Generalized Linear Mixed Models

Model

Definition 1.1.1. Let Y_{ki} denote the ith response in the kth cluster. Suppose that, conditional on the covariates $(\boldsymbol{x}_{ki}, \boldsymbol{z}_{ki})$ and the **random effect** $\boldsymbol{\gamma}_k$, Y_{ki} has a distribution in the exponential dispersion family. A **generalized linear mixed model** (GLMM) takes the following form:

$$g(\mu_{ki}) = \eta_{ki} \equiv \boldsymbol{x}'_{ki}\boldsymbol{\beta} + \boldsymbol{z}'_{ki}\gamma_k. \tag{1.1.1}$$

Here g is the link function, $\mu_{ki} = E(Y_{ki}|\boldsymbol{\gamma}_k, \boldsymbol{x}_{ki}, \boldsymbol{z}_{ki})$ is the conditional expectation of Y_{ki}, η_{ki} is the linear predictor, and $\boldsymbol{\beta}$ is described as the **fixed effect**. The random effects are taken as IID across clusters, with mean $E(\boldsymbol{\gamma}_k) = \boldsymbol{0}$, and variance $\mathrm{Var}(\boldsymbol{\gamma}_k) = \boldsymbol{G}(\boldsymbol{\alpha})$. Parameters $\boldsymbol{\alpha}$ for the random effect model are described as **variance components**. If a dispersion parameter ϕ is required, absorb this into $\boldsymbol{\alpha}$.

Define $\boldsymbol{y}_k = \mathrm{vec}(Y_{k1}, \cdots, Y_{kn_k})$ as the response vector for cluster k. Let $\mathcal{D}_k$ collect the covariates relevant to cluster k:

$$\mathcal{D}_k = \bigcup_{i=1}^{n_k} \{\boldsymbol{x}_{ki}, \boldsymbol{z}_{ki}\}.$$

Observations belonging to distinct clusters $\boldsymbol{y}_k$ and $\boldsymbol{y}_l$ are independent given covariates:

$$\big(\boldsymbol{y}_k \perp \boldsymbol{y}_l\big)\big|\big(\mathcal{D}_k, \mathcal{D}_l\big).$$

Within a given cluster k, observations Y_{ki} and Y_{kj} are conditionally independent given covariates and the cluster's random effect:

$$\big(Y_{ki} \perp Y_{kj}\big)\big|\big(\boldsymbol{\gamma}_k, \boldsymbol{x}_{ki}, \boldsymbol{x}_{kj}\big).$$

■

Remark 1.1.1. In contrast to LMMs, GLMMs generally assume that observations within a cluster are conditionally independent given the cluster's random effect. ♦

Remark 1.1.2. The components of a GLMM are summarized here: ♦

Structure	Dimension	Description
$\boldsymbol{\beta}$	$p \times 1$	Fixed effect.
$\boldsymbol{\gamma}_k$	$q \times 1$	Random effect.
η_{ki}	1×1	Linear predictor.
$\boldsymbol{\alpha}$	$r \times 1$	Variance components.
$\boldsymbol{G}(\boldsymbol{\alpha})$	$q \times q$	Random effect covariance.

Objective Function

2.1 Likelihood

Result 1.2.1 (Model Likelihood). Suppose there are K clusters of size n_k. Let $\mathcal{D}_k$ denote the covariates relevant to the kth cluster. The GLMM likelihood is:

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{k=1}^{K} \int \left\{ \prod_{i=1}^{n_k} f(y_{ki} | \mathcal{D}_{ki}, \boldsymbol{\gamma}_k; \boldsymbol{\beta}, \boldsymbol{\alpha}) \right\} f(\boldsymbol{\gamma}_k | \boldsymbol{\alpha}) d\boldsymbol{\gamma}_k. \tag{1.2.2}$$

♣

Proof. Note that the random effects $\{\boldsymbol{\gamma}_k\}$ are not among of the observe data. The likelihood for the observed data factors as:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \equiv f(\boldsymbol{y}|\mathcal{D}; \boldsymbol{\beta}, \boldsymbol{\alpha}) &= \prod_{k=1}^{K} f(\boldsymbol{y}_k | \mathcal{D}_k; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{k=1}^{K} \int f(\boldsymbol{y}_k, \boldsymbol{\gamma}_k | \mathcal{D}_k; \boldsymbol{\beta}, \boldsymbol{\alpha}) d\boldsymbol{\gamma}_k \\
&= \prod_{k=1}^{K} \int f(\boldsymbol{y}_k | \mathcal{D}_k, \boldsymbol{\gamma}_k; \boldsymbol{\beta}, \boldsymbol{\alpha}) f(\boldsymbol{\gamma}_k | \boldsymbol{\alpha}) d\boldsymbol{\gamma}_k.
\end{aligned}$$

Since observations within a cluster are conditionally independent given covariates $\mathcal{D}_k$ and the cluster's random effect $\boldsymbol{\gamma}_k$:

$$\prod_{k=1}^{K} \int f(\boldsymbol{y}_k | \mathcal{D}_k, \boldsymbol{\gamma}_k; \boldsymbol{\beta}, \boldsymbol{\alpha}) f(\boldsymbol{\gamma}_k | \boldsymbol{\alpha}) d\boldsymbol{\gamma}_k = \prod_{k=1}^{K} \int \left(\prod_{i=1}^{n_k} f(y_{ki} | \mathcal{D}_{ki}, \boldsymbol{\gamma}_k; \boldsymbol{\beta}, \boldsymbol{\alpha}) \right) f(\boldsymbol{\gamma}_k | \boldsymbol{\alpha}) d\boldsymbol{\gamma}_k.$$

■

2.2 Quasi-Likelihood

Remark 1.2.1. To perform estimation and inference in the GLMM framework, it suffices to specify the following components:

i. A mean model: $g(\mu_{ki}) = \eta_{ki}$.

ii. The mean-variance relationship: $\text{Var}(Y_{ki}) = \phi\nu(\mu_{ki})$.

