	(0.09) ^c	(0.19)	(0.24)	(0.23)	(0.22)	(0.15)	(0.06)
	0	0	0	1	0	0	0
	(0.02)	(0.02)	(0.07)	(0.24)	(0.29)	(0.29)	(0.04)
4	(0.09)	(0.08)	(0.21)	(0.35)	(0.19)	(0.07)	(0.00)
	(0.08)	(0.35)	(0.77)	(0.92)	(1.04)	(0.38)	(0.11)
	0	0	3	6	4	0	0
5	(0.23)	(0.23)	(1.06)	(3.59)	(3.92)	(3.22)	(0.46)
	(0.31)	(0.42)	(1.74)	(4.51)	(3.82)	(2.08)	(0.12)
	(0.06)	(0.31)	(1.85)	(4.73)	(3.61)	(0.99)	(0.19)
6	0	0	1	4	7	4	0
	(0.27)	(0.28)	(1.11)	(3.73)	(5.20)	(4.54)	(0.54)
	(0.08)	(0.16)	(1.04)	(4.18)	(5.50)	(4.65)	(0.41)
7	(0.12)	(0.38)	(1.55)	(4.32)	(6.49)	(3.20)	(0.94)
	1	0	0	2	6	10	1
	(0.36)	(0.35)	(1.29)	(4.03)	(5.85)	(7.05)	(0.72)
7	(0.02)	(0.05)	(0.51)	(3.19)	(6.52)	(8.56)	(1.15)
	(0.17)	(0.28)	(0.68)	(1.82)	(5.19)	(4.60)	(2.38)
	0	0	0	0	0	2	1
7	(0.05)	(0.05)	(0.21)	(0.67)	(0.84)	(0.96)	(0.11)
	(0.00)	(0.00)	(0.03)	(0.25)	(0.79)	(1.60)	(0.33)
	(0.06)	(0.09)	(0.15)	(0.32)	(1.39)	(2.23)	(2.02)

^aCell-wise smoothing of independence base model using SAS PROC NLMIXED.

^bCell-wise smoothing of $L \times L$ base model using SAS PROC NLMIXED.

^cAICC-chosen smoothing based on local linear likelihood.

Table 2 presents results for the two GLLMMs, each assuming uncorrelated random effects, and local likelihood. The estimated random effects standard deviations from the two GLLMMs are $\sigma_u = 0.25$ for the independence base model and $\sigma_u = 0.0$ for the uniform association model, suggesting that the $L \times L$ model does not smooth the data too severely. The estimated association is $\beta = 0.44$ (SE = 0.14), and a likelihood ratio test for $H_0: \beta = 0$ provides strong evidence of an association between the two responses, with the likelihood ratio statistic equaling 13.2 on 1df. A comparison of the GLLMM fit with unspecified β and the local likelihood approach shows that the GLLMM fit suggests more strongly that the cells corresponding to responses (2,6) and (6,1) are unusual relative to the general association of the table. This agrees with comments made by Simonoff (1996, p. 231), who noted that these cells are outliers compared to the others. This model-based approach also suggests that the large (6,6) cell is perhaps not as unusual as local likelihood would suggest.

5.2 Random association models for stratified ordinal data

Table 3 shows preliminary results for eight centers from a double-blind, parallel-group clinical study. The purpose of this study was to compare an active drug with placebo in

Table 3 Clinical trial relating treatment to response for eight centers

Center	Treatment	Response		
		Much better	Better	Unchanged/worse
1	Drug	13	7	6
	Placebo	1	1	10
2	Drug	2	5	10
	Placebo	2	2	1
3	Drug	11	23	7
	Placebo	2	8	2
4	Drug	7	11	8
	Placebo	0	3	2
5	Drug	15	3	5
	Placebo	1	1	5
6	Drug	13	5	5
	Placebo	4	0	1
7	Drug	7	4	13
	Placebo	1	1	11
8	Drug	15	9	2
	Placebo	3	2	2

the treatment of patients suffering from asthma. Patients were randomly assigned to treatment. At the end of the study, investigators described their perception of the patients' change in condition, using the ordinal scale (much better, better, unchanged or worse). Hartzel *et al.* (2001b) presented results from fixed-effect association models based on both local and cumulative odds ratios. However, these authors noted that it is natural to view the eight centers as a sample from the population of clinics of interest.

