

Generalized Linear Mixed Models

Outline

- Subject-specific Models
- Conditional Inference
- Random-effects Approaches
- Beta-Binomial Model
- Poisson-Gamma Model
- The Generalized Linear Mixed Model (GLMM)

Subject-Specific Models

Assumptions

- Given the subject-specific effects \mathbf{b}_i (a q -vector), the responses Y_{ij} are *independent* and follow a distribution from the exponential family.

$$Y_{ij} | \mathbf{b}_i \sim f(y_{ij} | \mathbf{b}_i, \theta).$$

Let

$$E(Y_{ij} | \mathbf{b}_i) = \mu_{ij}$$

then

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_{ij}^T \mathbf{b}_i,$$

where η_i is the linear predictor and g is the link function. \mathbf{X}_{ij} and \mathbf{Z}_{ij} are p - and q -vector of covariates, with \mathbf{Z} often being a subset of \mathbf{X} .

Three Ways to Handle Subject-specific Parameters

- Treated as fixed unknown parameters. Neyman and Scott (1948) showed that the ML estimates may be inconsistent due to the fact that the number of unknown parameters increases with the sample size.
- Conditional inference.
 - appropriate when only interested in regression coefficients that do not vary across subjects;
 - subject-specific effects $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m$ are treated as nuisance **parameters**;
 - estimate $\boldsymbol{\beta}$ using the conditional likelihood given the “sufficient” (more or less) statistics for \mathbf{b}_i .
- Random-effect approaches
 - appropriate when subject-specific coefficients are of interest or conditioning discards too much information.
 - treat \mathbf{b}_i as unobserved random variables and integrate them out to get the marginal likelihood of $\boldsymbol{\beta}$.
 - The random effects \mathbf{b}_i are independent and identically distributed with mean $\mathbf{0}$ and variance $D(\boldsymbol{\alpha})$. Its distribution G is completely specified with parameters $\boldsymbol{\alpha}$. That is, G does not depend on any covariates.
 - Estimation of inference for $\boldsymbol{\beta}$ is obtained from ML estimation, based on the marginal density for \mathbf{Y}_i .
 - Examples include hierarchical generalized linear model (beta-binomial model, poisson-gamma model) and generalized linear mixed model.

Hierarchical Generalized Linear Model

Overdispersion

- For independent count (Poisson or Binomial) data, there is often overdispersion. One possible cause is heterogeneity, for example, due to an unobserved important covariate.
- When the data is clustered, ignoring that the correlation can also result in overdispersion which, when ignored, leads to underestimate of the standard errors of the regression parameters.
- Example: In teratology experiments, pregnant rats are randomized to receive a teratogenic or a control agent, then the total number of animals in a litter and the number of birth defects are recorded. Because in a litter, all births have the same mother and hence their outcomes (having birth defect or not) are correlated.
- Example: Hospitals are randomized to use a new treatment program for alcoholics or stay with the current program. The response variable is the number of hospitalization in the year following enrollment in the program. One might expect a “hospital” effect such that the hospitalization events at the same hospital are correlated.
- Previously we used quasi-likelihood methods to take into account of the overdispersion by using a scale parameter ϕ .
- For likelihood-based methods, it is natural to model the “litter” or “hospital” effects as random effects. Early examples are beta-binomial and Poisson-Gamma (negative binomial) models.

Beta-Binomial Model

- For the teratology experiment, let $Y_{ijk} = 1$ if animal k from litter j in treatment group i has a birth defect, and 0 otherwise. Then we can assume:

$$Y_{ijk} \mid \pi_{ij} \stackrel{iid}{\sim} \text{Bin}(\pi_{ij})$$

$$\pi_{ij} \stackrel{iid}{\sim} \text{Beta}(\alpha_i, \beta_i)$$

- The reason to choose Beta distribution for π_{ij} is that it is the *conjugate* distribution for Binomial distribution and hence is computational convenient (important in early years!)
- Under this model, Y_{ijk} has a marginal Bernoulli distribution with mean

$$\mu_i = E(Y_{ijk}) = \frac{\alpha_i}{\alpha_i + \beta_i}.$$

- Note that $Y_{ij+} = \sum_k Y_{ijk}$ does not have a Binomial distribution, but a Beta-Binomial distribution with density:

$$\Pr(Y_{ij+} = y; \alpha_i, \beta_i) = \binom{n_{ij}}{y} \frac{B(y + \alpha_i, n_{ij} - y + \beta_i)}{B(\alpha_i, \beta_i)},$$

where B is the Beta function.

- The mean and variances are:

$$E(Y_{ij+}) = n_{ij}\mu_i$$

$$\text{Var}(Y_{ij+}) = n_{ij}\mu_i(1 - \mu_i) \{1 + \rho_i(n_{ij} - 1)\}$$

where

$$\rho_i = \frac{1}{\alpha_i + \beta_i + 1} = \frac{1}{\theta_i + 1}$$

$$\mu_i = \frac{\alpha_i}{\alpha_i + \beta_i}.$$

- The Beta-Binomial model is *not* in the exponential family. Even though it is adequate for simple experiments but is more difficult to extend, for example, to model covariates or to allow more complicated correlation structure.

- In contrast, in a quasi-likelihood model we may use

$$\text{Var}(Y_{ij+}) = \phi n_{ij} \mu_i (1 - \mu_i). \quad (1)$$

or use a variance function motivated by Beta-Binomial distribution with fixed ρ .

$$\text{Var}(Y_{ij+}) = n_{ij} \mu_i (1 - \mu_i) \{1 + (n_{ij} - 1) \rho\}. \quad (2)$$

There is no clear conclusion which model is better (Liang and McCullagh, 1993).

Moore's Teratology Data

This data example is from the website of Agresti (2002). Female rats were put on iron-deficient diets and then randomized to receive placebo (group 1) and three different iron supplements (groups 2, 3, and 4). They are sacrificed 3 weeks after pregnant.

n is the total number of fetuses in a litter and y is the number of dead fetuses.

```
> tera <- read.table("../data/tera.dat",
+                   col.names = c("litter", "group", "n", "y"))
>
> tera$group <- factor(tera$group)
> tera[1:5,]
  litter group  n  y
1      1     1 10  1
2      2     1 11  4
3      3     1 12  9
4      4     1  4  4
5      5     1 10 10
```

First we fit a Binomial model. From the residual plot, there is clearly overdispersion that depends on the litter size.

```
> tera.glm <- glm(cbind(y, n - y) ~ group, data = tera,
+                family = binomial)
> summary(tera.glm)
```

