Generalized Linear Mixed Models

A GLMM is defined by

1. *Random Component*: $Y_{ij}|\theta_{ij}, \alpha \sim p(\cdot)$ where $p(\cdot)$ is a member of the exponential family, that is

$$p(y_{ij}|\theta_{ij}, \alpha) = \exp[\{y_{ij}\theta_{ij} - b(\theta_{ij})\})/a(\alpha) + c(y_{ij}, \alpha)],$$

   for $i = 1, ..., m$ units, and $j = 1, ..., n_i$, measurements per unit.

2. *Systematic Component*: If $\mu_{ij} = \mathrm{E}[Y_{ij}|\theta_{ij}, \alpha]$ then we have a link function $g(\cdot)$, with

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i,$$

   so that we have introduced random effects into the linear predictor. The above defines the *conditional* part of the model. The random effects are then assigned a distribution, and in a GLMM this is assumed to be

$$\boldsymbol{b}_i \sim_{iid} N(\boldsymbol{0}, \boldsymbol{D}).$$

   We also have

$$\mathrm{var}(Y_{ij}|\theta_{ij}, \alpha) = \alpha v(\mu_{ij}).$$

Marginal Moments

Mean:

$$\begin{aligned} \mathrm{E}[Y_{ij}] &= \mathrm{E}\{\mathrm{E}[Y_{ij}|\boldsymbol{b}_i]\} \\ &= \mathrm{E}[\mu_{ij}] = \mathrm{E}_b[g^{-1}(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i)]. \end{aligned}$$

Variance:

$$\begin{aligned} \mathrm{var}(Y_{ij}) &= \mathrm{E}[\mathrm{var}(Y_{ij}|\boldsymbol{b}_i)] + \mathrm{var}(\mathrm{E}[Y_{ij}|\boldsymbol{b}_i]) \\ &= \alpha \mathrm{E}_b[v\{g^{-1}(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i)\}] + \mathrm{var}_b[g^{-1}(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i)]. \end{aligned}$$

Covariance:

$$\begin{aligned} \mathrm{cov}(Y_{ij}, Y_{ik}) &= \mathrm{E}[\mathrm{cov}(Y_{ij}, Y_{ik}|\boldsymbol{b}_i)] + \mathrm{cov}(\mathrm{E}[Y_{ij}|\boldsymbol{b}_i], \mathrm{E}[Y_{ik}|\boldsymbol{b}_i]) \\ &= \mathrm{cov}\{g^{-1}(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i), g^{-1}(\boldsymbol{x}_{ik}\boldsymbol{\beta} + \boldsymbol{z}_{ik}\boldsymbol{b}_i)\} \\ &\neq 0, \end{aligned}$$

for $j \neq k$ due to shared random effects, and

$$\mathrm{cov}(Y_{ij}, Y_{lk}) = 0,$$

for $i \neq l$, as there are no shared random effects.

Example: Log-Linear Regression for Seizure Data

Data on seizures were collected on 59 epileptics.

For each patient the number of epileptic seizures were recorded during a baseline period of eight weeks, after which patients were randomized to treatment with the anti-epileptic drug progabide, or to placebo.

The number of seizures was then recorded in four consecutive two-week periods.

The age of the patient was also available.

Figures 31-33 contain summaries.

