

Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models

Oliver Schabenberger, SAS Institute Inc., Cary, NC

ABSTRACT

This paper describes a new SAS/STAT[®] procedure for fitting models to non-normal or normal data with correlations or nonconstant variability. The GLIMMIX procedure is an add-on for the SAS/STAT product in SAS[®] 9.1 on the Windows platform. PROC GLIMMIX extends the SAS mixed model tools in a number of ways. For example, it

- models data from non-Gaussian distributions
- implements low-rank smoothing based on mixed models
- provides new features for LS-means comparisons and display
- enables you to use SAS programming statements to compute model effects, or to define link and variance functions
- fits models to multivariate data in which observations do not all have the same distribution or link

Applications of the GLIMMIX procedure include estimating trends in disease rates, modeling counts or proportions over time in a clinical trial, predicting probability of occurrence in time series and spatial data, and joint modeling of correlated binary and continuous data.

This paper describes generalized linear mixed models and how to use the GLIMMIX procedure for estimation, inference, and prediction.

INTRODUCTION

The GLIMMIX procedure is a new procedure in SAS/STAT software. It is an add-on for the SAS/STAT product in SAS 9.1 on the Windows platform. It is currently downloadable for the SAS 9.1 release from Software Downloads at support.sas.com.

PROC GLIMMIX performs estimation and statistical inference for generalized linear mixed models (GLMMs). A generalized linear mixed model is a statistical model that extends the class of generalized linear models (GLMs) by incorporating normally distributed random effects. A GLM can be defined in terms of several model components:

- a linear predictor η that is a linear combination of regression coefficients: $\eta_i = \mathbf{x}'_i\boldsymbol{\beta}$
- a link function $g(\cdot)$ that relates the mean of the data to the linear predictor, $g(E[Y_i]) = \eta_i$
- a response distribution for Y_i from the exponential family of distributions

The exponential family of distributions is very broad and contains many important distributions. For example, the binary, binomial, Poisson, negative binomial, normal, beta, gamma, and inverse Gaussian distribution are members of this family. A special case of the generalized linear model arises when the Y_i are normally distributed and the link function is the identity function. The resulting models are linear regression and analysis of variance models with normal errors.

Generalized linear models apply when the data are uncorrelated. In many studies, observations exhibit some form of dependency. For example, measurements of different attributes are taken from the same patient, observations are collected over time, sampling or randomization is carried out hierarchically, and so forth.

The models fit by the GLIMMIX procedure extend the GLM by incorporating correlations among the responses. This can be accomplished by including random effects in the linear predictor and/or by modeling the correlations among the data directly. The GLIMMIX procedure distinguishes the two approaches as “G-side” and “R-side” random effects. This terminology draws on a common specification of the linear mixed model,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}$$

where the random effects $\boldsymbol{\gamma}$ have a normal distribution with mean $\mathbf{0}$ and variance matrix \mathbf{G} . The distribution of the errors \mathbf{e} is normal with mean $\mathbf{0}$ and variance \mathbf{R} . Modeling with G-side effects, you specify the columns of the \mathbf{Z} matrix and the structure of \mathbf{G} . Modeling with R-side effects, you directly specify the covariance structure of the \mathbf{R} matrix.

In a generalized linear mixed model, G-side random effects are constructed similarly by adding random effects to the linear predictor. This leads to a model of the form

$$g(E[Y|\boldsymbol{\gamma}]) = \mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\boldsymbol{\gamma}$$

where $\boldsymbol{\gamma}$ is again normally distributed with mean $\mathbf{0}$ and variance matrix \mathbf{G} . Instead of specifying a distribution for Y , as in the case of a GLM, you now specify a distribution for the conditional response, $Y|\boldsymbol{\gamma}$. This formulation is also known as the conditional model specification.

A model with only R-side random effects is also known as a marginal model because there are no random effects on which you can condition the response. In a marginal model, you specify the mean

$$g(E[Y]) = g(\mu) = \mathbf{x}'\boldsymbol{\beta}$$

and a model for the covariances among the Y_i . The distributional assumption remains relevant because the mean and the variance are functionally related for most distributions in the exponential family. For example, for Poisson data, $\text{var}[Y] = E[Y] = \mu$. If \mathbf{A} is a diagonal matrix containing the variance functions, then the variance matrix in a model with only R-side random components is

$$\text{var}[Y] = \mathbf{A}^{1/2}\mathbf{R}\mathbf{A}^{1/2}$$

You can also combine G-side and R-side random effects in the same model. For example, in a study in which randomly sampled patients are measured repeatedly over time, you can model the patient effects by including random intercepts in the model; these are G-side components because they contribute to the linear predictor. For a given patient the correlations over time can be modeled with an R-side autoregressive structure, for example.

Combining these elements, the models fit by the GLIMMIX procedure can be represented as follows:

$$\begin{aligned} E[Y|\boldsymbol{\gamma}] &= g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}) = g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu} \\ \text{var}[\boldsymbol{\gamma}] &= \mathbf{G} \\ \text{var}[Y|\boldsymbol{\gamma}] &= \mathbf{A}^{1/2}\mathbf{R}\mathbf{A}^{1/2} \end{aligned}$$

where $g^{-1}(\cdot)$ is the inverse link function.

The class of generalized linear mixed models thus contains several other important types of statistical models. For example,

- Linear models: no random effects, identity link function, and normal distribution
- Generalized linear models: no random effects present

- Linear mixed models: random effects, identity link function, and normal distribution

Another class of mixed models with nonlinear components is known as the hierarchical linear mixed models (HLMM). One of the characteristics of these models is that the random effects effects are not necessarily normally distributed. GLMMs, on the other hand, only consider normally distributed random effects. The normal random effects can have a hierarchical (nested) structure, however. In other words, the GLIMMIX procedure can fit models for non-normal data with hierarchical random effects, provided that the random effects have a normal distribution.