Call:

```
glm(formula = cbind(y, n - y) ~ group, family = binomial, data = tera)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-4.42952	-0.97500	-0.02846	1.40242	2.78260

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.1440	0.1292	8.855	< 2e-16 ***
group2	-3.3225	0.3308	-10.043	< 2e-16 ***
group3	-4.4762	0.7311	-6.122	9.22e-10 ***
group4	-4.1297	0.4762	-8.672	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

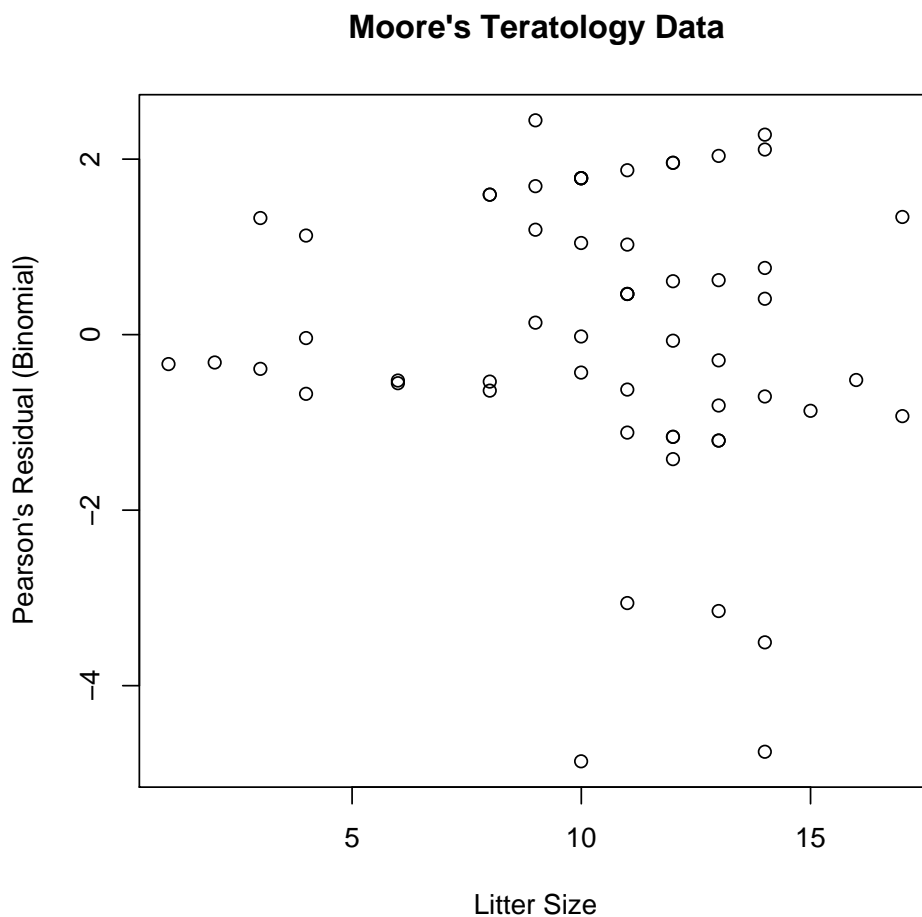
Null deviance: 509.43 on 57 degrees of freedom

Residual deviance: 173.45 on 54 degrees of freedom

AIC: 252.92

Number of Fisher Scoring iterations: 5

```
> plot (tera$n, resid (tera.glm, type = "p"), xlab = "Litter Size",  
+       ylab = "Pearson's Residual (Binomial)",  
+       main = "Moore's Teratology Data")
```



Quasi-likelihood Model 1

Next we fit a quasi-likelihood model with variance model in (1).

```
> tera.quasi <- glm (cbind(y, n - y) ~ group, data = tera,
+                    family = quasibinomial)
> summary (tera.quasi)
```

Call:

```
glm(formula = cbind(y, n - y) ~ group, family = quasibinomial,
     data = tera)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-4.42952	-0.97500	-0.02846	1.40242	2.78260

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.1440	0.2187	5.231	2.81e-06	***
group2	-3.3225	0.5600	-5.933	2.18e-07	***
group3	-4.4762	1.2375	-3.617	0.000656	***
group4	-4.1297	0.8061	-5.123	4.14e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 2.864945)

Null deviance: 509.43 on 57 degrees of freedom
 Residual deviance: 173.45 on 54 degrees of freedom
 AIC: NA

Number of Fisher Scoring iterations: 5

Note the point estimates are the same. The scale parameter does not affect the point estimates.

GEE and Empirical Standard Errors

The GEE model is the same as the quasi-likelihood model above but empirical (robust) standard errors are calculated.

```
> tera.gee <- gee (cbind(y, n - y) ~ group, data = tera, id = litter,
+                 family = binomial)
[1] "Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27"
[1] "running glm to get initial regression estimate"
[1] 1.143981 -3.322513 -4.476185 -4.129663
> summary (tera.gee)
```

```
GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)
```

Model:

```
Link:                               Logit
Variance to Mean Relation: Binomial
Correlation Structure:              Independent
```

Call:

```
gee(formula = cbind(y, n - y) ~ group, id = litter, data = tera,
    family = binomial)
```

Summary of Residuals:

	Min	1Q	Median	3Q	Max
	-0.10169492	0.03453548	2.74159021	8.24159021	13.24159021

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	1.143981	0.2186717	5.231499	0.2758667	4.146861
group2	-3.322513	0.5599915	-5.933149	0.4400582	-7.550168
group3	-4.476185	1.2375175	-3.617068	0.6104577	-7.332507
group4	-4.129663	0.8060677	-5.123221	0.5763810	-7.164814

Estimated Scale Parameter: 2.864945

Quasi-likelihood Model 2

This R function by Patrick Heagerty at UW fits quasi-likelihood model with the variance model in (2), that is,

$$\text{Var}(Y_{ij}) = n_{ij}\mu_{ij}(1 - \mu_{ij})(1 + \rho(n_{ij} - 1)).$$

Note that the point estimates are different now.

```
> source ("bod_regn.R")
> x <- cbind (1, tera$g2, tera$g3, tera$g4)
> tera.bod <- bod.regn (y = tera$y, n = tera$n, x = x,
+                       dispersion = "correlation",
+                       beta = c(1, -3, -4, -4), alpha = 2)
```

Z is NULL. Common alpha assumed.

```
> tera.bod
```

```
Regression for Binomial Overdispersed Data
dispersion = correlation
```

```
number of observations = 58
```

Mean parameters:

	estimate	mod. s.e.	emp. s.e.	Z	emp. p	emp.
beta1	1.212	0.222	0.270	4.496		0
beta2	-3.369	0.559	0.430	-7.827		0
beta3	-4.584	1.294	0.624	-7.350		0
beta4	-4.249	0.843	0.605	-7.021		0

Dispersion parameters:

	estimate	emp. s.e.
alpha	-1.669719	0.3082800

In this model $\hat{\rho} = \exp(-1.67) = 0.188$.

Beta-Binomial Model: Maximum Likelihood

The estimation of the “Beta-Binomial” regression models above are actually done using estimating equations (quasi-likelihood). The maximum likelihood estimation based on the Beta-Binomial model is not trivial. To use ML, we use the reparametrization: $(\alpha_i, \beta_i) \rightarrow (\mu_i, \theta_i)$.

In the next model we assume:

$$\begin{array}{ll} \text{location} & \text{logit}(\mu_i) = \beta_1 + \beta_2 g_2 + \beta_3 g_3 + \beta_4 g_4 \\ \text{shape (constant)} & \theta_i = \theta \end{array}$$

```
> library (rmutil)
> library (gnlm) % for 'gnlr' function
> tera$g2 <- as.numeric (tera$group == 2)
> tera$g3 <- as.numeric (tera$group == 3)
> tera$g4 <- as.numeric (tera$group == 4)
> library (boot) % for 'inv.logit' function
> attach (tera)
> tera.gnlr2 <- gnlr (cbind (y, n - y), distribution = "beta binomial",
+ mu = function (beta) \{
+ inv.logit (beta[1] + beta[2] * g2 +
+ beta[3] * g3 + beta[4] * g4)
+ \}, pmu = c(1, -3, -4, -4),
+ pshape = 2)
> tera.gnlr2
```

Call:

```
gnlr(cbind(y, n - y), distribution = "beta binomial",
mu = function(beta) \{
inv.logit(beta[1] + beta[2] * g2 + beta[3] * g3 + beta[4] *
g4)
\}, pmu = c(1, -3, -4, -4), pshape = 2)
```

beta binomial distribution

Log likelihood function:

```
\{
m <- mu1(p)
s <- exp(sh1(p))
```

```

t <- s * m
u <- s * (1 - m)
-sum(wt * (lbeta(y[, 1] + t, y[, 2] + u) - lbeta(t, u)))
\}

```

Location function:

```

\{
  inv.logit(beta[1] + beta[2] * g2 + beta[3] * g3 + beta[4] *
    g4)
\}

```

Log shape function:

```
p[1] * rep(1, n)
```

```

-Log likelihood      93.45675
Degrees of freedom  53
AIC                  98.45675
Iterations           7

```

Location parameters:

	estimate	se
beta[1]	1.346	0.2482
beta[2]	-3.114	0.5018
beta[3]	-3.868	0.8081
beta[4]	-3.923	0.6675

Shape parameters:

	estimate	se
p[1]	1.146	0.3299

Correlations:

	1	2	3	4	5
1	1.0000	-0.5338	-0.3383	-0.4025	0.1720
2	-0.5338	1.0000	0.2211	0.2548	-0.3143
3	-0.3383	0.2211	1.0000	0.1679	-0.2350
4	-0.4025	0.2548	0.1679	1.0000	-0.2436
5	0.1720	-0.3143	-0.2350	-0.2436	1.0000

Here $\hat{\rho} = 1/(\exp(1.146) + 1) = 0.242$.