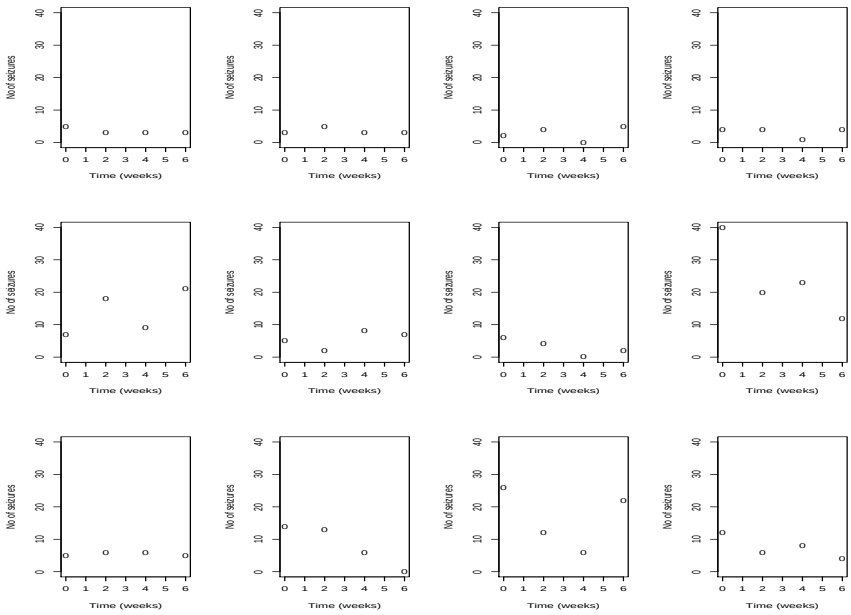

Figure 31: Number of seizures for selected individuals over time for placebo group.

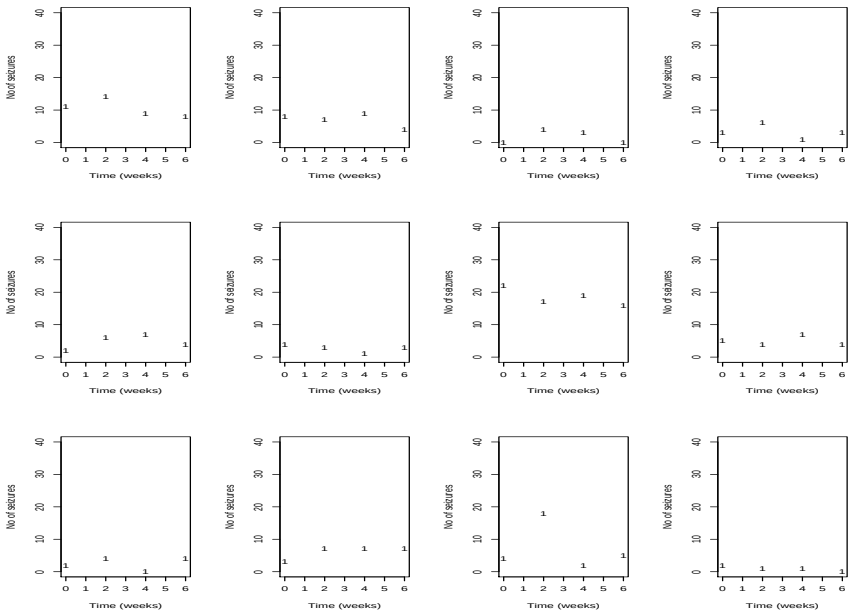

Figure 32: Number of seizures for selected individuals over time for progabide group.

