Introducing the BGLIMM Procedure for Bayesian Generalized Linear Mixed Models

Amy Shi and Fang Chen
SAS
amy.shi@sas.com

Bayes-Pharma, May, 2019

SAS Convention of Mixed Models

A mixed model contains fixed and random effects.

First consider a normal linear mixed model,

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \\ \boldsymbol{\gamma} &\sim N(\mathbf{0}, \mathbf{G}) \\ \boldsymbol{\epsilon} &\sim N(\mathbf{0}, \mathbf{R}) \end{aligned}$$

where $\boldsymbol{\beta}$ is fixed effects and $\boldsymbol{\gamma}$ is random effects.

- G-side matrix, $\mathbf{G}$, is the covariance matrix of the random effects.
- R-side matrix, $\mathbf{R}$, is the covariance matrix of the residuals.

GLMM Models

A generalized linear mixed model (GLMM) consists of the following:

- a linear predictor $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}$
- a response distribution in the exponential family (binary, binomial, Poisson, normal, gamma, negative binomial, ...).
- a link function

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

PROC BGLIMM: Simple Syntax

If you are somewhat familiar with PROC MIXED and PROC GLIMMIX, transition to PROC BGLIMM is straightforward.

The essential statements:

- MODEL statement: specifies Y, X, distribution, and link
- RANDOM statement: specifies Z and the G-side matrix
- REPEATED statement: specifies repeats and the R-side matrix
- ESTIMATE statement: computes linear combination of parameters

Syntax: MODEL Statement

MODEL ***response = fixed-effects*** **/ dist= link= ...;**

- Response
- Fixed effects
- 9 response distributions:
  - Binary, Binomial, Exponential, Gamma, Geometric, Inverse Gaussian, Negative binomial, Normal, Poisson.
- 8 link functions:
  - Log, Logit, Probit, Inverse, Identity, Pow(-2), Loglog, Complementary loglog

Simple Linear Regression with Class Variable

Example program:

```
proc bglimm data=Sashelp.Class seed=436792;
   class sex;
   model Weight = Height Age Sex / dist=normal
         coeffprior=normal(var=100);
run;
```

- Normal is the default distribution for a continuous response.
- COEFFPRIOR= specifies the prior of $\boldsymbol{\beta}$. The default is constant.
- SCALEPRIOR= specifies the prior of the scale parameter ϕ. The default is inverse gamma.

Poisson Regression

Example program:

```
proc bglimm data=LipCancer seed=10571042;
   class County;
   model Observed = x County / dist=poisson
         link=loglog offset=LogN;
run;
```

- The default link function for Poisson is the log function.
- OFFSET= adds an offset variable to the linear predictor.

Syntax: RANDOM statement

RANDOM ***random-effects*** **/ sub= group= type= ...;**

Defines **Z** and the **G**-side matrix.

- SUB= identifies the subjects for the random effects. A set of random effects is estimated for each subject level.
- GROUP= specifies groups by which to vary the covariance parameters; each level of the grouping effect produces a new set of covariance parameters.
- TYPE= defines the type of the **G**-side matrix.
  - 13 choices: VC, CS, AR, ARMA, TOEP, UN, ...

Logistic Random-Effects Model

Example program:

```
proc bglimm data=MultiCenter seed=976352;
   class Center Group;
   model SideEffect/N = Group / noint;
   random int / sub = Center;
run;
```

The random effects are assumed normally distributed:

$$\gamma_i \sim N(0, \mathbf{G_i})$$

Multiple RANDOM Statements

```
proc bglimm data=a;
   class Analyst Run Plate;
   model log_assay = Analyst;
   random int conc / sub=run(analyst)
     covprior=uniform(lower=0, upper=2);
   random int conc / sub=plate(run*analyst)
     covprior=halfnormal(var=4);
run;
```

The random effects can be nested or nonnested.

The COVPRIOR= option specifies the prior for the G-side covariance matrix.

Syntax: REPEATED statement

REPEATED ***repeated-effect*** **/ sub= group= type= ...;**

Specifies the **R**-side matrix in the model.

- A repeated-effect is required to define the proper location of the repeated responses.
- SUB= groups repeated measures together for the same subject.
- GROUP= specifies groups by which to vary the covariance parameters; each level of the grouping effect produces a new set of covariance parameters.
- TYPE= defines the type of the **R**-side matrix.
  - 13 choices: AR, ARMA, CS, TOEP, UN, VC, ...

Repeated Measures Model

```
proc bglimm data=Fev nmc=10000 seed=44672057
      outpost=FevOut;
   class Drug Patient Hour;
   model FEV = BaseVal Drug Hour;
   random int / sub=Patient;
   repeated Hour / sub=Patient(Drug) type=AR(1);
run;
```

- Repeated measurements can be balanced or unbalanced.