In the next model we assume,

$$\log(\theta_i) = \gamma_0 + \gamma_1 n_{ij}.$$

```
> tera.gnlr <- gnlr (cbind (y, n - y), distribution = "beta binomial",
+                   mu = function (beta) \{
+                       inv.logit (beta[1] + beta[2] * g2 +
+                                   beta[3] * g3 +
+                                   beta[4] * g4)
+                   \}, pmu = c(1, -3, -4, -4),
+                   shape = finterp (~ n),
+                   pshape = c(2, 0.5))
> tera.gnlr
```

Call:

```
gnlr(cbind(y, n - y), distribution = "beta binomial",
     mu = function(beta) \{
       inv.logit(beta[1] + beta[2] * g2 + beta[3] * g3 + beta[4] * g4)
     \}, pmu = c(1, -3, -4, -4), shape = finterp(~n), pshape = c(2,
     0.5))
```

beta binomial distribution

Log likelihood function:

```
\{
  m <- mu1(p)
  s <- exp(sh1(p))
  t <- s * m
  u <- s * (1 - m)
  -sum(wt * (lbeta(y[, 1] + t, y[, 2] + u) - lbeta(t, u)))
\}
```

Location function:

```
\{
  inv.logit(beta[1] + beta[2] * g2 + beta[3] * g3 +
            beta[4] * g4)
\}
```

Log shape function:

```
~n
```

```

-Log likelihood    93.44569
Degrees of freedom 52
AIC                99.44569
Iterations        22

```

Location parameters:

	estimate	se
beta[1]	1.335	0.2587
beta[2]	-3.105	0.5061
beta[3]	-3.859	0.8103
beta[4]	-3.921	0.6673

Shape parameters:

	estimate	se
(Intercept)	0.93694	1.4375
n	0.01776	0.1192

Correlations:

	1	2	3	4	5	6
1	1.0000	-0.5448	-0.34467	-0.39068	0.31051	-0.27987
2	-0.5448	1.0000	0.22836	0.25494	-0.19686	0.12883
3	-0.3447	0.2284	1.00000	0.16874	-0.12470	0.07299
4	-0.3907	0.2549	0.16874	1.00000	-0.06879	0.01326
5	0.3105	-0.1969	-0.12470	-0.06879	1.00000	-0.97347
6	-0.2799	0.1288	0.07299	0.01326	-0.97347	1.00000

From AIC the simpler Beta-Binomial is preferable and fit the data much better than the simple Binomial model (98 vs 253).

Poisson-Gamma Model

For clustered count data, we can assume

$$Y_{ij} | \mu_{ij} \stackrel{\text{iid}}{\sim} \text{Poisson}(\mu_{ij})$$

$$\mu_{ij} \stackrel{\text{iid}}{\sim} \text{Gamma}(\theta_i, \lambda_i)$$

The marginal distribution of Y_{ij} is *negative-Binomial* with

$$E(Y_{ij}) = E\{E(Y_{ij} | \mu_{ij})\} = \theta_i$$

$$\text{Var}(Y_{ij}) = \theta_i + \theta_i/\lambda_i^2$$

Note that we use a somewhat unusual parameterization for Gamma distribution such that $E(\mu_{ij}) = \theta_i$ and $\text{Var}(\mu_{ij}) = \theta_i^2/\lambda_i$.

We can then model the marginal mean response as

$$\log(E(Y_{ij})) = \mathbf{X}_{ij}^T \boldsymbol{\beta}.$$

For fixed λ_i , the negative binomial is in the exponential family. Therefore to find the MLE for the negative binomial model, we can alternate the two iterative steps until convergence.

- For fixed λ , fit the GLM to solve for $\boldsymbol{\beta}$ in a regression model.
- For fixed μ , estimate λ using Newton-Raphson method.

The `glm.nb` function in R `MASS` library implements this iterative procedure while `gnlr` can estimate all parameters simultaneously.

Conjugated Mixture Models

- The Beta-Binomial and Poisson-Gamma models are examples of **conjugated mixture models** where the marginal distribution has closed form.
- These models are mainly used to account for overdispersion or simple clustered data and are less suitable to study longitudinal data where the correlation structure may be more complicated.
- In Lee and Nelder (1996), they extended this class of model to what they called **hierarchical generalized linear models** which are like *generalized linear mixed models* but the distribution of the random effects need not be normal.

Conditional Inference

Sufficiency

Suppose a random vector \mathbf{Y} has density indexed by parameter θ , and $s = s(\mathbf{y})$ is a statistic. s is said to be *sufficient* for θ if

$$f(\mathbf{y}; \theta) \propto f(s; \theta)f(\mathbf{y} | s).$$

The inference for θ can be based on the marginal density of s and no information is lost. The conditional density $f(\mathbf{y} | s)$ is useful for model checking but not in inference for θ .

If

$$f(\mathbf{y}; \theta) \propto f(s | t; \theta)f(t),$$

then $t = t(\mathbf{y})$ is said to be *ancillary* for θ . In this case the conditional density $f(s | t, \theta)$ is used for inference about θ .

When there is a nuisance parameter λ , an extension of the factorization is

$$f(s; \theta, \lambda) = f(s_1 | s_2; \theta)f(s_2; \lambda).$$

- s_2 is sufficient for λ and is ancillary for θ .
- s_1 is conditionally or partially sufficient for θ .

We can use the conditional density $f(s_1 | s_2; \theta)$ for inference about θ and the marginal density $f(s_2; \lambda)$ for inference about λ .

Often we only have

$$f(\mathbf{y}; \theta, \lambda) = f(\mathbf{y} | s_2; \theta)f(s_2; \theta, \lambda).$$

We can still use $f(\mathbf{y} | s_2; \theta)$ for inference (for more pragmatic reasons) about θ but there is potential information in $f(s_2; \theta, \lambda)$ about θ .

Conditional Likelihood

- We will consider the binary (Bernoulli) and count (Poisson) data.
- Assume $a(\phi) = 1$ (no overdispersion) to simplify the discussion.
- Restrict to the canonical link, thus

$$\theta_{ij} = \eta_{ij} = g(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i.$$

The likelihood function for $\boldsymbol{\beta}$ for individual i is

$$\begin{aligned} \prod_{j=1}^{n_i} f(y_{ij} | \boldsymbol{\beta}, \mathbf{b}_i) &\propto \prod_{j=1}^{n_i} \exp \{ \theta_{ij} y_{ij} - \psi(\theta_{ij}) \} \\ &= \exp \left\{ \boldsymbol{\beta}^T \sum_{j=1}^{n_i} \mathbf{x}_{ij} y_{ij} + \mathbf{b}_i^T \sum_{j=1}^{n_i} \mathbf{z}_{ij} y_{ij} - \sum_{j=1}^{n_i} \psi(\theta_{ij}) \right\}. \quad (3) \end{aligned}$$

Hence the **sufficient statistics** for \mathbf{b}_i is

$$\mathbf{T}_i = \sum_j^{n_i} \mathbf{z}_{ij} Y_{ij}.$$

Let

$$\mathbf{S}_i = \sum_j^{n_i} \mathbf{x}_{ij} Y_{ij},$$

(which is sufficient for $\boldsymbol{\beta}$.)

