Generalized Linear Mixed Models

Introduction

Generalized linear models (GLMs) represent a class of fixed effects regression models for several types of dependent variables (i.e., continuous, dichotomous, counts). McCullagh and Nelder [32] describe these in great detail and indicate that the term 'generalized linear model' is due to Nelder and Wedderburn [35] who described how a collection of seemingly disparate statistical techniques could be unified. Common Generalized linear models (GLMs) include linear regression, **logistic regression**, and Poisson regression.

There are three specifications in a GLM. First, the linear predictor, denoted as η_i, of a GLM is of the form

$$\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}, \tag{1}$$

where $\boldsymbol{x}_i$ is the vector of regressors for unit i with fixed effects $\boldsymbol{\beta}$. Then, a link function $g(\cdot)$ is specified which converts the expected value μ_i of the outcome variable Y_i (i.e., $\mu_i = E[Y_i]$) to the linear predictor η_i

$$g(\mu_i) = \eta_i. \tag{2}$$

Finally, a specification for the form of the variance in terms of the mean μ_i is made. The latter two specifications usually depend on the distribution of the outcome Y_i, which is assumed to fall within the exponential family of distributions.

Fixed effects models, which assume that all observations are independent of each other, are not appropriate for analysis of several types of correlated data structures, in particular, for clustered and/or longitudinal data (*see* **Clustered Data**). In clustered designs, subjects are observed nested within larger units, for example, schools, hospitals, neighborhoods, workplaces, and so on. In longitudinal designs, repeated observations are nested within subjects (*see* **Longitudinal Data Analysis** and **Repeated Measures Analysis of Variance**). These are often referred to as multilevel [16] or hierarchical [41] data (*see* **Linear Multilevel Models**), in which the level-1 observations (subjects or repeated observations) are nested within the higher level-2 observations (clusters or subjects). Higher levels are also possible, for example, a three-level design could have repeated observations (level-1) nested within subjects (level-2) who are nested within clusters (level-3).

For analysis of such multilevel data, random cluster and/or subject effects can be added into the regression model to account for the correlation of the data. The resulting model is a mixed model including the usual fixed effects for the regressors plus the random effects. Mixed models for continuous normal outcomes have been extensively developed since the seminal paper by Laird and Ware [28]. For nonnormal data, there have also been many developments, some of which are described below. Many of these developments fall under the rubric of generalized linear mixed models (GLMMs), which extend GLMs by the inclusion of random effects in the predictor. Agresti et al. [1] describe a variety of social science applications of GLMMs; [12], [33], and [11] are recent texts with a wealth of statistical material on GLMMs.

Let i denote the level-2 units (e.g., subjects) and let j denote the level-1 units (e.g., nested observations). The focus will be on longitudinal designs here, but the methods apply to clustered designs as well. Assume there are $i = 1, \ldots, N$ subjects (level-2 units) and $j = 1, \ldots, n_i$ repeated observations (level-1 units) nested within each subject. A random-intercept model, which is the simplest mixed model, augments the linear predictor with a single random effect for subject i,

$$\eta_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \nu_i, \tag{3}$$

where ν_i is the random effect (one for each subject). These random effects represent the influence of subject i on his/her repeated observations that is not captured by the observed covariates. These are treated as random effects because the sampled subjects are thought to represent a population of subjects, and they are usually assumed to be distributed as $\mathcal{N}(0, \sigma_\nu^2)$. The parameter σ_ν^2 indicates the variance in the population distribution, and therefore the degree of heterogeneity of subjects.

Including the random effects, the expected value of the outcome variable, which is related to the linear

Reproduced from the *Encyclopedia of Statistics in Behavioral Science*. © John Wiley & Sons, Ltd.
ISBN: 0-470-86080-4.

predictor via the link function, is given as

$$\mu_{ij} = E[Y_{ij}|\nu_i, \boldsymbol{x}_{ij}]. \quad (4)$$

This is the expectation of the conditional distribution of the outcome given the random effects. As a result, GLMMs are often referred to as conditional models in contrast to the marginal generalized estimating equations (GEE) models (*see* **Generalized Estimating Equations (GEE)**) [29], which represent an alternative generalization of GLMs for correlated data (*see* **Marginal Models for Clustered Data**).

The model can be easily extended to include multiple random effects. For example, in longitudinal problems, it is common to have a random subject intercept and a random linear time-trend. For this, denote $\boldsymbol{z}_{ij}$ as the $r \times 1$ vector of variables having random effects (a column of ones is usually included for the random intercept). The vector of random effects $\boldsymbol{\nu}_i$ is assumed to follow a multivariate normal distribution with mean vector $\mathbf{0}$ and variance–covariance matrix $\boldsymbol{\Sigma}_v$ (*see* **Catalogue of Probability Density Functions**). The model is now written as

$$\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\nu}_i. \quad (5)$$

Note that the conditional mean μ_{ij} is now specified as $E[Y_{ij}|\boldsymbol{\nu}_i, \boldsymbol{x}_{ij}]$, namely, in terms of the vector of random effects.

