

PROC BGLIMM: The Smooth Transition to Bayesian Analysis

SESUG 2024



SESUG 2024

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.

β is the vector of regression parameters.

Z is the design matrix of random variables.

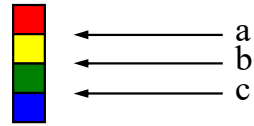
γ is the vector of random-effect parameters.

ε is the vector of random errors.



Toy Example

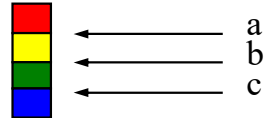
toy 1 adhesives



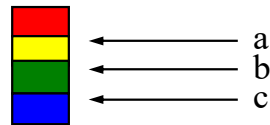
breaking strength



toy 2



toy 7



Time: 1 minute

General Linear Mixed Model

adhesive
(treatment) effects,
fixed

toy (block) effects,
random

$$y_{ij} = \mu + \beta_i + \gamma_j + \varepsilon_{ij}$$

random error

$$\gamma_j \sim N(0, \sigma_\gamma^2)$$

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



Time: 1 minute

PROC MIXED Program

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



PROC MCMC Program

```
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;
```



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 γ) of the data belongs to the exponential family of distributions.



GzLMM Formulation and PROC GLIMMIX

$$g(\mu | \gamma) = \mathbf{X}\beta + \mathbf{Z}\gamma$$

Diagram showing the components of the equation:

- LINK= option points to $g(\mu | \gamma)$
- MODEL statement points to $\mathbf{X}\beta$
- RANDOM statement points to $\mathbf{Z}\gamma$

$Y|\gamma \sim \text{exponential family}$ ← DIST= option

$\text{var}(\gamma) = \mathbf{G}$ ← Options in the RANDOM statement

$\text{Var}(y | \gamma) = \mathbf{A}_\mu \mathbf{R} \mathbf{A}_\mu$

RANDOM_RESIDUAL_
statement

*REPEATED in BGLIMM



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 | - Negative binomial |
| • Exponential | - Normal |
| • Gamma | - Poisson |
| • Geometric | - Binary |
| • Inverse Gaussian | |

- 8 link functions:

- | | |
|-----------|------------------------|
| • Log | - Identity |
| • Logit | - Loglog |
| • Probit | - Complementary loglog |
| • Inverse | - 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.



Thanks for attending.
Questions?

Business Card



Copyright © SAS Institute Inc. All rights reserved.