Linear Mixed Models

Introduction

1.1 Setting

Suppose that N total observations are grouped into K clusters. Let y_{ik} denote the ith outcome in the kth cluster. Group the N_k outcomes in th kth cluster to form the response vector $\boldsymbol{y}_k = (y_{1k}, \cdots, y_{kn_k})$. Associate with y_{ik} a $J \times 1$ vector $\boldsymbol{x}_{ik}$ of *fixed effect covariates*, and an $L \times 1$ vector $\boldsymbol{z}_{ik}$ of *random effect covariates*. Structure the covariates into design matrices $\boldsymbol{X}_k$ and $\boldsymbol{Z}_k$. Observations y_{ik_1} and y_{ik_2} belonging to distinct clusters are independent. However, observations y_{i_1k} and y_{i_2k} within a given cluster are potentially dependent.

1.2 Model

Definition 1.1.1. A **linear mixed effect model** for $\boldsymbol{y}_k$ takes the form:

$$\boldsymbol{y}_k = \boldsymbol{X}_k\boldsymbol{\beta} + \boldsymbol{Z}_k\gamma_k + \boldsymbol{\epsilon}_k,$$
$$\gamma_k \sim \big(\boldsymbol{0}, \boldsymbol{G}\big) \perp \boldsymbol{\epsilon}_k \sim \big(\boldsymbol{0}, \boldsymbol{R}_k\big).$$

Here $\boldsymbol{\beta}$ is *fixed effect* in the sense that its value is constant across clusters. $\boldsymbol{\gamma}_k$ is a *random effect* whose value varies across clusters according to a distribution with mean zero and covariance $\boldsymbol{G}$. $\boldsymbol{\epsilon}_k$ is a *residual* whose distribution has mean zero and cluster-specific covariance $\boldsymbol{R}_k$. ■

1.3 Notation

The components of an LMM are summarized here:

Structures	Dimension	Description
$\boldsymbol{\beta}$	$J \times 1$	Fixed effect
$\boldsymbol{\gamma}_k$	$L \times 1$	Cluster-specific random effect
$\boldsymbol{\epsilon}_k$	$N_k \times 1$	Residual
$\boldsymbol{\alpha}$	$M \times 1$	Covariance parameters
$\boldsymbol{G}(\boldsymbol{\alpha})$	$L \times L$	Random effect covariance
$\boldsymbol{R}_k(\boldsymbol{\alpha})$	$N_k \times N_k$	Residual covariance

Let $N = \sum_{k=1}^{K} N_k$. Define the following data structures:

Structure	Dimension
$\boldsymbol{y} = \text{vec}(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_K)'$	$N \times 1$
$\boldsymbol{X} = \text{rbind}(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_K)$	$N \times J$
$\boldsymbol{Z} = \text{diag}(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_K)$	$N \times KL$
$\mathcal{G} = \boldsymbol{I}_{K \times K} \otimes \boldsymbol{G}(\boldsymbol{\alpha})$	$KL \times KL$
$\mathcal{R} = \text{diag}\big\{\boldsymbol{R}_1(\boldsymbol{\alpha}), \cdots, \boldsymbol{R}_m(\boldsymbol{\alpha})\big\}$	$N \times N$

In compact notation, the LMM is expressible as:

$$\begin{gathered} \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \\ \boldsymbol{\gamma} \sim (\boldsymbol{0}, \mathcal{G}) \perp \boldsymbol{\epsilon} \sim (\boldsymbol{0}, \mathcal{R}). \end{gathered} \tag{1.1.1}$$

Here $\boldsymbol{\gamma}$ is the $KL \times 1$ vector $\text{vec}(\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_K)$ and $\boldsymbol{\epsilon}$ is the $N \times 1$ vector $\text{vec}(\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_K)$.

