Likelihood Inference

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

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

Unlike the LMEM 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).

Likelihood Inference fo loglinear GLMM

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 first approach. For simplicity we assume the canonical link function,

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

to give 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 we have sufficient statistics $\boldsymbol{t}_1 = \sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}\boldsymbol{x}_{ij} = \sum_{i=1}^{m} \boldsymbol{t}_{1i}$ for $\boldsymbol{\beta}$ and $\boldsymbol{t}_{2i} = \sum_{j=1}^{n_i} y_{ij}\boldsymbol{z}_{ij}$ for $\boldsymbol{b}_i$.

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

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

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

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

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 $\boldsymbol{t}_{1i}$.

Example: Matched Binary Pairs Data

Conditional Likelihood: Binary Longitudinal Data

Consider individual i with binary observations $y_{i1}, ..., y_{in_i}$ and assume the model $Y_{ij} \mid \gamma_i, \boldsymbol{\beta} \sim \text{Bernoulli}(p_{ij})$, where

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \gamma_i + \boldsymbol{x}_{ij}\boldsymbol{\beta}$$

with $\gamma_i = \beta_0 + b_i$ and $\boldsymbol{x}_{ij}\boldsymbol{\beta} = x_{ij1}\beta_1 + ... + x_{ijp}\beta_p$ (a slight change from our usual notation).

We have

$$\begin{aligned}
\Pr(y_{i1}, ..., y_{in_i} \mid \gamma_i, \boldsymbol{\beta}) &= \prod_{j=1}^{n_i} \frac{\exp(\gamma_i y_{ij} + \boldsymbol{x}_{ij}\boldsymbol{\beta} y_{ij})}{1 + \exp(\gamma_i + \boldsymbol{x}_{ij}\boldsymbol{\beta})} \\
&= \frac{\exp\left(\gamma_i \sum_{j=1}^{n_i} y_{ij} + \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} y_{ij} \boldsymbol{\beta}\right)}{\prod_{j=1}^{n_i}\left[1 + \exp(\gamma_i + \boldsymbol{x}_{ij}\boldsymbol{\beta})\right]} \\
&= \frac{\exp(\gamma_i t_{2i} + \boldsymbol{t}_{1i}\boldsymbol{\beta})}{\prod_{j=1}^{n_i}\left[1 + \exp(\gamma_i + \boldsymbol{x}_{ij}\boldsymbol{\beta})\right]} \\
&= \frac{\exp(\gamma_i t_{2i} + \boldsymbol{t}_{1i}\boldsymbol{\beta})}{k(\gamma_i, \boldsymbol{\beta})} \\
&= p(t_{1i}, \boldsymbol{t}_{2i} \mid \gamma_i, \boldsymbol{\beta})
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{t}_{1i} &= \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} y_{ij}, \quad t_{2i} = \sum_{j=1}^{n_i} y_{ij} \\
k(\gamma_i, \boldsymbol{\beta}) &= \prod_{j=1}^{n_i}\left[1 + \exp(\gamma_i + \boldsymbol{x}_{ij}\boldsymbol{\beta})\right].
\end{aligned}$$

We have

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

where

$$p(\boldsymbol{t}_{2i} \mid \gamma_i, \boldsymbol{\beta}) = \frac{\sum_{l=1}^{\binom{n_i}{y_{i+}}} \exp\left(\gamma_i \sum_{j=1}^{n_i} y_{ij} + \sum_{k=1}^{n_i} \boldsymbol{x}_{ik} y_{ik}^{l} \boldsymbol{\beta}\right)}{k(\gamma_i, \boldsymbol{\beta})},$$

where the summation is over the $\binom{n_i}{y_{i+}}$ ways of choosing y_{i+} ones out of n_i, and $\boldsymbol{y}_i^l = (y_{i1}^l, ..., y_{in_i}^l)$, $l = 1, ..., \binom{n_i}{y_{i+}}$ is the collection of these ways.