EXAMPLE: LOGISTIC REGRESSION WITH RANDOM INTERCEPTS

The following example demonstrates how to fit a generalized linear mixed model for binomial data with the GLIMMIX procedure.

Researchers investigated the performance of two medical procedures in a multicenter study. They randomly selected 15 centers for inclusion. One of the study goals was to compare the occurrence of side effects for the procedures. In each center n_A patients were randomly selected and assigned to procedure “A,” and n_B patients were randomly assigned to procedure “B.”

The following DATA step creates the data set for the analysis.

```
data multicenter;
  input center group$ n SideEffect @@;
  datalines;
  1 A 32 14    1 B 33 18    2 A 30 4    2 B 28 8
  3 A 23 14    3 B 24 9    4 A 22 7    4 B 22 10
  5 A 20 6     5 B 21 12    6 A 19 1    6 B 20 3
  7 A 17 2     7 B 17 6     8 A 16 7    8 B 15 9
  9 A 13 1     9 B 14 5    10 A 13 3   10 B 13 1
  11 A 11 1    11 B 12 2   12 A 10 1   12 B 9 0
  13 A 9 2     13 B 9 6    14 A 8 1    14 B 8 1
  15 A 7 1     15 B 8 0
  ;
```

The variable group identifies the two procedures, n is the number of patients who received a given procedure in a particular center, and SideEffect is the number of patients who reported side effects.

If Y_{iA} and Y_{iB} denote the number of patients in center i who report side effects for procedures A and B , respectively, then, for a given center, these are independent binomial random variables. To model the probability of side effects for the two drugs, π_{iA} and π_{iB} , you need to account for the fixed group effect and the random selection of centers. One possibility is to assume a model that relates group and center effects linearly to the logit of the probabilities:

$$\log \left\{ \frac{\pi_{iA}}{1 - \pi_{iA}} \right\} = \eta_{iA} = \beta_0 + \beta_A + \gamma_i$$

$$\log \left\{ \frac{\pi_{iB}}{1 - \pi_{iB}} \right\} = \eta_{iB} = \beta_0 + \beta_B + \gamma_i$$

The parameters β_0 , β_A , and β_B are fixed effects and the γ_i are the center effects. The γ_i are random effects because the centers participating in the study are a random sample of all possible centers. Note that these are G-side random effects, since they are part of the linear predictors. The formulation of this model is completed by specifying the conditional distribution of the data and the distribution of the random effects

$$Y_{iA} | \gamma_i \sim \text{binomial}(n_{iA}, \pi_{iA})$$

$$Y_{iB}|\gamma_i \sim \text{binomial}(n_{iB}, \pi_{iB})$$

$$\gamma_i \sim \text{iid } N(0, \sigma_c^2)$$

The following statements request that PROC GLIMMIX fit this model.

```
proc glimmix data=multicenter;
  class center group;
  model SideEffect/n = group / solution;
  random intercept / subject=center;
run;
```

The PROC GLIMMIX statement invokes the procedure. The CLASS statement instructs the procedure to treat the variables center and group as classification variables. The MODEL statement specifies the response variable as a sample proportion using the *events/trials* syntax. In terms of the previous formulas, SideEffect/n corresponds to Y_{iA}/n_{iA} for observations from Group A and to Y_{iB}/n_{iB} for observations from Group B. The SOLUTION option in the MODEL statement requests a listing of the fixed-effects parameter estimates. Because of the *events/trials* syntax, the GLIMMIX procedure defaults to the binomial distribution, with the default logit link. If these defaults are not appropriate, you can change the distribution with the DIST= option and the link function with the LINK= option of the MODEL statement. The RANDOM statement specifies that the linear predictor contains an intercept term that randomly varies at the level of the center effect.

The results of this analysis are shown in the following tables.

The GLIMMIX Procedure	
Model Information	
Data Set	WORK.MULTICENTER
Response Variable (Events)	SideEffect
Response Variable (Trials)	n
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	center
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Figure 1. Model Information Table

PROC GLIMMIX recognizes the variables SideEffect and n as the numerator and denominator in the *events/trials* syntax, respectively. The distribution, conditional on the random center effects, is binomial (Figure 1). The marginal variance matrix is block-diagonal, and observations from the same center form the blocks. The default estimation technique in generalized linear mixed models is residual pseudo-likelihood with a subject-specific expansion (METHOD=RSPL; see the following section).

The “Dimensions” table lists the size of relevant matrices; the “Optimization Information” table provides information about the methods and size of the optimization problem (Figure 2).

Dimensions	
G-side Cov. Parameters	1
Columns in X	3
Columns in Z per Subject	1
Subjects (Blocks in V)	15
Max Obs per Subject	2

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	1
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Profiled
Starting From	Data

Figure 2. Dimensions and Optimization Information Tables

There are three columns in the X matrix, corresponding to an intercept and the two levels of the group variable. For each subject (center), the Z matrix contains only an intercept column. The default optimization technique for generalized linear mixed models is the Quasi-Newton method. Because a residual likelihood technique is used to compute the objective function, only the covariance parameters are participating in the optimization. A lower boundary constraint is placed on the variance component for the random center effect. The solution for this variance cannot be less than zero. With pseudo-likelihood methods, optimization begins with an initial set of pseudo-data. By default the GLIMMIX procedure creates the initial data by applying the link function to adjusted data values. This is indicated in the last row of the “Optimization Information” table.

The “Fit Statistics” table in Figure 3 lists information about the fitted model.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	81.44
Generalized Chi-Square	30.69
Gener. Chi-Square / DF	1.10

Figure 3. Fit Statistics

The likelihood is preceded by the word “Pseudo” to indicate that it is computed from a pseudo-likelihood, rather than the true likelihood (explained in the following section). The ratio of the generalized chi-square statistic and its degrees of freedom is close to 1. This indicates that the variability in these data has been properly modeled, and that there is no residual overdispersion. The generalized chi-square statistic is a quadratic form in the marginal residuals that takes correlations among the data into account.