The objective function induced by these two components is the *quasi (log) likelihood*. ♦

Definition 1.2.1. The **quasi likelihood** contribution of y_{ki} is defined as:

$$\ell_{ki} = \ell(y_{ki}|\mathcal{D}_{ki}, \boldsymbol{\gamma}_k; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \int_{Y_{ki}}^{\mu_{ki}} \frac{y_{ki} - u}{\phi\nu(u)} du. \tag{1.2.3}$$

See [2] for the definition of quasi-likelihood. ■

Example 1.2.1. From (1.2.2), the GLMM log likelihood is:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{k=1}^{K} \ln \int \exp\left\{\sum_{i=1}^{n_k} \ln f(y_{ki}|\mathcal{D}_{ki}, \boldsymbol{\gamma}_k; \boldsymbol{\beta}, \boldsymbol{\alpha})\right\} f(\boldsymbol{\gamma}_k|\boldsymbol{\alpha}) d\boldsymbol{\gamma}_k.$$

Suppose the random effects $\boldsymbol{\gamma}_k \sim N(\mathbf{0}, \boldsymbol{G})$, then:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}) \propto \sum_{k=1}^{K} \ln \int \exp\left\{\sum_{i=1}^{n_k} \ln f(y_{ki}|\mathcal{D}_{ki}, \boldsymbol{\gamma}_k; \boldsymbol{\beta}, \boldsymbol{\alpha})\right\} \det(\boldsymbol{G})^{-1/2} e^{-\gamma_k' \boldsymbol{G}^{-1}\gamma_k} d\boldsymbol{\gamma}_k.$$

Substituting the quasi likelihood ℓ_{ki} for $\ln f(y_{ki}|\mathcal{D}_{ki}, \boldsymbol{\gamma}_k; \boldsymbol{\beta}, \boldsymbol{\alpha})$ and reducing gives:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}) \propto -\frac{K}{2} \ln \det(\boldsymbol{G}) - \sum_{k=1}^{K} \ln \int \exp\left\{-\frac{1}{2}\gamma_k' \boldsymbol{G}^{-1}\gamma_k + \sum_{i=1}^{n_k} \ell_{ki}\right\} d\gamma_k \tag{1.2.4}$$

♠

2.3 Laplace Approximation

Remark 1.2.2. The integral appearing in (1.2.4) is typically intractable. In the *Laplace approximation* to the GLMM quasi likelihood, the integrand is approximated via a second order Taylor expansion, then evaluated analytically to obtain an approximate objective function. ♦

Proposition 1.2.1. Suppose $\boldsymbol{\gamma}_k \overset{\text{IID}}{\sim} N(\mathbf{0}, \boldsymbol{G})$. Define $s_k : \mathbb{R}^q \to \mathbb{R}$:

$$s_k(\boldsymbol{\gamma}) = \frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{G}^{-1}\boldsymbol{\gamma} - \sum_{i=1}^{n_k} \ell_{ki}(\boldsymbol{\gamma}).$$

Taking $\boldsymbol{\gamma}_k^* = \arg\max_{\boldsymbol{\gamma}} s_k(\boldsymbol{\gamma})$ as the expansion point, the Laplace approximation to the quasi likelihood is:

$$\ell_q(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}^*) \propto -\frac{K}{2}\ln\det(\boldsymbol{G}) - \sum_{k=1}^{K} s_k(\boldsymbol{\gamma}_k^*) - \frac{1}{2}\sum_{k=1}^{K}\ln\det\left\{\frac{\partial^2 s_k}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}(\boldsymbol{\gamma}_k^*)\right\}. \tag{1.2.5}$$

♦

Proof. Define $I_k \in \mathbb{R}$ as:

$$I_k = \int \exp\left\{-\frac{1}{2}\boldsymbol{\gamma}_k'\boldsymbol{G}^{-1}\boldsymbol{\gamma}_k + \sum_{i=1}^{n_k}\ell_{ki}(\boldsymbol{\gamma}_k)\right\} d\boldsymbol{\gamma}_k = \int e^{-s_k(\boldsymbol{\gamma})} d\boldsymbol{\gamma}. \tag{1.2.6}$$

Take the 2nd order Taylor expansion of $s_k(\boldsymbol{\gamma})$ about $\gamma_k^* = \arg\max s_k(\boldsymbol{\gamma})$:

$$s_k(\boldsymbol{\gamma}) = s_k(\boldsymbol{\gamma}_k^*) + \frac{\partial s_k}{\partial\boldsymbol{\gamma}}(\boldsymbol{\gamma}_k^*)\cdot(\boldsymbol{\gamma} - \boldsymbol{\gamma}_k^*) + \frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_k^*)\cdot\frac{\partial^2 s_k}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}(\boldsymbol{\gamma}_k^*)\cdot(\boldsymbol{\gamma} - \boldsymbol{\gamma}_k^*) + \mathcal{O}\left(||\boldsymbol{\gamma} - \boldsymbol{\gamma}_k^*||^3\right).$$

Since $\boldsymbol{\gamma}_k^*$ is the maximum of $s_k(\boldsymbol{\gamma})$, the gradient of s_k at $\boldsymbol{\gamma}_k^*$ vanishes:

$$\frac{\partial s_k}{\partial\boldsymbol{\gamma}}(\boldsymbol{\gamma}_k^*) = \mathbf{0}.$$

Substitute the Taylor expansion for $s_k(\boldsymbol{\gamma})$ into (1.2.6) to obtain:

$$\hat{I}_k = \int_{\boldsymbol{\gamma}} \exp\left\{-s_k(\boldsymbol{\gamma}_k^*) - \frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_k^*)\cdot\frac{\partial^2 s_k}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}(\boldsymbol{\gamma}_k^*)\cdot(\boldsymbol{\gamma} - \boldsymbol{\gamma}_k^*)\right\} d\boldsymbol{\gamma}. \tag{1.2.7}$$

To evaluate $\hat{I}_k$, note that for $\boldsymbol{z}_{q\times 1} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$:

$$\int_{\mathbb{R}^q} (2\pi)^{-q/2}\det(\boldsymbol{\Sigma})^{-1/2}\exp\big\{-\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}\big\}d\boldsymbol{z} = 1.$$

Rearranging gives the identities:

$$\int_{\mathbb{R}^q} \exp\big\{-\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}\big\}d\boldsymbol{z} = (2\pi)^{q/2}\det(\boldsymbol{\Sigma})^{1/2}.$$

Replacing $\boldsymbol{\Sigma}^{-1}$ by $\boldsymbol{\Sigma}$:

$$\int_{\mathbb{R}^q} \exp\big\{-\boldsymbol{z}'\boldsymbol{\Sigma}\boldsymbol{z}\big\}d\boldsymbol{z} = (2\pi)^{q/2}\det(\boldsymbol{\Sigma})^{-1/2}.$$

Therefore, (1.2.7) evaluates to:

$$\hat{I}_k(\boldsymbol{\gamma}_k^*) = e^{-s_k(\boldsymbol{\gamma}_k^*)}(2\pi)^{q/2}\det\left\{\frac{\partial^2 s_k}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}(\boldsymbol{\gamma}_k^*)\right\}^{-1/2}.$$

Taking the logarithm:

$$\ln \hat{I}_k(\boldsymbol{\gamma}_k^*) \propto -s_k(\boldsymbol{\gamma}_k^*) - \frac{1}{2}\ln\det\left\{\frac{\partial^2 s_k}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}(\boldsymbol{\gamma}_k^*)\right\}.$$

Finally, the approximate log integrated quasi likelihood is:

$$\ell(\boldsymbol{\beta},\boldsymbol{\alpha}) = -\frac{K}{2}\ln\det(\boldsymbol{G}) + \sum_{k=1}^{K}\ln\hat{I}_k(\boldsymbol{\gamma}_k^*).$$

■

Remark 1.2.3. To evaluate the Laplace likelihood (1.2.5), it remains to:

i. Provide an expression for the cluster-specific expansions points $\boldsymbol{\gamma}_k^*$.

ii. Provide the form of the Hessian of $s_k(\boldsymbol{\gamma})$.