Example 1: Logistic Regression with Random Intercepts

- The response is sample proportions of side effects as binomial ratios.
- The fixed effect is Group.
- The random effect cluster is Center.

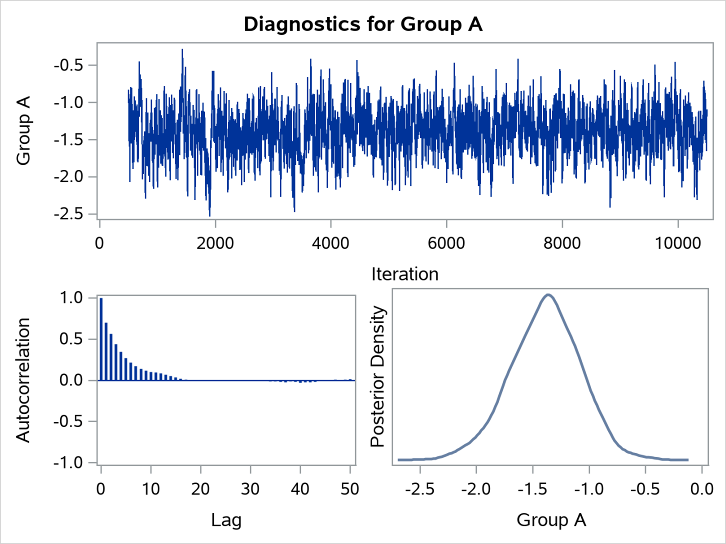
```
proc bglimm data=MultiCenter nmc=10000 thin=2 seed=976352
   outpost=CenterOut plots=all;
   class Center Group;
   model SideEffect/N = Group / noint;
   random int / sub = Center;
run;
```

Posterior Summary Statistics

Posterior Summaries and Intervals					
Parameter	**N**	**Mean**	**Standard Deviation**	**95% HPD Interval**	
Group A	5000	-1.3895	0.3102	-2.0071	-0.7956
Group B	5000	-0.8839	0.2968	-1.4819	-0.3186
Random Var	5000	0.9184	0.4198	0.3024	1.7515

- Group A vs. Group B
- 'Random Var' measures variability of center-level intercepts.
- Each center's intercept can be printed with MONITOR option in the RANDOM statement.

TAD Plots (Trace, Auto-correlation, Density plots)



Functions of Parameters

Example: probability difference between A and B:

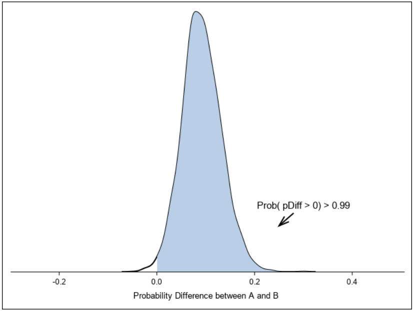
$$\text{pDiff} = \frac{\exp(\beta_b)}{1+\exp(\beta_b)} - \frac{\exp(\beta_a)}{1+\exp(\beta_a)}$$

```
data prob;
   set CenterOut;
   pDiff = logistic(group_b)
          - logistic(group_a);
run;

%sumint(data=prob, var=pDiff)
```

Posterior Summaries and Intervals

Parameter	N	Mean	Standard Deviation	95% HPD Interval	
pDiff	5000	0.0920	0.0395	0.0195	0.1750



Example 2: Repeated Measurements with Heterogeneity

- A two-treatment trial for patients with rheumatoid arthritis
- 67 subjects enrolled
- Subjects were followed up three times
- A grip strength measurement was taken at each follow-up visit
- A baseline grip strength (in mmHg) was measured at the start

```
data GripData;
input Subject Baseline Treat Gender$ Time Grip;
datalines;
26   175   1   M   1   161
26   175   1   M   2   210
26   175   1   M   3   230
27   165   1   M   1   215
27   165   1   M   2   245
27   165   1   M   3   265
...

71   104   2   F   1   107
71   104   2   F   2   .
71   104   2   F   3   .
72    60   2   F   1    60
72    60   2   F   2    55
72    60   2   F   3    58
;
```

Within-Subject Heterogeneity: R-Side

The initial model has fairly general specification for both the mean and the covariance structure (Littell et al. 2006).

