SAS® GLOBAL FORUM 2021

Bayesian Analysis of GLMMs Using PROC BGLIMM

Walter W. Stroup, University of Nebraska-Lincoln

ABSTRACT

Over the past two decades, generalized linear mixed models (GLMMs) have become standard tools for statistical analysis. Since its introduction, PROC GLIMMIX has been SAS®/STAT's primary GLMM procedure. The most recent edition of *SAS for Mixed Models* includes three chapters on using GLIMMIX for GLMMs.

PROC GLIMMIX is an excellent frequentist tool. However, Bayesian approaches are becoming increasingly important. Many academic journals prefer - some even require - Bayesian analysis. Even when not required, Bayesian methods allow you to use what you know prior to, or in the early stages of, an investigation.

PROC BGLIMM is a new SAS/STAT procedure that makes Bayesian implementation of GLMMs relatively easy. BGLIMM uses syntax similar to PROC GLIMMIX, but there are some differences. This tutorial presents what you need to know to get started using PROC BGLIMM. We use GLMM examples from *SAS for Mixed Models*, but with a Bayesian twist.

INTRODUCTION

Mixed models are important tools for analyzing data from many types of studies, including longitudinal or repeated measures, multi-level or split-plot experiments, blocked designs with incomplete blocks or missing data and multi-location studies. Linear mixed models (LMMs) accommodate response variables assumed to follow a normal (hereafter referred to as a Gaussian) distribution. Generalized linear mixed models (GLMMs) extend mixed model theory and methods to accommodate non-Gaussian responses such as categorical or count data. Because the LMM is a special case of the GLMM – a GLMM with Gaussian data – the acronym GLMM is used for the rest of this paper. In the SAS® system, PROC GLIMMIX is the preeminent mixed model procedure, allowing users to work with both LMMs and GLMMs. PROC GLIMMIX uses a **frequentist** approach to estimation and inference. The next section, "GLMM Basics" gives an overview of what this entails, but the basic idea is that frequentist statistics focus on obtaining estimates and standard errors that are used to construct test statistics for significance testing or to construct confidence intervals.

Bayesian statistics provide an alternative approach. Bayesian statistics focus on incorporating prior information to obtain posterior distributions of the statistics of interest. Posterior distributions combine what you know in advance – the prior – with the data you observe. Bayesian statistics are becoming increasingly important for data analysis. One reason is that many academic journals now discourage classical significance testing in favor of Bayesian analysis. Some journals have even banned significance tests, in effect requiring Bayesian analysis. The second important reason is that the statistics obtained from frequentist methods are essentially equivalent to Bayesian analysis with a "non-informative" prior. In other words, frequentist methods implicitly assume that you know nothing until you analyze the data. In reality, this is rarely, if ever, true (if you really know nothing before analyzing the data, you probably should not be doing the study!). Many data sets come from studies that may be part of a series – phases of clinical trials leading to the licensing of a pharmaceutical product – or similar experiments by graduate students with the same advisor.

PROC MCMC is the SAS system's original and all-purpose Bayesian procedure. Although you can use PROC MCMC for mixed models, the syntax is similar to PROC NLMIXED, which can be inconvenient for less-than-simple models or models with CLASS variables. PROC BGLIMM was developed to allow you to use syntax similar to PROC GLIMMIX, making Bayesian analysis more accessible for GLMMs.

This tutorial introduces PROC BGLIMM using three examples from *SAS for Mixed Models: Introduction and Basic Applications* (Stroup, et al. 2018).

- A multi-clinic trial with binomial data from Chapter 11 of *SAS for Mixed Models*.
- A multi-level (split-plot) experiment with count data from Chapter 13.
- Repeated measures (longitudinal) data from Chapter 8.

The next section called "GLMM Basics," covers the GLMM setting, essential definitions and terminology, and an overview and comparison of PROC GLIMMIX and PROC BGLIMM estimation and inference.

GLMM BASICS

The classical format for a statistical model is the equation

Response variable = systematic component + random component

Think of the systematic component as the fixed effects, e.g., $\mu + \tau_i$ in ANOVA-type models with treatment effects or $\beta_0 + \beta_i X$ in linear regression models. In mixed models, the random component is really two distinct components, random model effects such as block effects, and residual effects, such as those that account for serial covariance in a longitudinal mixed model. The classical response = fixed + random format works well if you can assume that the data follow a Gaussian distribution, but in many cases this is not true. Table 1 shows common types of response variables in the left-hand column and types of model effects across the top row.

Response Variable	Fixed		Random	
	Categorical (CLASS) $\mu + \tau_i$	Continuous $\beta_0 + \beta_i X$	Model effect	Residual / Covariance structure
Gaussian				"R-side"
Categorical binomial multinomial				
Continuous proportion beta				"G-side"
Count Poisson negative binomial				
time-to-event				
etc...				

Table 1. Response variable and model effect types covered by LMM and GLMM

Replacing the classical model format, the defining elements of the GLMM are as follows:

- Distributions:
  - $y|b \sim \mathcal{D}(\mu, \Sigma)$; $\mathcal{D}$ denotes some distribution, e.g., one of those listed above
  - $b \sim N(0, G)$
  - $f(y|b) \propto exp\left(\frac{y\theta - b(\theta)}{\phi}\right)$, $\mu = E(y|b) = \frac{\partial b(\theta)}{\partial \theta}$, $v(\mu) = \frac{\partial^2 b(\theta)}{\partial \theta^2}$, $Var(y|b) = \phi v(\mu)$

- $\Sigma = V_\mu^{1/2} R V_\mu^{1/2}$, $V_\mu^{1/2} = diag\left[\sqrt{v(\mu)}\right]$, R is scale matrix, e.g., $R = I\phi$ or residual covariance
- Link function: $\eta = g(\mu)$
- Linear Predictor: $\eta = X\beta + Zb$

PROC MIXED allows you to work in the first row of Table 1. PROCs BGLIMM, GLIMMIX, MCMC and NLMIXED allow you to work with all of the rows.

FREQUENTIST GLMM ESTIMATION AND INFERENCE

The guiding principle of GLMM estimation is maximum likelihood (ML). You obtain estimates of the fixed effects by maximizing the log-likelihood, $log[f(y;\beta)]$, where $f(y;\beta) = \int f(y;\beta|b)f(b)db$. In general, this integral is intractable.

SAS software uses two approximation strategies: linearization (pseudo-likelihood) and integral approximation (quadrature and Laplace). The mixed model equations for Gaussian data, the PROC MIXED default used to obtain REML estimates of the variance components and ML solutions for β and b, are special cases of pseudo-likelihood (PL). The REML version of PL is the PROC GLIMMIX default. Laplace and quadrature are PROC GLIMMIX options. PROC NLMIXED uses quadrature only.

Estimates and standard errors computed from PL or integral approximation are used to compute test statistics, *p*-values and confidence intervals. Classical frequentist inference makes extensive use of significance testing.

BAYESIAN ESTIMATION AND INFERENCE

The primary tool of Bayesian inference is the posterior distribution,

$$f(\beta,\sigma|y,b) = \frac{f(y,b|\beta,\sigma)f(\beta,\sigma)}{\iint f(y,b|\beta,\sigma)f(\beta,\sigma)d\beta d\sigma}$$

where σ denotes the vector of covariance parameters and $f(\beta,\sigma)$ denotes the prior distribution of the fixed effects and covariance parameters. The function $f(y,b|\beta,\sigma)$ is the same likelihood used in frequentist estimation, defined by the GLMM distributions $f(y|b)$ and $f(b)$.

As with the GLMM likelihood, the integrals required for the posterior distribution are generally intractable. Unlike frequentist estimation, neither linearization nor integral approximation are viable options. Instead, Bayesian methods approximate the posterior distribution by simulation.

PROC BGLIMM involves the following steps:

- Specify the GLMM (i.e., the defining elements listed above). This step specifies $f(y|b)$ and $f(b)$.
- Specify the prior distributions $f(\beta,\sigma)$. The BGLIMM procedure uses pre-programmed default priors. Depending on the model and data, you may or may not need to replace them with more appropriate prior distributions.
- PROC BGLIMM implements three computational steps. It is beyond the scope of this tutorial to provide technical description of the algorithms used to implement these steps. The following is simply a list of the steps and their purpose:
    - Tuning. This step uses the data, model and priors to come up with an approximation of the posterior distributions to be sampled.
    - Burn-in. This is initial sampling the posterior distributions. A coin flip exercise in introductory statistics class provides simple analogy: the proportion of flips

resulting in heads vacillates early in the sampling, but eventually settles down to a reasonable approximation of the probability of a head.

- o Sampling the posterior distribution. As the name implies, the posterior distribution of each parameter is sampled with the goal of getting a sufficiently accurate and detailed characterization.

The end result is a mean, standard deviation and quantiles of interest for the posterior distribution of each element of β and each covariance parameter. You can use BGLIMM's ESTIMATE statement to define estimable functions of the β that address objectives of your data analysis. Two forms of credible intervals are commonly used for Bayesian inference:

- the highest posterior density (HPD) interval. This is the narrowest interval between the lower and upper bound that contains a given percent of the posterior distribution.