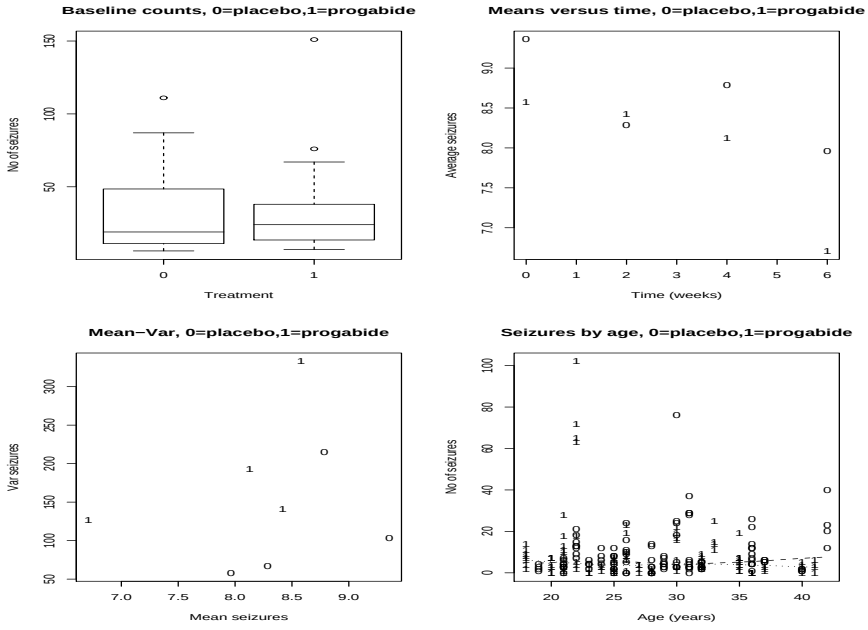

Figure 33: Summaries for seizure data.

A model for the seizure data

Let

$$\begin{aligned} Y_{ij} &= \text{number of seizures on patient } i \text{ at occasion } j \\ t_{ij} &= \text{observation period on patient } i \text{ at occasion } j \\ x_{i1} &= 0/1 \text{ if patient } i \text{ was assigned placebo/progabide} \\ x_{ij2} &= 0/1 \text{ if } j = 0/1, 2, 3, 4 \end{aligned}$$

with $t_{ij} = 8$ if $j = 0$ and $t_{ij} = 2$ if $j = 1, 2, 3, 4$, $i = 1, ..., 59$.

The question of primary scientific interest here is whether progabide reduces the number of seizures.

A marginal mean model is given by

$$E[Y_{ij}] = t_{ij} \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{ij2} + \beta_3 x_{i1} x_{ij2})$$

Group	$j = 0$ period	$j = 1, 2, 3, 4$ period
Placebo	β_0	$\beta_0 + \beta_2$
Progabide	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

Table 9: Parameter interpretation.

Precise definitions:

- $\exp(\beta_0)$ is the expected number of seizures in the placebo group in time period 0;
- $\exp(\beta_1)$ is the ratio of the expected seizure rate in the progabide group, compared to the placebo group, in time period 0;
- $\exp(\beta_2)$ is the ratio of the expected seizure rate at times $j = 1, 2, 3, 4$, as compared to $j = 0$, in the placebo group;
- $\exp(\beta_3)$ is the ratio of the expected seizure rates in the progabide group in the $j = 1, 2, 3, 4$ period, as compared to the placebo group, in the same period. Hence $\exp(\beta_3)$ is the parameter of interest.

More colloquially:

- β_0 `INTERCEPT`
- β_1 `BASELINE TREATMENT EFFECT`
- β_2 `PERIOD EFFECT`
- β_3 `TREATMENT` $\times$ `PERIOD EFFECT`

Mixed Effects Model for Seizure Data

Stage 1: $Y_{ij}|\boldsymbol{\beta}, b_i \sim_{ind}$ Poisson(μ_{ij}), with

$$g(\mu_{ij}) = \log \mu_{ij} = \log t_{ij} + \boldsymbol{x}_{ij}\boldsymbol{\beta} + b_i,$$

where

$$\boldsymbol{x}_{ij}\boldsymbol{\beta} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{i2} + \beta_3 x_{ij1} x_{i2}.$$

Hence

$$\mathrm{E}[Y_{ij}|b_i] = \mu_{ij} = t_{ij} \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + b_i), \quad \mathrm{var}(Y_{ij}|b_i) = \mu_{ij}.$$

Stage 2: $b_i \sim_{iid} N(0, \sigma^2)$.

The marginal mean is given by

$$\mathrm{E}[Y_{ij}] = t_{ij} \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \sigma^2/2),$$

and the marginal median by

$$t_{ij} \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}).$$

The marginal variance is given by

$$\begin{aligned} \mathrm{var}(Y_{ij}) &= \mathrm{E}[\mu_{ij}] + \mathrm{var}(\mu_{ij}) \\ &= \mathrm{E}[Y_{ij}]\{1 + \mathrm{E}[Y_{ij}](\mathrm{e}^{\sigma^2} - 1)\} = \mathrm{E}[Y_{ij}](1 + \mathrm{E}[Y_{ij}] \times \kappa) \end{aligned}$$

where $\kappa = \mathrm{e}^{\sigma^2} - 1 > 0$ illustrating excess-Poisson variation which increases as σ^2 increases.