The variance of the random center intercepts on the logit scale is estimated as $\hat{\sigma}_c^2 = 0.6176$ and can be found in the “Covariance Parameter Estimates” table (Figure 4). The solutions for the fixed intercept and group effect are listed under the heading “Solutions for Fixed Effects.”

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	center	0.6176	0.3181

Solutions for Fixed Effects						
Effect	group	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-0.8071	0.2514	14	-3.21	0.0063
group	A	-0.4896	0.2034	14	-2.41	0.0305
group	B	0

Figure 4. Covariance Parameter Estimates and Fixed Effects Solutions

Because of the fixed-effect parameterization used in the GLIMMIX procedure, the “Intercept” effect is an estimate of $\beta_0 + \beta_B$, and the “A” group effect is an estimate of $\beta_A - \beta_B$, the log-odds ratio. There is a significant difference between the two groups ($p=0.0305$).

You can produce the estimates of the average logits in the two groups and their predictions on the scale of the data with the LSMEANS statement in PROC GLIMMIX.

```
proc glimmix data=multicenter;
  class center group;
  model SideEffect/n = group / solution;
  random intercept / subject=center;
  lsmeans group / cl ilink;
  ods select lsmeans;
run;
```

The LSMEANS statement requests the least-squares means of the group effect on the logit scale. The CL option requests their confidence limits. The ILINK option adds estimates, standard errors, and confidence limits on the mean (probability) scale. The table in Figure 5 displays the results.

The GLIMMIX Procedure								
group Least Squares Means								
group	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
A	-1.2966	0.2601	14	-4.99	0.0002	0.05	-1.8544	-0.7388
B	-0.8071	0.2514	14	-3.21	0.0063	0.05	-1.3462	-0.2679
group Least Squares Means								
group	Mean	Standard Error Mean	Lower Mean	Upper Mean				
A	0.2147	0.04385	0.1354	0.3233				
B	0.3085	0.05363	0.2065	0.4334				

Figure 5. Least-squares Means

The “Estimate” column displays the least-squares mean estimate on the logit scale, and the “Mean” column represents its mapping onto the probability scale. These estimated probabilities are obtained by applying the inverse link function to the least-squares means:

$$\hat{\pi}_A = \frac{1}{1 + \exp\{1.2966\}} = 0.2147$$

$$\hat{\pi}_B = \frac{1}{1 + \exp\{0.8071\}} = 0.3085$$

The standard errors of the estimate of the mean are obtained by applying the delta method. The “Lower” and “Upper” columns are 95% confidence limits for the logits in the two groups. The “Lower Mean” and “Upper Mean” columns are the corresponding confidence limits for the probabilities of side effects. These limits are obtained by inversely linking the confidence bounds on the linear scale, and thus are not symmetric about the estimate of the probabilities.

To produce predicted values of the probability of side effects for each observation, you can use the OUTPUT statement in PROC GLIMMIX.

```
proc glimmix data=multicenter;
  class center group;
  model SideEffect/n = group;
  random intercept / subject=center solution;
  output out=glimmixout pred( blup ilink)=PredProb
    pred(noblup ilink)=PredProb_PA;
run;
proc print data=glimmixout;
run;
```

Two types of predicted vales are requested in the OUTPUT statement. The first type uses the solutions for the random effects (the estimated best linear unbiased predictors (BLUPs) in the final linearized model), and the second type uses predictions based on the fixed effects alone. The ILINK suboption of the PRED keyword requests that the inverse link function be applied to the resulting predictions. This yields predictions of probabilities. The results of these calculations are stored in the variables PredProb and PredProb_PA in the

data set glimmixout. The acronym “PA” is used here to identify a “population-averaged” prediction. Figure 6 displays the contents of this data set.

Obs	center	group	n	Side Effect	Pred Prob	Pred Prob_PA
1	1	A	32	14	0.40773	0.21474
2	1	B	33	18	0.52902	0.30852
3	2	A	30	4	0.17452	0.21474
4	2	B	28	8	0.25647	0.30852
5	3	A	23	14	0.39731	0.21474
6	3	B	24	9	0.51821	0.30852
7	4	A	22	7	0.31129	0.21474
8	4	B	22	10	0.42445	0.30852
9	5	A	20	6	0.35163	0.21474
10	5	B	21	12	0.46946	0.30852
11	6	A	19	1	0.10750	0.21474
12	6	B	20	3	0.16425	0.30852
13	7	A	17	2	0.19633	0.21474
14	7	B	17	6	0.28500	0.30852
15	8	A	16	7	0.40894	0.21474
16	8	B	15	9	0.53027	0.30852
17	9	A	13	1	0.18769	0.21474
18	9	B	14	5	0.27378	0.30852
19	10	A	13	3	0.14662	0.21474
20	10	B	13	1	0.21895	0.30852
21	11	A	11	1	0.13470	0.21474
22	11	B	12	2	0.20254	0.30852
23	12	A	10	1	0.10189	0.21474
24	12	B	9	0	0.15619	0.30852
25	13	A	9	2	0.33255	0.21474
26	13	B	9	6	0.44841	0.30852
27	14	A	8	1	0.14126	0.21474
28	14	B	8	1	0.21160	0.30852
29	15	A	7	1	0.11501	0.21474
30	15	B	8	0	0.17494	0.30852

Figure 6. Predicted Values

For an observation from group A, the two predicted values can be written as

$$\text{PredProb} = E[\widehat{Y}_{iA}|\gamma_i] = \frac{1}{1 + \exp\{-\widehat{\beta}_0 - \widehat{\beta}_A - \widehat{\gamma}_i\}}$$

$$\text{PredProb_PA} = E[\widehat{Y}_{iA}] = \frac{1}{1 + \exp\{-\widehat{\beta}_0 - \widehat{\beta}_A\}}$$

Because the random intercepts vary at the center level, the predicted probabilities are different for each center. For example, the solution for the random effect in center 1 is $\widehat{\gamma}_1 = 0.9233$, and is $\widehat{\gamma}_2 = -0.2573$ in center 2 (Figure 7).