- quantile-based interval. For example, you can construct a 95% credible interval using the 2.5 and 97.5 percentiles of the posterior distribution as the lower and upper bounds.

There are a number of diagnostics you should check before using PROC BGLIMM results. Diagnostics are introduced in the context of the examples in the following sections. These examples cover problems you are likely to encounter, how to interpret the relevant diagnostics, and PROC BGLIMM options you can use to address these problems in order to obtain a useable analysis.

EXAMPLE 1: MULTI-CLINIC TRIAL WITH BINOMIAL DATA

The first example is the Beitler and Landis (*Biometrics*, 1985) multi-clinic data set that appears in *SAS for Mixed Models* (2018), Chapter 11, Section 11.3. Two treatments, CNTL and DRUG, are compared at eight clinics sampled from a target population. At the jth clinic, n_{ij} patients are assigned to the ith treatment. The response variable, denoted y_{ij}, is the number of patients having a favorable outcome.

The first step in the analysis of these data is to identify an appropriate statistical model. You can do this by following a process presented in *SAS for Mixed Models*, Chapter 5. List the sources of variation separately for the "study" design and the treatment design. The study design, also called the "experiment design," describes the components of the design before being assigned treatments or treatment levels. In this case, the study design consists of clinics and groups of patients within clinics. The treatment design is simply CNTL and DRUG. Groups are randomly assigned to treatments. Table 2 shows the sources of variation.

STUDY DESIGN		TREATMENT DESIGN		COMBINED	
SOURCE	**DF**	**SOURCE**	**DF**	**SOURCE**	**DF**
clinic	7			clinic	7
		treatment	1	treatment	1
group(clinic)	8			group(clinic) | treatment a.k.a. clinic x treatment a.k.a. unit of observation	8-1=7
TOTAL	**15**			**TOTAL**	**15**

Table 2. Sources of Variation for Multi-Clinic Example Data

In the combined column, read "group(clinic) | treatment" as "group within clinic after accounting for treatment." List the rows so that treatment appears in the row immediately above the unit to which it is assigned – in this case, group(clinic).

The next step is to write model effects associated with each source of variation and the assumed probability distribution of any effects considered to be random. Tables 3 and 4

show two scenarios that lead to plausible mixed models. Table 3 describes the elements of a logit-normal GLMM, so-called because it the uses a logit link function, $log[\pi_{ij}/(1 - \pi_{ij})]$, where π_{ij} denotes the probability of a favorable outcome for the ij^{th} clinic-treatment combination, and all random effects are assumed to follow a normal (Gaussian) distribution. Note that clinics effects are random because the clinics in the study are a representative sample. Also note that there is a random effect for the group level source of variation. You should include this term to account for variation in π_{ij} at this level – if you don't, over-dispersion may cause misleading results.

Note that group(clinic) is the unit on which the observations (y_{ij}) are taken. Hence, you can refer to it as the "unit level" effect, unit being shorthand for unit of observation. Specifying the model requires stating the assumed distribution of the observations, conditional on the random effects, at the unit level, and relating its expected value to the linear predictor, e.g., through the inverse link function as shown here.

SOURCE	EFFECT DISTRIBUTION	OBSERVATION AND MODEL
clinic	$c_j \sim N(0, \sigma_c^2)$	
treatment	τ_i	
group(clinic) | treatment a.k.a. clinic x treatment a.k.a. unit of observation	$ct_{ij} \sim N(0, \sigma_{ct}^2)$	$y_{ij}\|c_j, ct_{ij} \sim \text{Binomial}(N_{ij}, \pi_{ij})$ $\pi_{ij} = 1/\{1 + exp[-(\eta + \tau_i + c_j + ct_{ij})]\}$

Table 3. Model Effects and Distributions Defining Logit-normal GLMM

Table 4 gives the elements of a beta-binomial mixed model, a commonly used alternative to the logit-normal GLMM for binomial data. The difference between the two models is the way group-level variation in the probability of a favorable outcome is modeled. The logit-normal GLMM includes a random effect in the linear predictor. The beta-binomial models the probability as a random variable, assuming that it follows a beta distribution. Table 4 gives the Ferrari and Cribari-Neto (2004) parameterization of the beta in terms of its mean (μ_{ij}) and scale parameter (φ). Alternatively, you can write the distribution in its conventional math-stat form, $p_{ij} \sim \text{Beta}(\alpha_{ij}, \beta_{ij})$, where $\alpha_{ij} = \varphi\mu_{ij}$ and $\beta_{ij} = \varphi(1 - \mu_{ij})$.

SOURCE	EFFECT DISTRIBUTION	OBSERVATION AND MODEL
clinic	$c_j \sim N(0, \sigma_c^2)$	
treatment	τ_i	
group(clinic) | treatment a.k.a. clinic x treatment a.k.a. unit of observation	$p_{ij} \sim \text{Beta}(\mu_{ij}, \varphi)$	$y_{ij}\|c_j, p_{ij} \sim \text{Binomial}(N_{ij}, p_{ij})$ $\mu_{ij} = 1/\{1 + exp[-(\eta + \tau_i + c_j)]\}$

Table 4. Model Effects and Distributions Defining Beta-Binomial Mixed Model

You can use PROC BGLIMM (or PROC GLIMMIX) to implement the logit-normal GLMM, but not the beta-binomial. This is because BGLIMM and GLIMMIX are limited to models with Gaussian random effects, whereas the beta-binomial has a non-Gaussian random effect, p_{ij}. (Although beyond the scope of this tutorial, you can implement the beta-binomial with PROC MCMC)

The beta-binomial is shown here to illustrate the difference between a "sensible" model and an improperly specified model. What constitutes a "sensible" model? You must have a one-to-one correspondence between sources of variation and model effects or parameters that account for them. Table 5 illustrates the difference.

	LOGIT-NORMAL	BETA-BINOMIAL	NAIVE GLMM	BETA-BIN W/ ct_{ij}
SOURCE	sensible	sensible	Over-Dispersion	Unit Confounding
clinic	c_j	c_j	c_j	c_j
treatment	τ_i	τ_i	τ_i	τ_i
unit	ct_{ij}	φ		ct_{ij}, φ

Table 5. Sensible and Improperly Specified Mixed Models for Multi-Clinic Binomial Data

Both the logit-normal and beta-binomial meet the sensible model criterion. The "naive GLMM" lacks any term that accounts for variation at the unit level, making the model vulnerable to over-dispersion. The ct_{ij} effect is thus needed for the logit-normal model, but if you include it in the beta-binomial model, it is confounded with the beta distribution's scale parameter (φ), which is not good.

The remainder of this section focuses on implementing the logit-normal model with PROC BGLIMM.

BASIC PROC BGLIMM STATEMENTS FOR THE LOGIT-NORMAL BGLIMM

PROC BGLIMM uses syntax borrowed from PROC GLIMMIX and, for repeated measures, from PROC MIXED. To illustrate, here are the basic GLIMMIX and BGLIMM statements. First, the GLIMMIX statements for the logit-normal GLMM from *SAS for Mixed Models*:

```
proc glimmix data=multi_clinic;
 class clinic treatment;
 model fav/nij =  treatment;
 random intercept treatment / subject=clinic;
 lsmeans treatments / ilink diff oddsratio cl;
run;
```

Now the BGLIMM statements:

```
proc bglimm data=multi_clinic plots=(trace autocorr density)
  diagnostics=all outpost=cout;
 class clinic treatment;
 model fav/nij =  treatment / init=pinit;
 random intercept treatment / subject=clinic;
 estimate "CNTL" intercept 1 treatment 1 0;
 estimate "DRUG" intercept 1 treatment 0 1;
 estimate "log_odds_ratio" treatment 1 -1;
 ods output estimates=model_scale;
run;
```

Notice that the CLASS, MODEL and RANDOM statements for the two procedures are identical. The FAV/NIJ (events/trials) syntax is specific to the binomial distribution. Also notice that there are no statements in the basic PROC BGLIMM program specifying prior distributions. BGLIMM has pre-programmed priors that it uses by default. These may or may not be appropriate, depending on the model and data. The additional options in the PROC BGLIMM statement for PLOT and DIAGNOSTICS should be considered standard operating procedure to either verify that the results are useable or to identify problems that need to be fixed, such as inappropriate default priors, before the output can be regarded as useable. The INIT=PINIT option in the MODEL statement causes the default starting values used for the fixed effects to be included is the SAS listing. Consider these to be part of the diagnostics. PROC BGLIMM does not have an LSMEANS statement. You must write the ESTIMATE statements that correspond to least squares means and treatment differences. PROC BGLIMM also lacks an ILINK option. PROC BGLIMM only computes model-scale (in this case, logit scale) estimates. The OUTPOST option in the PROC statement and the ODS OUTPUT statement give you two ways to obtain data-scale statistics. These are explained below in the subsection entitled "Post-Processing to Obtain Data-Scale Statistics."