For the marginal covariance

$$\begin{aligned} \mathrm{cov}(Y_{ij}, Y_{ik}) &= \mathrm{cov}\{t_{ij} \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + b_i), t_{ij} \exp(\boldsymbol{x}_{ik}\boldsymbol{\beta} + b_i)\} \\ &= t_{ij} \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{x}_{ik}\boldsymbol{\beta}) \times \mathrm{e}^{\sigma^2}\{\mathrm{e}^{\sigma^2} - 1\} = \mathrm{E}[Y_{ij}]\mathrm{E}[Y_{ik}]\kappa. \end{aligned}$$

Hence for individual i we have variance-covariance matrix

$$\begin{bmatrix} \mu_{i1} + \mu_{i1}^2\kappa & \mu_{i1}\mu_{i2}\kappa & \ldots & \mu_{i1}\mu_{in_i}\kappa \\ \mu_{i2}\mu_{i1}\kappa & \mu_{i2} + \mu_{i2}^2\kappa & \ldots & \mu_{i2}\mu_{in_i}\kappa \\ \ldots & \ldots & \ldots & \ldots \\ \mu_{in_i}\mu_{i1}\kappa & \mu_{in_i}\mu_{i2}\kappa & \ldots & \mu_{in_i} + \mu_{in_i}^2\kappa \end{bmatrix},$$

where $\kappa = \mathrm{e}^{\sigma^2} - 1 > 0$. A deficiency of this model is that we only have a single parameter (σ^2) to control both excess-Poisson variability and dependence.

Likelihood Inference

In general there are two approaches to inference from a likelihood perspective:

1. Carry out conditional inference in order to eliminate the random effects.
2. Make a distributional assumption for $\boldsymbol{b}_i$, and then carry out likelihood inference (using some form of approximation to evaluate the required integrals).

We first consider the conditional likelihood approach.

Conditional Likelihood

Recall the definition of conditional likelihood. Suppose the distribution of the data may be factored as

$$p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}) = h(\boldsymbol{y}) \times p(\boldsymbol{t}_1, \boldsymbol{t}_2 \mid \boldsymbol{\beta}, \boldsymbol{\gamma}) = h(\boldsymbol{y}) \times p(\boldsymbol{t}_1 \mid \boldsymbol{t}_2, \boldsymbol{\beta}) \times p(\boldsymbol{t}_2 \mid \boldsymbol{\beta}, \boldsymbol{\gamma}),$$

where we choose to ignore the second term and consider the conditional likelihood

$$L_c(\boldsymbol{\beta}) = p(\boldsymbol{t}_1 \mid \boldsymbol{t}_2, \boldsymbol{\beta}) = \frac{p(\boldsymbol{t}_1, \boldsymbol{t}_2 \mid \boldsymbol{\beta}, \boldsymbol{\gamma})}{p(\boldsymbol{t}_2 \mid \boldsymbol{\beta}, \boldsymbol{\gamma})}.$$

Maximizing the conditional likelihood yields an estimator, $\widehat{\boldsymbol{\beta}}_c$ with the usual properties, for example

$$\boldsymbol{I}_c(\boldsymbol{\beta})^{1/2}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}) \rightarrow_d \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}),$$

and $\boldsymbol{I}_c(\boldsymbol{\beta})$ is the expected information derived from the conditional likelihood.