Dichotomous Outcomes

Development of GLMMs for dichotomous data has been an active area of statistical research. Several approaches, usually adopting a logistic or probit regression model (*see* **Probits**) and various methods for incorporating and estimating the influence of the random effects, have been developed. A review article by Pendergast et al. [37] discusses and compares many of these developments.

The mixed-effects logistic regression model is a common choice for analysis of multilevel dichotomous data and is arguably the most popular GLMM. In the GLMM context, this model utilizes the logit link, namely

$$g(\mu_{ij}) = \text{logit}(\mu_{ij}) = \log\left[\frac{\mu_{ij}}{1-\mu_{ij}}\right] = \eta_{ij}. \quad (6)$$

Here, the conditional expectation $\mu_{ij} = E(Y_{ij}|\boldsymbol{\nu}_i, \boldsymbol{x}_{ij})$ equals $P(Y_{ij} = 1|\boldsymbol{\nu}_i, \boldsymbol{x}_{ij})$, namely, the conditional probability of a response given the random effects (and covariate values).

This model can also be written as

$$P(Y_{ij} = 1|\boldsymbol{\nu}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) = g^{-1}(\eta_{ij}) = \Psi(\eta_{ij}), \quad (7)$$

where the inverse link function $\Psi(\eta_{ij})$ is the logistic cumulative distribution function (cdf), namely $\Psi(\eta_{ij}) = [1 + \exp(-\eta_{ij})]^{-1}$. A nicety of the logistic distribution, that simplifies parameter estimation, is that the probability density function (pdf) is related to the cdf in a simple way, as $\psi(\eta_{ij}) = \Psi(\eta_{ij})[1 - \Psi(\eta_{ij})]$.

The probit model, which is based on the standard normal distribution, is often proposed as an alternative to the logistic model [13]. For the probit model, the normal cdf and pdf replace their logistic counterparts. A useful feature of the probit model is that it can be used to yield **tetrachoric correlations** for the clustered binary responses, and **polychoric correlations** for ordinal outcomes (discussed below). For this reason, in some areas, for example familial studies, the probit formulation is often preferred to its logistic counterpart.

Example

Gruder et al. [20] describe a smoking-cessation study in which 489 subjects were randomized to either a control, discussion, or social support conditions. Control subjects received a self-help manual and were encouraged to watch twenty segments of a daily TV program on smoking cessation, while subjects in the two experimental conditions additionally participated in group meetings and received training in support and relapse prevention. Here, for simplicity, these two experimental conditions will be combined. Data were collected at four telephone interviews: postintervention, and 6, 12, and 24 months later. Smoking abstinence rates (and sample sizes) at these four timepoints were 17.4% (109), 7.2% (97), 18.5% (92), and 18.2% (77) for the placebo condition. Similarly, for the combined experimental condition it was 34.5% (380), 18.2% (357), 19.6% (337), and 21.7% (295) for these timepoints.

Two logistic GLMM were fit to these data: a random intercept and a random intercept and linear trend of time model (*see* **Growth Curve Modeling**). These models were estimated using SAS PROC NLMIXED with adaptive quadrature. For these, it is the probability of smoking abstinence, rather than smoking, that

Table 1 Smoking cessation study: smoking status (0 = smoking, 1 = not smoking) across time (N = 489), GLMM logistic parameter estimates (Est.), standard errors (SE), and P values

	Random intercept model			Random int and trend model		
Parameter	Est.	SE	P value	Est.	SE	P value
Intercept	−2.867	.362	.001	−2.807	.432	.001
Time	.113	.122	.36	−.502	.274	.07
Condition (0 = control; 1 = experimental)	1.399	.379	.001	1.495	.415	.001
Condition by Time	−.322	.136	.02	−.331	.249	.184
Intercept variance	3.587	.600		3.979	1.233	
Intercept Time covariance				.048	.371	
Time variance				1.428	.468	
−2 log likelihood		1631.0			1594.7	

Note: P values not given for variance and covariance parameters (see [41]).

is being modeled. Fixed effects included a condition term (0 = control, 1 = experimental), time (coded 0, 1, 2, and 4 for the four timepoints), and the condition by time interaction. Results for both models are presented in Table 1. Based on a likelihood-ratio test, the model with random intercept and linear time trend is preferred over the simpler random intercept model ($\chi^2_2 = 36.3$). Thus, there is considerable evidence for subjects varying in both their intercepts and time trends. It should be noted that the test statistic does not have a chi-square distribution when testing variance parameters because the null hypothesis is on the border of the parameter space, making the P value conservative. Snijders and Bosker [46] elaborate on this issue and point out that a simple remedy, that has been shown to be reasonable in simulation studies, is to divide the P value based on the likelihood-ratio chi-square test statistic by two. In the present case, it doesn't matter because the P value is $<.001$ for $\chi^2_2 = 36.3$ even without dividing by two.