The GLIMMIX Procedure						
Solution for Random Effects						
Effect	Subject	Estimate	Std Err Pred	DF	t Value	Pr > t
Intercept	center 1	0.9233	0.3198	14	2.89	0.0119
Intercept	center 2	-0.2573	0.3591	14	-0.72	0.4854
Intercept	center 3	0.8799	0.3442	14	2.56	0.0228
Intercept	center 4	0.5025	0.3553	14	1.41	0.1791
Intercept	center 5	0.6847	0.3579	14	1.91	0.0764
Intercept	center 6	-0.8199	0.4377	14	-1.87	0.0821
Intercept	center 7	-0.1128	0.4040	14	-0.28	0.7842
Intercept	center 8	0.9283	0.3819	14	2.43	0.0291
Intercept	center 9	-0.1685	0.4327	14	-0.39	0.7029
Intercept	center 10	-0.4647	0.4584	14	-1.01	0.3278
Intercept	center 11	-0.5634	0.4806	14	-1.17	0.2607
Intercept	center 12	-0.8798	0.5351	14	-1.64	0.1224
Intercept	center 13	0.5999	0.4471	14	1.34	0.2010
Intercept	center 14	-0.5082	0.5233	14	-0.97	0.3479
Intercept	center 15	-0.7440	0.5501	14	-1.35	0.1977

Figure 7. Solutions for Random Center Effects

The population-averaged prediction, on the contrary, is a prediction for an average center and thus varies only by the level of the group variable. The values of PredProb_PA in Figure 6 are 0.21474 for observations from group A and 0.30852 for observations from group B; these values do not depend on the center. The center-specific predictions in variable PredProb, on the other hand, take the center effects into account. For the first center, for example, the predicted probabilities of side effects are

$$\frac{1}{1 + \exp\{0.8071 + 0.4896 - 0.9233\}} = \frac{1}{1 + \exp\{0.3734\}} = 0.4077$$

for group A and

$$\frac{1}{1 + \exp\{0.8071 - 0.9233\}} = \frac{1}{1 + \exp\{-0.1162\}} = 0.5290$$

for group B, respectively.

ESTIMATION METHODS FOR GLMMS

As in the case of linear mixed models, most estimation approaches for GLMMs rest on some form of likelihood principle. For example, to obtain maximum likelihood estimates, you would maximize the marginal likelihood

$$L(\beta, \theta, y) = \int f(y|\gamma)p(\gamma) d\gamma$$

where $f(y|\gamma)$ is the conditional distribution of the data, and $p(\gamma)$ is the distribution of the random effects. In the linear case, this integral can be solved in closed form, and the resulting likelihood or restricted likelihood can be maximized directly. When random effects enter in nonlinear form, the integral is not known except in special cases. The two basic solutions to the estimation problem in the nonlinear case can be categorized as methods based on integral approximation and methods based on linearization.

Integral approximation methods approximate the log likelihood function of the GLMM and numerically optimize the approximated function. These methods provide an actual objective function for optimization and enable you to compare nested models with true likelihood ratio tests. These methods are not suited, however, to accommodate a large number of random effects, crossed random effects, multiple subject effects, or complex R-side covariance structures. Integral approximation methods (adaptive Gaussian quadrature, Laplace approximation, and importance sampling) are implemented in the NLMIXED procedure of SAS/STAT.

Linearization methods employ expansions to approximate the model by one based on pseudo-data with fewer nonlinear components. The process of computing the linear approximation must be repeated several times until some criterion stabilizes. The fitting methods based on linearizations are usually doubly iterative. The generalized linear mixed model is approximated by a linear mixed model based on current values of the covariance parameter estimates. The resulting linear mixed model is then fit, which is itself an iterative process. On convergence, the new parameter estimates are used to update the linearization, which results in a new linear mixed model. The process stops when parameter estimates between successive linear mixed model fits change within a specified tolerance only.

The advantages of linearization-based methods include a relatively simple form of the linearized model that typically can be fit based on only the mean and variance in the linearized form. Models for which the joint distribution is difficult—or impossible—to ascertain can be fit with linearization-based approaches. Models with correlated errors, a large number of random effects, crossed random effects, and multiple types of subjects are thus excellent candidates for linearization methods. Furthermore, because the structure of the model fit at each stage is a linear mixed model, REML estimation is possible. The disadvantages of this approach include the absence of a true objective function for the overall optimization process and potentially biased estimates of the covariance parameters, especially for binary data. The objective function to be optimized after each linearization update is dependent on the current pseudo-data. The process can fail at both levels of the double iteration scheme.

This version of the GLIMMIX procedure fits generalized linear mixed models based on linearizations. The default estimation method in PROC GLIMMIX for models containing random effects is a technique known as restricted pseudo-likelihood (RPL) estimation (Wolfinger and O'Connell 1993). This technique employs an expansion around the current estimate of the best linear unbiased predictors of the random effects (METHOD=RSPL). Other methods implemented in the procedure enable you to change from a restricted (residual) pseudo-likelihood to a maximum pseudo-likelihood criterion and to change the locus of the expansion.