The **conditional distribution** of $\mathbf{y}_i \mid \mathbf{T}_i = \mathbf{t}$ is

$$\begin{aligned} \Pr(\mathbf{y}_i \mid \mathbf{T}_i = \mathbf{t}) &= \frac{\Pr(\mathbf{y}_i, \mathbf{T}_i = \mathbf{t})}{\Pr(\mathbf{T}_i = \mathbf{t})} \\ &= \frac{\exp(\boldsymbol{\beta}^T \mathbf{s}_i + \mathbf{b}_i^T \mathbf{t})}{\sum_{\mathbf{y}_i^* \in \mathcal{R}_{t,i}} \exp(\boldsymbol{\beta}^T \mathbf{s}_i^* + \mathbf{b}_i^T \mathbf{t})} \\ &= \frac{\exp(\boldsymbol{\beta}^T \mathbf{s}_i)}{\sum_{\mathbf{y}_i^* \in \mathcal{R}_{t,i}} \exp(\boldsymbol{\beta}^T \mathbf{s}_i^*)} \end{aligned}$$

where

$$\mathcal{R}_{t,i} = \{(y_{i1}, \dots, y_{in_i}) : \mathbf{T}_i = \mathbf{t}\},$$

that is, the set of outcomes for which the statistic \mathbf{T}_i takes the value \mathbf{t} (Note: \mathbf{T}_i depends on i only through covariates \mathbf{Z}_i).

For all the data, the conditional likelihood is proportional to

$$\prod_{i=1}^m \frac{\exp(\boldsymbol{\beta}^T \mathbf{s}_i)}{\sum_{\mathbf{y}_i^* \in \mathcal{R}_{t,i}} \exp(\boldsymbol{\beta}^T \mathbf{s}_i^*)}.$$

- The conditional likelihood uses part of the data that does not contain information about $(\mathbf{b}_1, \dots, \mathbf{b}_m)$ to estimate $\boldsymbol{\beta}$.
- For simple cases such as the random intercept model, the conditional likelihood is relatively easy to maximize.
- It is not necessary to specify the distribution of \mathbf{b}_i .
- When the distribution of \mathbf{b}_i depends on covariates, an important assumption for random effects models is violated. In the case of a random intercept model, using the conditional likelihood will still give a consistent estimate of $\boldsymbol{\beta}$.

Random Intercept Model

In the random intercept model, the linear predictor is

$$\eta_{ij} = \gamma_i + \mathbf{x}_{ij}^T \boldsymbol{\beta} \quad [\text{+offsets if necessary}],$$

where $\gamma_i = \beta_0 + b_i$ and \mathbf{x}_{ij} does not include an intercept term. The sufficient statistic for γ_i is

$$T_i = \sum_{j=1}^{n_i} Y_{ij} = Y_{i+}.$$

The joint likelihood function for $\boldsymbol{\beta}$ and the γ_i is proportional to

$$\prod_{i=1}^m \exp \left[\gamma_i \sum_{j=1}^{n_i} y_{ij} + \left(\sum_{j=1}^{n_i} y_{ij} \mathbf{x}_{ij}^T \right) \boldsymbol{\beta} - \sum_{j=1}^{n_i} \log \{ 1 + \exp(\gamma_i + \mathbf{x}_{ij}^T \boldsymbol{\beta}) \} \right].$$

The conditional likelihood for $\boldsymbol{\beta}$ is proportional to

$$\prod_i^m \frac{\exp(\sum_{j=1}^{n_i} y_{ij} \mathbf{x}_{ij}^T \boldsymbol{\beta})}{\sum_{\{L \in R_i\}} \exp(\sum_{\{l \in L\}} \mathbf{x}_{il}^T \boldsymbol{\beta})}. \quad (4)$$

For example, if for subject i , there were $y_{i+} = 2$ out of $n_i = 10$ defected pregnancies were observed, then R_i contains all ways of choosing 2 y_{ij} 's from the 10 outcomes $\{y_{ij}; j = 1, \dots, 10\}$. There are 10-choose-2 (=45) possible ways in total. $\sum_{\{L \in R_i\}}$ is the sum across all 45 ways.

Logistic Regression for Binary Responses

2×2 Crossover Trial

$$n_i = 2, i = 1, \dots, m$$

Compare two treatments

- A: active drug.
- B: placebo.

Responses:

- 1 for a normal electrocardiogram reading.
- 0 for a abnormal reading.

	$y_1 = 1$	$y_1 = 0$	$y_1 = 1$	$y_1 = 0$
Group	$y_2 = 1$	$y_2 = 1$	$y_2 = 0$	$y_2 = 0$
AB	22	0	6	6
BA	18	4	2	9

(This setting looks familiar. Do we know how to analyze this data set to determine the treatment effect?)

If $y_{i+} = 2$ or 0 (2 successes or no successes in two trials), R_{i2} has a single element

$$(y_{i1}, y_{i2}) = (1, 1) \text{ or } (0, 0)$$

and the contribution to (4) is 1. Therefore, and we do not need to consider responses (1,1) or (0,0) in calculating the conditional likelihood.

Let

$$\begin{aligned}
 x_1 &= \begin{cases} 1 & \text{if A (active)} \\ 0 & \text{if B (placebo)} \end{cases} \\
 x_2 &= \begin{cases} 1 & \text{if period 2} \\ 0 & \text{if period 1} \end{cases} \\
 x_3 &= x_1 \times x_2
 \end{aligned}$$

The conditional likelihood for β is

$$\begin{aligned}
 \mathcal{L}(\beta) &= \prod_{i:y_{i1}=1} \frac{\exp(\mathbf{x}_{i1}^T \beta)}{\exp(\mathbf{x}_{i1}^T \beta) + \exp(\mathbf{x}_{i2}^T \beta)} \\
 &\quad \times \prod_{i:y_{i2}=1} \frac{\exp(\mathbf{x}_{i2}^T \beta)}{\exp(\mathbf{x}_{i1}^T \beta) + \exp(\mathbf{x}_{i2}^T \beta)}. \quad (5)
 \end{aligned}$$

Write

		Group AB				Group BA	
		period 1 (A)				period 1 (B)	
		1	0			1	0
period 2	1	a_1	c_1	period 2	1	a_2	b_2
(B)	0	b_1	d_1	(A)	0	c_2	d_2

a_1, d_1, a_2 and d_2 do not contribute to the conditional likelihood.

Values of $\sum_{j=1}^2 \mathbf{x}_{ij}^T \mathbf{y}_{ij}$

- We can ignore calculating these for $(y_{i1}, y_{i2}) = (1,1)$ or $(0,0)$
- For $i \in$ group AB

$$\mathbf{X}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

So,

$$\mathbf{x}_{i1}^T \boldsymbol{\beta} = \beta_1$$

$$\mathbf{x}_{i2}^T \boldsymbol{\beta} = \beta_2$$

and there are c_1 and b_1 of those terms.

- For $i \in$ group BA

$$\mathbf{X}_i = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

So,

$$\mathbf{x}_{i1}^T \boldsymbol{\beta} = 0$$

$$\mathbf{x}_{i2}^T \boldsymbol{\beta} = \beta_1 + \beta_2 + \beta_3$$

and there are c_2 and b_2 of those terms.