DIAGNOSTICS

The PLOT, DIAGNOSTICS=ALL and INIT=PINIT produce items in the SAS listing the help you identify problems and to decide if the results are useable. This section has two parts: a

list of the items and a brief description of their purpose, and diagnostics produced by the default PROC BGLIMM statements given above and their interpretation.

List of Diagnostics

The PLOT option gives you three plots: autocorrelation, density and trace. Trace plots are also called "caterpillar" plots, because ideally they should look like a "fuzzy caterpillar," which indicates the sampling algorithm has produced a reasonable approximation of the posterior distribution. The algorithms used to sample the posterior distributions are vulnerable to autocorrelation. The autocorrelation plot shows how quickly autocorrelation decreases as lag increases. The density plot, as the name implies, shows the posterior density produced by the sampling algorithm.

DIAGNOSTICS=ALL produces the following:

- Effective Sample Size (ESS). Autocorrelation reduces posterior sampling efficiency. ESS gives a measure of the sample size after adjusting for autocorrelation. Low efficiency, per se, is not a problem, but ESS should be, at the very least, >1000 in order to have an accurate approximation of the posterior distribution.

- Heidelberger-Welch. These are stationarity tests – is the posterior distribution sampling giving consistent results from beginning (immediately after burn-in) to end? The key columns in the SAS listing will say "passed" or "failed."

- Raftery-Lewis. Addresses the accuracy of the posterior distribution's percentile estimates. The key statistic is the dependence factor. Ideally, it should be $\cong 1$.

- Geweke. Compares estimates early and late in the sampling process. Use the Geweke statistics to test the null hypothesis that they are acceptably similar. You want to see p-values consistent with failing to reject the null hypothesis.

The listing also includes the number of burn-in iterations, the posterior density sample size and the priors the BGLIMM procedure uses by default.

Results of Preliminary PROC BGLIMM Run with Default Settings

Figure 1 shows two diagnostic plots. On the left is the plot for the logit-normal GLMM's intercept parameter produced by BGLIMM program given above. On the right is an example of what the plot should look like in a run whose results are useable. The plots for the other parameters are not shown, but they are similar to these.
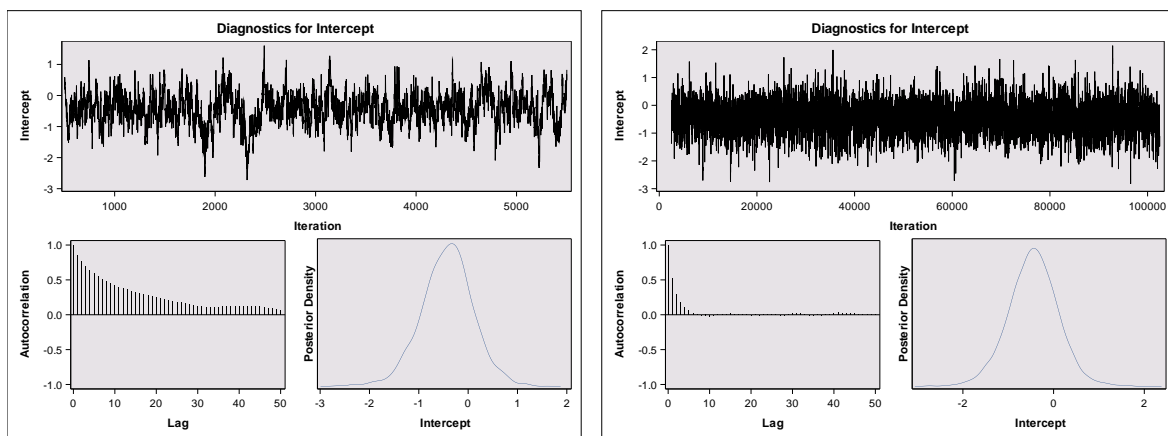
Figure 1. Example Diagnostic Plots

The left-hand plot is what you do not want to see. The trace plot shows erratic and inconsistent sampling of the posterior distribution. It looks more like a seismograph during an earthquake than the "fuzzy caterpillar" plot on the right. The right-hand plot shows

consistent sampling from beginning to end. The autocorrelation plot on the left also signals trouble. Autocorrelation should drop to zero quickly, as it does by lag 10 in the right-hand plot. The BGLIMM run with default settings produces autocorrelation that never goes to zero, even at lag 50. If autocorrelation does not drop to zero at least by lag 25, consider the results unusable.

Output 1 shows results produced by the DIAGNOSTICS=ALL option. Statistics that signal trouble are highlighted in **bold**.

Output 1. Diagnostic Statistics from PROC BGLIMM Logit-normal Program with Default Settings

	Effective Sample Sizes		
Parameter	**ESS**	**Autocorrelation Time**	**Efficiency**
Intercept	**174.1**	28.7273	**0.0348**
trt cntl	**301.6**	16.5782	**0.0603**
trt drug	.	.	.
Random VC(1)	**886.7**	5.6389	**0.1773**
Random VC(2)	**722.3**	6.9225	**0.1445**

Geweke Diagnostics		
Parameter	**z**	**Pr > |z|**
Intercept	-0.4815	0.6302
trt cntl	-0.2990	0.7649
trt drug	.	.
Random VC(1)	0.5412	0.5884
Random VC(2)	0.9604	0.3369

Raftery-Lewis Diagnostics				
Quantile=0.025 Accuracy=+/-0.005 Probability=0.95 Epsilon=0.001				
	Number of Samples			Dependence Factor
Parameter	**Burn-In**	**Total**	**Minimum**	
Intercept	29	32909	3746	**8.7851**
trt cntl	18	19114	3746	**5.1025**
trt drug	.	.	.	.
Random VC(1)	4	4636	3746	1.2376
Random VC(2)	4	4714	3746	1.2584

Heidelberger-Welch Diagnostics								
	Stationarity Test				Half-Width Test			
Parameter	**Cramer-von Mises Stat**	**p-Value**	**Test Outcome**	**Iterations Discarded**	**Half-Width**	**Mean**	**Relative Half-Width**	**Test Outcome**
Intercept	0.0951	0.6092	Passed	0	0.0875	-0.4288	-0.2041	**Failed**
trt cntl	0.2343	0.2098	Passed	0	0.0630	-0.9815	-0.0641	Passed
trt drug	.	.			.	.	.	

Heidelberger-Welch Diagnostics								
	Stationarity Test				Half-Width Test			
Parameter	Cramer-von Mises Stat	p-Value	Test Outcome	Iterations Discarded	Half-Width	Mean	Relative Half-Width	Test Outcome
Random VC(1)	0.2008	0.2659	Passed	2000	0.0655	1.8446	0.0355	Passed
Random VC(2)	0.0432	0.9160	Passed	0	0.0855	1.2928	0.0661	Passed

The "Effective Sample Sizes" table ESS and "Efficiency" results indicate trouble. All ESS values are less than 1000. The Geweke statistics look okay here, but you should always check them. The Raftery-Lewis dependence factors for "Random VC(1)" and "VC(2)," the clinic and clinic x treatment variance (σ_c^2 and σ_{ct}^2) respectively, are acceptably close to one, but those for the intercept and "trt cntl" effects (η and τ_{CNTL}) are much too high. Rules of thumb being at best approximate, if you see dependence factors greater than 3 or 4, consider your results to be unusable. Finally, the Heidelberger-Welch half-width test for intercept failed.

ADDRESSING PROBLEMS

The problems identified by the above diagnostics can occur for several reasons. The number of burn-in iterations or the size of the posterior distribution sampling may be inadequate. You can use thinning, that is, only taking every 5th or 10th or even 50th posterior density sample instead of every sample, to reduce autocorrelation. The default starting values for the fixed effect parameters may not be appropriate. The default priors may be too diffuse, including values that are technically in the parameter space, but highly implausible. Or the default priors may be where the parameters are not, that is, the most likely values of the parameter may be in an extreme tail of the prior distribution. Output 2 shows the defaults used by PROC BGLIMM for the logit-normal GLMM.

Output 2. Sampling, Thinning, Starting Values and Priors for Default Logit-normal BGLIMM Run

Model Information	
Burn-In Size	500
Simulation Size	5000
Thinning	1

Initial Values for Fixed Effects	
Parameter	Value
Intercept	-0.3102
trt cntl	-0.4040

Priors for Fixed Effects	
Parameter	Prior
Intercept	Constant
trt cntl	Constant

Priors for Scale and Covariance Parameters	
Parameter	Prior
Random Cov (Diag)	Inverse Gamma (Shape=2, Scale=2)

The "Model Information" table shows the default burn-in size (number of burn-in iterations), the "simulation size" (number of samples of the posterior density) and thinning. Thinning equal one means no thinning was done. The "Initial Values for Fixed Effects" table shows the starting values for η and τ_{CNTL}. Compare these to their estimates from PROC GLIMMIX, $\hat{\eta} = -0.4571$ and $\hat{\tau}_{CNTL} = -0.7462$. You can use the estimates from GLIMMIX as starting values instead of letting BGLIMM's algorithm choose starting values.