♦

Proposition 1.2.2. Recall the following matrices defined in the context of GLMs:

$$\boldsymbol{\Delta}_k = \mathrm{diag}\big\{\dot{g}(\mu_{ki})\big\}, \qquad \boldsymbol{W}_k = \mathrm{diag}\left\{\frac{1}{\phi\nu(\mu_{ki})\dot{g}^2(\mu_{ki})}\right\}. \tag{1.2.8}$$

The expansion point $\boldsymbol{\gamma}_k^* = \arg\max_{\gamma} s_k(\boldsymbol{\gamma})$ is a solution to:

$$\boldsymbol{G}^{-1}\boldsymbol{\gamma} - \boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{\Delta}_k(\boldsymbol{y}_k - \boldsymbol{\mu}_k) = \boldsymbol{0}. \tag{1.2.9}$$

♦

Proof. Recall that $s_k(\boldsymbol{\gamma})$ was defined as:

$$s_k(\boldsymbol{\gamma}) = \frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{G}^{-1}\boldsymbol{\gamma} - \sum_{i=1}^{n_k}\ell_{ki}(\boldsymbol{\gamma}).$$

Finding the gradient of $s_k(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$:

$$\begin{aligned}\frac{\partial s_k}{\partial\boldsymbol{\gamma}} &= \boldsymbol{G}^{-1}\boldsymbol{\gamma} - \sum_{i=1}^{n_k}\frac{\partial\ell_{ki}}{\partial\mu_{ki}}\frac{\partial\mu_{ki}}{\partial\eta_{ki}}\frac{\partial\eta_{ki}}{\partial\boldsymbol{\gamma}} \\ &= \boldsymbol{G}^{-1}\boldsymbol{\gamma} - \sum_{i=1}^{n_k}\frac{y_{ki}-\mu_{ki}}{\phi\nu(\mu_{ki})}\frac{1}{\dot{g}(\mu_{ki})}\boldsymbol{z}_{ki}.\end{aligned}$$

■

Proposition 1.2.3. Under the canonical link, the Hessian of $s_k(\boldsymbol{\gamma})$ w.r.t. $\boldsymbol{\gamma}$ is:

$$\ddot{s}(\boldsymbol{\gamma}) = \frac{\partial^2 s_k}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} = \boldsymbol{G}^{-1} + \boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{Z}_k.$$

♦

Proof. Since $\dot{g}(\mu_{ki})\nu(\mu_{ki}) = 1$ under the canonical link, the gradient of $s_k(\boldsymbol{\gamma})$ is:

$$\dot{s}_k(\boldsymbol{\gamma}) = \boldsymbol{G}^{-1}\boldsymbol{\gamma} - \sum_{i=1}^{n_k}\frac{y_{ki}-\mu_{ki}}{\phi}\boldsymbol{z}_{ki}$$

Taking the partial of $\dot{s}_k(\boldsymbol{\gamma})$ w.r.t. $\boldsymbol{\gamma}'$:

$$\begin{aligned}
\ddot{s}(\boldsymbol{\gamma}) &= \frac{\partial}{\partial\boldsymbol{\gamma}'}\left\{\boldsymbol{G}^{-1}\boldsymbol{\gamma} - \sum_{i=1}^{n_k}\frac{y_{ki}-\mu_{ki}}{\phi}\boldsymbol{z}_{ki}\right\} \\
&= \boldsymbol{G}^{-1} + \sum_{i=1}^{n_k}\frac{\boldsymbol{z}_{ki}}{\phi}\frac{\partial\mu_{ki}}{\partial\eta_{ki}}\frac{\partial\eta_{ki}}{\partial\boldsymbol{\gamma}'} \\
&= \boldsymbol{G}^{-1} + \sum_{i=1}^{n_k}\frac{\boldsymbol{z}_{ki}}{\phi}\frac{1}{\dot{g}(\mu_{ki})}\boldsymbol{z}_{ki}' \\
&= \boldsymbol{G}^{-1} + \sum_{i=1}^{n_{ki}}\boldsymbol{z}_{ki}\frac{1}{\phi\nu(\mu_{ki})\dot{g}^2(\mu_{ki})}\boldsymbol{z}_{ki}'.
\end{aligned}$$

■

Proposition 1.2.4. Under the canonical link, the Laplace likelihood is expressible as:

$$\ell_q(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\gamma}^*) \propto \sum_{k=1}^{K}\sum_{i=1}^{n_k}\ell_{ki}(\boldsymbol{\gamma}_k^*) - \frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{\gamma}_k^*)'\boldsymbol{G}^{-1}\boldsymbol{\gamma}_k^* - \frac{1}{2}\sum_{k=1}^{K}\ln\det\left(\boldsymbol{I} + \boldsymbol{G}\boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{Z}_k\right). \tag{1.2.10}$$

♦

Proof. Recall that $s_k(\boldsymbol{\gamma})$ was defined as:

$$s_k(\boldsymbol{\gamma}) = \frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{G}^{-1}\boldsymbol{\gamma} - \sum_{i=1}^{n_k}\ell_{ki}(\boldsymbol{\gamma})$$

Writing out the approximate quasi likelihood:

$$\begin{aligned}
\ell(\boldsymbol{\beta},\boldsymbol{\alpha}) &\propto -\frac{K}{2}\ln\det(\boldsymbol{G}) - \sum_{k=1}^{K}s_k(\boldsymbol{\gamma}_k^*) - \frac{1}{2}\sum_{k=1}^{K}\ln\det\left\{\frac{\partial^2 s_k}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}(\boldsymbol{\gamma}_k^*)\right\} \\
&= -\frac{K}{2}\ln\det(\boldsymbol{G}) - \frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{\gamma}_k^*)'\boldsymbol{G}^{-1}\boldsymbol{\gamma}_k^* + \sum_{k=1}^{K}\sum_{i=1}^{n_k}\ell_{ki}(\boldsymbol{\gamma}_k^*) - \frac{1}{2}\sum_{k=1}^{K}\ln\det\left(\boldsymbol{G}^{-1} + \boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{Z}_k\right).
\end{aligned}$$

