GLMM: Estimation of \( \alpha \)

Recall our GLMM looks like

\[
Y_i = X_i \alpha + Z_i \beta_i + \delta_i
\]

\( \beta_i \sim N(0, D) \)

\( \delta_i \sim N(0, R_i) \) (usually \( R_i = \sigma^2 I \))

Therefore:

\[
E[Y_i] =
\text{Var}[Y_i] =
\]

\( Y_i \sim \)

Note that our variances are an explicit function of what’s in \( Z_i \) (e.g., longitudinal times \( t_1, \cdots, t_J \)).

**ESTIMATION of \( \Sigma \):**

\( \Sigma_i \) now is formulated to have two unknown pieces in it, \( D \) and \( R_i \). These are estimated as we have seen before using numerical algorithms to carry out REML estimation.

**“ESTIMATION” of \( \beta_i \):**

The \( \beta_i \) are not parameters, like \( \alpha \) is; they are random effects. Thus the usual terminology is that we predict the \( \beta_i \), not that we estimate them.

How do we do that?

With conditional means.

Suppose we have normality for the \( Y_i \), then we can again write down a big multivariate normal likelihood for each cluster:

\[
\left( \frac{1}{\sqrt{2\pi}} \right)^J |\Sigma_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(Y_i-X_i \alpha)\Sigma_i^{-1}(Y_i-X_i \alpha)}
\]

and our maximum likelihood estimate of \( \alpha \) is

\[
\hat{\alpha} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y.
\]

\( \hat{\alpha} \) is still just a big complicated linear function of the \( Y \)-values, so

\[
\hat{\alpha} \sim N(\alpha, (X'\hat{\Sigma}^{-1}X)^{-1})
\]

where we must again estimate the variance

\[
\text{Var}[\hat{\alpha}] = (X'\hat{\Sigma}^{-1}X)^{-1}.
\]
Mixed models are different from what we have seen before because we can compute marginal means (or variances) and conditional means (or variances):

- **MARGINAL** (population-average)
  \[ E[Y_i] = \text{Var}[Y_i] = \]

- **CONDITIONAL** (cluster-specific)
  \[ E[Y_i | \beta_i] = \text{Var}[Y_i | \beta_i] = \]

Thus we have:

- **MARGINAL LIKELIHOOD** for \( Y_i \):
  \[
  \left( \frac{1}{\sqrt{2\pi}} \right)^J |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(Y_i - X_i \alpha)^T \Sigma^{-1} (Y_i - X_i \alpha)}
  \]

- **CONDITIONAL LIKELIHOOD** for \( Y_i | \beta_i \):
  \[
  \left( \frac{1}{\sqrt{2\pi}} \right)^J |R|^{-\frac{1}{2}} e^{-\frac{1}{2}(Y_i - X_i \alpha - Z_i \beta_i)^T R^{-1} (Y_i - X_i \alpha - Z_i \beta_i)}
  \]

- **RANDOM EFFECTS LIKELIHOOD** for \( \beta_i \) (presuming normally distributed random effects):
  \[
  \left( \frac{1}{\sqrt{2\pi}} \right)^q |D|^{-\frac{1}{2}} e^{-\frac{1}{2}(\beta_i - 0)^T D^{-1} (\beta_i - 0)}
  \]

Using some properties of the multivariate normal distribution, we can write down a joint likelihood for \( \beta_i \) and \( Y_i | \beta_i \):

\[
\left( \frac{1}{\sqrt{2\pi}} \right)^J \left| \begin{bmatrix} D & 0 \\ 0 & R_i \end{bmatrix} \right|^{-\frac{1}{2}} e^{-\frac{1}{2} \left[ \begin{bmatrix} Y_j - X_{ij} \alpha \\ Y_i - (X_i \alpha + Z_i \beta_i) \end{bmatrix} \right]^T \left[ \begin{bmatrix} D & 0 \\ 0 & R_i \end{bmatrix} \right]^{-1} \left[ \begin{bmatrix} Y_j - X_{ij} \alpha \\ Y_i - (X_i \alpha + Z_i \beta_i) \end{bmatrix} \right]}
\]

We can maximize this joint likelihood to both estimate \( \alpha \) and predict \( \beta_i \) simultaneously.