1.4 Likelihood

Proposition 1.1.1. Let $\mathcal{D}_k$ denote the collection covariates relevant to $\boldsymbol{y}_k$. For any LMM, the likelihood is expressible as:

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

♦

Proof. Factoring the likelihood across clusters:

$$L(\boldsymbol{\beta},\boldsymbol{\alpha}) = f(\boldsymbol{y}|\mathcal{D};\boldsymbol{\beta},\boldsymbol{\alpha}) = \prod_{k=1}^{K} f(\boldsymbol{y}_k|\mathcal{D}_k;\boldsymbol{\beta},\boldsymbol{\alpha})$$

Introducing the random effect:

$$L(\boldsymbol{\beta},\boldsymbol{\alpha}) = \prod_{k=1}^{K}\int f(\boldsymbol{y}_k,\gamma_k|\mathcal{D}_k;\boldsymbol{\beta},\boldsymbol{\alpha})d\gamma_k = \prod_{k=1}^{K}\int f(\boldsymbol{y}_k|\mathcal{D}_k,\gamma_k;\boldsymbol{\beta},\boldsymbol{\alpha})f(\gamma_k|\boldsymbol{\alpha})d\gamma_k.$$

■

Marginal Model Approach

Assumption 1.2.1. Hereafter, independent normal distributions are assumed for the random effects and residuals:

$$\begin{gathered}\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\epsilon}, \\ \boldsymbol{\gamma} \sim N\big(\mathbf{0},\mathcal{G}\big) \perp \boldsymbol{\epsilon} \sim N\big(\mathbf{0},\mathcal{R}\big).\end{gathered} \tag{1.2.2}$$

■

Proposition 1.2.1. The **induced marginal model** for $\boldsymbol{y}$ from (1.1.1) is:

$$\boldsymbol{y}\big|(\boldsymbol{X},\boldsymbol{Z}) \sim N\big(\boldsymbol{X\beta},\boldsymbol{\Sigma}\big), \tag{1.2.3}$$

where $\boldsymbol{\Sigma} \equiv \mathrm{Var}(\boldsymbol{y}|\mathcal{D}) = \mathcal{R} + \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'$. ♦

Proof. Since $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$ are normally distributed, the distribution of $\boldsymbol{y}$ integrated w.r.t. $\boldsymbol{\gamma}$ is again normal. By iterated expectation, the mean is:

$$E(\boldsymbol{y}|\mathcal{D}) = E\big\{E(\boldsymbol{y}|\gamma,\mathcal{D})|\mathcal{D}\big\} = E\big\{\boldsymbol{X\beta} + \boldsymbol{Z\gamma}\big\} = \boldsymbol{X\beta}.$$

By law of total variance, the covariance of $\boldsymbol{y}$ is:

$$\boldsymbol{\Sigma} \equiv \mathrm{Var}(\boldsymbol{y}|\mathcal{D}) = E\big\{\mathrm{Var}(\boldsymbol{y}|\gamma,\mathcal{D})|\mathcal{D}\big\} + \mathrm{Var}\big\{E(\boldsymbol{y}|\gamma,\mathcal{D})|\mathcal{D}\big\} = \mathcal{R} + \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'.$$

■

Corollary 1.2.1. The log likelihood of the induced marginal model is:

$$\ell(\boldsymbol{\beta},\boldsymbol{\alpha}) \propto -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\big(\boldsymbol{y} - \boldsymbol{X\beta}\big)'\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{y} - \boldsymbol{X\beta}\big). \tag{1.2.4}$$

♣

2.1 Estimation of $\boldsymbol{\beta}$

Result 1.2.1 (Score for $\boldsymbol{\beta}$). The score equation for $\boldsymbol{\beta}$ is:

$$\mathcal{U}_\beta = \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).$$

Solving the score equation $\mathcal{U}_\beta \overset{\text{Set}}{=} \mathbf{0}$ gives the *generalized least squares* (GLS) estimator:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} = \mathcal{I}_{\beta\beta'}^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}. \tag{1.2.5}$$

♣

Result 1.2.2 (Information for $\boldsymbol{\beta}$). The Hessian of the log likelihood w.r.t. $\boldsymbol{\beta}$ is:

$$\mathcal{H}_{\beta\beta'} = \frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = -\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$$

The expected information for $\boldsymbol{\beta}$ is:

$$\mathcal{I}_{\beta\beta'} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}. \tag{1.2.6}$$

♣

2.2 Estimation of $\boldsymbol{\alpha}$

2.2.1 Profile Likelihood

Remark 1.2.1. Since the estimator for $\boldsymbol{\beta}$ is available in closed form, we proceed by forming the profile log likelihood for $\boldsymbol{\alpha}$, and differentiating to obtain the *efficient score.*

♦

Definition 1.2.1. Define the *error projection* $\boldsymbol{Q}$ as:

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

■

Result 1.2.3 (Error Projection Properties). The error projection has the following properties:

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

ii. $\boldsymbol{Q}\Sigma\boldsymbol{Q} = \boldsymbol{Q}$.