Combining the log determinant terms gives the result. ■

2.4 Penalized Quasi Likelihood

Remark 1.2.4. In *penalized quasi likelihood* (PQL), the Laplace objective is simplified by supposing that the weight matrix $\boldsymbol{W}_k$ changes only slowly with respect to μ_{ki}. For fixed $\boldsymbol{\alpha}$, a simple iteratively reweighted least squares (IRLS) procedure is available for solving the PQL score equations for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. For additional details, see [1]. ♦

Definition 1.2.2. Suppose that the weight matrix in (1.2.8) changes slowly w.r.t.the conditional mean:

$$\frac{\partial \boldsymbol{W}_k}{\partial \mu_{ki}} \approx \boldsymbol{0}. \tag{A1}$$

Under the PQL assumption (A1), the Laplace objective (1.2.10) is proportional to the **PQL objective** ℓ_{PQL}, which is *defined* as:

$$\ell_{\text{PQL}}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\gamma}) \propto \sum_{k=1}^{K}\sum_{i=1}^{n_k}\ell_{ki}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\gamma}_k) - \frac{1}{2}\sum_{k=1}^{K}\boldsymbol{\gamma}_k'\boldsymbol{G}^{-1}(\boldsymbol{\alpha})\boldsymbol{\gamma}_k. \tag{1.2.11}$$

In the PQL framework, $\boldsymbol{\gamma}_k$ is treated as a nuisance parameter that requires estimation. ■

Proposition 1.2.5 (PQL Score Equations). The PQL score equations are:

$$\mathcal{U}_\beta = \frac{\partial \ell_{\text{PQL}}}{\partial \boldsymbol{\beta}} = \sum_{k=1}^{K}\boldsymbol{X}_k'\boldsymbol{W}_k\boldsymbol{\Delta}_k\big(\boldsymbol{y}_k - \boldsymbol{\mu}_k\big),$$
$$\mathcal{U}_{\gamma_k} = \frac{\partial \ell_{\text{PQL}}}{\partial \boldsymbol{\gamma}_k} = -\boldsymbol{G}^{-1}\boldsymbol{\gamma}_k + \boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{\Delta}_k\big(\boldsymbol{y}_k - \boldsymbol{\mu}_k\big).$$

♦

Proof. The score for $\boldsymbol{\beta}$ is:

$$\mathcal{U}_\beta = \frac{\partial \ell_{\text{PQL}}}{\partial \boldsymbol{\beta}} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\frac{\partial \ell_{ki}}{\partial \mu_{ki}}\frac{\partial \mu_{ki}}{\partial \eta_{ki}}\frac{\partial \eta_{ki}}{\partial \boldsymbol{\beta}} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\frac{y_{ki}-\mu_{ki}}{\phi\nu(\mu_{ki})}\frac{1}{\dot{g}(\mu_{ki})}\boldsymbol{x}_{ki}.$$

The score for $\boldsymbol{\gamma}_k$ was obtained in (1.2.9). ■

Proposition 1.2.6. Under the canonical link, the PQL information matrices are:

- For $\boldsymbol{\beta}$:

$$\mathcal{J}_{\beta\beta'} = \sum_{k=1}^{K} \boldsymbol{X}_k' \boldsymbol{W}_k \boldsymbol{X}_k.$$

- For $\boldsymbol{\gamma}_k$:

$$\mathcal{J}_{\gamma_k\gamma_k'} = \boldsymbol{G}^{-1} + \boldsymbol{Z}_k' \boldsymbol{W}_k \boldsymbol{Z}_k.$$

- Cross information:

$$\mathcal{J}_{\beta\gamma_k'} = \boldsymbol{X}_k' \boldsymbol{W}_k \boldsymbol{Z}_k.$$

♦

Proof. Recall that under the canonical link $\nu(\mu_{ki})\dot{g}(\mu_{ki}) = 1$. The PQL score for $\boldsymbol{\beta}$ simplifies to:

$$\mathcal{U}_\beta = \sum_{k=1}^{K}\sum_{i=1}^{n_k} \frac{y_{ki} - \mu_{ki}}{\phi} \boldsymbol{x}_{ki}.$$

The Hessian w.r.t. $\boldsymbol{\beta}$ is:

$$\mathcal{H}_{\beta\beta'} = \frac{\partial \mathcal{U}_\beta}{\partial \boldsymbol{\beta}'} = -\sum_{k=1}^{K}\sum_{i=1}^{n_k} \boldsymbol{x}_{ki} \frac{1}{\phi \dot{g}(\mu_{ki})} \boldsymbol{x}_{ki}' = -\sum_{k=1}^{K} \boldsymbol{X}_k' \boldsymbol{W}_k \boldsymbol{X}_k.$$

The PQL score equation for $\boldsymbol{\gamma}_k$ is:

$$\mathcal{U}_{\gamma_k} = -\boldsymbol{G}^{-1}\boldsymbol{\gamma}_k + \sum_{i=1}^{n_k} \frac{y_{ki} - \mu_{ki}}{\phi} \boldsymbol{z}_{ki}$$

The Hessian w.r.t. $\boldsymbol{\gamma}_k$ is:

$$\mathcal{H}_{\gamma_k\gamma_k'} = \frac{\partial \mathcal{U}_{\gamma_k}}{\partial \boldsymbol{\gamma}_k'} = -\boldsymbol{G}^{-1} - \sum_{i=1}^{n_k} \boldsymbol{z}_{ki} \frac{1}{\phi \dot{g}(\mu_{ki})} \boldsymbol{z}_{ki}' = -\boldsymbol{G}^{-1} - \boldsymbol{Z}_k' \boldsymbol{W}_k \boldsymbol{Z}_k.$$

Finally, the cross Hessian matrix is:

$$\mathcal{H}_{\beta\gamma_k'} = \frac{\partial \mathcal{U}_\beta}{\partial \boldsymbol{\gamma}_k'} = -\sum_{i=1}^{n} \boldsymbol{x}_{ki} \frac{1}{\phi \dot{g}(\mu_{ki})} \boldsymbol{z}_{ki}' = -\boldsymbol{X}_k' \boldsymbol{W}_k \boldsymbol{Z}_k.$$

■

Estimation

3.1 Working Vector

Remark 1.3.1. Let $n = \sum_{k=1}^{m} n_k$. To simplify notation, define the following structures:

Structure	Dimension	Description
$\boldsymbol{y} = \text{vec}(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_K)$	$n \times 1$	Response vector.
$\boldsymbol{W} = \text{diag}(\boldsymbol{W}_1, \cdots, \boldsymbol{W}_K)$	$n \times n$	Weight matrix.
$\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\Delta}_1, \cdots, \boldsymbol{\Delta}_K)$	$n \times n$	Delta matrix.
$\boldsymbol{\mu} = \text{vec}(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_K)$	$n \times 1$	Mean vector.
$\boldsymbol{X} = \text{rbind}(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_K)$	$n \times p$	Fixed effect covariates.
$\boldsymbol{\gamma} = \text{vec}(\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_K)$	$Kq \times 1$	Random effect.
$\boldsymbol{Z} = \text{diag}(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_K)$	$n \times Kq$	Random effect covariates.
$\mathcal{G}^{-1} = \boldsymbol{I}_{K\times K} \otimes \boldsymbol{G}^{-1}$	$Kq \times Kq$	Random effect covariance.