In terms of the fixed effects, both models indicate a nonsignificant time effect for the control condition, and a highly significant condition effect at time 0 (e.g., $z = 1.495/.415 = 3.6$ in the second model). This indicates a positive effect of the experimental conditions on smoking abstinence relative to control at postintervention. There is also some evidence of a negative condition by time interaction, suggesting that the beneficial condition effect diminishes across time. Note that this interaction is not significant ($P < .18$) in the random intercept and trend model, but it is significant in the random intercept model ($P < .02$). Since the former is preferred by the likelihood-ratio test, we would conclude that the interaction is not significant.

This example shows that the significance of model terms can depend on the structure of the random effects. Thus, one must decide upon a reasonable model for the random effects as well as for the fixed effects. A commonly recommended approach for this is to perform a sequential procedure for model selection. First, one includes all possible covariates of interest into the model and selects between the possible models of random effects using likelihood-ratio tests and model fit criteria. Then, once a reasonable random effects structure is selected, one trims model covariates in the usual way.

IRT Models

Because the logistic model is based on the logistic response function, and the random effects are assumed normally distributed, this model and models closely related to it are often referred to as logistic/normal models, especially in the latent trait model literature [4]. Similarly, the probit model is sometimes referred to as a normal/normal model. In many respects, latent trait or **item response theory** (IRT) models, developed in the educational testing and psychometric literatures, represent some of the earliest GLMMs. Here, item responses ($j = 1, 2, \ldots, n$) are nested within subjects ($i = 1, 2, \ldots, N$). The simplest IRT model is the **Rasch model** [40] which posits the probability of a correct response to the dichotomous item j ($Y_{ij} = 1$) conditional on the random effect or 'ability' of subject i (θ_i) in terms of the logistic cdf as

$$P(Y_{ij} = 1|\theta_i) = \Psi(\theta_i - b_j), \tag{8}$$

where b_j is the threshold or difficulty parameter for item j (i.e., item difficulty). Subject's ability is commonly denoted as θ in the IRT literature (i.e., instead of ν). Note that the Rasch model is simply a random-intercepts model that includes item dummies as fixed regressors. Because there is only one parameter per item, the Rasch model is also called the *one-parameter IRT model*. A more general IRT model, the two-parameter model [5], also includes a parameter for the discrimination of the item in terms of ability.

Though IRT models were not originally cast as GLMMs, formulating them in this way easily allows covariates to enter the model at either level (i.e., items or subjects). This and other advantages of casting IRT models as mixed models are described by Rijmen et al. [43], who provide a comprehensive overview and bridge between IRT models, mixed models, and GLMMs. As they point out, the Rasch model, and variants of it, belong to the class of GLMMs. However, the more extended two-parameter model is not within the class of GLMMs because the predictor is no longer linear, but includes a product of parameters.

Ordinal Outcomes

Extending the methods for dichotomous responses to ordinal response data has also been actively pursued; Agresti and Natarajan [2] review many of these developments. Because the proportional odds model described by McCullagh [31], which is based on the logistic regression formulation, is a common choice for analysis of ordinal data, many of the GLMMs for ordinal data are generalizations of this model, though models relaxing this assumption have also been described [27]. The proportional odds model expresses the ordinal responses in C categories ($c = 1, 2, \ldots, C$) in terms of $C - 1$ cumulative category comparisons, specifically, $C - 1$ cumulative logits (i.e., log odds). Here, denote the conditional cumulative probabilities for the C categories of the outcome Y_{ij} as $P_{ijc} = P(Y_{ij} \leq c|\boldsymbol{\nu}_i, \boldsymbol{x}_{ij}) = \sum_{c=1}^{C} p_{ijc}$, where p_{ijc} represents the conditional probability of response in category c. The logistic GLMM for the conditional cumulative probabilities $\mu_{ijc} = P_{ijc}$ is given in terms of the cumulative logits as

$$\log\left[\frac{\mu_{ijc}}{1-\mu_{ijc}}\right] = \eta_{ijc} \quad (c = 1, \ldots, C-1), \quad (9)$$

where the linear predictor is now

$$\eta_{ijc} = \gamma_c - [\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\nu}_i], \quad (10)$$

with $C - 1$ strictly increasing model thresholds γ_c (i.e., $\gamma_1 < \gamma_2 \cdots < \gamma_{C-1}$). The thresholds allow the cumulative response probabilities to differ. For identification, either the first threshold γ_1 or the model intercept β_0 is typically set to zero. As the regression coefficients $\boldsymbol{\beta}$ do not carry the c subscript, the effects of the regressors do not vary across categories. McCullagh [31] calls this assumption of identical odds ratios across the $C - 1$ cutoffs the proportional odds assumption.