The linearization method first constructs a pseudo-model and pseudo-data. Recall that

$$E[\mathbf{Y}|\boldsymbol{\gamma}] = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}) = g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu}$$

where $\boldsymbol{\gamma} \sim N(\mathbf{0}, \mathbf{G})$ and $\text{var}[\mathbf{Y}|\boldsymbol{\gamma}] = \mathbf{A}^{1/2}\mathbf{R}\mathbf{A}^{1/2}$. A first-order Taylor series of $\boldsymbol{\mu}$ about $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\gamma}}$ yields

$$g^{-1}(\boldsymbol{\eta}) \doteq g^{-1}(\tilde{\boldsymbol{\eta}}) + \tilde{\boldsymbol{\Delta}}\mathbf{X}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \tilde{\boldsymbol{\Delta}}\mathbf{Z}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}})$$

where

$$\tilde{\boldsymbol{\Delta}} = \left(\frac{\partial g^{-1}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right)_{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}}$$

is a diagonal matrix of derivatives of the conditional mean evaluated at the expansion locus. Rearranging terms yields the expression

$$\tilde{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\mu} - g^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\boldsymbol{\gamma}} \doteq \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}$$

The left-hand side is the expected value, conditional on γ , of

$$\tilde{\Delta}^{-1}(\mathbf{Y} - g^{-1}(\tilde{\eta})) + \mathbf{X}\tilde{\beta} + \mathbf{Z}\tilde{\gamma} \equiv \mathbf{P}$$

and

$$\text{var}[\mathbf{P}|\gamma] = \tilde{\Delta}^{-1} \mathbf{A}^{1/2} \mathbf{R} \mathbf{A}^{1/2} \tilde{\Delta}^{-1}$$

You can thus consider the model

$$\mathbf{P} = \mathbf{X}\beta + \mathbf{Z}\gamma + \epsilon$$

which is a linear mixed model with pseudo-response \mathbf{P} , fixed effects β , random effects γ , and $\text{var}[\epsilon] = \text{var}[\mathbf{P}|\gamma]$. The GLIMMIX procedure assumes that ϵ has a normal distribution and estimates the parameters of this linear mixed (pseudo-) model. At convergence the pseudo-response \mathbf{P} and the matrices Δ and \mathbf{A} are updated and another linear mixed pseudo-model is fit. The process continues until changes in the fixed-effects and covariance parameter estimates are sufficiently small.

EXAMPLE: JOINT MODELS FOR BINARY AND POISSON DATA

The GLIMMIX procedure can fit models to data where the distribution and/or the link function varies across observations. This is accomplished through the DIST=BYOBS or LINK=BYOBS specification of the MODEL statement.

The data set joint, created in the DATA step below, provides an example of a bivariate outcome variable. It reflects the condition and length of hospital stay for 32 herniorrhaphy patients. These data are produced from data given by Mosteller and Tukey (1977) and reproduced in Hand et al. (1994, p. 390, 391).

```
data joint;
  length dist $7;
  input d$ patient age OKstatus response @@;
  if d = 'B' then dist='Binary'; else dist='Poisson';
  datalines;
B 1 78 1 0 P 1 78 1 9 B 2 60 1 0 P 2 60 1 4
B 3 68 1 1 P 3 68 1 7 B 4 62 0 1 P 4 62 0 35
B 5 76 0 0 P 5 76 0 9 B 6 76 1 1 P 6 76 1 7
B 7 64 1 1 P 7 64 1 5 B 8 74 1 1 P 8 74 1 16
B 9 68 0 1 P 9 68 0 7 B 10 79 1 0 P 10 79 1 11
B 11 80 0 1 P 11 80 0 4 B 12 48 1 1 P 12 48 1 9
B 13 35 1 1 P 13 35 1 2 B 14 58 1 1 P 14 58 1 4
B 15 40 1 1 P 15 40 1 3 B 16 19 1 1 P 16 19 1 4
B 17 79 0 0 P 17 79 0 3 B 18 51 1 1 P 18 51 1 5
B 19 57 1 1 P 19 57 1 8 B 20 51 0 1 P 20 51 0 8
B 21 48 1 1 P 21 48 1 3 B 22 48 1 1 P 22 48 1 5
B 23 66 1 1 P 23 66 1 8 B 24 71 1 0 P 24 71 1 2
B 25 75 0 0 P 25 75 0 7 B 26 2 1 1 P 26 2 1 0
B 27 65 1 0 P 27 65 1 16 B 28 42 1 0 P 28 42 1 3
B 29 54 1 0 P 29 54 1 2 B 30 43 1 1 P 30 43 1 3
B 31 4 1 1 P 31 4 1 3 B 32 52 1 1 P 32 52 1 8
;
```

For each patient, two responses were recorded. A binary response takes on the value one if a patient experienced a routine recovery and the value zero if post-operative intensive care was required. The second response variable is a count variable measuring the length of hospital stay after the operation (in days). The

binary variable OKstatus is a regressor variable that distinguishes patients based on their post-operative physical status ("1" implies better status) and the variable age is a patient's age.

You could model these data with a separate logistic model for the binary outcome and, for example, a Poisson model for the count outcome. Such separate analyses would not take into account the correlation between the two outcomes, however. It is reasonable to assume that the length of post-operative hospitalization is correlated with whether a patient requires intensive care.

In the following analysis, the correlation between the two types of outcomes for a patient is modeled with shared (G-side) random effects. The data set variable dist identifies the distribution for each observation. For those observations that follow a binary distribution the response variable option "(event='1')" determines which value of the binary variable is modeled as the event. Because a LINK= option is not specified, the link is also chosen observation by observation as the default link for the respective distribution.

```
proc glimmix data=joint;
  class patient dist;
  model response(event='1') = dist dist*age dist*OKstatus /
    noint s dist=byobs(dist);
  random int / subject=patient;
run;
```

Results of this analysis are shown in Figure 8.