The default priors for η and τ_{CNTL} are "constant" – essentially uniform distributions between plus and minus infinity. Overly diffuse priors that include highly implausible parameter values can cause problems – given that the actual values of η and τ_{CNTL} are unlikely to be very far from their GLIMMIX estimates, the "constant" prior may be an issue. The prior for both variance components is an inverse gamma with shape and scale parameters equal to two. Figure 2 shows a plot of the inverse gamma(2,2) distribution with vertical lines showing the GLIMMIX estimates of σ_c^2 and σ_{ct}^2.
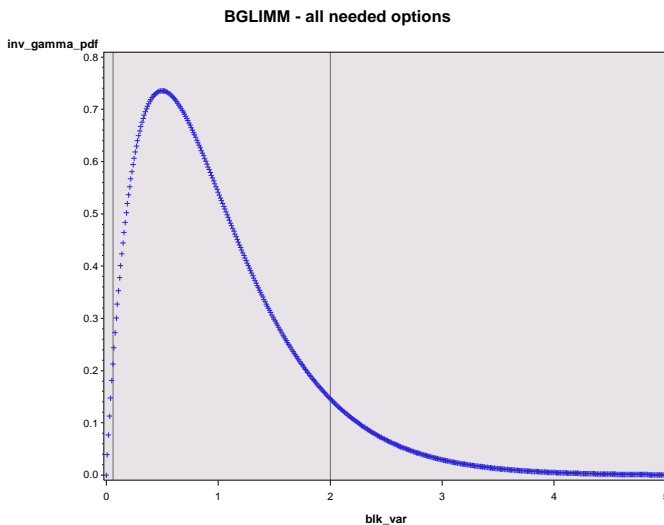
Figure 2. Plot of PROC BGLIMM Default Prior for Variance Components

The lower vertical bar is the GLIMMIX estimate of the clinic x treatment variance, $\hat{\sigma}_{ct}^2 = 0.06$. The upper vertical bar is the GLIMMIX estimate $\hat{\sigma}_c^2 \cong 2$. You can see that the most likely value of the clinic x treatment variance is in the extreme lower tail of the default prior. Most of the probability mass of the inverse gamma(2,2) distribution is located over values that are implausibly high for σ_{ct}^2. The clinic variance shows the opposite problem, although to a lesser extent – its most likely value is in the upper tail of the default prior.

Christiansen, et al. (2011) suggest a strategy for selecting an appropriate prior for a variance component. They suggest using precision, defined as the inverse of the variance, denoted here as $\gamma = 1/\sigma^2$, and identifying a gamma distribution whose mode is approximately equal to the most likely value of the parameter being estimated (you can use the estimates from PROC GLIMMIX as a logical most likely value) and whose upper and lower extreme quantiles (say the 1st and 99th) correspond to values below and above which are thought to be highly unlikely. The mode of a gamma distribution with shape parameter α and scale parameter β is $\beta(\alpha - 1)$. You can specify the mode, try several values of α, solve for β and evaluate the resulting distribution.

The following is an example a program you can use for prior hunting. The example is for the clinic x treatment variance. The PROC GLIMMIX estimate using the pseudo-likelihood default is $\hat{\sigma}^2_{ct} = 0.06$. The estimate using GLIMMIX's METHOD=QUADRATURE option is $\hat{\sigma}^2_{ct} = 0.01$. These translate to precisions of 17 and 100 respectively. The program uses a mode between the two, in this example 40. Use the PDF and CDF functions to create the distribution for each α, β combination. The PROC PRINT statement allows you to examine the lower and upper quantiles of the candidate distributions. The PROC GPLOT statement allows you to visualize the candidate distributions.

The program statements are as follows:

```
data gamma_prior;
 mode=40;
 /* A=shape parameter */
 /* B=scale parameter */
 do a=1.01,1.05,1.1,1.25,1.5,2,3,5,10,20;
  b=mode/(a-1);
  do precision=0 to 200;
   pdf_gamma=pdf("gamma",precision,a,b);
   cdf_gamma=cdf("gamma",precision,a,b);
  output;
 end;
 end;
proc sort data=gamma_prior; by a;
proc print;
 where (0.009<cdf_gamma<0.011 or 0.989<cdf_gamma<0.991);
run;
proc gplot data=gamma_prior;
 by a;
 plot pdf_gamma*precision/href=17,40,100;
 plot cdf_gamma*precision/vref=0.01,0.99;
run;
```

Some trial and error, and art along with science, are involved at this point. The distribution selected is gamma with shape=3 and scale=20. Figure 3 shows plots of the p.d.f. and c.d.f.
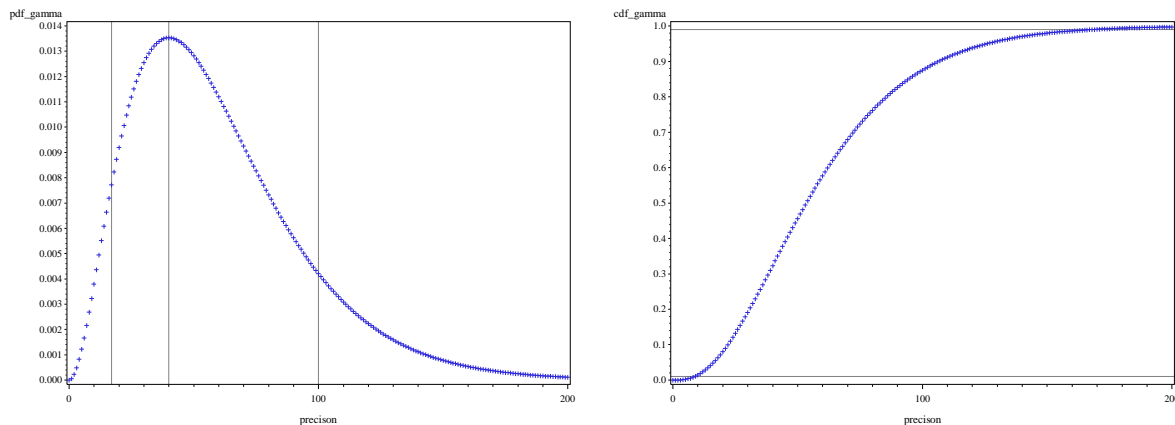
Figure 3. Plots of the Prior Distribution Selected for the Clinic x Treatment Variance

You can see that p.d.f. has a mode of 40 and contains the precision values 17 and 100 well within its range. The c.d.f. plot shows that the lower and upper extreme quantiles are just under 10 and just under 200, respectively. Translated to the variance scale, this prior covers a range between $\sigma^2_{ct} = 0.005$ to σ^2_{ct} just over 0.1, the range considered plausible. Using

the same strategy, a suitable prior for the clinic precision is gamma with shape=1.5 and scale=1.

REVISED PROC BGLIMM PROGRAM

The following shows the PROC BGLIMM program modified to include the options needed to obtain useable results. These options appear in **bold**. The program:

```
proc bglimm data=clinics seed=81152097
   nbi=2500 nmc=100000 thin=10
   plots=(trace autocorr density) diagnostics=all
   statistics(percent=(2.5 50 97.5))
   outpost=cout;
 class clinic trt;
 model fav/Nij=trt /
   init=(list=(-0.46 -0.75) pinit) coeffprior=normal(var=9);
 random  intercept  / subject=clinic covprior=igamma(shape=1.5, scale=1);
 random  trt / subject=clinic covprior=igamma(shape=3, scale=0.05);
 estimate "CNTL" intercept 1 trt 1 0;
 estimate "DRUG" intercept 1 trt 0 1;
 estimate "log odds-ratio" trt 1 -1;
 ods output estimates=modelscale;
 title "BGLIMM - all needed options";
 run;
```

The NBI, NMC and THIN options increase burn-in iterations, posterior distribution sample size and thinning, respectively. The INIT=(LIST=(option allows you to specify starting values for the fixed effect parameters η and τ_{CNTL}. The COEFFPRIOR option specifies a $N(\mu = 0, \sigma^2 = 9)$ prior distribution for these effects. Note that your options for COEFFPRIOR are either "constant" or "normal." VAR allows you to specify the variance, but the normal prior is always centered at a mean of zero. The COVPRIOR option specifies a prior for the variance of random model effects. Notice that you can have multiple RANDOM statements, which allows you to use different priors for each variance. The COVPRIOR must be specified in terms of the inverse gamma distribution. The required syntax is IGAMMA (SHAPE=< $value$ >, SCALE=< $value$ >). Use the following result: $\gamma \sim Gamma(\alpha, \beta)$ is equivalent to $1/\gamma \sim inverse - gamma(\alpha, 1/\beta)$. For example, the $gamma(3,20)$ prior for the clinic x treatment precision translates to an $inverse - gamma(3,0.05)$ prior for the clinic x treatment variance. Output 3 shows selected results.

