Biometry 971
Biometrical Modelling
GLMM Lectures

S. D. Kachman
Department of Biometry
University of Nebraska–Lincoln

An Introduction to Generalized Linear Mixed Models

Outline

- Motivation
- Model
- Estimation
  - Joint versus Marginal
- Testing

Motivation

Clinical Trial

- 2 drug treatments
- 8 clinics
- n_{ij} subjects assigned to treatment i at clinic j
- y_{ij} "favorable" responses for treatment i at clinic j
- p_{ij} proportion of favorable responses for treatment i at clinic j

Clinical Model

Treatment Design:

- $\mu + \tau_i$

Experimental Design:

- c_j clinic effect, iid $\mathrm{N}(0, \sigma_c^2)$
- $(\tau c)_{ij}$ ij clinic by treatment interaction, iid $\mathrm{N}(0, \sigma_{rc}^2)$
- e_{ij} error, iid $\mathrm{N}(0, \sigma^2)$???

Range Example

- Factorial Experiment (management x mix)
- 7 Management levels assigned in (4) randomized blocks
- 4 Seed Mix assigned to sub-units within the block
- Response: Number of Plants

Range Model

Treatment Design:

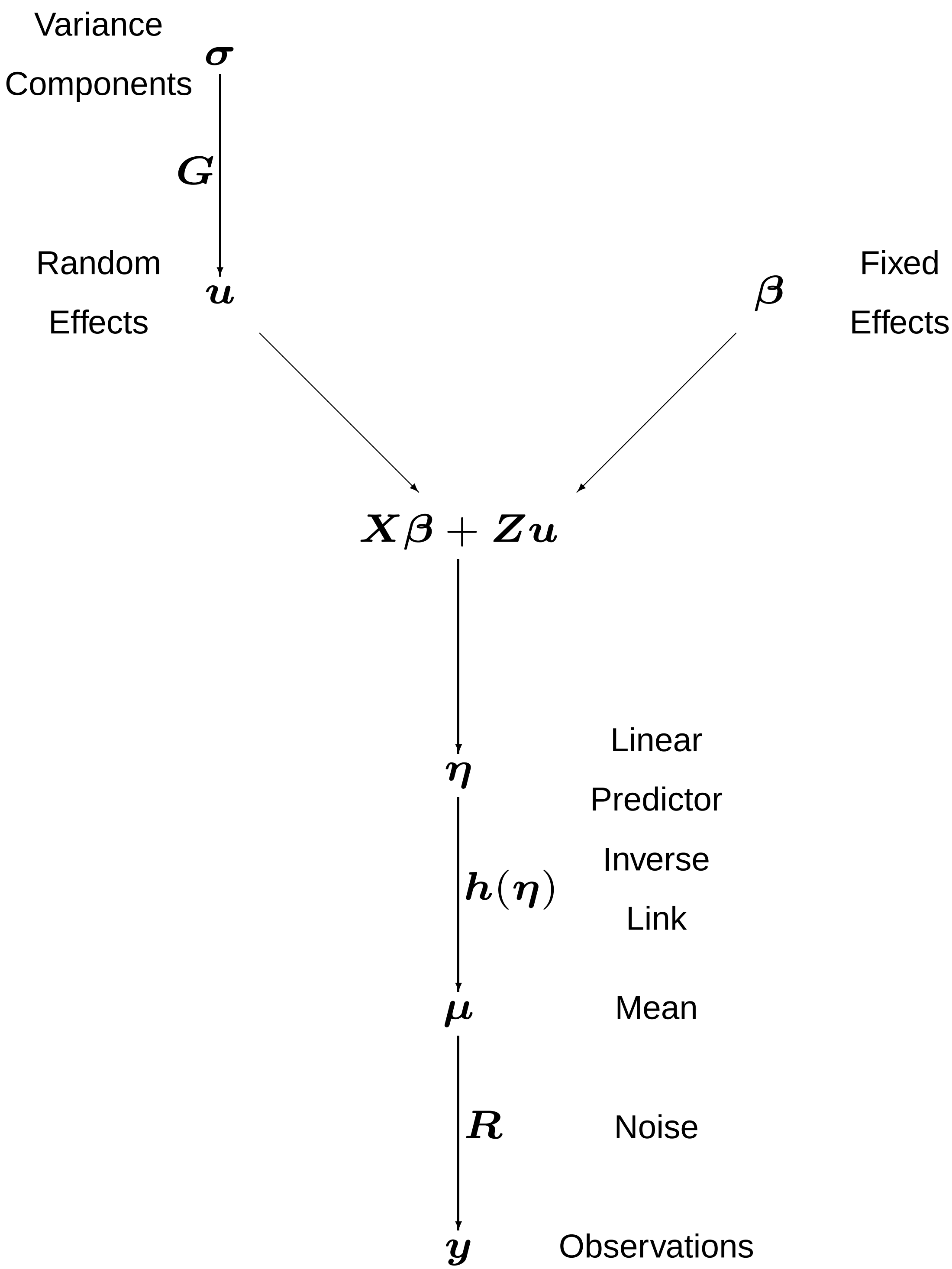
- $\mu + \tau_j + \delta_k + (\tau\delta)_{jk}$