**Aside:** Useful multivariate normal relations:

Suppose we have

\[
\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)
\]

then

\[
E[Z_2 | Z_1] = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (Z_1 - \mu_1)
\]

\[
\text{Var}[Z_2 | Z_1] = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}
\]

Also, the lower right hand block of \( \Sigma^{-1} \) is

\[
\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}
\]

which does not equal \( \Sigma_{22}^{-1} \) unless \( \Sigma_{21} = \Sigma_{12} = 0 \).
Regarding estimation, maximizing that likelihood turns out to be equivalent to solving this set of matrix equations for $\alpha$ and $\beta$ together:

\[
\begin{bmatrix}
X'R^{-1}X & X'R^{-1}Z \\
Z'R^{-1}X & Z'R^{-1}Z + D^{-1}
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix} = \begin{bmatrix}
X'R^{-1}Y \\
Z'R^{-1}Y
\end{bmatrix}.
\]

These are Henderson’s mixed model equations and from them we can simultaneously estimate:

\[
\hat{\alpha}(\Sigma) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y
\]
\[
\hat{\beta}(\Sigma) = DZ'\Sigma^{-1}(Y - X\hat{\alpha}).
\]

$\hat{\alpha}(\Sigma)$ is the Best Linear Unbiased Estimator (BLUE) and $\hat{\beta}(\Sigma)$ is the Best Linear Unbiased Predictor (BLUP) when $\Sigma$ is known. As usual, we plug in $\hat{\Sigma}$ based on $\hat{D}$ and $\hat{R}$.

\[
\hat{\Sigma}_i = Z_i\hat{D}Z_i' + \hat{R}_i
\]

Notes:
- See also Robinson (1991) “That BLUP is a good thing” in Statistical Science.
- This same equation for $\hat{\beta}$ can be derived from a Bayesian approach:

\[
\hat{\beta} = E[\beta | Y] = \text{posterior mean of } \beta
\]
\[
\text{Var}[\hat\beta] = \hat{D}Z'\left[\Sigma^{-1} - \Sigma^{-1}X(X'\Sigma X)^{-1}X'\Sigma^{-1}\right]Z\hat{D}
\]
\[
\hat{\beta} \sim N
\]

RANDOM EFFECTS —>$\Sigma$ CONNECTION

Suppose in the Orthodontic Study we needed random intercepts but not slopes:

\[
Y_{ij} = X_{ij}\alpha + \beta_i + \delta_{ij}
\]
where $i =$ child, $j =$ age

\[
\beta_i \sim N(0, D) \text{ indep.}
\]
\[
\delta_{ij} \sim N(0, \sigma^2)
\]

What do the marginal variance and covariance look like?

\[
\text{Var}[Y_{ij}] = \text{Var}[X_{ij}\alpha + \beta_i + \delta_{ij}]
\]
\[
= \text{Var}[X_{ij}\alpha] + \text{Var}[\beta_i] + \text{Var}[\delta_{ij}]
\]
\[
= \text{Var}[X_{ij}\alpha] + \text{Var}[\beta_i] + \text{Var}[\delta_{ij}]
\]

\[
\text{Cov}[Y_{ij}, Y_{i\ell}] =
\]

\[
\text{Corr}[Y_{ij}, Y_{i\ell}]
\]
There are natural relationships between GLM and GLMM models. With link function $g(\cdot)$ as identity link:

<table>
<thead>
<tr>
<th>GLM</th>
<th>GLMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{ij} \sim \text{Normal}$</td>
<td>$Y_{ij} \sim \text{Normal}$</td>
</tr>
<tr>
<td>$E[Y_{ij}] = X_{ij}\alpha$</td>
<td>$E[Y_{ij}] = X_{ij}\alpha$</td>
</tr>
<tr>
<td>$E[Y_{ij}</td>
<td>\beta_i] = X_{ij}\alpha + Z_{ij}\beta_i$</td>
</tr>
<tr>
<td>$\text{Var}[Y_i] = \Sigma_i$</td>
<td>$\text{Var}[Y_i] = \Sigma_i = Z_iDZ_i' + R_i$</td>
</tr>
<tr>
<td>$\text{Var}[Y_i</td>
<td>\beta_i] = R_i$</td>
</tr>
<tr>
<td>$\beta_i \sim \mathcal{N}(0, D)$</td>
<td>$\beta_i \sim \mathcal{N}(0, D)$</td>
</tr>
</tbody>
</table>

Note that $E[Y_{ij}] = X_{ij}\alpha$ in both models. Unconditional (marginal) mean

We will also look at allowing $g(\cdot)$ to be something else. Then things change a little.