Effective Sample Sizes			
Parameter	**ESS**	**Autocorrelation Time**	**Efficiency**
Intercept	2952.3	3.3872	0.2952
trt cntl	10000.0	1.0000	1.0000
trt drug	.	.	.
Random1 Var	7644.9	1.3081	0.7645
Random2 Var	8756.1	1.1421	0.8756

Raftery-Lewis Diagnostics				
Quantile=0.025 Accuracy=+/-0.005 Probability=0.95 Epsilon=0.001				
	Number of Samples			Dependence Factor
Parameter	**Burn-In**	**Total**	**Minimum**	
Intercept	6	6350	3746	1.6951
trt cntl	2	3834	3746	1.0235
trt drug	.	.	.	.
Random1 Var	2	3803	3746	1.0152
Random2 Var	2	3650	3746	0.9744

Posterior Summaries and Intervals					
Parameter	**N**	**Mean**	**Standard Deviation**	**95% HPD Interval**	
Intercept	10000	-0.4491	0.5504	-1.6003	0.5728
trt cntl	10000	-0.7454	0.3053	-1.3201	-0.1303
trt drug	0	.	.	.	.
Random1 Var	10000	2.1165	1.5208	0.3744	4.6572
Random2 Var	10000	0.0261	0.0281	0.00444	0.0639

Posterior Summaries						
				Percentiles		
Parameter	**N**	**Mean**	**Standard Deviation**	**2.5**	**50**	**97.5**
Intercept	10000	-0.4491	0.5504	-1.5574	-0.4418	0.6304
trt cntl	10000	-0.7454	0.3053	-1.3539	-0.7410	-0.1619
trt drug	0	.	.	.	.	.
Random1 Var	10000	2.1165	1.5208	0.6293	1.7386	5.7644
Random2 Var	10000	0.0261	0.0281	0.00701	0.0193	0.0845

Results from ESTIMATE Statements				
Label	**Mean**	**Standard Deviation**	**95% HPD Interval**	
CNTL	-1.1946	0.5612	-2.3136	-0.0881
DRUG	-0.4491	0.5504	-1.6003	0.5728
log odds-ratio	-0.7454	0.3053	-1.3201	-0.1303

Output 3. Selected Results from Useable PROC BGLIMM Run for Beitler-Landis Data

The "Effective Samples Sizes" table shows acceptable ESS and efficiency numbers. All Raftery-Lewis dependence factors are <2. In the interest of space, the other diagnostics are not shown here, but they are similarly acceptable, indicating that this run has produced useable results. The "Posterior Summaries and Intervals" table shows the mean, standard deviation and 95% highest posterior density (HPD) intervals for the model fixed effects and the variance components. The estimates in the "Mean" column are similar to those obtained with PROC GLIMMIX. Note that the fixed effect component of the linear predictor, $\eta + \tau_i$ is not full rank. PROC BGLIMM uses the same convention – setting the last effect, in this case

τ_{DRUG}, to zero – as other SAS linear model procedures. The "Posterior Summaries" table gives percentiles specified by the STATISTICS(PERCENT= option instead of HPD intervals. The "Results from ESTIMATE Statements" table give the posterior distribution mean, standard deviation and 95% HPD intervals from the ESTIMATE statements.

Note that all of these statistics are on the model scale – in this case the logit scale. Because PROC BGLIMM has no ILINK option, you must output these results and use post-processing steps to obtain data scale estimates. The most obvious of these are the probabilities of a favorable outcome for each treatment, $\hat{\pi}_{CNTL}$ and $\hat{\pi}_{DRUG}$, the difference between the two, $\hat{\pi}_{DRUG} - \hat{\pi}_{CNTL}$, and the odds-ratio, $\frac{\hat{\pi}_{CNTL}}{1-\hat{\pi}_{CNTL}} / \frac{\hat{\pi}_{DRUG}}{1-\hat{\pi}_{DRUG}}$. There are two ways to do this. These are shown in the next section.

POST-PROCESSING TO OBTAIN DATA-SCALE ESTIMATE

You can either use the data set produced by the OUTPOST option in the PROC statement or the ESTIMATES from the ODS OUTPUT statement to obtain data-scale estimates.

Data Scale Estimates Using OUTPOST and the %SUMINT Macro

After you run the PROC BGLIMM program that produces useable output, use the following statements:

```
data datasc;
 set cout;
 pr_cntl=logistic(cntl);
 pr_drug=logistic(drug);
 ProbDiff=logistic(drug)-logistic(cntl);
 OddsRatio=exp(log_odds_ratio);
run;
%sumint(data=datasc, var=pr_cntl: pr_drug: ProbDiff: OddsRatio)
```

The OUTPOST option creates a data set of posterior samples for the model fixed effects and ESTIMATES. In this example, the data set is called COUT. The DATA step creates a new data set, DATASC, with the favorable outcome probabilities, labelled PR_CNTL and PR_DRUG, their difference, PROBDIFF, and the odds-ratio. The LOGISTIC function implements the logit's inverse link, $\pi = 1/[1 + exp(-\eta)]$. The difference $\hat{\tau}_{CNTL} - \hat{\tau}_{DRUG}$ estimates the log of the odds-ratio, so $exp(\hat{\tau}_{CNTL} - \hat{\tau}_{DRUG})$ gives you the estimated odds-ratio. The %SUMIT macro computes and prints inferential statistics for the terms listed after VAR= for the data set given by DATA=<*data set name*>. Output 4 shows the results.

Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
OddsRatio	10000	0.4970	0.1532	0.2251	0.8004
pr_cntl	10000	0.2461	0.0990	0.0750	0.4475
pr_drug	10000	0.3965	0.1224	0.1679	0.6394
ProbDiff	10000	0.1504	0.0667	0.0248	0.2800

Output 4. Data Scale Estimates from OUTPOST and %SUMINT

Data Scale Estimates Using ODS OUTPUT

Instead of using the OUTPOST data set and the %SUMINT macro, you can use the following statements:

```
data lsm; set modelscale;
```

```
    if label='CNTL' or label='DRUG';
    prob=1/(1+exp(-Mean));
    lower=1/(1+exp(-HPDLower));
    upper=1/(1+exp(-HPDUpper));
  data oddsratio; set modelscale;
    if label='log odds ratio';
    OddsRatio=exp(Mean);
    lower=exp(HPDLower);
    upper=exp(HPDUpper);
  proc print data=lsm;
  proc print data=oddsratio;
  run;
```

MODELSCALE is the data set produced by the ODS OUTPUT statement. These DATA steps produce two data sets. The one called LSM uses the logit's inverse link to compute the estimates and HPD interval lower and bounds for π_{CNTL} and π_{DRUG}. The data set called ODDSRATIO computes the estimates and HPD bounds for the odds-ratio. Use the IF LABEL= statement to select items to be included in each data set. The LABEL names are the labels used in the ESTIMATE statements. Both data sets use the results of the ESTIMATE statements instead of the posterior sample data set, so the results differ from the %SUMINT listing above. If you want the estimated difference between the probabilities, $\hat{\pi}_{DRUG} - \hat{\pi}_{CNTL}$ use OUTPOST and %SUMINT – it is more convenient. Output 5 shows the results using the ODS OUTPUT approach.

Obs	Label	Mean	StdDev	HPDLower	HPDUpper	prob	lower	upper
1	CNTL	-1.1946	0.5612	-2.3136	-0.0881	0.23245	0.09001	0.47798
2	DRUG	-0.4491	0.5504	-1.6003	0.5728	0.38957	0.16793	0.63942

Obs	Label	Mean	StdDev	HPDLower	HPDUpper	OddsRatio	lower	upper
1	log odds-ratio	-0.7454	0.3053	-1.3201	-0.1303	0.47452	0.26712	0.87780

Output 5. Data Scale Estimates from ODS OUTPUT and Follow-up DATA Steps

The Mean, StdDev, HPDLower and HPDUpper values are model scale values from the BGLIMM ESTIMATE statements. The three columns on the right, labelled "prob" or "OddsRatio" and "lower" and "upper" are the data scale estimates.

EXAMPLE 2: MULTI-LEVEL DESIGN WITH COUNT DATA

This example appears in *SAS for Mixed Models* (2018) in Chapter 13, Section 13.3. The data are from an agricultural experiment comparing two cultivation methods and two seed mixes. Each of six fields is divided into two sections, called whole plots. Each whole plot is randomly assigned to a method so that both methods appear at each field. Each whole plot is divided into two split plots to which mixes are randomly assigned. Table 6 shows the sources of variation.

EXPERIMENT DESIGN		TREATMENT DESIGN		COMBINED	
SOURCE	DF	SOURCE	DF	SOURCE	DF
field	5			field	5
		method	1	method	1
whole-plot(field)	6			field x method (wp)	6-1=5
		mix	1	mix	1
		method x mix	1	method x mix	1

15

split-plot(wp)	12			sp (wp) | method, mix	12-2=10
TOTAL	**23**			**TOTAL**	**23**

Table 6. Sources of Variation for Multi-Level Field Trial

The response variable is a discrete count. *SAS for Mixed Models* describes two "sensible" models, the Poisson-normal GLMM and the negative binomial GLMM. This section presents the latter, as it is the model of choice for these data, and the negative binomial illustrates PROC BGLIMM options that would not be used with the Poisson-normal model.