- The fixed effects contain 12 cell means: 2 treatments by 2 genders by 3 visits.
- The repeated measurements are taken over the Time variable and are grouped according to the Subject variable.

```
proc bglimm data=GripData seed=475193;
   class Subject Treat Gender Time;
   model Grip = Gender*Treat*Time Baseline / noint;
   repeated Time / sub=Subject type=un r rcorr;
   run;
```

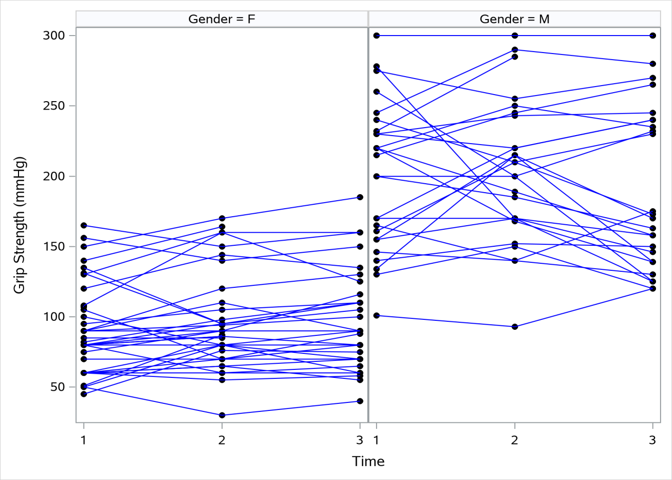
Posterior R-Side Covariance and Correlation Matrices

Requested via the options `r` and `rcorr`.

Estimated R Matrix			
Row	**Col 1**	**Col 2**	**Col 3**
1	604.96	308.00	288.96
2	308.00	950.48	885.65
3	288.96	885.65	1304.71

Estimated R Correlation Matrix			
Row	**Col 1**	**Col 2**	**Col 3**
1	1.0000	0.4062	0.3252
2	0.4062	1.0000	0.7953
3	0.3252	0.7953	1.0000

Grip Strength Measurements over Time by Gender



Between-Subject Heterogeneity by Gender

To account for distinct covariance structures of the two gender groups, you can fit the model by adding the option GROUP=GENDER to the REPEATED statement:

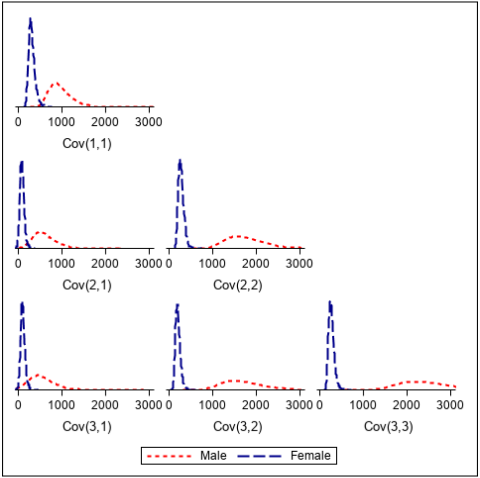
```
proc bglimm data=GripData seed=475193;
   class Subject Treat Gender Time;
   model Grip = Gender*Treat*Time Baseline / noint;
   repeated Time / sub=Subject type=un group=Gender r rcorr;
run;
```

R-Side Covariance Matrices by Gender

The BGLIMM Procedure

Estimated R Matrix

Group	Row	Col 1	Col 2	Col 3
Gender F	1	300.08	77.2769	95.2165
Gender F	2	77.2769	267.23	195.20
Gender F	3	95.2165	195.20	257.48
Gender M	1	960.37	591.43	528.93
Gender M	2	591.43	1773.63	1710.94
Gender M	3	528.93	1710.94	2504.11



Between-Subject Heterogeneity in Random effects

You can account for more between-subject heterogeneity by adding a random statement.

```
proc bglimm data=GripData seed=475193 nmc=20000 thin=4;
   class Subject Treat Gender Time;
   model Grip = Gender*Treat*Time Baseline / noint;
   random int / sub=Subject group=Gender covprior=uniform;
   repeated Time / sub=Subject type=un group=Gender r rcorr
            covprior=iw(scale=500);
run;
```

Summary

PROC BGLIMM is a Bayesian procedure that is designed specifically for fitting generalized linear mixed models.

- The procedure adopts familiar SAS syntax in specifying GLMMs.
- A key enhancement over the existing MCMC procedure is its simplicity.
- Efficient sampling algorithms are parallelized for performance
- PROC BGLIMM models missing data, nested or nonnested multilevel models, and repeated-measures data.
- It provides several built-in prior distributions for regression coefficients and covariance parameters.

For More Information

See the 2019 SAS Global Forum paper 'Introducing the BGLIMM Procedure for Bayesian Generalized Linear Mixed Models'.

PROC BGLIMM requires SAS/STAT 15.1 (SAS 9.4M6). Complete documentation of the procedure can be found at `http://support.sas.com/documentation/onlinedoc/stat/151/bglimm.pdf`.

You can find additional coding examples at `http://support.sas.com/rnd/app/examples/STATexamples.html`.