Because the ordinal model is defined in terms of the cumulative probabilities, the conditional probability of a response in category c is obtained as the difference of two conditional cumulative probabilities:

$$P(Y_{ij} = c|\boldsymbol{\nu}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) = \Psi(\eta_{ijc}) - \Psi(\eta_{ij,c-1}). \quad (11)$$

Here, $\gamma_0 = -\infty$ and $\gamma_C = \infty$, and so $\Psi(\eta_{ij0}) = 0$ and $\Psi(\eta_{ijC}) = 1$ (*see* **Ordinal Regression Models**).

Example

Hedeker and Gibbons [25] described a random-effects ordinal probit regression model, examining longitudinal data collected in the NIMH Schizophrenia Collaborative Study on treatment related changes in overall severity. The dependent variable was item 79 of the Inpatient Multidimensional Psychiatric Scale (IMPS; [30]), scored as: (a) normal or borderline mentally ill, (b) mildly or moderately ill, (c) markedly ill, and (d) severely or among the most extremely ill. In this study, patients were randomly assigned to receive one of four medications: placebo, chlorpromazine, fluphenazine, or thioridazine. Since previous analyses revealed similar effects for the three antipsychotic drug groups, they were combined

Table 2 Experimental design and weekly sample sizes

	Sample size at week						
Group	0	1	2	3	4	5	6
Placebo (n = 108)	107	105	5	87	2	2	70
Drug (n = 329)	327	321	9	287	9	7	265

Note: Drug = Chlorpromazine, Fluphenazine, or Thioridazine.

in the analysis. The experimental design and corresponding sample sizes are listed in Table 2.

As can be seen from Table 2, most of the measurement occurred at weeks 0, 1, 3, and 6, with some scattered measurements at the remaining timepoints.

Here, a logistic GLMM with random intercept and trend was fit to these data using SAS PROC NLMIXED with adaptive quadrature. Fixed effects included a dummy-coded drug effect (placebo = 0 and drug = 1), a time effect (square root of week; this was used to linearize the relationship between the cumulative logits and week) and a drug by time interaction. Results from this analyses are given in Table 3.

The results indicate that the treatment groups do not significantly differ at baseline (drug effect), the placebo group does improve over time (significant negative time effect), and the drug group has greater improvement over time relative to the placebo group (significant negative drug by time interaction). Thus, the analysis supports use of the drug, relative to placebo, in the treatment of schizophrenia.

Comparing this model to a simpler random-intercepts model (not shown) yields clear evidence of significant variation in both the individual intercept and time-trends (likelihood-ratio $\chi^2_2 = 77.7$). Also, a moderate negative association between the intercept and linear time terms is indicated, expressed as a correlation it equals $-.40$, suggesting that those patients with the highest initial severity show the greatest improvement across time (e.g., largest negative time-trends). This latter finding could be a result of a 'floor effect', in that patients with low initial severity scores cannot exhibit large negative time-trends due to the limited range in the ordinal outcome variable. Finally, comparing this model to one that allows nonproportional odds for all model covariates (not shown) supports the proportional odds assumption ($\chi^2_6 = 3.63$). Thus, the three covariates (drug, time, and drug by time) have similar effects on the three cumulative logits.

Survival Analysis Models

Connections between ordinal regression and survival analysis models (*see* **Survival Analysis**) have led to developments of discrete and grouped-time survival analysis GLMMs [49]. The basic notion is that the time to the event can be considered as an ordinal variable with C possible event times, albeit with right-censoring accommodated. Vermunt [50] also describes related log-linear mixed models for survival analysis or **event history analysis**.

Nominal Outcomes

Nominal responses occur when the categories of the response variable are not ordered. General regression models for multilevel nominal data have been considered, and Hartzel et al. [22] synthesizes much of the work in this area, describing a general mixed-effects model for both clustered ordinal and nominal responses.

Table 3 NIMH Schizophrenia Collaborative Study: severity of illness (IMPS79) across time ($N = 437$), GLMM logistic parameter estimates (Est.), standard errors (SE), and P values

Parameter	Est.	SE	P value
Intercept	7.283	.467	.001
Time (sqrt week)	−.879	.216	.001
Drug (0 = placebo; 1 = drug)	.056	.388	.88
Drug by Time	−1.684	.250	.001
Threshold 2	3.884	.209	.001
Threshold 3	6.478	.290	.001
Intercept variance	6.847	1.282	
Intercept-time covariance	−1.447	.515	
Time variance	1.949	.404	
−2 log likelihood		3326.5	

Note: Threshold 1 set to zero for identification. P values not given for variance and covariance parameters (see [41]). NIMH = National Institute of Mental Health; IMPS79 = Inpatient Multidimensional Psychiatric Scale, Item 79.