Experimental Design:

- R_i is the effect of block i, iid $\mathrm{N}(0, \sigma_R^2)$
- $(R\tau)_{ij}$ is the whole plot error, iid $\mathrm{N}(0, \sigma_{RT}^2)$
- e_{ijk} is the sublpot error, iid $\mathrm{N}(0, \sigma^2)$

Issues

- Normal $\Rightarrow$ Mixed Model
  - Transformation?
- Only Fixed effects $\Rightarrow$ GLM
  - Clinic: Binomial with a logit link
  - Count: Poisson or Negative Binomial
- Combine $\Rightarrow$ Generalized Linear Mixed Model


Variance
Components
$\boldsymbol{\sigma}$
$\boldsymbol{G}$
Random
Effects
$\boldsymbol{u}$
$\boldsymbol{\beta}$
Fixed
Effects
$\boldsymbol{X\beta} + \boldsymbol{Zu}$
$\boldsymbol{\eta}$
Linear
Predictor
$\boldsymbol{h}(\boldsymbol{\eta})$
Inverse
Link
$\boldsymbol{\mu}$
Mean
$\boldsymbol{R}$
Noise
$\boldsymbol{y}$
Observations

Distribution

$y_i|\boldsymbol{u}$ independent $f(y_i; \eta_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{z}_i\boldsymbol{u})$

$$f(\boldsymbol{y}|\boldsymbol{u}) = f(\boldsymbol{y}; \boldsymbol{\eta}) = \Pi_i f(y_i; \eta_i)$$

$$\boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{G})$$

$$f(\boldsymbol{y}, \boldsymbol{u}) = f(\boldsymbol{y}; \boldsymbol{\eta}) \left(\frac{1}{\sqrt{2\pi}} \right)^q |\boldsymbol{G}|^{-1} e^{-\frac{\boldsymbol{u}'\boldsymbol{G}^{-1}\boldsymbol{u}}{2}}$$

Estimation

Alternatives

1. Marginal
   - Focus on the fixed effects
   - $f(\boldsymbol{y};\boldsymbol{X\beta})$ difficult to obtain
2. Joint
   - Both fixed and random effects
   - Asymptotics
3. Bayesian approaches
   - Gibbs sampling
   - "Random" fixed effects

Marginal

Review:

1. $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$ with the usual assumptions
2. $\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X\beta}, \boldsymbol{V} = \boldsymbol{ZGZ}' + \boldsymbol{R})$
3. $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$
4. $\widehat{\boldsymbol{u}} = \boldsymbol{GZ}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$

The reason this was “easy”, was 2

Challenge:

- Obtain $f(\boldsymbol{y}; \boldsymbol{X\beta}) = \int \cdots \int f(\boldsymbol{y}, \boldsymbol{u}) d\boldsymbol{u}$
  1. Exact: difficult at best unless only a single set of random effects. May still be difficult/impossible to obtain a closed form solution.
  2. Approximate: Generally based on a Taylor’s series expansion of the log likelihood

Joint

Review:

- $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$ with the usual assumptions

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{pmatrix}$$

- GLM with usual assumptions

$$(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}^*$$

where

$$\boldsymbol{W} = \boldsymbol{D}'\boldsymbol{R}^{-1}\boldsymbol{D}$$

$$\boldsymbol{D} = \left[\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}'}\right]$$

$$\boldsymbol{y}^{*} = \boldsymbol{\eta} + \boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$

Combined:

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}^{*} \\ \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y}^{*} \end{pmatrix}$$

These equations can be obtained in a variety of ways:

- Bayesian posterior mode estimate using a flat prior for $\boldsymbol{\beta}$
- Taylor's series expansion of the log likelihood

Testing:

- As in mixed models.

Challenge:

- Unlike GLMs, increasing number of replications also increases number of effects in the model
- Unlike Mixed Models, link function is nonlinear.
- Asymptotic properties are questionable

Variance Component Estimation

- Variances of random effects
  - As in a standard mixed model
- Over-dispersion parameter(s)
  - $\phi v(\mu_i)$
    - Same a residual variance
  - General case $v(\mu_i, \phi)$

$$\widehat{e}_i = y_i - \widehat{\mu}_i$$

$$PEV(\widehat{e}_i) = \boldsymbol{d}_i \begin{pmatrix} x_i' & z_i' \end{pmatrix} LHS^{-} \begin{pmatrix} x_i \\ z_i \end{pmatrix} \boldsymbol{d}_i'$$

Solve

$$\boldsymbol{Q} = v(\mu_i, \phi)^{-1} \frac{\partial v(\mu_i, \phi)}{\partial \phi} v(\mu_i, \phi)^{-1}$$

$$\sum_i \widehat{e}_i' \boldsymbol{Q} \widehat{e}_i =$$

$$\mathrm{tr}\big[\boldsymbol{Q}\, \{v(\mu_i, \phi) - PEV(\widehat{e}_i)\}\big]$$

  - Computationally expensive