Conditional Likelihood for GLMMs

In the context of GLMMs we have

$$L_c(\boldsymbol{\beta}) = \prod_{i=1}^{m} p(\boldsymbol{t}_{1i} \mid \boldsymbol{t}_{2i}, \boldsymbol{\beta}) = \prod_{i=1}^{m} \frac{p(\boldsymbol{t}_{1i}, \boldsymbol{t}_{2i} \mid \boldsymbol{\beta}, \boldsymbol{b}_i)}{p(\boldsymbol{t}_{2i} \mid \boldsymbol{\beta}, \boldsymbol{b}_i)}$$

where

$$p(\boldsymbol{t}_{1i}, \boldsymbol{t}_{2i} \mid \boldsymbol{\beta}, \boldsymbol{b}_i) \propto p(\boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{b}_i)$$

and

$$p(\boldsymbol{t}_{2i} \mid \boldsymbol{\beta}, \boldsymbol{b}_i) = \sum_{S_{2i}} p(\boldsymbol{u}_{1i}, \boldsymbol{t}_{2i} \mid \boldsymbol{\beta}, \boldsymbol{b}_i),$$

and S_{2i} is the set of values of $\boldsymbol{y}_i$ such that $\boldsymbol{T}_{2i} = \boldsymbol{t}_{2i}$, a set of disjoint events.

The different notation is to emphasize that $\boldsymbol{T}_{1i}$ takes on values different to $\boldsymbol{t}_{1i}$.

For simplicity we assume the canonical link function,

$$g(\mu_{ij}) = \theta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i$$

and assume $\alpha = 1$. Viewing $\boldsymbol{b}_i$ as fixed effects we have the likelihood

$$L(\boldsymbol{\beta}, \boldsymbol{b}) = \exp\left\{\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}\boldsymbol{x}_{ij}\boldsymbol{\beta} + y_{ij}\boldsymbol{z}_{ij}\boldsymbol{b}_i - b(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i)\right\},$$

so that

$$\boldsymbol{t}_1 = \sum_{i=1}^{m} \boldsymbol{t}_{1i} = \sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}\boldsymbol{x}_{ij}$$

and

$$\boldsymbol{t}_{2i} = \sum_{j=1}^{n_i} y_{ij}\boldsymbol{z}_{ij}.$$

We emphasize that no distribution has been specified for the $\boldsymbol{b}_i$, and they are being viewed as fixed effects.

Conditional Likelihood for the Poisson GLMM

Assume for simplicity that $\boldsymbol{z}_{ij}\boldsymbol{b}_i = b_i$, so that we have the random intercepts only model. Also, in an obvious change in notation

$$\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{x}_i\boldsymbol{\beta}_1 + b_i = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \gamma_i$$

so that $\boldsymbol{\beta}$ are the regression associated with covariates that change within an individual.

Then

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \prod_{i=1}^m p(\boldsymbol{y}_i \mid \boldsymbol{\beta}, \gamma_i) = \prod_{i=1}^m \frac{\exp\left(-\sum_{j=1}^m \mu_{ij} + \sum_{j=1}^m y_{ij}\log\mu_{ij}\right)}{\prod_{j=1}^{n_i} y_{ij}!} \\ &= c_1 \prod_{i=1}^m \exp\left(-\mu_{i+} + y_{i+}\gamma_i + \sum_{j=1}^{n_i} y_{ij}\log\left(t_{ij}\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right)\right) \end{aligned}$$

where $c_1^{-1} = \prod_i \prod_j y_{ij}!$ and $\mu_{i+} = \sum_{j=1}^m t_{ij}\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})$.

In this case the distribution of the conditioning statistic is straightforward:

$$y_{i+} \mid \boldsymbol{\beta}, \gamma_i \sim \text{Poisson}(\mu_{i+})$$

so that

$$\begin{aligned} p(y_{i+} \mid \boldsymbol{\beta}, \gamma_i) &= c_2 \prod_{i=1}^m \exp(-\mu_{i+} + y_{i+}\log\mu_{i+}) \\ &= c_2 \prod_{i=1}^m \exp\left(-\mu_{i+} + y_{i+}\gamma_i + y_{i+}\log\left(\sum_{j=1}^{n_i} t_{ij}\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right)\right) \end{aligned}$$

where $c_2^{-1} = y_{i+}!$

Hence

$$p(\boldsymbol{y} \mid y_{1+}, ..., y_{n_i+}, \boldsymbol{\beta}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma})}{p(y_{1+}, ..., y_{n_i+} \mid \boldsymbol{\beta}, \boldsymbol{\gamma})}$$

which is given by

$$\frac{c_1 \prod_i \exp\left(-\mu_{i+} + y_{i+}\gamma_i + \sum_{j=1}^{n_i} y_{ij}\log(t_{ij}\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right)}{c_2 \prod_{i=1}^m \exp\left(-\mu_{i+} + y_{i+}\gamma_i + y_{i+}\log\left(\sum_{j=1}^{n_i} t_{ij}\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right)\right)}$$