From their fixed effects analysis, Hartzel *et al.* (2001b) concluded that association models containing stratum effects only describe the data well. We consider analogous mixed association models of the form

$$\log \theta_{ij(k)} = b_k \quad (5.1)$$

where $b_k \stackrel{iid}{\sim} N(\beta, \sigma_b^2)$. As noted in Section 2.1, this model is the special case of model (2.4) with $\sigma_u^2 = 0$. From results discussed in Section 3, one can use some software for GLMMs to implement ML fitting of model (5.1) for both local and global log odds ratios. Table 4 presents predicted odds ratios from both the local and global versions of the model. Because this example has only two treatments, the model for global odds ratios is equivalent to a model for cumulative odds ratios. Table 5 shows the predicted cell counts for the two cases.

The estimated mean for the center-specific global log odds ratios is $\hat{\beta} = 0.84$ (SE = 0.48), providing moderate evidence of a positive mean treatment effect. The estimated standard deviation of $\{b_k\}$ is $\hat{\sigma}_b = 1.03$ (SE = 0.48). The likelihood ratio statistic for testing $H_0: \sigma_b = 0$ is 3.7, yielding strong evidence of heterogeneity in the global log odds ratios across centers with approximate p -value $p = P(\chi_1^2 > 3.7) / 2 = 0.027$ (Self and Liang, 1987). The model fit for the local odds ratios yields similar results, with estimated mean log local odds ratio 0.59 (SE = 0.24), but less evidence of heterogeneity across strata $\hat{\sigma}_b = 0.34$ (SE = 0.36).

Table 4 Estimated center-specific local and global log odds ratios for clinical trial data (Table 3)

Center	Odds ratio	Model	
		Fixed (GLLM)	Random (GLLMM)
1	Local	1.73	0.93
	Global	2.74	2.00
2	Local	-1.15	0.23
	Global	-1.67	-0.49
3	Local	0.24	0.49
	Global	0.30	0.45
4	Local	0.72	0.62
	Global	0.81	0.82
5	Local	1.33	0.81
	Global	2.26	1.63
6	Local	-0.42	0.35
	Global	-0.94	0.0
7	Local	0.95	0.70
	Global	1.55	1.25
8	Local	0.69	0.62
	Global	0.88	0.66

Table 4 also presents estimated center-specific log odds ratios from the fixed effects version of model (5.1). The random effects models shrink the fixed effects estimates towards the overall mean log odds ratios. For instance, four of the fixed effects estimates are negative, whereas only one of the random effects estimates is. In all centers, the global log odds ratio estimates are larger than their local counterparts in both the fixed and random GLLMs. This is usually the case for association between ordinal categorical variables.

In this example, with ML estimation, shrinkage was greater for the local log odds ratios than their global counterparts. This can be seen by comparing estimated standard deviations of the random effects (0.34 for local versus 1.03 for global), or by comparing shrinkage of the individual fixed effects (Table 4). This difference arises because, for each stratum, the local odds ratios use less data than the global odds ratios. As a result, the overall mean plays a more important role in the local predictions. Because each global odds ratio uses all the data in each stratum, the data from that stratum are weighted more heavily than for the local version using only a subset of the data.

Hartzel *et al.* (2001b) considered an analogous approach with random effects in proportional odds models. In their model the random effects enter directly on the cumulative logit scale rather than the log odds ratio scale. This model is

$$\text{logit}[P(Y \leq j)] = \alpha_j + u_k + b_k x_i$$

where the indicator $x_1 = 0$ and $x_2 = 1$ and where (u_k, b_k) are correlated random effects with possibly different variances. This model implies a center-specific cumulative log odds ratio b_k having a mean of 0.92 with standard error of 0.53. This model is not equivalent to the model (5.1) for the cumulative log odds ratios, because that model results from the more complex intercept structure α_{jk} for the logit. The generalized

Table 5 Predicted cell counts from random association model (5.1)