♦

Definition 1.3.1. Define the **working vector** $\tilde{\boldsymbol{y}}$ as:

$$\tilde{\boldsymbol{y}} = \boldsymbol{\eta} + \boldsymbol{\Delta}(\boldsymbol{y} - \boldsymbol{\mu}). \tag{1.3.12}$$

Here the linear predictor is $\boldsymbol{\eta} \equiv \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma}$ and the mean vector is $\boldsymbol{\mu} = h(\boldsymbol{\eta})$. ■

3.2 Normal Equations

Proposition 1.3.1 (Fisher Scoring Algorithm). Fix the covariance parameters $\boldsymbol{\alpha}$. The solutions $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ to the PQL score equations satisfy the GLMM **normal equations**:

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z} + \mathcal{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{W}\tilde{\boldsymbol{y}} \\ \boldsymbol{Z}'\boldsymbol{W}\tilde{\boldsymbol{y}} \end{pmatrix}, \tag{1.3.13}$$

where $\tilde{\boldsymbol{y}}$ is the working vector (1.3.12). ♦

Proof. The PQL score equations are compactly expressible as:

$$\begin{aligned} \mathcal{U}_\beta &= \boldsymbol{X}'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{y} - \boldsymbol{\mu}), \\ \mathcal{U}_\gamma &= -\mathcal{G}^{-1}\boldsymbol{\gamma} + \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{y} - \boldsymbol{\mu}). \end{aligned} \tag{1.3.14}$$

Let $\mathcal{U} = (\mathcal{U}_\beta, \mathcal{U}_\gamma)'$ denote the stacked score equations, and let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$ group the parameters. Take the first order Taylor expansion of $\mathcal{U}$ around the PQL estimate $\hat{\boldsymbol{\theta}}$:

$$\mathbf{0} = \mathcal{U}(\hat{\boldsymbol{\theta}}) = \mathcal{U}(\boldsymbol{\theta}) + \mathcal{H}(\boldsymbol{\theta})\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big) + \mathcal{O}_p(1).$$

Under the canonical link, the PQL Hessian is:

$$\mathcal{H}(\boldsymbol{\theta}) = \frac{\partial \mathcal{U}}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}) = -\begin{pmatrix} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z} + \mathcal{G}^{-1} \end{pmatrix}.$$

From the first order Taylor expansion:

$$\mathcal{U}(\boldsymbol{\theta}) \approx -\mathcal{H}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{y}-\boldsymbol{\mu}) \\ \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{y}-\boldsymbol{\mu}) - \mathcal{G}^{-1}\gamma \end{pmatrix} \approx \begin{pmatrix} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z} + \mathcal{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\gamma} - \gamma \end{pmatrix}$$

Rearranging gives the PQL *normal equations*:

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z} + \mathcal{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{W}\tilde{\boldsymbol{y}} \\ \boldsymbol{Z}'\boldsymbol{W}\tilde{\boldsymbol{y}} \end{pmatrix},$$

where $\tilde{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\Delta}(\boldsymbol{y}-\boldsymbol{\mu}) = \boldsymbol{\eta} + \boldsymbol{\Delta}(\boldsymbol{y}-\boldsymbol{\mu})$ is the working vector. ■

Remark 1.3.2. Comparing the PQL normal equations (1.3.13) with *mixed model equations* (see notes on LMMs) demonstrates that the working-vector $\tilde{\boldsymbol{y}}$ follows a LMM:

$$\tilde{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{1.3.15}$$
$$\boldsymbol{\gamma} \sim N(0, \mathcal{G}) \perp \boldsymbol{\epsilon} \sim N(0, \boldsymbol{W}^{-1})$$

The induced marginal model is:

$$\tilde{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ with:

$$\boldsymbol{\Sigma} = \boldsymbol{W}^{-1} + \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'. \tag{1.3.16}$$

From the results for LMMs, the best linear unbiased estimator of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{y}}. \tag{1.3.17}$$

Define the error projection:

$$\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}.$$

The difference between the working response $\tilde{\boldsymbol{y}}$ and $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ is expressible as:

$$\tilde{\boldsymbol{y}} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{\Sigma}\big\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\big\}\tilde{\boldsymbol{y}} = \boldsymbol{\Sigma}\boldsymbol{Q}\tilde{\boldsymbol{y}}.$$

The best linear predictor of $\boldsymbol{\gamma}$ is:

$$\hat{\gamma} = \hat{E}(\gamma|\tilde{\boldsymbol{y}}) = \mathcal{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{y}} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \mathcal{G}\boldsymbol{Z}'\boldsymbol{Q}\tilde{\boldsymbol{y}}. \tag{1.3.18}$$

♦

3.3 Variance Component Estimation

Remark 1.3.3. For fixed $\boldsymbol{\alpha}$, the PQL normal equations (1.3.13) suggest the following procedure for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

Algorithm 1 PQL Estimation

Require: Covariance parameters $\boldsymbol{\alpha}$; observed data $\mathcal{D} = \{\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}\}$.

1: Initialize $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$.
2: **repeat**
3: Using the current $(\boldsymbol{\beta}^{(r)}, \boldsymbol{\gamma}^{(r)})$, construct:
4: **substeps**
5: $\boldsymbol{\Delta}^{(r)} = \text{diag}\{\dot{g}(\mu_{ki}^{(r)})\}$.
6: $\boldsymbol{W}^{(r)} = \text{diag}\{\phi\nu(\mu_{ki}^{(r)})\dot{g}^2(\mu_{ki}^{(r)})\}^{-1}$.
7: $\tilde{\boldsymbol{y}}^{(r)} = \boldsymbol{\eta}^{(r)} + \boldsymbol{\Delta}^{(r)}(\boldsymbol{y} - \boldsymbol{\mu}^{(r)})$.
8: $\boldsymbol{\Sigma}^{(r)} = \boldsymbol{W}^{(r),-1} + \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}$.
9: **end substeps**
10: Update: $\boldsymbol{\beta}^{(r+1)} \leftarrow \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{(r),-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{(r),-1}\tilde{\boldsymbol{y}}^{(r)}$.
11: Update: $\boldsymbol{\gamma}^{(r+1)} \leftarrow \mathcal{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{(r),-1}\left(\tilde{\boldsymbol{y}} - \boldsymbol{X}\boldsymbol{\beta}^{(r+1)}\right)$.
12: **until** $(\boldsymbol{\beta}^{(r+1)}, \boldsymbol{\gamma}^{(r+1)})$ stabilize.