After simplification:

$$\begin{aligned} p(\boldsymbol{y} \mid y_{1+}, ..., y_{n_i+}, \boldsymbol{\beta}) &= \frac{c_1 \prod_{i=1}^m \prod_{j=1}^{n_i} (t_{ij} \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}))^{y_{ij}}}{c_2 \prod_{i=1}^m \left(\sum_{j=1}^{n_i} t_{ij} \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})\right)^{y_{i+}}} \\ &= \begin{pmatrix} y_{i+} \\ y_{i1} ... y_{in_i} \end{pmatrix} \prod_{i=1}^m \prod_{j=1}^{n_i} \left(\frac{t_{ij} \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})}{\sum_{l=1}^{n_i} t_{il} \exp(\boldsymbol{x}_{il}\boldsymbol{\beta})} \right)^{y_{ij}} \end{aligned}$$

which is a multinomial likelihood (we have conditioned a set of Poisson counts on their total so obvious!):

$$y_{ij} \mid y_{i+}, \boldsymbol{\beta} \sim \text{Mult}_{n_i}(y_{i+}, \boldsymbol{\pi}_i)$$

where $\boldsymbol{\pi}_i^{\text{T}} = (\pi_{i1}, ..., \pi_{in_i})$ and

$$\pi_{ij} = \frac{t_{ij} \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta})}{\sum_{l=1}^{n_i} t_{il} \exp(\boldsymbol{x}_{il}\boldsymbol{\beta})}.$$

Conditional Likelihood for the Seizure Data

Recall

$$\begin{aligned} Y_{ij} &= \text{number of seizures on patient } i \text{ at occasion } j \\ t_{ij} &= \text{observation period on patient } i \text{ at occasion } j \\ x_{i1} &= 0/1 \text{ if patient } i \text{ was assigned placebo/progabide} \\ x_{ij2} &= 0/1 \text{ if } j = 0/1, 2, 3, 4 \end{aligned}$$

with $t_{ij} = 8$ if $j = 0$ and $t_{ij} = 2$ if $j = 1, 2, 3, 4$, $i = 1, ..., 59$.

A log-linear random intercept model is given by

$$\log \text{E}[Y_{ij} \mid b_i] = \log t_{ij} + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{ij2} + \beta_3 x_{i1} x_{ij2} + b_i$$

Precise definitions:

- $\exp(\beta_0)$ is the expected number of seizures for a typical individual in the placebo group in time period 0;
- $\exp(\beta_1)$ is the ratio of the expected seizure rate in the progabide group, compared to the placebo group, for a typical individual, i.e. one with $b_i = 0$, in time period 0;
- $\exp(\beta_2)$ is the ratio of the expected seizure rate at times $j = 1, 2, 3, 4$, as compared to $j = 0$, for a typical individual in the placebo group;
- $\exp(\beta_3)$ is the ratio of the expected seizure rates in the progabide group in the $j = 1, 2, 3, 4$ period, as compared to the placebo group, in the same period for a typical individual. Hence $\exp(\beta_3)$ is the parameter of interest.

In the conditional likelihood notation:

$$\log \mathrm{E}[Y_{ij} \mid \gamma_i] = \log t_{ij} + \gamma_i + \beta_2 x_{ij2} + \beta_3 x_{i1} x_{ij2}$$

where $\gamma_i = \beta_0 + \beta_1 x_{i1} + b_i$ so that we cannot estimate β_1, which is not a parameter of primary interest.