♣

Proof. (i.) Expanding the GLS estimator in the residual $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ gives:

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

To establish the second point, expand the right hand error projection of $\boldsymbol{Q}\Sigma\boldsymbol{Q}$:

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

The conclusion holds if the second term vanishes. Expanding the error projection in the second term gives:

$$\begin{aligned}\boldsymbol{Q}\boldsymbol{X}(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\Sigma^{-1} =& \Sigma^{-1}\boldsymbol{X}(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\Sigma^{-1} \\ &- \Sigma^{-1}\boldsymbol{X}(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\Sigma^{-1}\boldsymbol{X}(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\Sigma^{-1} = \boldsymbol{0}.\end{aligned}$$

■

Remark 1.2.2. The following is an identity for differentiation of the error projection $\boldsymbol{Q}$ w.r.t. a variance component α_p.

$$\frac{\partial \boldsymbol{Q}}{\partial \alpha_p} = -\boldsymbol{Q}\frac{\partial \Sigma}{\partial \alpha_p}\boldsymbol{Q}.$$

♦

Proposition 1.2.2. The profile log likelihood for $\boldsymbol{\alpha}$ is:

$$\ell_p(\boldsymbol{\alpha}) \propto -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}. \tag{1.2.7}$$

♦

Proof. The marginal log likelihood for $\boldsymbol{y}$ is:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}) \propto -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).$$

Substituting the GLS estimator (1.2.5) for $\boldsymbol{\beta}$ into the marginal log likelihood:

$$\ell_p(\boldsymbol{\alpha}) \propto -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}).$$

Expressing profile log likelihood in terms of the error projection:

$$\begin{aligned}\ell_p(\boldsymbol{\alpha}) &\propto -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\big\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})\big\}'\boldsymbol{\Sigma}\big\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})\big\} \\ &= -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}\boldsymbol{y} \\ &= -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}.\end{aligned}$$

■

2.2.2 Restricted Likelihood

Definition 1.2.2. A **restricted likelihood** is formed by applying a Jeffreys' to the fixed effects:

$$\pi(\boldsymbol{\beta}) \propto \det(\mathcal{I}_{\beta\beta'})^{-1/2}.$$

The restricted log likelihood for $\boldsymbol{\alpha}$ is:

$$\begin{aligned}\ell_r(\boldsymbol{\alpha}) &\equiv \ell_p(\boldsymbol{\alpha}) + \ln \pi(\boldsymbol{\beta}) = \ell_p(\boldsymbol{\alpha}) - \frac{1}{2}\ln\det(\mathcal{I}_{\beta\beta'}) \\ &\propto -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y} - \frac{1}{2}\ln\det(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}). \end{aligned} \tag{1.2.8}$$

■

Remark 1.2.3. Maximum likelihood estimates (MLEs) of the variance components $\boldsymbol{\alpha}$ are downward biased, whereas the restricted MLEs (ReMLs) are unbiased. ♦

Remark 1.2.4. The following are identities for differentiation of a matrix $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ w.r.t. a variance component α_p:

- Derivative of inverse:

$$\frac{\partial}{\partial\alpha_p}\boldsymbol{\Sigma}^{-1} = -\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{\Sigma}^{-1}.$$

- Derivative of log determinant:

$$\frac{\partial}{\partial\alpha_p}\ln\det(\boldsymbol{\Sigma}) = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\right).$$

♦

Result 1.2.4 (Restricted Score for $\boldsymbol{\alpha}$). The restricted score for α_p is:

$$\mathcal{U}_{\alpha_p} = \frac{\partial\ell_r}{\partial\alpha_p} = -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\right) + \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\boldsymbol{y}. \tag{1.2.9}$$

♣

Proof. The derivative of the restricted log likelihood (1.2.8) w.r.t. α_p is:

$$\begin{aligned}\frac{\partial\ell_r}{\partial\alpha_p} = &-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\right) + \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\boldsymbol{y} \\ &+ \frac{1}{2}\mathrm{tr}\left((\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right).\end{aligned}$$

