Bayesian Analysis: A Deep Dive into BGLIMM

Danny Modlin
Sr. Analytical Training Consultant – Cary, NC

Objectives

• Discuss current incorporation methods for random effects.
• Review syntax of PROC BGLIMM.
• Explore code examples of PROC BGLIMM.
Mixed Modeling Procedures in SAS

Classical
- GENMOD
- MIXED
- GLIMMIX
- NLMIXED

Bayesian
- GENMOD w/BAYES Statement
- BGLIMM
- MCMC

General Linear Mixed Model

$$y = X\beta + Z\gamma + \varepsilon$$

where
- $$y$$ is the vector of observed responses.
- $$X$$ is the design matrix of predictor variables.
- $$\beta$$ is the vector of regression parameters.
- $$Z$$ is the design matrix of random variables.
- $$\gamma$$ is the vector of random-effect parameters.
- $$\varepsilon$$ is the vector of random errors.
General Linear Mixed Model

\[ y_{ij} = \beta_0 + \beta_i X_i + \gamma_j + \varepsilon_{ij} \]

\( \gamma_j \sim N(0, \sigma^2_\gamma) \)

\( \varepsilon_{ij} \sim N(0, \sigma^2) \)

Generalized Linear Mixed Models (GzLMMs)

- GzLMMs enable modeling random effects and correlated errors for nonnormal data.
- A linear predictor can contain random effects.
  \[ \eta = X\beta + Z\gamma \]
- The random effects are normally distributed.
- The conditional mean, \( \mu | \gamma \), relates to the linear predictor through a link function.
  \[ g(\mu | \gamma) = \eta \]
- The conditional distribution (given \( \gamma \)) of the data belongs to the exponential family of distributions.
GzLMM Formulation and PROC GLIMMIX

\[ g(\mu | \gamma) = X\beta + Z\gamma \]

- LINK= option
- MODEL statement
- RANDOM statement
- Y|\gamma \sim \text{exponential family}
- DIST= option
- \text{var}(\gamma) = G
- \text{var}(y|\gamma) = A \mu A'\mu
- \text{RANDOM}_{\text{RESIDUAL}}\_statement
- \text{RANDOM}_{\text{RESIDUAL}}\_statement
- \text{*REPEATED in BGLIMM}

PROC MIXED Program

```sas
proc mixed data=sasuser.toy;
  class adhesive toy;
  model pressure=adhesive / solution ddfm=kr;
  random toy;
run;
```
PROC MCMC Program

```sas
proc mcmc data=toy seed=27513 diag=all dic outpost=mixed
  propcov=quanew thin=25 nbi=5000 ntu=5000 nmc=500000
  plots(smooth)=all mchistory=brief stats=all;
array beta[3];
parms beta: 0;
parms s2t 1;
parms s2g 1;
prior beta: ~ normal(0, var = 1e5);
prior s2: ~ igamma(2.001, scale = 1.001);
random gamma ~ normal(0, var=s2g) subject=toy
  monitor=(gamma) namesuffix=position;
mu = beta[adhesivebeta] + gamma;
model pressure ~ normal(mu, var = s2t);
title "Bayesian Analysis of the Toy Data Set";
run;
```

PROC BGLIMM Program

```sas
proc bglimm data=sasuser.toy seed=8675309;
  class adhesive toy;
  model pressure=adhesive / dist=normal;
  random int / sub=toy;
run;
```
Features of BGLIMM

• Suite of covariance structures (for both G- and R-side)
• Covariance heterogeneity modeling
• Built-in priors
• Model Comparison via DIC statistic
• Multi-threading of optimal sampling

PROC BGLIMM Syntax
PROC BGLIMM Statement

• DATA= names the input data set
• SEED= random seed for simulation
• OUTPOST= output a data set to contain posterior samples
• NBI= number of burn-in iterations
• NMC= number of Markov chain iterations
• NTHREADS= number of CPUs to run simulations simultaneously
• STATS= posterior statistics
• DIAG= convergence diagnostics
• PLOTS= plotting
PROC BGLIMM Syntax

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

- 9 response distributions:
  - Binomial
  - Exponential
  - Gamma
  - Geometric
  - Inverse Gaussian
  - Negative binomial
  - Normal
  - Poisson
  - Binary

- 8 link functions:
  - Log
  - Logit
  - Probit
  - Inverse
  - Identity
  - Loglog
  - Complementary loglog
  - PowerMinus2

PROC BGLIMM Syntax

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

- SUB= option to identify the subjects for the random effects

- GROUP= option to identify groups by which to vary the covariance parameters; each new level of the grouping effect produces a new set of covariance parameters

- TYPE= option to define the covariance structure of G

  - 13 choices: AR, ARMA, CS, TOEP, UN, VC, ...

- *Multiple RANDOM statements can be used.
PROC BGLIMM Syntax
REPEATED repeated-effect / sub= group= type= ...;

• A repeated-effect is required to define the proper location of the repeated responses.

• SUB= option to group repeated measures together for the same subject

• GROUP= option to identify groups by which to vary the covariance parameters

• TYPE= option to define the covariance structure
  - 13 choices: AR, ARMA, CS, TOEP, UN, VC, ...

Sampling

• PROC BGLIMM updates parameters conditionally and sequentially through Gibbs sampling.

• The fixed-effect parameters are drawn together first at each iteration.
• The random-effect parameters are updated by subjects.
• The G-side covariance parameters are then sampled.
• Lastly, the R-side covariance parameters are updated.

• If present, missing response values are treated as parameters and are thus sampled as well.
Prior Distributions

- Fixed-effect parameters (Betas)
  - Flat/constant; normal
- Scale parameter
  - Inverse gamma; gamma; improper

- G-side Covariance parameters
  - Inverse wishart; inverse gamma; uniform; halfcauchy; halfnormal; siwishart

- R-side Covariance parameters
  - Inverse wishart; inverse gamma

Examples using BGLIMM

This demonstration illustrates the concepts discussed previously.
Questions?