Towards obtaining an estimation procedure for $\boldsymbol{\alpha}$, the PQL estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are substituted into the PQL objective (1.2.11) to obtain a *profile objective* for the variance components. ♦

3.3.1 Total Quasi Likelihood

Proposition 1.3.2. The difference between the working vector $\tilde{\boldsymbol{y}}$ and the fitted linear predictor $\hat{\boldsymbol{\eta}}$ is expressible as:

$$\tilde{\boldsymbol{y}} - \hat{\boldsymbol{\eta}} = \boldsymbol{W}^{-1}\boldsymbol{Q}\tilde{\boldsymbol{y}}. \tag{1.3.19}$$

♦

Proof. The fitted linear predictor is:

$$\hat{\boldsymbol{\eta}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\hat{\boldsymbol{\gamma}}.$$

Recall that $\tilde{\boldsymbol{y}} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{\Sigma}\boldsymbol{Q}\tilde{\boldsymbol{y}}$ and that $\boldsymbol{Z}\hat{\boldsymbol{\gamma}} = \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'\boldsymbol{Q}\tilde{\boldsymbol{y}}$, thus:

$$\tilde{\boldsymbol{y}} - \hat{\boldsymbol{\eta}} = \left(\tilde{\boldsymbol{y}} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right) - \boldsymbol{Z}\hat{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}\boldsymbol{Q}\tilde{\boldsymbol{y}} - \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'\boldsymbol{Q}\tilde{\boldsymbol{y}} = \left(\Sigma - \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'\right)\boldsymbol{Q}\tilde{\boldsymbol{y}}.$$

The conclusion follows since $\boldsymbol{\Sigma} = \boldsymbol{W}^{-1} + \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'$, or $\boldsymbol{\Sigma} - \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}' = \boldsymbol{W}^{-1}$. ■

Definition 1.3.2. Let T denote the *total quasi (log) likelihood*:

$$T(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) \equiv \sum_{k=1}^{K}\sum_{i=1}^{n_k} \ell_{ki} = \sum_{k=1}^{K}\sum_{i=1}^{n_k} \int_{y_{ki}}^{\mu_{ki}} \frac{(y_{ki}-u)}{\phi\nu(u)}du.$$

The **quasi deviance** is -2ϕ times the total quasi likelihood:

$$D(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k) \equiv -2\phi \cdot T = -2\sum_{k=1}^{K}\sum_{i=1}^{n_k} \int_{y_{ki}}^{\mu_{ki}} \frac{(y_{ki}-u)}{\nu(u)}du.$$

The Pearson χ^2 approximation to the model deviance is:

$$D \approx \sum_{k=1}^{K}\sum_{i=1}^{n_k} \frac{(y_{ki}-\hat{\mu}_{ki})^2}{\nu(\hat{\mu}_{ki})}.$$

■

Proposition 1.3.3. The Pearson approximation to the total quasi likelihood is:

$$\tilde{T} = -\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\boldsymbol{W}^{-1}\boldsymbol{Q}\tilde{\boldsymbol{y}}. \tag{1.3.20}$$

♦

Proof. Using the Pearson χ^2 approximation to the deviance :

$$D = -2\phi \cdot T \approx \sum_{k=1}^{K}\sum_{i=1}^{n_k} \frac{(y_{ki}-\hat{\mu}_{ki})^2}{\nu(\hat{\mu}_{ki})}.$$

Define the Pearson approximation to the total quasi likelihood as:

$$\tilde{T} = -\frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n_k} \frac{(y_{ki}-\hat{\mu}_{ki})^2}{\phi\nu(\hat{\mu}_{ki})}.$$

Expressed in matrix notation:

$$\tilde{T} = -\frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{y}_k - \hat{\boldsymbol{\mu}}_k)'\boldsymbol{\Delta}_k\boldsymbol{W}_k\boldsymbol{\Delta}_k(\boldsymbol{y}_k - \hat{\boldsymbol{\mu}}_k) = -\frac{1}{2}(\boldsymbol{y}-\hat{\boldsymbol{\mu}})'\boldsymbol{\Delta}\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{y}-\hat{\boldsymbol{\mu}}).$$

Recall that the working vector was defined as $\tilde{\boldsymbol{y}} = \hat{\boldsymbol{\eta}} + \boldsymbol{\Delta}(\boldsymbol{y}-\hat{\boldsymbol{\mu}})$. Replacing $\boldsymbol{\Delta}(\boldsymbol{y}-\hat{\boldsymbol{\mu}})$ by $\big(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\eta}}\big)$ gives $\tilde{T} = -\frac{1}{2}(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\eta}})'\boldsymbol{W}(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\eta}})$. Finally, applying (1.3.19):

$$\tilde{T} = -\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\boldsymbol{W}^{-1}\boldsymbol{W}\boldsymbol{W}^{-1}\boldsymbol{Q}\tilde{\boldsymbol{y}}.$$

■

3.3.2 Sylvester's Identity

Proposition 1.3.4. Consider the matrices $\boldsymbol{A}_{n\times q}$ and $\boldsymbol{B}_{q\times n}$ with $q<n$. The matrices $(\boldsymbol{AB})_{n\times n}$ and $(\boldsymbol{BA})_{q\times q}$ have the same non-zero eigenvalues. ♦

Proof. Suppose $\lambda\neq 0$ is an eigenvalue of $\boldsymbol{BA}$. There exists an eigenvector $\boldsymbol{u}_\lambda$, distinct from the null vector $\boldsymbol{u}_\lambda\neq\boldsymbol{0}$, s.t. $\boldsymbol{BAu}_\lambda=\lambda\boldsymbol{u}_\lambda$. Left multiply by $\boldsymbol{A}$ to obtain:

$$\boldsymbol{ABAu}_\lambda=\lambda\boldsymbol{Au}_\lambda \qquad \boldsymbol{AB}(\boldsymbol{Au}_\lambda)=\lambda(\boldsymbol{Au}_\lambda).$$

Thus, $\lambda\neq 0$ is an eigenvalue of $\boldsymbol{BA}$. Analogous reasoning shows that if $\lambda\neq 0$ is an eigenvalue of $\boldsymbol{AB}$, then $\lambda\neq 0$ is an eigenvalue of $\boldsymbol{BA}$. ■

Proposition 1.3.5 (Sylvester). Consider the $\boldsymbol{A}_{n\times q}$ and $\boldsymbol{B}_{q\times n}$ with $q<n$, then:

$$\det(\boldsymbol{I}_{n\times n}+\boldsymbol{AB})=\det(\boldsymbol{I}_{q\times q}+\boldsymbol{BA}). \tag{1.3.21}$$