Center	Treatment	Response		
		Much better	Better	Unchanged/worse
1	Drug	(12.1) ^a (12.6) ^b	(5.8) (6.2)	(8.0) (7.2)
	Placebo	(1.8) (1.4)	(2.2) (1.8)	(7.9) (8.8)
2	Drug	(3.3) (2.8)	(5.5) (5.2)	(8.1) (9.0)
	Placebo	(0.7) (1.2)	(1.5) (1.8)	(2.8) (2.0)
3	Drug	(10.9) (10.8)	(23.9) (23.9)	(6.0) (6.3)
	Placebo	(2.0) (2.2)	(7.1) (7.1)	(3.0) (2.7)
4	Drug	(6.3) (6.4)	(12.0) (12.0)	(7.5) (7.6)
	Placebo	(0.6) (0.6)	(2.0) (2.0)	(2.4) (2.4)
5	Drug	(14.0) (14.3)	(3.1) (3.1)	(5.8) (5.6)
	Placebo	(1.9) (1.7)	(0.9) (0.9)	(4.1) (1.1)
6	Drug	(14.4) (14.0)	(4.0) (4.1)	(4.4) (4.9)
	Placebo	(2.5) (3.0)	(1.0) (0.9)	(1.5) (1.1)
7	Drug	(6.7) (6.7)	(3.6) (3.9)	(13.5) (13.4)
	Placebo	(1.3) (1.3)	(1.4) (1.1)	(10.3) (11.0)
8	Drug	(15.2) (15.3)	(8.3) (8.1)	(2.4) (2.6)
	Placebo	(2.7) (2.7)	(2.7) (2.9)	(1.5) (1.4)

^aRandom local odds ratio model.

^bRandom global odds ratio model.

log-linear formulation for the global odds ratios is more general, as the GLLMMs do not require cutpoint parameters to be the same across stratum. Likewise, expressing the model directly in terms of adjacent categories logits and using the simpler cutpoint structure, Hartzel *et al.* (2001b) reported an estimated mean log local odds ratio of 0.63 (SE = 0.34).

Q4

For tables with more than two rows, the cumulative odds model and the GLLMM for global odds ratios relate to different sets of odds ratios. In this case, the choice of one model over the other would depend on the scientific questions of interest. Global odds ratios are useful when rows and columns use the same ordered categories, in which case they provide a natural way of specifying a joint distribution for a bivariate response (Plackett, 1965). Cumulative logits are relevant if interest focuses on a stochastic ordering among rows on a column response variable.

6 Extensions and challenges

This paper has proposed a new class of generalized log-linear mixed models for categorical data. Special cases include many existing models. In addition, this unifying framework suggests ways to extend certain methods for categorical data, such as the penalized likelihood smoothing of Simonoff (1983; 1987) and association models for stratified contingency tables.

We considered specific cases for bivariate responses. Then, model fitting is relatively straightforward in that one can usually write the conditional distribution of the data given the random effects as a function of the fixed and random effects. Given a closed form for this distribution, one can subsequently use existing methods for GLMMs to fit the model. When the dimension of the random effects is large in this setting, the Monte Carlo EM algorithm is an effective algorithm for model fitting.

A topic for future research is the development of general model-fitting strategies when a GLLMM does not yield closed-form expressions for this conditional distribution and the dimension of the random effects is large. Unfortunately, MCEM algorithms are computationally prohibitive in this case, because computation of the simulated complete data log-likelihood, which requires iteration, would then be necessary for all Monte Carlo samples within all EM iterations. A potential alternative approach to model fitting is the modified EM scheme of Steele (1996), in which the expectation of the complete data score equations is approximated with a second-order Laplace approximation. In the GLMM setting, Steele (1996) showed through both analytical arguments and simulation that the second-order correction decreased the bias incurred by the PQL approach of Breslow and Clayton (1993). Other fitting approaches may also be possible. For instance, for GLLMs, Haber (1985a,b), Lang and Agresti (1994), and Lang *et al.* (1999) have used Lagrange multipliers to maximize the log-likelihood subject to constraints implied by the model. Overall, finding a general mechanism for fitting the class of GLLMMs is a strong challenge for future research.

A referee has pointed out that several interesting questions occur relating to properties of GLLMMs. For instance, when used for smoothing, what are the asymptotic properties of the cell probability estimates? For a fixed number of cells in a contingency table, presumably these estimates give relatively more weight to the sample proportions as the sample size increases. Thus, we conjecture that they are consistent even when a simpler model that is the basis of the smoothing does not hold. As support for this, in the simulation study, the GLLMMs behave as well as the sample proportions in very large samples ($n = 1000$). Compared to other smoothing methods, it is difficult to study this analytically because of the lack of a closed form for the estimates.