The GLIMMIX Procedure						
Model Information						
Data Set			WORK.JOINT			
Response Variable			response			
Response Distribution			Multivariate			
Link Function			Multiple			
Variance Function			Default			
Variance Matrix Blocked By			patient			
Estimation Technique			Residual PL			
Degrees of Freedom Method			Containment			
Covariance Parameter Estimates						
	Cov Parm	Subject	Estimate	Standard Error		
	Intercept	patient	0.2990	0.1116		
Solutions for Fixed Effects						
Effect	dist	Estimate	Standard Error	DF	t Value	Pr > t
dist	Binary	5.7783	2.9048	29	1.99	0.0562
dist	Poisson	0.8410	0.5696	29	1.48	0.1506
age*dist	Binary	-0.07572	0.03791	29	-2.00	0.0552
age*dist	Poisson	0.01875	0.007383	29	2.54	0.0167
OKstatus*dist	Binary	-0.4697	1.1251	29	-0.42	0.6794
OKstatus*dist	Poisson	-0.1856	0.3020	29	-0.61	0.5435

Figure 8. Bivariate Analysis

The “Model Information” table shows that the distribution of the data is multivariate and that there are possibly multiple link functions involved; by default PROC GLIMMIX uses a logit link for the binary observations and a log link for the Poisson data.

The estimate of the variance of the random patient intercept is 0.2990, and the estimated standard error of this variance component estimate is 0.1116. There appears to be significant patient-to-patient variation in the intercepts. The estimates of the fixed effects as well as their estimated standard errors are shown in the “Solutions for Fixed Effects” table. Notice that these estimates are different from estimates you would obtain in separate analyses of the two outcomes, because of the induced correlation. In particular, when the length of hospital stay and the post-operative condition are modeled jointly, the pre-operative health status (variable OKStatus) does not appear significant. In separate analyses the initial health status is a significant predictor of the length of hospital stay.

You can also model the dependence between the outcomes directly by specifying an R-side covariance structure. In the following analysis, a marginal model is fit where the (2×2) covariance matrix for a subject is modeled with the Cholesky parameterization of an unstructured covariance matrix. This parameterization is an alternative to an unconstrained symmetric covariance matrix. It ensures that the estimated covariance matrix is at least positive semidefinite and the parameterization stabilizes the numerical optimizations.

```
proc glimmix data=joint;
  class patient dist;
  model response(event='1') = dist dist*age dist*OKstatus /
    noint s dist=byobs(dist);
  random _residual_ / subject=patient type=chol;
run;
```

ANALYSIS EXAMPLES

The following examples illustrate some typical applications of PROC GLIMMIX as well as the range of problems that can be approached with this new procedure. Several of these examples are discussed in greater detail in the PROC GLIMMIX documentation.

Generalized Linear Models

The following code fits the model for a randomized complete block design to binomial proportions.

```
ods html;
ods graphics on;

proc glimmix;
  class block entry;
  model y/n = block entry;
  lsmeans entry / plots=(diffplot) adjust=tukey;
run;

ods graphics off;
ods html close;
```

PROC GLIMMIX defaults to the binomial distribution for these data because of the *events/trials* syntax for the response. The LSMEANS statement requests least-squares means for the entry effect and a graphical display of the pairwise least-squares means differences. The comparisons in the display are adjusted for multiplicity.

Overdispersed Generalized Linear Models

To add a multiplicative overdispersion component to a generalized linear model, use a simple R-side residual effect. Continuing the previous example, the statements

```

proc glimmix;
  class block entry;
  model y/n = block entry;
  random _residual_;
run;

```

replace the variance function $\mu(1 - \mu)$ with $\phi\mu(1 - \mu)$, and ϕ is estimated based on the Pearson chi-square statistic.

Models with Group-Dependent Overdispersion

On occasion, you may want to vary the degree of overdispersion depending on some group membership. An R-side random statement

```

random _residual_;

```

in a GLM would add a single overdispersion parameter. Since the GROUP= option of the RANDOM statement enables you to vary covariance parameters by the levels of a factor, the following statements produce overdispersion specific to the levels of factor A.

```

proc glimmix;
  class A;
  model Count = A / ddfm=kr dist=Poisson;
  random _residual_ / group=A sub=A;
  lsmeans A / diff;
run;

```

Quasi-Likelihood Estimation for Independent Data

If your model is a GLM-type model for which only the link and variance function are known, PROC GLIMMIX estimates the parameters by quasi-likelihood methods. This case arises, for example, if you supply your own variance function to PROC GLIMMIX.

Assume that the response variable prop takes continuous values between 0 and 1. A binary or binomial distribution does not seem appropriate because the former cannot take on values except for 0 and 1, and the latter requires a binomial denominator to form a proportion. Furthermore, the variance function of the binary or binomial distribution may not be appropriate. If you wish to change this variance function, you can supply your own function by expressing the reserved symbol `_VARIANCE_` as a function of the reserved symbol `_MU_`.

```

proc glimmix;
  class site variety;
  _variance_ = _mu_**2 * (1-_mu_)**2;
  model prop = site variety / link=logit;
  lsmeans variety / diff=control('1');
  random _residual_;
run;

```

Since a user-defined variance function specification overrides the variance function that is implied by a chosen or default distribution, the GLIMMIX procedure treats the response distribution in this case as unknown. The fixed-effects are estimated by quasi-likelihood based only on the mean and variance of the data.

Computed Variables

In the following example, several variables are computed within PROC GLIMMIX. These are the variables `x`, `logn`, and `SMR_pred`. Notice that the computed variable `x` is used as an effect in the MODEL statement

and that `logn` serves as the offset variable for this particular model (an offset variable is a regressor whose coefficient is known to be 1; offsets arise frequently in modeling Poisson data, if, for example, the counts refer to time intervals of different lengths, or in modeling ratios).

```
proc glimmix;
  class county;
  x          = employment / 10;
  logn       = log(expCount);
  SMR_pred = 100*exp(_zgamma_ + _xbeta_);
  model observed = x / dist=poisson offset=logn s ddfm=none;
  random county;
  id county observed employment SMR SMR_pred;
  output out=glimmixout;
run;
```

The variables `x` and `logn` are constructed from variables in the input data set. The variable `SMR_pred` on the other hand is computed from symbols internal to PROC GLIMMIX. `_ZGAMMA_` refers to the G-side random component of the linear predictor, `_XBETA_` refers to the fixed effect component of the linear predictor.