In the nominal GLMM, the probability that $Y_{ij} = c$ (a response occurs in category c) for a given individual i, conditional on the random effects $\boldsymbol{v}$, is given by:

$$p_{ijc} = P(Y_{ij} = c|\boldsymbol{v}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) = \frac{\exp(\eta_{ijc})}{1 + \sum_{h=1}^{C} \exp(\eta_{ijh})} \text{ for } c = 2, 3, \ldots C, \quad (12)$$

$$p_{ij1} = P(Y_{ij} = 1|\boldsymbol{v}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) = \frac{1}{1 + \sum_{h=1}^{C} \exp(\eta_{ijh})}, \quad (13)$$

with the linear predictor $\eta_{ijc} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_c + \boldsymbol{z}'_{ij}\boldsymbol{v}_{ic}$. Both the regression coefficients $\boldsymbol{\beta}_c$ and the random-effects carry the c subscript; the latter allows the variance–covariance matrix $\boldsymbol{\Sigma}_{v_c}$ to vary across categories. In the model above, these parameters represent differences relative to the first category. The nominal model can also be written to allow for any possible set of $C - 1$ contrasts, see [24] for an example of this.

Ranks

In ranking data, individuals are asked to rank C distinct options with respect to some criterion. If the individuals are only asked to provide the option with the highest (or lowest) rank of the C categories, then the resulting data consist of either an ordinal outcome (if the C options are ordered) or a nominal outcome (if the C options are not ordered), and analysis can proceed using the models described above. In the more general case, individuals are asked for, say, the top three options, or to fully rank the C options from the 'best' to the 'worst' (i.e., all options receive a rank from 1 to C). The former case consists of partial ranking data, while the latter case represents full ranking data. As these data types are generalizations of nominal and ordinal data types, it is not surprising that statistical models for ranking data are generalizations of the models for ordinal and nominal models described above. In particular, since the C options are usually not ordered options, models for ranking data have close connections with models for nominal outcomes. GLMMs for ranking data are described in [6] and [45]. These articles show the connections between models for multilevel nominal and ranking data, as well as develop several extensions for the latter.

Counts

For count data, various types of Poisson mixed models have been proposed. A review of some of these methods applied to longitudinal Poisson data is given in [47]. For computational purposes, it is convenient for the univariate random effects to have a gamma distribution in the population of subjects [3]. However, as described in [11], adding multiple normally distributed random effects on the same scale as the fixed effects of the Poisson regression model provides a more general and flexible model.

Let Y_{ij} be the value of the count variable (where Y_{ij} can equal $0, 1, \ldots$) associated with individual i and timepoint j. If this count is assumed to be drawn from a Poisson distribution, then the mixed Poisson regression model indicates the expected number of counts as

$$\log \mu_{ij} = \eta_{ij}, \quad (14)$$

with the linear predictor $\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i$. In some cases the size of the time interval over which the events are counted varies. For example, McKnight and Van Den Eeden [34] describe a study in which the number of headaches in a week is recorded, however, not all individuals are measured for all seven days. For this, let t_{ij} represent the follow-up time associated with units i and j. The linear predictor is now augmented as

$$\eta_{ij} = \log t_{ij} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i, \quad (15)$$

which can also be expressed as

$$\mu_{ij} = t_{ij} \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i) \quad (16)$$

or $\mu_{ij}/t_{ij} = \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i)$ to reflect that it is the number of counts per follow-up period that is being modeled. The term $\log t)_{ij}$ is often called an *offset*.

Assuming the Poisson process for the count Y_{ij}, the probability that $Y_{ij} = y$, conditional on the random effects $\boldsymbol{v}$, is given as

$$P(Y_{ij} = y|\boldsymbol{v}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) = \exp(-\mu_{ij})\frac{(\mu_{ij})^y}{y!}. \quad (17)$$

It is often the case that count data exhibit more zero counts than what is consistent with the Poisson distribution. For such situations, zero-inflated Poisson (ZIP) mixed models, which contain a logistic (or probit) regression for the probability of a nonzero response and a Poisson regression for the zero and nonzero counts, have been developed [21]. A somewhat related model is described by Olsen and Schafer [36] who propose a two-part model that includes a logistic model for the probability of a nonzero response and a conditional linear model for the mean response given that it is nonzero.