The asymptotic properties become even more questionable under the sparse framework whereby the number of cells grows with the sample size. For cell probabilities $\{\pi_i\}$ and estimates $\{\hat{\pi}_i\}$, Simonoff (1983) showed that for penalized likelihood estimators consistency can occur in the sense that $\sup_i |\hat{\pi}_i/\pi_i - 1| \rightarrow 0$ in probability as the sample size and number of cells grow in such a way that the $\{\pi_i\}$ themselves approach 0. It is unclear whether such a property would apply to the estimates obtained using GLLMMs, but the connection mentioned in (2.2) with penalized likelihood estimation suggests that this may happen.

For the smoothing application, another interesting question relates to whether there is a theoretical advantage to using a good base model. Our simulations suggested that unless there is a strong association and large sample size, adding random effects to a baseline independence model worked essentially as well as adding random effects to a baseline uniform association model. For model (2.1) for smoothing odds ratios, is the relative size of $\hat{\sigma}_u$ for this model compared to $\hat{\sigma}_u$ for the independence version of the model ($\beta = 0$) related to the quality of estimation in the two cases?

Another area worthy of future research is development of specific GLLMMs that may be useful for various applications. For instance, Section 2 mentioned some models that generalize Goodman's association models that may be worth special study. In addition, it may be worth studying models that have random effects for describing both association and marginal distributions. For instance, in a study with matched pairs, let (Y_{1s}, Y_{2s}) denote the responses for cluster s , with $\pi_{j|rs} = P(Y_{rs} = j)$, and let $\gamma_{j|rs} = P(Y_{rs} \leq j)$. In repeated measurement studies, one could simultaneously model effects of covariates on the marginal distributions of Y_{rs} , $r = 1, 2$ and model odds ratios describing the (Y_{1s}, Y_{2s}) association. Let \mathbf{x}_{rs} and \mathbf{z}_{rs} represent covariates of interest associated with Y_{rs} , $r = 1, 2$, and let \mathbf{x}_{3s} and \mathbf{z}_{3s} represent the covariates associated with the odds ratios. Then, a GLLMM for (Y_{1s}, Y_{2s}) is

$$\log\left(\frac{\gamma_{j|1s}}{1 - \gamma_{j|1s}}\right) = \alpha_{1j} + \mathbf{x}_{1s}^\top \boldsymbol{\beta}_1 + \mathbf{z}_{1s}^\top \mathbf{u}_1 \quad (6.1)$$

$$\log\left(\frac{\gamma_{j|2s}}{1 - \gamma_{j|2s}}\right) = \alpha_{2j} + \mathbf{x}_{2s}^\top \boldsymbol{\beta}_2 + \mathbf{z}_{2s}^\top \mathbf{u}_2 \quad (6.2)$$

$$\log \theta_{ij|s} = \beta_0 + \mathbf{x}_{3s}^\top \boldsymbol{\beta}_3 + \mathbf{z}_{3s}^\top \mathbf{u}_3, \quad \text{for all } i, j \quad (6.3)$$

where $\{\alpha_{rj}\}$ are nondecreasing in j for $r = 1, 2$. Ten Have and Morabia (1999) considered models of this type for binary responses. Model (6.1)–(6.3) also contains as special cases mixed model representations (e.g., Brumback *et al.*, 1999) of generalized additive models for marginal and association parameters (Bustami *et al.*, 2001).

Lang and Agresti (1994) and Lang *et al.* (1999) provided conditions under which the simultaneous association and marginal models specified by a GLLM are compatible, and it is of interest to extend this work to GLLMMs. It is also important to determine when problems may exist with identifiability. Recent theoretical results given by Bergsma and Rudas (2002) on 'sets of ordered sets' should be useful for this purpose.

Other potentially useful extensions of the GLLMM involve relaxing particular assumptions made by (1.2). For instance, one may want to take a distribution-free approach for the random effects, assuming only an unspecified number and location of discrete mass points. This would adapt an approach described by Aitkin (1996). Overall, the general GLLMM class holds potential for a wide variety of applications, with general approaches to model fitting representing a significant challenge for future research.