Since $\boldsymbol{x}_{i1} = \boldsymbol{x}_{i2} = \boldsymbol{x}_{i3} = \boldsymbol{x}_{i4}$ and $t_{i0} = 8 = \sum_{j=1}^{4} t_{ij}$, we effectively have two observation periods which we label (slightly abusing our previous notation), $j = 0, 1$. Let $Y_{i1} = \sum_{j=1}^{4} Y_{ij}$.

For the placebo group:

$$Y_{i1} \sim_{ind} \text{Binomial}(Y_{i+}, \pi_{i1})$$

for $i = 1, ..., 29$, with

$$\pi_{i1} = \frac{\exp(\beta_2)}{1 + \exp(\beta_2)}.$$

For the progabide group:

$$Y_{i1} \sim_{ind} \text{Binomial}(Y_{i+}, \pi_{i1})$$

for $i = 30, ..., 59$, where

$$\pi_{i1} = \frac{\exp(\beta_2 + \beta_3)}{1 + \exp(\beta_2 + \beta_3)}.$$

.

This model is straightforward to fit in R:

```
> xcond <- c(rep(0,28),rep(1,31))
> condmod <- glm(cbind(y1,y0)~xcond,family=binomial)
> summary(condmod)
Call:
glm(formula = cbind(y1, y0) ~ xcond, family = binomial)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.11080    0.04689   2.363   0.0181 *
xcond       -0.10368    0.06505  -1.594   0.1110
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 306.50  on 58  degrees of freedom
Residual deviance: 303.96  on 57  degrees of freedom
```

Hence the treatment effect is $\exp(-.10) = 0.90$ so that the rate of seizures is estimated as 10% less in the progabide group, though this change is not statistically significant.

Conditional Likelihood for the Seizure Data

The overall fit of the random intercept model is poor (304 on 57 degrees of freedom).

Once possibility is to extend the model to allow a random slope for the effect of treatment x_{ij2}, i.e. $\beta_{2i} = \beta_2 + b_{2i}$, but a conditional likelihood approach for this model will condition away the information relevant for estimation of β_3.

We will examine such a model using a mixed effects approach.

Likelihood Inference in the Mixed Effects Model

As with the linear mixed effects model (LMEM) we maximize $L(\boldsymbol{\beta}, \boldsymbol{\alpha})$ where $\boldsymbol{\alpha}$ denote the variance components in $\boldsymbol{D}$, and

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{i=1}^{m} \int p(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{b}_i) \times p(\boldsymbol{b}_i|\boldsymbol{\alpha}) \, \mathrm{d}\boldsymbol{b}_i.$$

As with the NLMEM the required integrals are not available in closed form and so some sort of analytical or numerical approximation is required.

Example: Log-linear Poisson regression GLMM

With a single random effect we have $\boldsymbol{\alpha} = \sigma^2$.

$$\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \prod_{i=1}^{m} \int \prod_{j=1}^{n_i} \frac{\exp(-\mu_{ij})\mu_{ij}^{y_{ij}}}{y_{ij}!} \times (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2} b_i^2\right) \, \mathrm{d}b_i \\
&= \prod_{i=1}^{m} (2\pi\sigma^2)^{-1/2} \exp\left(\sum_{i=1}^{n_i} y_{ij} x_{ij} \boldsymbol{\beta}\right) \\
&\times \int \exp\left(-\mathrm{e}^{b_i} \sum_{j=1}^{n_i} \mathrm{e}^{\boldsymbol{x}_{ij}\boldsymbol{\beta}} + \sum_{j=1}^{n_i} y_{ij} b_i - \frac{1}{2\sigma^2} b_i^2\right) \, \mathrm{d}b_i \\
&= \prod_{i=1}^{m} \exp\left(\sum_{i=1}^{n_i} y_{ij} x_{ij} \boldsymbol{\beta}\right) \times \int h(b_i) \frac{\exp\{-b_i^2/(2\sigma^2)\}}{(2\pi\sigma^2)^{-1/2}} \, \mathrm{d}b_i,
\end{aligned}$$

an integral with respect to a normal random variable (which is analytically intractable).