Estimation

Parameter estimation in GLMMs typically involves **maximum likelihood** (ML) or variants of ML. Additionally, the solutions are usually iterative ones that can be numerically quite intensive. Here, the solution is merely sketched; further details can be found in [33] and [12].

For the models presented, (7), (11), (12)–(13), and (17), indicate the probability of a level-1 response Y_{ij} for a given subject i at timepoint j, conditional on the random effects $\boldsymbol{v}_i$. While the form of this probability depends on the form of the response variable, let $P(Y_{ij}|\boldsymbol{v}_i)$ represent the conditional probability for any of these forms. Here, for simplicity, we omit conditioning on the covariates $\boldsymbol{x}_{ij}$. Let $\boldsymbol{Y}_i$ denote the vector of responses from subject i. The probability of any response pattern $\boldsymbol{Y}_i$ (of size n_i), conditional on $\boldsymbol{v}_i$, is equal to the product of the probabilities of the level-1 responses:

$$\ell(\boldsymbol{Y}_i|\boldsymbol{v}_i) = \prod_{i=1}^{n_i} P(Y_{ij}|\boldsymbol{v}_i). \tag{18}$$

The assumption that a subject's responses are independent given the random effects (and therefore can be multiplied to yield the conditional probability of the response vector) is known as the **conditional independence** assumption. The marginal density of $\boldsymbol{Y}_i$ in the population is expressed as the following integral of the conditional likelihood $\ell(\cdot)$

$$h(\boldsymbol{Y}_i) = \int_{\boldsymbol{v}_i} \ell(\boldsymbol{Y}_i|\boldsymbol{v}_i) f(\boldsymbol{v}_i)\, \mathrm{d}\boldsymbol{v}_i, \tag{19}$$

where $f(\boldsymbol{v}_i)$ represents the distribution of the random effects, often assumed to be a multivariate normal density. Whereas (18) represents the conditional probability, (19) indicates the unconditional probability for the response vector of subject i. The marginal log-likelihood from the sample of N subjects is then obtained as $\log L = \sum_i^N \log h(\boldsymbol{Y}_i)$. Maximizing this log-likelihood yields ML estimates (which are sometimes referred to as maximum marginal likelihood estimates) of the regression coefficients $\boldsymbol{\beta}$ and the variance-covariance matrix of the random effects $\boldsymbol{\Sigma}_{v_i}$.

Integration over the random-effects distribution

In order to solve the likelihood solution, integration over the random-effects distribution must be performed. As a result, estimation is much more complicated than in models for continuous normally distributed outcomes where the solution can be expressed in closed form. Various approximations for evaluating the integral over the random-effects distribution have been proposed in the literature; many of these are reviewed in [44]. Perhaps the most frequently used methods are based on first- or second-order Taylor expansions. Marginal quasi-likelihood (MQL) involves expansion around the fixed part of the model, whereas penalized or predictive quasi-likelihood (PQL) additionally includes the random part in its expansion [17]. Unfortunately, these procedures yield estimates of the regression coefficients and random effects variances that are biased towards zero in certain situations, especially for the first-order expansions [7].

More recently, Raudenbush et al. [42] proposed an approach that uses a combination of a fully multivariate Taylor expansion and a Laplace approximation. This method yields accurate results and is computationally fast. Also, as opposed to the MQL and PQL approximations, the deviance obtained from this approximation can be used for likelihood-ratio tests.

Numerical integration can also be used to perform the integration over the random-effects distribution. Specifically, if the assumed distribution is normal, Gauss–Hermite quadrature can approximate the above integral to any practical degree of accuracy. Additionally, like the Laplace approximation, the numerical quadrature approach yields a deviance that can be readily used for likelihood-ratio tests. The integration is approximated by a summation on a specified number of quadrature points for each dimension of the integration. An issue with the quadrature approach is that it can involve summation

over a large number of points, especially as the number of random-effects is increased. To address this, methods of adaptive quadrature have been developed which use a few points per dimension that are adapted to the location and dispersion of the distribution to be integrated [39].

More computer-intensive methods, involving iterative simulations, can also be used to approximate the integration over the random effects distribution. Such methods fall under the rubric of **Markov chain Monte Carlo** (MCMC; [15]) algorithms. Use of MCMC for estimation of a wide variety of models has exploded in the last 10 years or so; MCMC solutions for GLMMs are described in [9].