Acknowledgements

This work was supported by grants from NIH (BA Coull and A Agresti) and a grant from NSF (A Agresti). The authors thank Jim Hobert for providing computational advice, and the referees for helpful comments.

References

- Aitkin M (1996) A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, **6**, 251–62.
- Bergsma WP, Rudas T (2002) Marginal models for categorical data. *Annals of Statistics*, **30**, 140–59.
- Booth JG, Hobert JP (1999) Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B*, **61**, 265–85.
- Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
- Brumback BA, Ruppert D, Wand MP (1999) Comment to ‘Variable selection and function estimation in additive nonparametric regression using a data-based prior’. *Journal of the American Statistical Association*, **94**, 794–97.
- Bustami R, Lesaffre E, Molenberghs G, Loos R, Danckaerts M, Vlietinck, R (2001) Modelling bivariate ordinal responses smoothly with examples from ophthalmology and genetics. *Statistics in Medicine*, **20**, 1825–42.
- Chan KS, Ledolter J (1995) Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, **90**, 242–52.
- Clogg CC (1982) Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association*, **77**, 803–15.
- Coull BA, Agresti A (2000) Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution. *Biometrics*, **56**, 73–80.
- Dale J (1986) Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–17.
- Fitzmaurice GM, Laird NM, Rotnitzky AG (1993) Regression models for discrete longitudinal responses (with discussion). *Statistical Science*, **8**, 284–309.
- Ghosh M, Rao JNK (1994) Small area estimation: an appraisal (Discussion: p. 76–93). *Statistical Science*, **9**, 55–76.
- Glonek GFV, McCullagh P (1995) Multivariate logistic models. *Journal of the Royal Statistical Society, Series B, Methodological*, **57**, 533–46.
- Glonek GFV (1996) A class of regression models for multivariate categorical responses. *Biometrika*, **83**, 15–28.
- Goodman LA (1979) Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, **74**, 537–52.
- Green PJ (1987) Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, **55**, 245–59.
- Grizzle JE, Starmer CF, Koch GG (1969) Analysis of categorical data by linear models. *Biometrics*, **25**, 489–504.
- Haber M (1985a) Maximum likelihood methods for linear and log-linear models in categorical data. *Computational Statistics and Data Analysis*, **3**, 1–10.
- Haber M (1985b) Log-linear models for correlated marginal totals of a contingency table. *Communications in Statistics, Part A – Theory and Methods*, **14**, 2845–56.
- Hartzel J, Agresti A, Caffo B (2001a) Multinomial logit random effects models. *Statistical Modelling*, **1**, 81–102.
- Hartzel J, Liu I, Agresti A (2001b) Describing heterogeneous effects in stratified ordinal contingency tables, with application to multi-center clinical trials. *Computational Statistics and Data Analysis*, **35**, 429–49.
- Hedeker D, Gibbons RD (1994) A random-effects ordinal regression model for multilevel analysis. *Biometrics*, **50**, 933–44.

- Hobert JP, Wand MP (2000) *Automatic generalized nonparametric regression via maximum likelihood*. Technical Report, Department of Biostatistics, Harvard School of Public Health.
- Lang JB, Agresti A (1994) Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, **89**, 625–32.
- Lang JB, McDonald JW, Smith PWF (1994) Association-marginal modelling of multivariate categorical responses: a maximum likelihood approach. *Journal of the American Statistical Association*, **94**, 1161–71.
- Liu Q, Pierce DA (1994) A note on Gauss–Hermite quadrature. *Biometrika*, **81**, 624–29.
- McCulloch CE (1997) Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, **92**, 162–70.
- Molenberghs G, Lesaffre E (1994) Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, **89**, 633–44.
- Molenberghs G, Lesaffre E (1999) Marginal modelling of multivariate categorical data. *Statistics in Medicine*, **18**, 2237–55.
- Plackett RL (1965) A class of bivariate distributions. *Journal of the American Statistical Association*, **60**, 516–22.
- Quintana FA, Liu JS, del Pino GE (1999) Monte Carlo EM with importance reweighting and its applications in random effects models. *Computational Statistics & Data Analysis*, **29**, 429–44.
- Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605–10.
- Simonoff JS (1983) A penalty function approach to smoothing large sparse contingency tables. *Annals of Statistics*, **11**, 208–18.
- Simonoff JS (1987) Probability estimation via smoothing in sparse contingency tables with ordered categories. *Statistics & Probability Letters*, **5**, 55–63.
- Simonoff JS (1996) *Smoothing Methods in Statistics*. New York: Springer.
- Simonoff JS (1998) Three sides of smoothing: categorical data smoothing, non-parametric regression, and density estimation. *International Statistical Review*, **66**, 137–56.
- Steele BM (1996) A modified EM algorithm for estimation in generalized mixed models. *Biometrics*, **52**, 1295–1310.
- Ten Have TR, Morabia A (1999) Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data. *Biometrics*, **55**, 85–93.
- Tutz G, Hennevogl W (1996) Random effects in ordinal regression models. *Computational Statistics and Data Analysis*, **22**, 537–57.
- Wager CG, Coull BA, Lange N (2003) Modelling spatial intensity for replicated inhomogeneous point patterns in brain imaging. *Journal of the Royal Statistical Society, Series B*, in press.
- Williams OD, Grizzle JE (1972) Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association*, **67**, 55–63.