- The conditional likelihood is proportional to

$$\left(\frac{e^{\beta_2}}{e^{\beta_1} + e^{\beta_2}} \right)^{c_1} \left(\frac{e^{\beta_1}}{e^{\beta_1} + e^{\beta_2}} \right)^{b_1} \quad (\text{AB group})$$

$$\times \left(\frac{e^{\beta_1 + \beta_2 + \beta_3}}{1 + e^{\beta_1 + \beta_2 + \beta_3}} \right)^{b_2} \left(\frac{1}{1 + e^{\beta_1 + \beta_2 + \beta_3}} \right)^{c_2} \quad (\text{BA group})$$

- Special case: When $\beta_2 = \beta_3 = 0$, i.e. only consider the treatment effect, the conditional likelihood reduces to

$$\begin{aligned} \left(\frac{1}{1+e^{\beta_1}}\right)^{c_1} \left(\frac{e^{\beta_1}}{1+e^{\beta_1}}\right)^{b_1} \left(\frac{e^{\beta_1}}{1+e^{\beta_1}}\right)^{b_2} \left(\frac{1}{1+e^{\beta_1}}\right)^{c_2} \\ = \left(\frac{e^{\beta_1}}{1+e^{\beta_1}}\right)^{b_1+b_2} \left(\frac{1}{1+e^{\beta_1}}\right)^{c_1+c_2} \end{aligned}$$

Therefore

$$\begin{aligned} \hat{\beta}_1 &= \log\left(\frac{b_1+b_2}{c_1+c_2}\right) = \log\left(\frac{6+4}{0+2}\right) = 1.61 \\ \widehat{\text{s.e.}}(\hat{\beta}_1) &= \sqrt{\frac{1}{b_1+b_2} + \frac{1}{c_1+c_2}} = 0.77 \end{aligned}$$

- When $\beta_3 = 0$, the conditional likelihood is

$$\begin{aligned} \left(\frac{e^{\beta_2}}{e^{\beta_1}+e^{\beta_2}}\right)^{c_1} \left(\frac{e^{\beta_1}}{e^{\beta_1}+e^{\beta_2}}\right)^{b_1} \left(\frac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}}\right)^{b_2} \left(\frac{1}{1+e^{\beta_1+\beta_2}}\right)^{c_2} \\ = (1-p_1)^{c_1} p_1^{b_1} p_2^{b_2} (1-p_2)^{c_2} \end{aligned}$$

where

$$\text{logit}(p_1) = \beta_1 - \beta_2, \quad \text{logit}(p_2) = \beta_1 + \beta_2.$$

Then

$$\begin{aligned} \hat{p}_1 &= \frac{b_1}{b_1+c_1}, \quad \hat{p}_2 = \frac{b_2}{b_2+c_2} \\ \hat{\beta}_1 &= \frac{1}{2}(\text{logit } \hat{p}_1 + \text{logit } \hat{p}_2) = \frac{1}{2} \log\left(\frac{b_1 b_2}{c_1 c_2}\right) \\ \widehat{\text{s.e.}}(\hat{\beta}_1) &= \frac{1}{2} \sqrt{b_1^{-1} + c_1^{-1} + b_2^{-1} + c_2^{-1}}. \end{aligned}$$

Since $c_1 = 0$, use *ad hoc* convention of adding 0.5 to give

$$\begin{aligned} \hat{\beta}_1 &= \frac{1}{2} \log\left(\frac{6 \cdot 4}{0.5 \cdot 2}\right) = 1.59 \\ \widehat{\text{s.e.}}(\hat{\beta}_1) &= 0.85 \end{aligned}$$

- Note:

1. In both models, $\hat{\beta}_1 \approx 1.6$ is marginally significant at a (2-sided) 5% level.
2. $\exp(1.6) \approx 5$ indicates the odds of a normal result for a treated patient is about 5 times of the odds for a non-treated patient.
3. In conditional logistic regression we can only estimate effects for within-cluster covariates and any interactions between a within-cluster covariate and a cluster-level covariate.

Conditional Logistic Regression Results

The `clogit` function in package `survival` can fit conditional logistic regression models.

```
> xover <- read.table ("C:/xover1.data",
+                       col.names = c("id", "class", "y", "intercept",
+                                     "trt", "period", "xover", "BA"))
> with (xover, ftable (BA, trt, y))
      y  0  1
BA trt
0  0      6 28
   1     12 22
1  0     11 22
   1     13 20

> library (survival)
Loading required package: splines
> xover.cl <- clogit (y ~ trt + strata (id), data = xover)
> summary (xover.cl)
Call: coxph(formula = Surv(rep(1, 134), y) ~ trt + strata(id), data
= xover, method = "exact")

      n= 134
      coef exp(coef) se(coef)      z      p
trt -1.61      0.2   0.775 -2.08 0.038

      exp(coef) exp(-coef) lower .95 upper .95
trt      0.2         5     0.0438     0.913

Rsquare= 0.043 (max possible= 0.117 )
Likelihood ratio test= 5.82 on 1 df,  p=0.0158
Wald test = 4.32 on 1 df,  p=0.0377
Score (logrank) test = 5.33 on 1 df, p=0.0209
```

GEE Results

```
> library (gee)
> summary (gee (y ~ trt, data = xover, cor = "exchangeable",
+             id = id, family = binomial, scale.fix = TRUE))
[1] "Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27"
[1] "running glm to get initial regression estimate"
[1] 1.0788097 -0.5600159
```

```
GEE: GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)
```

Model:

```
Link:                               Logit
Variance to Mean Relation: Binomial
Correlation Structure:               Exchangeable
```

```
Call: gee(formula = y ~ trt, id = id, data = xover, family =
binomial, corstr = "exchangeable", scale.fix = TRUE)
```

Summary of Residuals:

	Min	1Q	Median	3Q	Max
	-0.7462687	-0.6268657	0.2537313	0.3731343	0.3731343

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	1.0788097	0.2807553	3.842527	0.2807553	3.842527
trt	-0.5600159	0.2328349	-2.405205	0.2356943	-2.376026

```
Estimated Scale Parameter: 1 Number of Iterations: 1
```

Working Correlation

	[,1]	[,2]
[1,]	1.0000000	0.6233823
[2,]	0.6233823	1.0000000

Note

- The marginal coefficients are smaller in absolute value.

$$| 0.56 | \leq | 1.61 |$$

(consistent with the theoretical inequality in Neuhaus Theorem).

- The marginal and conditional coefficients have different interpretations.
- The Z statistics are similar for the two regressions.
- See DHLZ sections 9.2 and 9.3 for detail.

Poisson Regression for Count Responses

- $(y_{i1}, \dots, y_{in_i})$ are independent Poisson r.v.s with

$$\log E(y_{ij} | \gamma_i, \beta) = \gamma_i + \mathbf{x}_{ij}^T \beta + \log(t_{ij})$$

where $\gamma_i = \beta_0 + b_i$ and \mathbf{x}_{ij} does not include the intercept term.

- The likelihood contributed by the i -th individual is proportional to

$$\exp \left\{ \gamma_i \sum_j y_{ij} + \beta^T \sum_j \mathbf{x}_{ij} y_{ij} + \sum_j y_{ij} \log(t_{ij}) - \sum_j t_{ij} \exp(\gamma_i + \mathbf{x}_{ij}^T \beta) \right\} \frac{1}{\prod_j y_{ij}!}$$

Then $y_{i+} = \sum_j y_{ij}$ is a sufficient statistic for γ_i .