Notice that the variable `SMR_pred` is not used in the MODEL or RANDOM statement of this PROC GLIMMIX run. Because it is listed in the ID statement, however, its value will be saved to the OUTPUT data set for each observation. This provides a convenient mechanism to save estimated quantities that depend on the converged estimates of the model.

Spatial Smoothing

The following code smooths the minimum perfect match value `min_pm` on a microarray chip to estimate the background image surface. A regular grid (lattice) was placed over the chip and the variables `blk_x` and `blk_y` identify the lattice positions.

```
proc glimmix;
  model min_pm = blk_x blk_y;
  random blk_x blk_y / type=rsmooth;
  output out=glimmixout pred(blup)=min_pm_pred;
run;
```

The TYPE=RSMOOTH covariance structure requests that a radial smoother with random spline coefficients be used to estimate the image surface. The OUTPUT statement saves the predicted values to a data set for plotting purposes. The predicted values are computed using the BLUPs, that is, the solutions for the random spline coefficients.

The radial smoother in PROC GLIMMIX is a low-rank smoother, that is, the number of knots is typically much less than the number of observations. PROC GLIMMIX uses by default knots constructed from the vertices of a *kd*-tree to determine the number and placement of the knots. This provides an efficient method to construct smoothing knots in one or more dimensions while taking into account the configuration of the data in the (random) regressor space.

Spatial models fit with the TYPE=RSMOOTH structure are usually more flexible than models obtained with traditional parametric covariance structures, such as

```
proc glimmix;
  class obsno;
  model min_pm = blk_x blk_y;
  random obsno / type=sp(exp) (x y);
run;
```

TYPE=RSMOOTH models are also computationally much more expedient.

Multinomial Data

PROC GLIMMIX can fit models to ordinal and nominal multinomial data. For ordered outcomes, the models are based on cumulative probabilities. For nominal outcomes, the procedure fits generalized logit models. For example, the next statements fit a model to the multinomial variable Style.

```
proc glimmix;
  class School Program;
  model Style(ref=first) = School Program / solution
                        dist = multinomial
                        link = glogit;

  freq count;
run;
```

Because the link function is chosen as the GLOGIT, the response variable is assumed to be nominal (un-ordered) and a generalized logit model is fit. In a generalized logit model one of the outcome categories serves as the reference category. The "ref=first" response variable option selects the first level of the Style variable as the reference level.

The FREQ statement specifies that the data are grouped. The variable count specifies for each record in the data set how many times the particular combination of Style, School, and Program occurred in the study.

In order to fit a model for ordinal data, select one of the link functions for cumulative probabilities; for example, the cumulative logit:

```
proc glimmix;
  class sub variety;
  model rating = variety / dist=multinomial
                        link=cumlogit;

  freq count;
run;
```

Notice that not all options and statements are compatible with multinomial distributions in PROC GLIMMIX. For example, an analysis of least-squares means or R-side random effects is not available for multinomial data. For more details, refer to the PROC GLIMMIX documentation.

CONTRASTING THE GLIMMIX AND MIXED PROCEDURES

Since the normal distribution is one of the possible response distributions in PROC GLIMMIX, you can fit standard linear mixed models for normally distributed data with PROC GLIMMIX. These models can also be fit with the MIXED procedure. The following list contrasts some important differences between PROC MIXED and the GLIMMIX procedure in SAS 9.1.

Syntax Differences

The GLIMMIX procedure models all random components of the model through the RANDOM statement. You use the `_RESIDUAL_` keyword or the RESIDUAL option of the RANDOM statement to instruct the GLIMMIX procedure that a random effect models an R-side component. For example, the following sets of statements fit the same spatial model with R-side exponential covariance structure.

```
proc glimmix;
  model z = x y;
  random _residual_ / subject=intercept type=sp(exp)(x y);
run;
```

```

proc mixed;
  model z = x y;
  repeated / subject=intercept type=sp(exp) (x y);
run;

```

In the following statements a G-side random intercept is combined with an R-side auto-regressive correlation structure. You use the RESIDUAL option of the RANDOM statement in PROC GLIMMIX if you need to specify an effect before the option slash.

```

proc glimmix;
  class id time gender;
  model z = gender age gender*age;
  random int / subject=id;
  random time / subject=id type=ar(1) residual;
run;

proc mixed;
  class id time gender;
  model z = gender age gender*age;
  random int / subject=id;
  repeated time / subject=id type=ar(1);
run;

```

The GLIMMIX procedure enables you to define model terms and variables with programming statements. For example, in the following code example, a scaled version of the x variable is used in the RANDOM statement:

```

proc glimmix;
  x2 = x / 10;
  model y(event='1') = x / dist=binary;
  random x2 / type=rsmooth;
run;

```

Functional Differences

The MIXED procedure implements one method of optimization, a ridge-stabilized Newton-Raphson algorithm. With the GLIMMIX procedure, you can choose from a variety of optimization methods via the NLOPTIONS statement. The default method for GLMMs is a Quasi-Newton algorithm. A ridge-stabilized Newton-Raphson algorithm, akin to the method in the MIXED procedure, is selected with the TECHNIQUE=NRRIDG option in the NLOPTIONS statement of the GLIMMIX procedures.

The GLIMMIX procedure does not support Type I —Type III (ANOVA) estimation methods for variance components. Also, the procedure does not have a METHOD=MIVQUE0 option. However, the NOITER option in the PARS statement can be used to obtain minimum variance quadratic unbiased estimates (with 0 priors) (MIVQUE0, Goodnight 1978) with the GLIMMIX procedure.