Q3

Appendix

Example of SAS code (Version 8) for using PROC NLMIXED to implement maximum likelihood fitting of random association model (5.1) using global odds ratios to Table 3.

```
data clin1;
input center y11 y12 y13 y21 y22 y23;
datalines;
1 13 7 6 1 1 10
2 2 5 10 2 2 1
```

```

3 11 23 7 2 8 2
4 7 11 8 0 3 2
5 15 3 5 1 1 5
6 13 5 5 4 0 1
7 7 4 13 1 1 11
8 15 9 2 3 2 2
;
run;

data clin2;
set clin1;
r1tot=y11 + y12 + y13;
r2tot=y21 + y22 + y23;
c1tot=y11 + y21;
c2tot=c1tot + y12 + y22;
total=r1tot + r2tot;
run;

proc nlmixed data=clin2 qpoints=15 cov;
  parms alpha=0.82 sd=1.0;
  global1=exp(alpha + u);
  global2=exp(alpha + u);
  etal=r1tot/total;
  zetal=c1tot/total;
  zeta2=c2tot/total;
  s1=((1 + (etal + zetal)*(global1 - 1))**2
    + 4*global1*(1 - global1)*etal*zetal)**.5;
  s2=((1 + (etal + zeta2)*(global2 - 1))**2
    + 4*global2*(1 - global2)*etal*zeta2)**.5;
  if (global1=1.0) then
    f1=etal*zetal;
  else
    f1=.5*((global1 - 1)**(-1))*
      (1 + (etal + zetal)*(global1 - 1) - s1);
  if (global2=1.0) then
    f2=etal*zeta2;
  else
    f2=.5*((global2 - 1)**(-1))*
      (1 + (etal + zeta2)*(global2 - 1) - s2);
  f11=f1;
  f12=etal - f11;
  f13=zetal - f11;
  f14=1 - (f11 + f12 + f13);
  f21=f2;
  f22=etal - f21;
  f23=zeta2 - f21;

```

```
f24 = 1 - (f21 + f22 + f23);  
pi11 = f11;  
pi12 = f21 - f11;  
pi13 = f22;  
pi21 = f13;  
pi22 = f23 - f13;  
pi23 = f24;
```

```
l1 = y11*log(pi11) + y12*log(pi12) + y13*log(pi13)  
    + y21*log(pi21) + y22*log(pi22) + y23*log(pi23);
```

```
model y11 ~ general(l1);  
random u ~ normal(0, sd*sd) subject=center;  
predict alpha + u out=pred;  
run;
```


Journals Offprint Order Form



A member of the
Hodder Headline
Group

Arnold Journals
338 Euston Road
London NW1 3BH
Tel: +44 (0)20 7873 6000
Fax: +44 (0)20 7873 6376

This form should be returned at once to the above address

- Title of Journal:
- 1st Author:

Statistical Modelling 3(4)
[provisional]

A. FREE OFFPRINTS - 25 offprints of your article will be supplied free of charge

Please indicate opposite, the name and full postal address to whom they should be sent. In the case of multi-author articles, free offprints are only sent to the corresponding author.

-

-

-

-

B. PURCHASE OF ADDITIONAL OFFPRINTS

Please note that if an article is by more than one author, only one offprint form is sent and all offprints should be ordered on that form in consultation with the co-authors.