- The distribution of y_{i+} is Poisson with mean

$$\sum_j e^{\gamma_i + \mathbf{x}_{ij}^T \beta + \log t_{ij}} = \lambda_i.$$

- Let $\mathbf{s}_i = \sum_j \mathbf{x}_{ij} y_{ij}$. The conditional distribution of \mathbf{y}_i given y_{i+}

$$\begin{aligned} & \Pr(\mathbf{y}_i | y_{i+}) \\ &= \frac{\exp \left\{ \gamma_i y_{i+} + \beta^T \mathbf{s}_i + \sum_j y_{ij} \log(t_{ij}) - \sum_j t_{ij} \exp(\gamma_i + \mathbf{x}_{ij}^T \beta) \right\} y_{ij}!}{\left(\prod_j y_{ij}! \right) \lambda_i^{y_{i+}} \exp(-\lambda_i) / y_{i+}!} \\ &= \binom{y_{i+}}{y_{i1}, \dots, y_{in_i}} \frac{\exp \left\{ \beta^T \mathbf{s}_i + \sum_j y_{ij} \log t_{ij} \right\}}{\left(\sum_j \exp \{ \mathbf{x}'_{ij} \beta + \log t_{ij} \} \right)^{y_{i+}}} \end{aligned}$$

Seizure Data

- $x_{ij1} = 1$ or 0 for progabide or placebo, respectively.
- $x_{ij2} = 1$ if $j = 1, 2, 3, 4$; 0 if $j = 0$.
- $x_{ij3} = x_{ij1}x_{ij2}$ (interaction term). Note that β_3 is the parameter of interest.

The model:

$$\log E(y_{ij} | \gamma_i, \boldsymbol{\beta}) = \gamma_i + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \log t_{ij}$$

Note e^{γ_i} is the expected baseline count for individual i .

- For the placebo group,

$$\mathbf{x}_{ij}^T = \begin{cases} (0, 0, 0) & \text{if } j = 0 \\ (0, 1, 0) & \text{if } j = 1, 2, 3, 4 \end{cases}$$

- For the treatment group,

$$\mathbf{x}_{ij}^T = \begin{cases} (1, 0, 0) & \text{if } j = 0 \\ (1, 1, 1) & \text{if } j = 1, 2, 3, 4 \end{cases}$$

Thus

$$\begin{aligned} \mathbf{s}_i &= \sum_{j=0}^4 \mathbf{x}_{ij} y_{ij} \\ &= \mathbf{x}_{i0} y_{i0} + \mathbf{x}_{i1} \sum_{j=1}^4 y_{ij} \\ &= \mathbf{x}_{i0} y_{i0} + \mathbf{x}_{i1} (y_{i+} - y_{i0}) \end{aligned}$$

If $i \in$ placebo group,

$$\mathbf{s}_i^T \boldsymbol{\beta} = \beta_2 (y_{i+} - y_{i0}),$$

otherwise it equals

$$y_{i0} \beta_1 + (y_{i+} - y_{i0})(\beta_1 + \beta_2 + \beta_3) = (y_{i+} - y_{i0})(\beta_2 + \beta_3) + y_{i+} \beta_1.$$

The conditional likelihood is proportional to

$$\prod_i \frac{\exp(\mathbf{s}_i^T \boldsymbol{\beta}) \exp(\sum_j y_{ij} \log(t_{ij}))}{\left(\sum_{j=0}^4 \exp\{\mathbf{x}_{ij} \boldsymbol{\beta} + \log t_{ij}\}\right)^{y_{i+}}}$$

Here we have $t_{i0} = 8$ and $t_{i1} = t_{i2} = t_{i3} = t_{i4} = 2$.

- For $i \in$ placebo group, the likelihood is proportional to

$$\frac{\exp\{\beta_2(y_{i+} - y_{i0})\} \prod_{j=0}^4 t_{ij}^{y_{ij}}}{(t_{i0} + \exp\{\beta_2 \sum_{j=1}^4 t_{ij}\})^{y_{i+}}} = \frac{e^{\beta_2(y_{i+} - y_{i0})}}{(1 + e^{\beta_2})^{y_{i+}}} \times \frac{\prod_{j=0}^4 t_{ij}^{y_{ij}}}{8^{y_{i+}}} \\ \propto (1 - \pi_1)^{y_{i+} - y_{i0}} \pi_1^{y_{i0}}$$

which is a Binomial (y_{i+}, π) , with

$$\pi_1 = \frac{1}{1 + e^{\beta_2}}.$$

- For $i \in$ progabide group, the likelihood is proportional to

$$\frac{e^{y_{i+}\beta_1} e^{(\beta_2 + \beta_3)(y_{i+} - y_{i0})}}{e^{y_{i+}\beta_1} (1 + e^{\beta_2 + \beta_3})^{y_{i+}}} = \pi_2^{y_{i0}} (1 - \pi_2)^{y_{i+} - y_{i0}}$$

where

$$\pi_2 = \frac{1}{1 + e^{\beta_2 + \beta_3}}$$

- In summary, the conditional likelihood is proportional to

$$\prod_{i=1}^{28} \pi_1^{y_{i0}} (1 - \pi_1)^{y_{i+} - y_{i0}} \prod_{i=29}^{59} \pi_2^{y_{i0}} (1 - \pi_2)^{y_{i+} - y_{i0}}$$

- π_1 and π_2 are the probabilities that an individual's seizure occurs before rather than after the randomization, for placebo and progabide groups, respectively.
- If progabide is helpful in reducing seizures we would observe $\pi_1 < \pi_2$, or equivalently, $1 + e^{\beta_2} > 1 + e^{\beta_2 + \beta_3} \Leftrightarrow \beta_3 < 0$.
- Results (patient 207 deleted)

	GEE	Cond. Like.
β_2	0.11 (0.12)	0.11 (0.047)
β_3	-0.30 (0.17)	-0.30 (0.07)

Notes:

1. Conditional likelihood inference leads to conclusion that progabide's effect is highly significant.
2. But the fitted model is grossly inadequate on a Pearson χ^2 statistic (see DHLZ 9.4.1).
3. Because of (2) the estimated s.e.'s in conditional likelihood approach may be too small in this example.
4. The homogeneity assumption that everyone's response to the treatment is the same is inadequate, that is, a random slope is needed.

Maximum Likelihood for GLMM

We need specify a distribution for \mathbf{b}_i to use maximum likelihood estimation. For *generalized linear mixed models*, we will assume

$$\mathbf{b}_i \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, D(\boldsymbol{\alpha})),$$

where $\boldsymbol{\alpha}$ is the variance parameter for the random effects.

The joint likelihood for $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is the marginal distribution of \mathbf{Y} ,

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{i=1}^m \int \prod_{j=1}^{n_i} f(y_{ij} | \boldsymbol{\beta}, \mathbf{b}_i) \Phi(\mathbf{b}_i | \boldsymbol{\alpha}) d\mathbf{b}_i.$$

where Φ is the multivariate normal density function. Note that there are q levels of integral where q is the dimension of \mathbf{b}_i .

For f in exponential family, we have

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = (2\pi)^{-q/2} |D|^{-1/2} \exp\{\mathbf{1}^T c(\mathbf{y})\} J(\boldsymbol{\beta}, \boldsymbol{\alpha})$$

where

$$J(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \int_{R^q} \exp\{\mathbf{y}^T (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) - \mathbf{1}^T b(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) - \frac{1}{2} \mathbf{b}^T D^{-1} \mathbf{b}\} d\mathbf{b}.$$

(Some messing up of the notation, here b refers to the $b(\theta)$ part of exponential family density.)

In general, the marginal likelihood has no closed form and the integration is quite difficult.