Estimation of random effects

In many cases, it is useful to obtain estimates of the random effects. The random effects $\boldsymbol{\nu}_i$ can be estimated using empirical Bayes methods (*see* **Random Effects in Multivariate Linear Models: Prediction**). For the univariate case, this estimator $\hat{v}_i$ is given by:

$$\hat{v}_i = E(\nu_i|\boldsymbol{Y}_i) = h_i^{-1} \int_{\nu_i} \nu_i \ell_i f(\nu_i)\, \mathrm{d}\nu_i \qquad (20)$$

where ℓ_i is the conditional probability for subject i under the particular model and h_i is the analogous marginal probability. This is simply the mean of the posterior distribution. Similarly, the variance of the posterior distribution is obtained as

$$V(\hat{v}_i|\boldsymbol{Y}_i) = h_i^{-1} \int_{\nu_i} (\nu_i - \hat{v}_i)^2 \ell_i f(\nu_i)\, \mathrm{d}\nu_i. \qquad (21)$$

These quantities may then be used, for example, to evaluate the response probabilities for particular subjects (e.g., person-specific trend estimates). Also, Ten Have [48] suggests how these empirical Bayes estimates can be used in performing residual diagnostics.

Discussion

Though the focus here has been on two-level GLMMs for nonnormal data, three-level (and higher) generalizations have also been considered in the literature [14]. Also, software for fitting GLMMs is readily available in the major statistical packages (i.e., SAS PROC NLMIXED, STATA) and in several independent programs (HLM, [8]; EGRET, [10]; MLwiN, [18]; LIMDEP, [19]; MIXOR, [26]; MIXNO, [23]; GLLAMM, [38]). Not all of these programs fit all of the GLMMs described here; some only allow random-intercepts models or two-level models, for example, and several vary in terms of how the integration over the random effects is performed. However, though the availability of these software programs is relatively recent, they have definitely facilitated application of GLMMs in psychology and elsewhere. The continued development of these models and their software implementations should only lead to greater use and understanding of GLMMs for analysis of correlated nonnormal data.

Acknowledgments

Thanks are due to Dr. Robin Mermelstein for use of the smoking-cessation study data, and to Drs. Nina Schooler and John Davis for use of the schizophrenia study data. This work was supported by National Institutes of Mental Health Grant MH56146.

References

[1] Agresti, A., Booth, J.G., Hobart, J.P. & Caffo, B. (2000). Random-effects modeling of categorical response data, *Sociological Methodology* **30**, 27–80.

[2] Agresti, A. & Natarajan, R. (2001). Modeling clustered ordered categorical data: a survey, *International Statistical Review* **69**, 345–371.

[3] Albert, J. (1992). A Bayesian analysis of a Poisson random effects model for home run hitters, *The American Statistician* **46**, 246–253.

[4] Bartholomew, D.J. & Knott, M. (1999). *Latent Variable Models and Factor Analysis*, 2nd Edition, Oxford University Press, New York.

[5] Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability, in *Statistical Theories of Mental Test Scores*, F.M. Lord & M.R. Novick, eds, Addison-Wesley, Reading.

[6] Böckenholt, U. (2001). Mixed-effects analyses of rank-ordered data, *Psychometrika* **66**, 45–62.

[7] Breslow, N.E. & Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion, *Biometrika* **82**, 81–91.

[8] Bryk, A.S., Raudenbush, S.W. & Congdon, R. (2000). *HLM Version 5*, Scientific Software International, Chicago.

[9] Clayton, D. (1996). Generalized linear mixed models, in *Markov Chain Monte Carlo Methods in Practice*, W.R. Gilks, S. Richardson & D.J. Spiegelhalter, eds, Chapman & Hall, New York, pp. 275–303.

[10] Corcoran, C., Coull, B. & Patel, A. (1999). *EGRET for Windows User Manual*, CYTEL Software Corporation, Cambridge.

[11] Diggle, P., Heagerty, P., Liang, K.-Y. & Zeger, S.L. (2002). *Analysis of Longitudinal Data*, 2nd Edition, Oxford University Press, New York.

[12] Fahrmeir, L. & Tutz, G.T. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd Edition, Springer-Verlag, New York.

[13] Gibbons, R.D. & Bock, R.D. (1987). Trend in correlated proportions, *Psychometrika* **52**, 113–124.

[14] Gibbons, R.D. & Hedeker, D. (1997). Random-effects probit and logistic regression models for three-level data, *Biometrics* **53**, 1527–1537.

[15] Gilks, W., Richardson, S. & Spiegelhalter, D.J. (1997). *Markov Chain Monte Carlo in Practice*, Chapman & Hall, New York.

[16] Goldstein, H. (1995). *Multilevel Statistical Models*, 2nd Edition, Halstead Press, New York.

[17] Goldstein, H. & Rasbash, J. (1996). Improved approximations for multilevel models with binary responses, *Journal of the Royal Statistical Society, Series B* **159**, 505–513.

[18] Goldstein, H. Rasbash, J. Plewis, I. Draper, D. Browne, W. & Wang, M. (1998). *A User's Guide to MLwiN*, University of London, Institute of Education, London.

[19] Greene, W.H. (1998). *LIMDEP Version 7.0 User's Manual*, (revised edition), Econometric Software, Plainview.