Offprint Price List (£ sterling UK and Europe; US\$ Rest of World)

	25		50		100		150		200	
	£	\$	£	\$	£	\$	£	\$	£	\$
1-4 pages	72	100	94	132	142	200	216	304	260	364
5-8 pages	95	134	125	175	190	267	263	368	342	480
9-16 pages	125	175	142	200	216	304	287	403	380	533
17-24 pages	142	200	165	230	238	333	340	477	432	605
Extra 8 pages	14	19	21	29	25	35	30	42	43	60

For larger quantities contact the publisher for a quotation.

Add 100% for any offprints including colour reproduction.

I wish to purchaseadditional offprints

ADDRESS FOR DELIVERY

(please print in capitals)

IMPORTANT

1. Cheques drawn on a UK or US bank should be made payable to Hodder Headline Group. We are unable to accept credit or debit card payments.
2. Orders will not normally be mailed until the publisher is in receipt of either the appropriate payment or an official purchase order.
3. The above are prepublication prices and apply only to orders received before the publication goes to press.
4. All despatches are by surface mail, normally within four weeks of publication.

ADDRESS FOR INVOICE (please print in capitals)

Payment enclosed

Please invoice

Official order follows

5. Claims cannot be considered more than three months after despatch.

VAT will be added to UK invoices. Members of the EU will be required to pay VAT unless a VAT number is provided with order.

VAT Number:

Official order attached

Order no.

Signed.....

Date...../...../.....

TRANSFER OF COPYRIGHT

Please return completed form to:

**Arnold Journals
Statistical Modelling
338 Euston Road
London NW1 3BH
UK.**



A member of the
Hodder Headline
PLC Group

STATISTICAL MODELLING

In consideration of the publication in the above journal *Statistical Modelling* the contribution entitled
..... ("the
contribution")

by (all authors' names):
.....

1. To be filled in if copyright belongs to you

I/we warrant that I am/we are the sole owner/s of the complete copyright in the Contribution and I/we hereby assign to Arnold (Publishers) Limited the complete copyright in the Contribution in all formats and media.

2. To be filled in if copyright does not belong to you

a) Name and address of copyright

holder.....
.....
.....

b) I/we warrant that I am/we are the sole owner/s of the complete copyright in the Contribution and I/we hereby grant Arnold (Publishers) Limited the non-exclusive right to publish the Contribution throughout the world in all formats and media and to deal with requests from third parties in the manner specified in paragraphs 2 and 4 overleaf.

3. To be filled in if US Government exemption applies

I/we certify that the Contribution was written in the course of employment by the United States Government, and therefore copyright protection is not available.

4. I/we warrant that I/we have full power to enter into this Agreement, and that the Contribution does not infringe any existing copyright, or contain any scandalous, defamatory, libellous or unlawful matter.

- Signed as (tick one) the sole author(s) of the Contribution
 one author authorised to execute this transfer on behalf of all the authors of the Contribution
 the copyright holder or authorised agent of the copyright holder of the Contribution

Name (block letters)
.....

Address

.....

Signature Date

.....

(Additional authors should provide this information on a separate sheet please)

Notes for Contributors

1. The Journal's policy is to acquire copyright in all contributions. There are two reasons for this: (a) ownership of copyright by one central organisation tends to make it easier to maintain effective international protection against unauthorised use; (b) it also allows for requests from third parties to reprint or reproduce a contribution, or part of it, to be handled in accordance with a general policy which is sensitive both to any relevant changes in international copyright legislation and to the general desirability of encouraging the dissemination of knowledge.
2. Arnold co-operates in various licensing schemes which allow organisations to copy material within agreed restraints (e.g. the CLA in the UK and the CCC in the USA).
3. All contributors retain the rights to reproduce their paper for their own purposes provided no sale is involved, and to reprint their paper in any volume of which they are editor or author. Permission will automatically be given to the publisher of such a volume, subject to the normal acknowledgement.
4. It is understood that in some cases copyrights will be held by the contributor's employer. If so, Arnold requires non-exclusive permission to deal with requests from third parties, on the understanding that any requests it receives will be handled in accordance with paragraph 3.
5. Arnold will provide each contributor with a complimentary copy of the issue of the Journal in which the Contribution appears.