Hence

$$\begin{aligned} L_c(\boldsymbol{\beta}) &= \prod_{i=1}^{m} \frac{\exp\left(\gamma_i \sum_{j=1}^{n_i} y_{ij} + \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} y_{ij} \boldsymbol{\beta}\right)}{\sum_{l=1}^{\binom{n_i}{y_{i+}}} \exp\left(\gamma_i \sum_{j=1}^{n_i} y_{ij} + \sum_{k=1}^{n_i} \boldsymbol{x}_{ik} y_{ik}^{l} \boldsymbol{\beta}\right)} \\ &= \prod_{i=1}^{m} \frac{\exp\left(\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} y_{ij} \boldsymbol{\beta}\right)}{\sum_{l=1}^{\binom{n_i}{y_{i+}}} \exp\left(\sum_{k=1}^{n_i} \boldsymbol{x}_{ik} y_{ik}^{l} \boldsymbol{\beta}\right)} \end{aligned}$$

Notes

- Can be computationally expensive to evaluate likelihood if n_i is large, e.g. if $n_i = 20$ and $y_{i+} = 10$, $\binom{n_i}{y_{i+}} = 184,756$.
- There is no contribution to the conditional likelihood from individuals:
  - With $n_i = 1$.
  - With $y_{i+} = 0$ or $y_{i+} = n_i$.
  - For those covariates with $x_{i1} = ... = x_{in_i} = x_i$. The conditional likelihood estimates β's that are associated with within-individual covariates. If a covariate only varies between individuals, then it cannot be estimated using conditional likelihood. For covariates that vary both between and within individuals, only the within-individual contrasts are used.
- The similarity to Cox's partial likelihood may be exploited to carry out computation.
- We have not made a distribution assumption for the γ_i's!

Examples:

If $n_i = 3$ and $\boldsymbol{y}_i = (0, 0, 1)$ so that $y_{i+} = 1$ then

$$\boldsymbol{y}_i^1 = (1, 0, 0), \quad \boldsymbol{y}_i^2 = (0, 1, 0), \quad \boldsymbol{y}_i^3 = (0, 0, 1),$$

and the contribution to the conditional likelihood is

$$\frac{\exp(\boldsymbol{x}_{i3}\boldsymbol{\beta})}{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta}) + \exp(\boldsymbol{x}_{i2}\boldsymbol{\beta}) + \exp(\boldsymbol{x}_{i3}\boldsymbol{\beta})}.$$

If $n_i = 3$ and $\boldsymbol{y}_i = (1, 0, 1)$ so that $y_{i+} = 2$ then

$$\boldsymbol{y}_i^1 = (1, 1, 0), \quad \boldsymbol{y}_i^2 = (1, 0, 1), \quad \boldsymbol{y}_i^3 = (0, 1, 1),$$

and the contribution to the conditional likelihood is

$$\frac{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta} + \boldsymbol{x}_{i3}\boldsymbol{\beta})}{\exp(\boldsymbol{x}_{i1}\boldsymbol{\beta} + \boldsymbol{x}_{i2}\boldsymbol{\beta}) + \exp(\boldsymbol{x}_{i1}\boldsymbol{\beta} + \boldsymbol{x}_{i3}\boldsymbol{\beta}) + \exp(\boldsymbol{x}_{i2}\boldsymbol{\beta} + \boldsymbol{x}_{i3}\boldsymbol{\beta})}.$$

Penalized Quasi-Likelihood

Breslow and Clayton (1993) introduced the method of Penalized Quasi-Likelihood (PQL) which was an attempt to extend quasi-likelihood to GLMMs. One justification of the method is a Laplace approximation.

If we write the required integration in the form

$$\log L(\boldsymbol{\beta}, \boldsymbol{\alpha}) = -\frac{1}{2} \log |\boldsymbol{D}| + \log \left(\int \exp\{-\alpha(\boldsymbol{b})\} \mathrm{d}\boldsymbol{b} \right),$$

and use a Laplace approximation to the second term to obtain

$$\log L(\boldsymbol{\beta}, \boldsymbol{\alpha}) \approx -\frac{1}{2} \log |\boldsymbol{D}| - \frac{1}{2} \log |\kappa''(\hat{\boldsymbol{b}})| - \kappa(\hat{\boldsymbol{b}}),$$

where $\hat{\boldsymbol{b}}$ maximizes $\alpha(\boldsymbol{b})$, and so satisfies $\kappa'(\hat{\boldsymbol{b}}) = 0$.

PQL is very poor for binary data but may be OK for binomial and Poisson data (as long as the counts are not too small).

Within the `MASS` package the `glmmPQL` function obtains PQL estimates, while within the `lme4` package the `GLMM` function obtains PQL estimates, or second-order approximations.