If you compute predicted values in the GLIMMIX procedure in a model with only R-side random components and missing values for the dependent variable, the predicted values will not be kriging predictions, as is the case with the MIXED procedure.

The GLIMMIX procedure does not support the broad array of covariance structures that you can draw on with the MIXED procedure. For example, the Kronecker structures and certain heterogeneous structures are not available in the GLIMMIX procedure.

The DDFM=BETWITHIN degrees of freedom method in the MODEL statement of the GLIMMIX procedure requires that the data be processed by subjects; see the “Dimensions” table in the default PROC GLIMMIX output.

In the GLIMMIX procedure, RANDOM statement options apply only to the RANDOM statement in which they are specified. For example, the statements

```
random a      / solution;
random a*b    / G;
random a*b*c  / alpha=0.04;
```

in the GLIMMIX procedure request that the solution vector be printed for the A and A*B*C random effects and that the G matrix corresponding to the A*B interaction random effect is printed. Confidence intervals with a 0.96 coverage probability are produced for the solutions of the A*B*C effect. In the MIXED procedure, the SOLUTION option, for example, applies to all RANDOM statements.

The GLIMMIX procedure provides new enhancements for least-squares means processing. You can obtain a LINES display to compare least-squares means. ODS statistical graphics are supported for LS-mean comparisons. For example, PROC GLIMMIX produces a LS-mean–LS-mean scatterplot (Diffogram). You can perform multiple comparisons among simple effects. This enables you to compare cell means in a classification model by holding other factors fixed. The *p*-values of these comparisons can be adjusted for multiplicity.

Converting Code from the %GLIMMIX Macro

For a number of years, the %GLIMMIX macro has been available through the SAS technical support site. This macro also implements the doubly-iterative linearization methods used by PROC GLIMMIX. Since the macro makes repeated calls to PROC MIXED as its linear mixed model engine, you can easily convert %GLIMMIX macro code to PROC GLIMMIX statements.

For example, to fit the logistic model with random intercepts discussed earlier, the macro invocation that is equivalent to the PROC GLIMMIX statements

```
proc glimmix data=multicenter;
  class center group;
  model SideEffect/n = group / solution;
  random intercept / subject=center;
run;
```

is as follows:

```
%include 'glimm800.sas';
%glimmix(data=multicenter,
  stmts=%str(
    class center group;
    model SideEffect/n = group / solution;
    random intercept / subject=center;
    parms (0.2) (1) / hold=2;
  ),
  error=binomial,
  link=logit
);
```

The statement block in the %str() macro is identical to the PROC GLIMMIX statements, with the following exceptions. The distribution and link function are specified as options in the MODEL statement in PROC GLIMMIX, and as macro options with %GLIMMIX. Since the binomial distribution does not have a free scale parameter, PROC GLIMMIX does not estimate this parameter by default. On the contrary, the MIXED procedure, which is invoked by the %GLIMMIX macro, adds a residual variance parameter regardless of the distribution selected with the ERROR= option. This variance acts as a multiplier on the variance function.

The PARMS statement in the %GLIMMIX invocation sets the value of the residual variance parameter to 1 and places a hold on the parameter. This effectively removes the overdispersion component.

If you write the %GLIMMIX statements without the PARMS statement, that is,

```
%include 'glmm800.sas';
%glimmix(data=multicenter,
  stmts=%str(
    class center group;
    model SideEffect/n = group / solution;
    random intercept / subject=center;
  ),
  error=binomial,
  link =logit
);
```

the equivalent PROC GLIMMIX statements are

```
proc glimmix data=multicenter;
  class center group;
  model SideEffect/n = group / solution;
  random intercept / subject=center;
  random _residual_;
run;
```

The second RANDOM statement adds an R-side overdispersion effect that acts as a multiplier on the variance function. It lifts the restriction that $\phi = 1$.

OBTAINING THE GLIMMIX PROCEDURE

PROC GLIMMIX is currently an add-on to the SAS/STAT product in SAS 9.1 on the (32-bit) Windows platform. The procedure does not ship with SAS 9.1. It is downloadable from Software Downloads at support.sas.com. Documentation is also available at this site. See sas.com/statistics for up-to-date information on the SAS/STAT product.

REFERENCES

- Breslow, N. E. and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25.
- Goodnight, J. H. (1978), SAS Technical Report R-105, *Computing MIVQUE0 Estimates of Variance Components*, Cary, NC: SAS Institute Inc.
- Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J., and Ostrowski, E. (1994), *A Handbook of Small Data Sets*, London: Chapman and Hall.
- McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, Second Edition, London: Chapman and Hall.
- Mosteller, F. and Tukey, J. W. (1977), *Data Analysis and Regression*, Reading, MA: Addison-Wesley.
- Ngo, L. and Wand, M. P. (2004) "Smoothing with Mixed Model Software," *Journal of Statistical Software*, 9:1 [<http://www.jstatsoft.org/>].
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge: Cambridge University Press.

- Schabenberger, O. and Gregoire, T. G. (1996), "Population-averaged and Subject-specific Approaches for Clustered Categorical Data," *Journal of Statistical Computation and Simulation*, 54, 231–253.
- Schall, R. (1991), "Estimation in Generalized Linear Models with Random Effects," *Biometrika*, 78:719–727.
- Wolfinger, R. and O'Connell, M. (1993), "Generalized Linear Mixed Models: A Pseudo-Likelihood Approach," *Journal of Statistical Computation and Simulation*, 4, 233–243.
- Wolfinger, R., Tobias, R., and Sall, J. (1994) "Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models," *SIAM Journal on Scientific Computing*, 15(6), 1294–1310.

CONTACT INFORMATION Oliver Schabenberger, SAS Institute Inc., SAS Campus Drive, Cary, NC 27513
Email: oliver.schabenberger@sas.com

SAS and SAS/STAT are registered trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.