[20] Gruder, C.L., Mermelstein, R.J., Kirkendol, S., Hedeker, D., Wong, S.C., Schreckengost, J., Warnecke, R.B., Burzette, R. & Miller, T.Q. (1993). Effects of social support and relapse prevention training as adjuncts to a televised smoking cessation intervention, *Journal of Consulting and Clinical Psychology* **61**, 113–120.

[21] Hall, D.B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study, *Biometrics* **56**, 1030–1039.

[22] Hartzel, J., Agresti, A. & Caffo, B. (2001). Multinomial logit random effects models, *Statistical Modelling* **1**, 81–102.

[23] Hedeker, D. (1999). MIXNO: a computer program for mixed-effects nominal logistic regression, *Journal of Statistical Software* **4**(5), 1–92.

[24] Hedeker, D. (2003). A mixed-effects multinomial logistic regression model, *Statistics in Medicine*, **22** 1433–1446.

[25] Hedeker, D. & Gibbons, R.D. (1994). A random-effects ordinal regression model for multilevel analysis, *Biometrics* **50**, 933–944.

[26] Hedeker, D. & Gibbons, R.D. (1996). MIXOR: a computer program for mixed-effects ordinal probit and logistic regression analysis, *Computer Methods and Programs in Biomedicine* **49**, 157–176.

[27] Hedeker, D. & Mermelstein, R.J. (1998). A multilevel thresholds of change model for analysis of stages of change data, *Multivariate Behavioral Research* **33**, 427–455.

[28] Laird, N.M. & Ware, J.H. (1982). Random-effects models for longitudinal data, *Biometrics* **38**, 963–974.

[29] Liang, K.-Y. & Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**, 13–22.

[30] Lorr, M. & Klett, C.J. (1966). *Inpatient Multidimensional Psychiatric Scale: Manual*, Consulting Psychologists Press, Palo Alto.

[31] McCullagh, P. (1980). Regression models for ordinal data (with discussion), *Journal of the Royal Statistical Society, Series B* **42**, 109–142.

[32] McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*, 2nd Edition, Chapman & Hall, New York.

[33] McCulloch, C.E. & Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*, Wiley, New York.

[34] McKnight, B. & Van Den Eeden, S.K. (1993). A conditional analysis for two-treatment multiple period crossover designs with binomial or Poisson outcomes and subjects who drop out, *Statistics in Medicine* **12**, 825–834.

[35] Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized linear models, *Journal of the Royal Statistical Society, Series A* **135**, 370–384.

[36] Olsen, M.K. & Schafer, J.L. (2001). A two-part random effects model for semicontinuous longitudinal data, *Journal of the American Statistical Association* **96**, 730–745.

[37] Pendergast, J.F., Gange, S.J., Newton, M.A., Lindstrom, M.J., Palta, M. & Fisher, M.R. (1996). A survey of methods for analyzing clustered binary response data, *International Statistical Review* **64**, 89–118.

[38] Rabe-Hesketh, S. Pickles, A. & Skrondal, A. (2001). GLLAMM Manual, Technical Report 2001/01, Institute of Psychiatry, King's College, University of London, Department of Biostatistics and Computing.

[39] Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature, *The Stata Journal* **2**, 1–21.

[40] Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*, Danish Institute of Educational Research, Copenhagen.

[41] Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical Linear Models in Social and Behavioral Research: Applications and Data-Analysis Methods*, 2nd Edition, Sage Publications, Thousand Oaks.

[42] Raudenbush, S.W., Yang, M.-L. & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation, *Journal of Computational and Graphical Statistics* **9**, 141–157.

[43] Rijmen, F., Tuerlinckx, F., De Boeck, P. & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory, *Psychological Methods* **8**, 185–205.

[44] Rodríguez, G. & Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses, *Journal of the Royal Statistical Society, Series A* **158**, 73–89.

[45] Skrondal, A. & Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings, *Psychometrika* **68**, 267–287.

[46] Snijders, T. & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, Sage Publications, Thousand Oaks.

[47] Stukel, T.A. (1993). Comparison of methods for the analysis of longitudinal interval count data, *Statistics in Medicine* **12**, 1339–1351.

[48] Ten Have, T.R. (1996). A mixed effects model for multivariate ordinal response data including correlated discrete failure times with ordinal responses, *Biometrics* **52**, 473–491.

[49] Ten Have, T.R. & Uttal, D.H. (1994). Subject-specific and population-averaged continuation ratio logit models for multiple discrete time survival profiles, *Applied Statistics* **43**, 371–384.

[50] Vermunt, J.K. (1997). *Log-linear Models for Event Histories*, Sage Publications, Thousand Oaks.

DONALD HEDEKER