♦

Proof. Let $\boldsymbol{BA}=\boldsymbol{U\Lambda U}^{-1}$ denote the eigenvalue decomposition of the $q\times q$ matrix $\boldsymbol{BA}$. Expressing the determinant of $\boldsymbol{I}+\boldsymbol{BA}$ as a product of eigenvalues:

$$\det(\boldsymbol{I}+\boldsymbol{BA})=\det(\boldsymbol{I}+\boldsymbol{U\Lambda U}^{-1})=\det\left\{\boldsymbol{U}(\boldsymbol{I}+\boldsymbol{\Lambda})\boldsymbol{U}^{-1}\right\}=\det(\boldsymbol{I}+\boldsymbol{\Lambda})=\prod_{j=1}^{q}(1+\lambda_j).$$

Observe that the set of null eigenvalues $\{j:\lambda_j=0\}$ contribute a one to the product, and hence leave the it unchanged. Thus:

$$\det(\boldsymbol{I}+\boldsymbol{BA})=\prod_{\{j:\lambda_j\neq 0\}}(1+\lambda_j).$$

Now let $\boldsymbol{AB}=\boldsymbol{VTV}^{-1}$ denotes the eigenvalue decomposition of the $n\times n$ matrix $\boldsymbol{AB}$. $\det(\boldsymbol{I}+\boldsymbol{AB})$ is likewise expressible as a product over the non-zero eigenvalues of $\boldsymbol{AB}$:

$$\det(\boldsymbol{I}+\boldsymbol{AB})=\prod_{\{j:\tau_j\neq 0\}}(1+\tau_j)$$

Since $\boldsymbol{AB}$ and $\boldsymbol{BA}$ have the same non-zero eigenvalues, conclude that:

$$\det(\boldsymbol{I}+\boldsymbol{AB})=\prod_{\{j:\tau_j\neq 0\}}(1+\tau_j)=\prod_{\{j:\lambda_j\neq 0\}}(1+\lambda_j)=\det(\boldsymbol{I}+\boldsymbol{BA}).$$

■

Proposition 1.3.6. Under the PQL assumption (A1):

$$\ln\det(\boldsymbol{I}+\boldsymbol{GZ}_k'\boldsymbol{W}_k\boldsymbol{Z}_k)\propto\ln\det(\boldsymbol{\Sigma}_k). \tag{1.3.22}$$

♦

Proof. Let $\boldsymbol{B} = \boldsymbol{G}\boldsymbol{Z}_k'$ and $\boldsymbol{A} = \boldsymbol{W}_k\boldsymbol{Z}_k$. Using Sylvester's identity (1.3.21):

$$\ln\det(\boldsymbol{I} + \boldsymbol{G}\boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{Z}_k) = \ln\det(\boldsymbol{I} + \boldsymbol{W}_k\boldsymbol{Z}_k\boldsymbol{G}\boldsymbol{Z}_k').$$

Recall from the working vector LMM (1.3.15) that:

$$\text{Var}\big(\tilde{\boldsymbol{y}}_k|\boldsymbol{X}_k, \boldsymbol{Z}_k\big) = \boldsymbol{W}_k^{-1} + \boldsymbol{Z}_k\boldsymbol{G}\boldsymbol{Z}_k'.$$

Expressing the log determinant in terms of $\boldsymbol{\Sigma}_k$:

$$\ln\det(\boldsymbol{I} + \boldsymbol{W}_k\boldsymbol{Z}_k\boldsymbol{G}\boldsymbol{Z}_k') = \ln\det(\boldsymbol{W}_k\boldsymbol{\Sigma}_k) = \ln\det(\boldsymbol{W}_k) + \ln\det(\boldsymbol{\Sigma}_k).$$

Now, assuming $\boldsymbol{W}_k$ changes only slowly w.r.t. μ_{ki}, such that $\partial_\alpha \ln\det(\boldsymbol{W}_k) \approx 0$:

$$\ln\det(\boldsymbol{I} + \boldsymbol{G}\boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{Z}_k) \propto \ln\det(\boldsymbol{\Sigma}_k).$$

■

3.3.3 Profile Quasi Likelihood

Result 1.3.1. Under the PQL assumption (A1), a profile quasi likelihood for the variance components $\boldsymbol{\alpha}$ is:

$$\ell_{pq}(\boldsymbol{\alpha}) \propto -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\tilde{\boldsymbol{y}}, \tag{1.3.23}$$

where $\boldsymbol{\Sigma} = \boldsymbol{W}^{-1} + \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'$ and $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}$ are defined as for the working vector LMM (1.3.15). ♣

Proof. Recall that the Laplace objective (1.2.10) took the form:

$$\ell_q(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \propto \sum_{k=1}^{K}\sum_{i=1}^{n_k}\ell_{ki} - \frac{1}{2}\sum_{k=1}^{K}(\gamma_k)'\boldsymbol{G}^{-1}\gamma_k - \frac{1}{2}\sum_{k=1}^{K}\ln\det\big(\boldsymbol{I} + \boldsymbol{G}\boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{Z}_k\big).$$

Proceed by substituting the PQL estimates of $\boldsymbol{\beta}$ (1.3.17) and of $\boldsymbol{\gamma}$ (1.3.18) into the Laplace objective to form a profile objective. The leading term is the total quasi likelihood T, which may be approximated as (1.3.20) using Pearson's statistic:

$$\sum_{k=1}^{K}\sum_{i=1}^{n_k}\ell_{ki}(\hat{\gamma}_k) \approx -\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\boldsymbol{W}^{-1}\boldsymbol{Q}\tilde{\boldsymbol{y}}.$$

The second term, evaluated at the BLUP $\hat{\boldsymbol{\gamma}} = \mathcal{G}\boldsymbol{Z}'\boldsymbol{Q}\tilde{\boldsymbol{y}}$, is expressible as:

$$-\frac{1}{2}\sum_{k=1}^{K}\hat{\gamma}_k'\boldsymbol{G}^{-1}\hat{\gamma}_k = -\frac{1}{2}\hat{\gamma}'\mathcal{G}^{-1}\hat{\gamma} = -\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\boldsymbol{Z}\mathcal{G}\mathcal{G}^{-1}\mathcal{G}\boldsymbol{Z}'\boldsymbol{Q}\tilde{\boldsymbol{y}}.$$

Combining the first and second terms gives:

$$-\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\boldsymbol{W}^{-1}\boldsymbol{Q}\tilde{\boldsymbol{y}} - \frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\boldsymbol{Z}\boldsymbol{\mathcal{G}}\boldsymbol{Z}'\boldsymbol{Q}\tilde{\boldsymbol{y}} = -\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\big(\boldsymbol{W}^{-1} + \boldsymbol{Z}\boldsymbol{\mathcal{G}}\boldsymbol{Z}'\big)\boldsymbol{Q}\tilde{\boldsymbol{y}} = -\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}\tilde{\boldsymbol{y}},$$

where $\boldsymbol{\Sigma}$ is the working vector covariance (1.3.16). Direct calculation demonstrates that $\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q} = \boldsymbol{Q}$ (see notes on ReML). Therefore:

$$\ell_q(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \hat{\boldsymbol{\gamma}}) = -\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\tilde{\boldsymbol{y}} - \frac{1}{2}\sum_{k=1}^{K}\ln\det\big(\boldsymbol{I} + \boldsymbol{G}\boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{Z}_k\big).$$