Estimation Methods for GLMM

- Generalized estimating equations (GEE)
- Numerical evaluation of the likelihood
 - Gaussian-Hermite quadrature
 - Adaptive quadrature (SAS PROC NLMIXED).
 - Importance sampling and Monte Carlo integration.
- Numerical maximization of the likelihood (without evaluating it)
 - Expectation-maximization (EM) algorithm
 - Monte Carlo EM.
 - Monte Carlo Newton-Raphson.
- Approximate likelihood
 - Penalized quasi-likelihood (PQL), Bias-corrected PQL.
 - Linearization.
- Bayesian Markov chain Monte Carlo (Zeger and Karim, 1991; Clayton, 1996)
 - Gibbs sampling
 - Metropolis-Hastings algorithm

Fitting GLMM: Software

- SAS:
 - PROC NL MIXED uses Gauss-Hermite or Adaptive Quadrature methods.
 - PROC GLIMMIX uses linearization idea to fit penalized/marginal quasi-likelihood models with Gaussian random effects.
- R
 - glmmPQL (MASS): penalized quasi-likelihood, allows the use of an additional correlation structure.
 - GLMM (lme4 a newer reimplementation of nlme): similar to (same as?) glmmPQL.
 - glmmML (glmmML): maximum likelihood using Gauss-Hermite quadrature, only can fit models with a random intercept. (work in progress)
 - glmm (repeated): Gauss-Hermite quadrature, models with random intercept.
 - gnlmix (repeated): non-linear regression with mixed random effects for the location parameters. Non-Gaussian mixing distributions are allowed.
 - glmm (GLMMGibbs): Gibbs sampling.
- BUGS (OpenBUGS/WinBUGS)

2 × 2 Crossover Trial

```
> library(MASS)
> xover.glmmPQL <- glmmPQL(fixed=y ~ trt, random = ~1 | id,
+                           family = "binomial", data = xover)
> summary(xover.glmmPQL)
```

Linear mixed-effects model fit by maximum likelihood

Data: xover

AIC BIC logLik

NA NA NA

Random effects:

Formula: ~1 | id

(Intercept) Residual

StdDev: 3.423162 0.5171149

Variance function:

Structure: fixed weights

Formula: ~invwt

Fixed effects: y ~ trt

	Value	Std.Error	DF	t-value	p-value
(Intercept)	2.626981	0.5395541	66	4.868799	0
trt	-1.888602	0.3940556	66	-4.792730	0

Correlation:

(Intr)

trt -0.449

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-3.2435553	-0.2443096	0.1640708	0.4219378	2.9639048

Number of Observations: 134 Number of Groups: 67

$\hat{\beta}(s.e.)$	GEE	CLR	GLMM
Intercept	1.1 (0.28)		2.63 (0.54)
Treatment	0.56 (0.23)	1.61 (0.78)	1.89 (0.39)
ρ	0.62 (0.1)		

2 × 3 Crossover Trial

```
> library (MASS)
> xover.glmm <- glmmPQL (fixed = relief ~ tx2 + tx3 + p2 + p3
>                        + ptx2 + ptx3,
+                        random = ~ 1 | id, family = "binomial",
+                        data = xover3)
```

```
iteration 1
```

```
iteration 2
```

```
> summary (xover.glmm)
```

Linear mixed-effects model fit by maximum likelihood

Data: xover3

	AIC	BIC	logLik
	1194.117	1226.094	-588.0585

Random effects:

x Formula: ~1 | id

(Intercept) Residual

StdDev: 0.04334977 1.002844

Variance function:

Structure: fixed weights

Formula: ~invwt

Fixed effects: relief ~ tx2 + tx3 + p2 + p3 + ptx2 + ptx3

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-1.0865137	0.3337615	166	-3.255359	0.0014
tx2	2.1056712	0.4084809	166	5.154882	0.0000
tx3	2.0683181	0.3895146	166	5.309989	0.0000
p2	0.4123220	0.4685381	166	0.880018	0.3801
p3	0.5865394	0.4830292	166	1.214294	0.2264
ptx2	-0.1282313	0.5101216	166	-0.251374	0.8018
ptx3	-0.9285876	0.4954788	166	-1.874122	0.0627

Correlation:

	(Intr)	tx2	tx3	p2	p3	ptx2
tx2	-0.539					
tx3	-0.599	0.472				
p2	-0.232	-0.232	-0.198			
p3	-0.215	-0.237	-0.205	0.699		
ptx2	-0.259	0.425	0.269	-0.574	-0.593	
ptx3	-0.174	0.059	0.347	-0.532	-0.550	0.511

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-2.2261231	-0.5793593	0.4702024	0.5989729	2.2217509

Number of Observations: 258

Number of Groups: 86

```
> xover.glmmML <- glmmML (relief ~ tx2 + tx3 + p2 + p3 +
+                          ptx2 + ptx3,
+                          cluster = xover3$id, family = binomial,
+                          data = xover3,
+                          n.points = 20)
> xover.glmmML
Call:  glmmML(formula = relief ~ tx2 + tx3 + p2 + p3 + ptx2 + ptx3,
              data = xover3, cluster = xover3$id, family = binomial,
              n.points = 20)
```

	coef	se(coef)	z	Pr(> z)
(Intercept)	-1.0865	0.3282	-3.3101	9.33e-04
tx2	2.1056	0.4017	5.2411	1.60e-07
tx3	2.0683	0.3831	5.3990	6.70e-08
p2	0.4124	0.4608	0.8949	3.71e-01
p3	0.5866	0.4751	1.2348	2.17e-01
ptx2	-0.1284	0.5017	-0.2560	7.98e-01
ptx3	-0.9286	0.4873	-1.9056	5.67e-02

Standard deviation in mixing distribution: 3.241e-05

Std. Error: 0.2163

Residual deviance: 283.6 on 250 degrees of freedom AIC: 299.6

- Note the difference in AIC, even though presumably they are maximizing the same likelihood.

Indonesian Children Health Study

```
> ICHS <- read.table("../data/ICHS.dat", header = TRUE)
> ichs.glmm <- glmmPQL (RESPONSE ~ VITA + AGE + I(AGE^2) +
+                       GENDER + TIME,
+                       random = ~ TIME | ID,
+                       data = ICHS, family = "binomial")
> summary (ichs.glmm)
```

Linear mixed-effects model fit by maximum likelihood

Data: ICHS

	AIC	BIC	logLik
	7918.793	7971.926	-3949.397

Random effects:

Formula: ~TIME | ID

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	2.8959344	(Intr)
TIME	0.1635745	-0.482
Residual	0.6537811	

Variance function:

Structure: fixed weights

Formula: ~invwt

Fixed effects: RESPONSE ~ VITA + AGE + I(AGE^2) + GENDER + TIME

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-2.0547139	0.7325637	1249	-2.804826	0.0051
VITA	0.4156394	0.3749237	245	1.108597	0.2687
AGE	0.6095973	0.4359468	245	1.398330	0.1633
I(AGE^2)	-0.0974293	0.0564336	245	-1.726441	0.0855
GENDER	-0.9629119	0.3613544	245	-2.664730	0.0082
TIME	0.0396433	0.0169233	1249	2.342525	0.0193

Correlation:

	(Intr)	VITA	AGE	I(AGE^2)	GENDER	TIME
VITA	-0.085					
AGE	-0.850	-0.130				
I(AGE^2)	0.770	0.128	-0.976			
GENDER	-0.189	0.028	-0.073	0.066		
TIME	-0.182	0.005	-0.009	0.008	-0.003	

Standardized Within-Group Residuals:

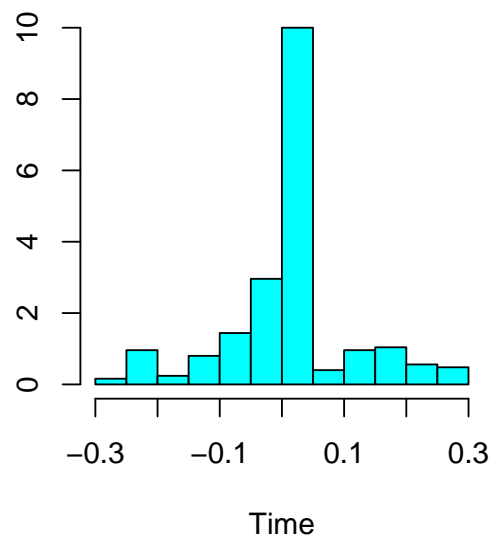
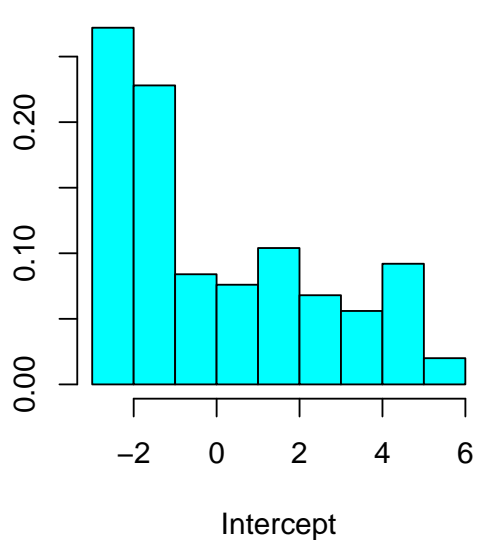
Min	Q1	Med	Q3	Max
-3.0848406	-0.3567323	-0.2277855	0.3358386	3.4221396

Number of Observations: 1500

Number of Groups: 250

```
> truehist (ranef (ichs.glmm)[,1], xlab = "Intercept")
```

```
> truehist (ranef (ichs.glmm)[,2], xlab = "Time")
```



- Empirical Bayes estimates of the random effects are natural by-product of PQL.

Software Differences

For GLMM, difference algorithms and software may give quite different answers.

```
> summary (glmmPQL (RESPONSE ~ VITA + AGE + I(AGE^2) +
+                 GENDER + TIME,
+                 random = ~ 1 | ID,
+                 data = ICHS, family = "binomial"))
```

Linear mixed-effects model fit by maximum likelihood

Data: ICHS

	AIC	BIC	logLik
	7709.842	7752.348	-3846.921

Random effects:

Formula: ~1 | ID

(Intercept) Residual

StdDev: 2.330407 0.729902

Variance function:

Structure: fixed weights

Formula: ~invwt

Fixed effects: RESPONSE ~ VITA + AGE + I(AGE^2) + GENDER + TIME

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-1.8854716	0.6745516	1249	-2.795148	0.0053
VITA	0.3995431	0.3479949	245	1.148129	0.2520
AGE	0.5629374	0.4046444	245	1.391190	0.1654
I(AGE^2)	-0.0890962	0.0523895	245	-1.700650	0.0903
GENDER	-0.8961807	0.3353563	245	-2.672324	0.0080
TIME	0.0327428	0.0114319	1249	2.864150	0.0043

Correlation:

	(Intr)	VITA	AGE	I(AGE^	GENDER
VITA		-0.084			
AGE		-0.859	-0.131		
I(AGE^2)		0.778	0.128	-0.976	
GENDER		-0.191	0.027	-0.071	0.064
TIME		-0.132	0.002	0.003	-0.006

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-----	----	-----	----	-----

-3.1088570 -0.5485138 -0.2382666 0.3486534 3.6635217

Number of Observations: 1500

Number of Groups: 250

```
> glmmML (RESPONSE ~ VITA + AGE + I(AGE^2) + GENDER + TIME,
+         cluster = ICHS$ID,
+         data = ICHS, family = binomial,
+         n.points = 20)
```

```
Call: glmmML(formula = RESPONSE ~ VITA + AGE + I(AGE^2) + GENDER +
TIME, family = binomial, data = ICHS, cluster = ICHS$ID, n.points =
20)
```

	coef	se(coef)	z	Pr(> z)
(Intercept)	-2.37305	0.84701	-2.802	0.00508
VITA	0.52672	0.43070	1.223	0.22100
AGE	0.75228	0.50129	1.501	0.13300
I(AGE^2)	-0.11906	0.06480	-1.837	0.06610
GENDER	-1.19115	0.41700	-2.856	0.00428
TIME	0.03369	0.01588	2.122	0.03380

Standard deviation in mixing distribution: 2.713

Std. Error: 0.1861

Residual deviance: 1341 on 1493 degrees of freedom AIC: 1355

```

> library (repeated)
> ichs.glmm2 <- glmm (RESPONSE ~ VITA + AGE + I(AGE^2) +
+                   GENDER + TIME, nest = ID,
+                   data = ICHS, family = "binomial",
+                   points = 20)
There were 17 warnings (use warnings() to see them)
> summary (ichs.glmm2)
Call:
glmm(RESPONSE ~ VITA + AGE + I(AGE^2) + GENDER + TIME, nest = ID,
     data = ICHS, family = "binomial", points = 20)

Deviance Residuals:
      Min       1Q   Median       3Q      Max
-1.732e+00 -1.233e-04 -4.726e-11  1.846e-13  1.842e+00

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.32668    0.37305  -6.237 4.46e-10 ***
VITA          0.60359    0.17127   3.524 0.000425 ***
AGE           0.79657    0.20543   3.878 0.000105 ***
I(AGE^2)     -0.12761    0.02686  -4.751 2.03e-06 ***
GENDER       -1.21289    0.16934  -7.163 7.91e-13 ***
TIME          0.03423    0.01594   2.147 0.031802 *
sd            2.84078    0.15638  18.166 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1822.3 on 1499 degrees of freedom
Residual deviance: 1328.7 on 1493 degrees of freedom
AIC: 1342.7

Number of Fisher Scoring iterations: 17
Normal mixing variance: 8.070035

```

	Random Intercept Model		
	glmmPQL	glmmML	glmm
Intercept	-1.88 (0.76)	-2.37 (0.85)	-2.33 (0.37)
vita	0.40 (0.35)	0.53 (0.43)	0.60 (0.17)
age	0.56 (0.40)	0.75 (0.50)	0.80 (0.21)
age ²	-0.089 (0.052)	-0.12 (0.06)	-0.13 (0.027)
gender	-0.90 (0.34)	-1.19 (0.42)	-1.21 (0.17)
time	0.033 (0.011)	0.034 (0.016)	0.034 (0.016)
τ	2.33	2.86	2.84 (0.16)

Further Reading

- Chapter 9 of DHLZ.
- Chapter 13 of Molenberghs and Verbeke (2005).

References

- Agresti A (2002) Categorical data analysis. John Wiley & Sons, 2nd edn.
- Breslow NE and Clayton DG (1993) Approximate inference in generalized linear mixed models. *Journal of American Statistician Association* 88:9-25.
- Breslow NE and Lin X (1995) Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* 82:81-91.
- Clayton DG (1996) Generalized linear mixed models. In *Markov Chain Monte Carlo in Practice* (edited by WR Gilks, S Richardson, and DJ Spiegelhalter), Chapman & Hall, London, pp. 259-273.
- Gelfand AE and Carlin BP (1993) Maximum likelihood estimation for constrained- or missing-data problems. *Canadian Journal of Statistics* 21:303-311.
- Geyer CJ and Thompson EA (1992) Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B* 54:657-699.
- Liang KY and McCullagh P (1993) Case studies in binary dispersion. *Biometrics* 49:623-630.
- Lin X and Breslow NE (1996) Bias correction in generalized linear mixed models with multiple component of dispersion. *Journal of American Statistician Association* 91:1007-1016.
- McCulloch CE (1997) Maximum likelihood algorithms for generalized linear mixed models. *Journal of American Statistician Association* 92:162-170.
- (2003) Generalized linear mixed models, vol. 7 of NSF-CBMS Regional Conference Series in Probability and Statistics. Institute of Mathematical Statistics.
- Zeger SL and Karim MR (1991) Generalized linear models with random effects: a Gibbs sampling approach. *Journal of American Statistician Association* 86:79-86.
- Zeger SL, Liang KY, and Albert PS (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44:1049-1060.