Table 7 uses the combined sources of variation column from Table 6 to show how the negative binomial arises in the context of this experiment.

SOURCE	EFFECT DISTRIBUTION	OBSERVATION AND MODEL
field	$f_k \sim N(0, \sigma_f^2)$	
method	α_i	
field x method (wp)	$w_{ik} \sim N(0, \sigma_w^2)$	
mix	β_j	
method x mix	$\alpha\beta_{ij}$	
sp(wp)|method,mix a.k.a. unit of obs.	$u_{ijk} \sim gamma(1/\varphi, \varphi)$	$y_{ijk}\|f_k, w_{ik}, u_{ijk} \sim Poisson(\lambda_{ijk}u_{ijk})$ $\lambda_{ijk} = exp(\eta + \alpha_i + \beta_j + \alpha\beta_{ij} + f_k + w_{ik})$

Table 7. Model Effects and Distribution Defining Poisson-Gamma Process

The method and mix effects are denoted by α and β, respectively. The negative binomial arises from a Poisson-gamma process. Variation at the unit of observation level is assumed to follow a gamma distribution with $E(u_{ijk}) = 1$. If you integrate out the unit level term, $y_{ijk}|f_k, w_{ik} \sim NB(\lambda_{ijk}, \varphi)$, where NB denotes negative binomial. Thus, the following elements define the GLMM:

- Distributions: $y_{ijk}|f_k, w_{ik} \sim NB(\lambda_{ijk}, \varphi)$, $f_k \sim N(0, \sigma_f^2)$, $w_{ik} \sim N(0, \sigma_w^2)$

- Link function: $\eta_{ijk} = log(\lambda_{ijk})$

- Linear predictor: $\eta + \alpha_i + \beta_j + \alpha\beta_{ij} + f_k + w_{ik}$

With PROC BGLIMM, replacing the effects part of the linear predictor, $\eta + \alpha_i + \beta_j + \alpha\beta_{ij}$, with its cell means, full rank equivalent makes it easier to use the ESTIMATE statement to compute terms of interest. The cells means form can be written $\eta_{ij} + f_k + w_{ik}$.

The CLASS, MODEL and RANDOM statements for this model are:

```
class field method mix;
model count=method*mix / noint distribution=negbin;
random intercept method/ subject=field;
```

As with Example 1, the default BGLIMM settings give unusable results. In the interest of space, not all diagnostics are shown. Output 6 shows two examples, the Geweke statistics and the plots for the negative binomial scale parameter, φ.

Geweke Diagnostics		
Parameter	**z**	**Pr > |z|**
method 1*mix 1	1.8922	0.0585
method 1*mix 2	2.3868	0.0170
method 2*mix 1	0.9584	0.3379
method 2*mix 2	1.8992	0.0575
Scale	-2.7547	0.0059

Geweke Diagnostics		
Parameter	**z**	**Pr > |z|**
Random VC(1)	1.4784	0.1393
Random VC(2)	-0.4486	0.6537
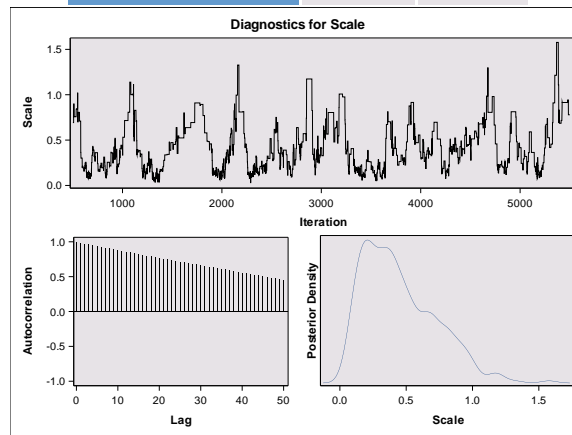
Output 6. Example diagnostics: PROC BGLIMM negative binomial GLMM with default settings

Four of the seven parameters on the "Geweke Diagnostics" table show low *p*-values. For the scale parameter, $p<0.01$. The trace and autocorrelation plots are unacceptable. To obtain useable results, you must work through the same set of steps as illustrated in the previous section. The following program statements are the result:

```
proc bglimm data=sp_counts plots=(trace autocorr density)
   nbi=2500 nmc=1000000 thin=100 seed=20210209
   diagnostics=all dic outpost=model_scale;
class field method mix;
model count=method*mix / noint distribution=negbin
   init=(list=(2.5, 2.9, 0.8, 2.0) Pinit)
   scaleprior=gamma(shape=20,iscale=25);
random intercept method/ subject=field
   covprior=igamma(shape=4,scale=3);
estimate "LSM_11" method*mix 1 0 0 0;
estimate "LSM_12" method*mix 0 1 0 0;
estimate "LSM_21" method*mix 0 0 1 0;
estimate "LSM_22" method*mix 0 0 0 1;
estimate "interaction" method*mix 1 -1 -1 1;
estimate "Method Main Effect" method*mix 1 1 -1 -1 / divisor=2;
estimate "Mix Main Effect" method*mix 1 -1 1 -1 / divisor=2;
ods output estimates=link_scale;
title "Negative Binomial GLMM";
run;
```

The burn-in, posterior sampling and thinning options may look extreme, but computing time is minimal, and they guarantee useable results. The INIT option uses PROC GLIMMIX results as starting values for the METHOD*MIX (η_{ij}) effects. Changing from the "constant" to the "normal" prior has little effect with this model.

The COVPRIOR in the RANDOM statement is based on using estimates from PROC GLIMMIX, $\hat{\sigma}_f^2 = 1.64$ and $\hat{\sigma}_w^2 = 1.38$, as most likely values. Given that they are similar in value, the same inverse gamma with shape=4 and scale =3 is used for both. The SCALEPRIOR option allows you to specify a prior for the negative binomial scale parameter. The search using the Christiansen, et al. approach described in the previous section resulted in selecting a

$gamma(shape = 20, scale = 0.04)$ prior. Note that BGLIMM requires the gamma distribution to be specified using the inverse scale, hence ISCALE=25 rather than SCALE=0.04.

The ESTIMATE statements define means, interaction and main effect differences. These are given on the model (log) scale in the SAS listing. You can use either the %SUMINT macro with the OUTPOST=MODEL_SCALE data set or the ODS OUTPUT data set to compute data-scale equivalents. In the interest of space, only the OUTPOST plus %SUMINT program statements and output are shown.

The SAS statements:

```
data data_scale;
 set model_scale;
 Lambda_11=exp(LSM_11);
 Lambda_12=exp(LSM_12);
 Lambda_21=exp(LSM_21);
 Lambda_22=exp(LSM_22);
 Method_Diff_Ratio=exp(Method_Main_Effect);
 Mix_Diff_Ratio=exp(Mix_Main_Effect);
%sumint(data=data_scale, var=Lambda_11: Lambda_12: Lambda_21: Lambda_22:
        interaction: Method_Diff_Ratio: Mix_Diff_Ratio)
```

Output 7 shows the listing.

Posterior Summaries and Intervals					
Parameter	**N**	**Mean**	**Standard Deviation**	**95% HPD Interval**	
Lambda_11	10000	19.2095	18.9908	0.7638	52.9988
Lambda_12	10000	31.6319	31.0823	1.4792	86.1046
Lambda_21	10000	4.2671	3.5069	0.4212	10.6353
Lambda_22	10000	13.5674	11.0600	1.0802	32.3851
interaction	10000	0.6695	0.8461	-0.9995	2.3477
Method_Diff_Ratio	10000	3.6754	2.7677	0.3665	8.3976
Mix_Diff_Ratio	10000	0.4728	0.2049	0.1595	0.8911

Output 7. Data-Scale Results of Negative Binomial Split-Plot GLMM Analysis

The LAMBDA_11, etc. terms give the rate parameter estimates, $\hat{\lambda}_{ij} = exp(\hat{\eta}_{ij})$ for the four method-mix combinations. The "interaction" term remains on the model scale. There is no need to express it on the data scale, as its only purpose is to assess the magnitude of the method x mix interaction. The HPD interval includes zero; using interval as the criterion, you could conclude that there is insufficient evidence to reject the hypothesis of no interaction. The METHOD_DIFF_RATIO" term estimates $\bar{\lambda}_{1.}/\bar{\lambda}_{2.}$ where $\bar{\lambda}_{i.}$ is the rate parameter for the ith method averaged over both mixes. The difference $\bar{\eta}_{1.} - \bar{\eta}_{2.} = log(\bar{\lambda}_{1.}) - log(\bar{\lambda}_{2.})$, hence $exp(\bar{\eta}_{1.} - \bar{\eta}_{2.}) = \bar{\lambda}_{1.}/\bar{\lambda}_{2.}$. The "MIX_DIFF_RATIO" is similarly defined, but for mix instead of method. Think of these as data scale main effect measures. Although not shown here, you could also obtain statistics for differences between selected LSM terms, similar to the way PROBDIFF was defined in the previous section.