By proportion (1.3.22), the third term is proportionate to:

$$-\frac{1}{2}\sum_{k=1}^{K}\ln\det\big(\boldsymbol{I} + \boldsymbol{G}\boldsymbol{Z}_k'\boldsymbol{W}_k\boldsymbol{Z}_k\big) \propto -\frac{1}{2}\sum_{k=1}^{K}\ln\det(\boldsymbol{\Sigma}_k) = -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}).$$

Overall, the profile log likelihood for $\boldsymbol{\alpha}$ is:

$$\begin{aligned}\ell_{pq}(\boldsymbol{\alpha}) \equiv \ell_q(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \hat{\boldsymbol{\gamma}}) &= -\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\tilde{\boldsymbol{y}} - \frac{1}{2}\ln\det(\boldsymbol{\Sigma}) \\ &= -\frac{1}{2}\sum_{k=1}^{K}\big\{\tilde{\boldsymbol{y}}_k'\boldsymbol{Q}_k\tilde{\boldsymbol{y}}_k + \ln\det(\boldsymbol{\Sigma}_k)\big\}.\end{aligned}$$

■

Remark 1.3.4. A restricted profile quasi likelihood is formed by applying a Jeffrey's prior $\pi(\boldsymbol{\beta}) \propto \det(\mathcal{I}_{\beta\beta'})^{-1/2} = \det(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1/2}$ to the fixed effects to obtain:

$$\ell_{rpq}(\boldsymbol{\alpha}) \propto \ell_{pq}(\boldsymbol{\alpha}) + \ln\pi(\boldsymbol{\beta}) = -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\tilde{\boldsymbol{y}} - \frac{1}{2}\ln\det(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}). \quad (1.3.24)$$

The ReML score equation for a variance component α_p is:

$$\frac{\partial\ell_{rpq}}{\partial\alpha_p} = -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\right) + \frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\tilde{\boldsymbol{y}}.$$

The ReML cross information between α_p and α_q is:

$$\mathcal{I}_{\alpha_p\alpha_q} = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\right).$$

Typically, the ReML score equations have no closed form and must be solved numerically. When this is so, a computationally efficient approximation to the ReML information matrix is the **average information matrix**:

$$\mathcal{A}_{\alpha_p\alpha_q} = \frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\boldsymbol{Q}\tilde{\boldsymbol{y}}.$$

♦

Inference

4.1 Fixed Effects

Proposition 1.4.1. Consider the GLMM:

$$g(\mu_{ki}) = \boldsymbol{x}'_{A,ki}\boldsymbol{\beta}_A + \boldsymbol{x}'_{B,ki}\boldsymbol{\beta}_B + \boldsymbol{z}'_{ki}\gamma_k, \tag{1.4.25}$$

where the fixed effect $\boldsymbol{\beta}$ has been partitioned into those parameters $\boldsymbol{\beta}_A$ that are restricted under the null, and those parameters $\boldsymbol{\beta}_B$ that are not. A score test of $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_{A0}$ is:

$$T_S = \tilde{\boldsymbol{y}}_0'\tilde{\boldsymbol{Q}}_0\boldsymbol{X}_A\big(\boldsymbol{X}_A'\tilde{\boldsymbol{Q}}_0\boldsymbol{X}_A\big)^{-1}\boldsymbol{X}_A'\tilde{\boldsymbol{Q}}_0\tilde{\boldsymbol{y}} \dot{\sim} \chi^2_{\dim(\boldsymbol{\beta}_A)}.$$

The components of this statistic are:

i. Null model working response:

$$\tilde{\boldsymbol{y}}_0 = \tilde{\boldsymbol{\eta}}_0 + \tilde{\boldsymbol{\Delta}}(\boldsymbol{y} - \tilde{\boldsymbol{\mu}}_0),$$

where $\tilde{\boldsymbol{\eta}}_0 = \boldsymbol{X}_A\boldsymbol{\beta}_{A0} + \boldsymbol{X}_B\tilde{\boldsymbol{\beta}}_B + \boldsymbol{Z}\tilde{\boldsymbol{\gamma}}$, $\tilde{\boldsymbol{\mu}}_0 = h(\tilde{\boldsymbol{\eta}}_0)$, and $(\tilde{\boldsymbol{\beta}}_B, \tilde{\boldsymbol{\gamma}})$ are estimates obtained by fitting (1.4.25) under $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_{A0}$.

ii. Null model error projection:

$$\tilde{\boldsymbol{Q}}_0 = \tilde{\boldsymbol{\Sigma}}_0^{-1} - \tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{X}_B(\boldsymbol{X}_B'\tilde{\boldsymbol{\Sigma}}_0^{-1}\boldsymbol{X}_B)^{-1}\boldsymbol{X}_B'\tilde{\boldsymbol{\Sigma}}_0^{-1},$$

where $\tilde{\boldsymbol{\Sigma}}_0 = \tilde{\boldsymbol{W}}_0^{-1} + \boldsymbol{Z}\tilde{\mathcal{G}}\boldsymbol{Z}'$ is the working vector covariance under $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_{A0}$.

♦

Proof. From (1.3.14), the PQL score vector for $\boldsymbol{\beta}_A$ is:

$$\mathcal{U}_{\beta_A} = \boldsymbol{X}_A'\boldsymbol{W}\boldsymbol{\Delta}(\boldsymbol{y} - \boldsymbol{\mu}_0) = \boldsymbol{X}_A\boldsymbol{Q}_0\tilde{\boldsymbol{y}}_0.$$

By direct calculation, the variance of the PQL score is:

$$\mathrm{Var}\big(\mathcal{U}_{\beta_A}|\boldsymbol{X},\boldsymbol{Z}\big) = \boldsymbol{X}_A\boldsymbol{Q}_0\boldsymbol{\Sigma}_0\boldsymbol{Q}_0\boldsymbol{X}_A = \boldsymbol{X}_A\boldsymbol{Q}_0\boldsymbol{X}_A.$$

■

References

[0] Chen, H. and Wang, C. and Conomos, M. P. and others. “Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models”. In: *American Journal of Human Genetics* 98.4 (2016), pp. 653–66.

[1] N. E. Breslow and D. G. Clayton. “Approximate Inference in Generalized Linear Mixed Models”. In: *Journal of the American Statistical Association* 88.421 (1993), pp. 9–25.

[2] R. Wedderburn. “Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method”. In: *Biometrika* 61.3 (1974), pp. 439–447.