EXAMPLE 3: REPEATED MEASURES DATA

This example is from Chapter 8 of *SAS for Mixed Models* (2018). It appears as "Respiratory Data" for a repeated measures example in Littell, Pendergast and Natarijan (2000), and in Chapter 4 of *SAS for Linear Models, 4th Edition* (Littell, et al., 2002). The data are from a

trial comparing a standard asthma treatment (A), a test drug (C) and a placebo (P). Treatments were randomly assigned to 24 patients (72 patients total participated in the trial). The response variable, FEV1, is a measure of breathing ability. A baseline measurement (BASEFEV1) was taken on each patient, then measurements every hour for eight hours starting at one hour after application. FEV1 can be assumed to follow a Gaussian distribution. Table 8 shows the experiment design, treatment design and combined sources of variation.

EXPERIMENT DESIGN		TREATMENT		COMBINED	
SOURCE	DF	SOURCE	DF	SOURCE	DF
		drug	2	drug	2
patient	72-1=71			patient(drug) a.k.a. between	71-2=69
		hour	7	hour	7
		drug x hour	14	drug x hour	14
occasion(patient)	72(8-1) = 504			occ(patient)|drug a.k.a. within	504-7-14 = 493
TOTAL	575			TOTAL	575

Table 8. Sources of Variation for Respiratory Repeated Measures Data

In repeated measures terminology, patient(drug) is called the **between subjects** effect and the occ(patient)|drug (occasion within patient after accounting for drug) residual term is called the **within subjects** effect. Following *SAS for Mixed Models*, the LMM of choice has the following elements:

- Distributions

  - $\mathbf{y}_{ij}|b_{ij} \sim N(\boldsymbol{\mu}_{ij}, \mathbf{R})$, where $\mathbf{y}_{ij}' = [y_{ij1} \quad y_{ij2} \quad y_{ij3} \quad \cdots \quad y_{ij7} \quad y_{ij8}]$, the vector of FEV1 measures on the jth patient assigned to the ith drug over the $k = 1,2,3,\ldots,7,8$ hours of observation, $\boldsymbol{\mu}_{ij}$ is the corresponding mean vector and $\mathbf{R}$ is the 8×8 covariance matrix modeling serial correlation among the within subjects effects. *SAS for Mixed Models* uses the AR(1) covariance model.

  - $b_{ij} \sim N(0, \sigma_b^2)$ is the between subjects effect

- Linear Predictor: $\mu_{ijk} = \eta + \delta_i + \tau_k + \delta\tau_{ik} + b_{ij}$, where δ and τ denote drug and time (hour) effects, respectively.

The covariance parameters for the for the R matrix are ρ for autocorrelation and σ_w^2 for residual (within subjects) variance. You can include the baseline covariance, BASEFEV1, by modifying the linear predictor to $\mu_{ijk} = \eta + \beta X_{ij} + \delta_i + \tau_k + \delta\tau_{ik} + b_{ij}$, where X_{ij} denotes BASEFEV1 for the jth patient assigned to the ith drug.

Figure 4 shows the interaction plot, which you can obtain using the PLOT=MEANPLOT option in the PROC GLIMMIX LSMEANS statement. The interaction plot is instructive, because it helps you visualize the objectives of this study and the requirements of the model needed to address the objectives.
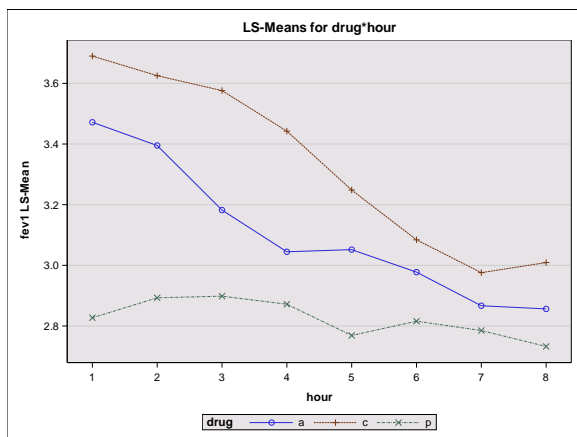
Figure 4. Interaction Plot showing Change in FEV1 over hours by treatment.

The treatment is initially effective for the A and C drugs but not the placebo. Over time, FEV1 decreases as the drugs' effectiveness wears off. The objective of the study is to see if there is a difference between the treatments as measured by initial effectiveness and by how long the effect lasts.

There are two ways to approach this objective. One is to model HOUR as a CLASS variable and track the simple effects $\mu_{Ak} - \mu_{Pk}$ and $\mu_{Ck} - \mu_{Pk}$ at each hour k, where $\mu_{ik} = \eta + \delta_i + \tau_k + \delta\tau_{ik}$. The other is to redefine μ_{ij} in terms of an unequal slopes linear regression model, i.e., $\mu_{ik} = \alpha_i + \beta_i H_k$, where H_k denotes the kth hour, and compare intercept and slope coefficients.

NOTE: The documentation for PROC BGLIMM uses this data set as an example. In one passage, the documentation gives what it calls "Model 2" and "Model 3." The MODEL statements for each are:

- Model 2: `fev1=basefev1 drug hour;`
- Model 3: `fev1=basefev1 drug|hour;`

The documentation characterizes the difference between model 2 and model 3 as "minor," and shows results for model 2 but not model 3. Perhaps the difference is "minor" from a programming point of view, but the difference is not "minor" from a statistical practice point of view. Model 2 is incapable of addressing the objectives of this study and therefore fails the "sensible model" criterion. To address this study's objectives, the drug x hour interaction must be included in the model in some form.

The PROC BGLIMM statements for simple effects and unequal slopes approaches are given as follows. For the simple effects approach, it is much more convenient to specify the cell-means form of the model. The statements are:

```
proc bglimm data=fev1uni nbi=1000 nmc=50000 thin=5 seed=97816352
     plots=(trace autocorr density) diagnostics=all;
   class Drug Patient Hour;
   model FEV1 = BaseFev1 Drug*Hour;
   random int / subject=Patient(drug);
   repeated Hour / subject=Patient(Drug) type=ar(1) r rcorr;
run;
```

Notice that PROC BGLIMM barrows the syntax of the PROC MIXED REPEATED statement to specify the covariance model for the within subjects effect. The default priors work well for this model, but increasing burn-in, posterior sampling and thinning improves the results. There are twenty-four DRUG*HOUR least squares means, and sixteen simple effects

comparing drug A and C to the placebo at each hour, for a total to forty ESTIMATE statements. In the interest of space, only one example of each type of statement is shown.

The example ESTIMATE statements are:

```
estimate "LSM A at hour 1"  intercept 1 basefev1 2.6493
   Drug*Hour 1 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0;
estimate "A vs P at hour 1"
   Drug*Hour 1 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0  -1 0 0 0 0 0 0 0;
```

The coefficients for least squares means (LSM) obtain the adjusted mean at the average BASEFEV1 covariate value of 2.6493. If you are in doubt, use the E option of the PROC GLIMMIX LSMEANS statement to get the coefficients needed for these statements.

The statements for the unequal slopes approach are as follows:

```
data fev1uni;
 set fev1uni;
 H=hour;
proc bglimm data=fev1uni nbi=1000 nmc=50000 thin=5 seed=44672057
      plots=(trace autocorr density) diagnostics=all;
   class Drug Patient Hour;
   model FEV1 = BaseFev1 drug H(drug);
   random int / subject=Patient(drug);
   repeated Hour / subject=Patient(Drug) type=ar(1) r rcorr;
   estimate "intercept - Drug A" intercept 1 basefev1 2.6493 drug 1 0 0;
   estimate "intercept - Drug C" intercept 1 basefev1 2.6493 drug 0 1 0;
   estimate "intercept - Drug P" intercept 1 basefev1 2.6493 drug 0 0 1;
   estimate "slope - Drug A" H(drug) 1 0 0;
   estimate "slope - Drug C" H(drug) 0 1 0;
   estimate "slope - Drug P" H(drug) 0 0 1;
   estimate "intercept A vs C" drug 1 -1 0;
   estimate "intercept A vs P" drug 1 0 -1;
   estimate "intercept C vs P" drug 0 1 -1;
   estimate "slope A vs C" H(drug) 1 -1 0;
   estimate "slope A vs P" H(drug) 1 0 -1;
   estimate "slope C vs P" H(drug) 0 1 -1;
 run;
```

You need to define a second variable for HOUR – in this case the DATA step with H=HOUR – and use one as the direct regression variable in the MODEL statement and the other as a CLASS variable for the REPEATED statement. The "INTERCEPT – DRUG A" etc. ESTIMATE statements obtain $\hat{\alpha}_i$ the estimated regression intercepts, adjusted for the baseline covariate, and the "SLOPE – DRUG A" etc. statements obtain $\hat{\beta}_i$, the slopes. The next six statements compare intercepts and slopes among the DRUG treatments.

Output 8 shows the covariance parameter estimates obtained using the cell means model. Output 9 shows the least squares means listing from the cell means model. Output 10 shows the simple effect difference results.

Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
Residual Var	10000	0.0846	0.0104	0.0664	0.1058

Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
Residual AR(1)	10000	0.5386	0.0550	0.4278	0.6407
Random Var	10000	0.2458	0.0453	0.1612	0.3341

Output 8. Covariance parameter estimates.

"Residual Var" gives the within subjects variance statistics. "Mean" gives you $\hat{\sigma}_w^2 = 0.0846$. "Residual AR(1)" give the autocorrelation parameter, $\hat{\rho} = 0.5386$. "Random Var" gives the between subjects variance, $\hat{\sigma}_b^2 = 0.2458$.

Results from ESTIMATE Statements				
Label	Mean	Standard Deviation	95% HPD Interval	
LSM A at hour 1	3.4739	0.1178	3.2501	3.7107
LSM C at hour 1	3.6903	0.1187	3.4705	3.9330
LSM P at hour 1	2.8264	0.1183	2.5956	3.0589
LSM A at hour 2	3.3963	0.1175	3.1685	3.6287
LSM C at hour 2	3.6251	0.1194	3.3849	3.8521
LSM P at hour 2	2.8916	0.1176	2.6606	3.1195
LSM A at hour 3	3.1839	0.1181	2.9470	3.4114
LSM C at hour 3	3.5764	0.1190	3.3390	3.8059
LSM P at hour 3	2.8974	0.1186	2.6571	3.1239
LSM A at hour 4	3.0460	0.1175	2.8098	3.2714
LSM C at hour 4	3.4424	0.1190	3.2178	3.6824
LSM P at hour 4	2.8713	0.1180	2.6237	3.0877
LSM A at hour 5	3.0533	0.1185	2.8278	3.2873
LSM C at hour 5	3.2487	0.1187	3.0172	3.4829
LSM P at hour 5	2.7686	0.1183	2.5390	2.9997
LSM A at hour 6	2.9800	0.1183	2.7533	3.2130
LSM C at hour 6	3.0845	0.1186	2.8482	3.3096
LSM P at hour 6	2.8160	0.1179	2.5817	3.0423
LSM A at hour 7	2.8687	0.1181	2.6414	3.0987
LSM C at hour 7	2.9764	0.1181	2.7273	3.1924
LSM P at hour 7	2.7859	0.1180	2.5595	3.0183
LSM A at hour 8	2.8581	0.1171	2.6377	3.0913
LSM C at hour 8	3.0093	0.1175	2.7724	3.2359
LSM P at hour 8	2.7330	0.1188	2.4999	2.9644

Output 9. Drug x Treatment Least Squares Means for Repeated Measures Data

Results from ESTIMATE Statements				
Label	**Mean**	**Standard Deviation**	**95% HPD Interval**	
A vs P at hour 1	0.6475	0.1652	0.3224	0.9630
C vs P at hour 1	0.8639	0.1683	0.5235	1.1894
A vs P at hour 2	0.5048	0.1648	0.1606	0.8070
C vs P at hour 2	0.7335	0.1678	0.4078	1.0694
A vs P at hour 3	0.2865	0.1651	-0.0478	0.6028
C vs P at hour 3	0.6789	0.1687	0.3453	1.0093
A vs P at hour 4	0.1747	0.1649	-0.1565	0.4920
C vs P at hour 4	0.5711	0.1680	0.2481	0.9090
A vs P at hour 5	0.2847	0.1658	-0.0489	0.6031
C vs P at hour 5	0.4801	0.1677	0.1611	0.8200
A vs P at hour 6	0.1641	0.1651	-0.1476	0.4993
C vs P at hour 6	0.2685	0.1669	-0.0668	0.5785
A vs P at hour 7	0.0828	0.1654	-0.2411	0.4083
C vs P at hour 7	0.1905	0.1673	-0.1365	0.5220
A vs P at hour 8	0.1251	0.1659	-0.1932	0.4553
C vs P at hour 8	0.2763	0.1667	-0.0641	0.5925

Output 10. Simple effect differences between drugs (A, C) and placebo (P)

Output 10 is the PROC BGLIMM analog to the results you would get with the SLICEDIFF=HOUR option in PROC GLIMMIX.

Output 11 gives the ESTIMATE statement results from the unequal slopes regression model.

Results from ESTIMATE Statements				
Label	**Mean**	**Standard Deviation**	**95% HPD Interval**	
intercept - Drug A	3.5169	0.1205	3.2852	3.7577
intercept - Drug C	3.8065	0.1197	3.5730	4.0403
intercept - Drug P	2.8859	0.1199	2.6516	3.1233
slope - Drug A	-0.0887	0.0117	-0.1115	-0.0651
slope - Drug C	-0.1057	0.0117	-0.1285	-0.0825
slope - Drug P	-0.0160	0.0119	-0.0390	0.00708
intercept A vs C	-0.2896	0.1690	-0.6066	0.0506
intercept A vs P	0.6310	0.1706	0.3002	0.9703
intercept C vs P	0.9206	0.1703	0.5897	1.2525
slope A vs C	0.0170	0.0166	-0.0166	0.0486
slope A vs P	-0.0727	0.0168	-0.1057	-0.0399
slope C vs P	-0.0897	0.0167	-0.1236	-0.0578

Output 11. Unequal slopes linear regression repeated measures model estimates and differences

You can see that the estimated intercepts for drug A and C and both greater than for the placebo (3.51 and 3.81 versus 2.89) indicating that initial breathing ability is greater for the two drugs than the placebo. On the other hand, the estimated slopes for drug A and C are -0.09 and -0.11 respectively versus -0.02 for the placebo. The placebo's HPD credible interval for slope includes zero, whereas the intervals for drugs A and C do not. The placebo's regression is essentially flat because its initial effect is negligible.

CONCLUSION

In the SAS system hierarchy, PROC BGLIMM occupies a space between PROC GLIMMIX and PROC MCMC. You can use PROC BGLIMM for any GLMM that PROC GLIMMIX can implement. The advantage of the BGLIMM procedure is that it allows to access to Bayesian estimation and inference using the same CLASS, MODEL and RANDOM syntax as PROC GLIMMIX. The main difference between BGLIMM and GLIMMIX is that because Bayesian analysis depends on complex simulation algorithms, there are more diagnostics that must be monitored and more options that you may need to use in order to get a useable analysis.

PROC MCMC uses syntax borrowed from PROC NLMIXED. This means that it does not have a CLASS statement, making it more tedious to specify models, especially effects models with factorial treatment structures. On the other hand, unlike PROC BGLIMM, you can use PROC MCMC to fit nonlinear models, semiparametric models, and mixed models with non-Gaussian random effects, such as beta-binomial and Poisson-gamma models. PROC MCMC also provides more flexibility in specifying prior distributions. For example, you can use priors centered at the starting value for fixed effects, where BGLIMM limits you to "constant" or zero-centered Gaussian priors.

Finally, when you use PROC BGLIMM with non-Gaussian response variables, these is an additional post-processing step, either using the %SUMINT macro or the ODS OUTPUT data set, in order to get data scale estimates inferential statistics. The lack of an ILINK option means an extra step, but as we see in Examples 1 and 2, it is not a difficult step.

PROC BGLIMM makes Bayesian analysis highly accessible to data analysts who have PROC GLIMMIX experience but are new to the Bayesian world.

REFERENCES

Beitler, P.J. and J.R. Landis. 1985. "A Mixed-Effects Model for Categorical Data." *Biometrics*, 41:991-1000.

Christiansen, R., W. Johnson, A. Branscum and T.E. Hanson. 2011. *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. Boca Raton, FL: CRC Press.

Ferrari, S. L. P., and Cribari-Neto, F. 2004. "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31:799–815

Littell, R.C, J. Pendergast and R. Natarajan. 2000. "Modeling Covariance Structure in the Analysis of Repeated Measures Data." *Statistics in Medicine*, 19:1793-1819.

Littell, R.C., W.W. Stroup and R.J. Freund. 2002. *SAS® for Linear Models, 4th edition*. Cary, NC: SAS Institute, Inc.

Stroup, W.W. 2013. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Boca Raton, FL: CRC Press.

Stroup, W.W., G.A. Milliken, E.A. Claassen and R.D. Wolfinger. 2018. *SAS® for Mixed Models: Introduction and Basic Applications*. Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

Acknowledge Oliver Schabenberger, who developed PROC GLIMMIX and was a valued partner in GLMM's early days. Fang Chen, who developed PROC BGLIMM and who provided helpful advice about the procedure. George Milliken, Russ Wolfinger and Elizabeth Claassen, my *SAS® for Mixed Models* co-authors and esteemed mixed model brain trust.

RECOMMENDED READING

- *SAS® for Mixed Models: Introduction and Basic Applications*

- *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Walt Stroup
wstroup@unl.edu