Combining the trace terms gives:

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

■

Proposition 1.2.3. $\boldsymbol{Q}\boldsymbol{y}$ is distributed as:

$$\boldsymbol{Q}\boldsymbol{y}|\mathcal{D} \sim N(\boldsymbol{0}, \boldsymbol{Q}).$$

♦

Proof. Since $\boldsymbol{y}|\mathcal{D} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is normally distributed, the linear function $\boldsymbol{Q}\boldsymbol{y}$ is again normally distributed with mean:

$$\begin{aligned} E(\boldsymbol{Q}\boldsymbol{y}|\mathcal{D}) = \boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta} &= \big\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\big\}\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}, \end{aligned}$$

and variance:

$$\mathrm{Var}(\boldsymbol{Q}\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{Q}\mathrm{Var}(\boldsymbol{y}|\boldsymbol{X})\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q} = \boldsymbol{Q}.$$

■

Proposition 1.2.4. Suppose $E(\boldsymbol{y}) = \boldsymbol{\mu}$ and $\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}$. The expectation of the quadratic form $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$ is:

$$E(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}) = \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}.$$

♦

Proof.

$$\begin{aligned} E(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}) &= E\big\{\mathrm{tr}(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y})\big\} = E\big\{\mathrm{tr}(\boldsymbol{y}\boldsymbol{y}'\boldsymbol{A})\big\} = \mathrm{tr}\big\{E(\boldsymbol{y}\boldsymbol{y}')\boldsymbol{A}\big\} \\ &= \mathrm{tr}\big\{(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')\boldsymbol{A}\big\} = \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}. \end{aligned}$$

■

Result 1.2.5 (Restricted Information for $\boldsymbol{\alpha}$). The restricted information between α_p and α_q is:

$$\mathcal{I}_{\alpha_p\alpha_q} = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\right). \tag{1.2.10}$$

♣

Proof. Differentiating the restricted score (1.2.9) to obtain the Hessian:

$$\begin{aligned} \mathcal{H}_{\alpha_p\alpha_q} \equiv& \frac{\partial^2\ell_r}{\partial\alpha_p\partial\alpha_q} \\ =& +\frac{1}{2}\mathrm{tr}\left(\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\right) - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{Q}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\alpha_q\partial\alpha_p}\right) \\ & -\frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\boldsymbol{y} + \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\alpha_q\partial\alpha_p}\boldsymbol{Q}\boldsymbol{y} - \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\boldsymbol{Q}\boldsymbol{y}. \end{aligned}$$

Taking the expectation:

$$\begin{aligned}E(\mathcal{H}_{\alpha_p\alpha_q}|\boldsymbol{X}) =& \frac{1}{2}\text{tr}\left(\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\right) - \frac{1}{2}\text{tr}\left(\boldsymbol{Q}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\alpha_q\partial\alpha_p}\right) \\ & - \frac{1}{2}\text{tr}\left(\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\right) + \frac{1}{2}\text{tr}\left(\boldsymbol{Q}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\alpha_p\partial\alpha_q}\right) - \frac{1}{2}\text{tr}\left(\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\right).\end{aligned}$$

Combining like terms gives the result. ■

Definition 1.2.3. Suppose the covariance matrix $\boldsymbol{\Sigma}$ is linear in the parameters $\boldsymbol{\alpha}$ s.t.

$$\frac{\partial^2\boldsymbol{\Sigma}}{\partial\alpha_p\partial\alpha_q} = \boldsymbol{0}.$$

In this setting, the observed information is:

$$\mathcal{J}_{\alpha_p\alpha_q} = -\frac{1}{2}\text{tr}\left(\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\right) + \boldsymbol{y}'\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\boldsymbol{Q}\boldsymbol{y}.$$

The **average information** is defined as:

$$\mathcal{A}_{\alpha_p\alpha_q} = \frac{1}{2}(\mathcal{I}_{\alpha_p\alpha_q} + \mathcal{J}_{\alpha_p\alpha_q}) = \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\boldsymbol{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_q}\boldsymbol{Q}\boldsymbol{y}. \tag{1.2.11}$$

■

Remark 1.2.5. The *Newton-Raphson* iteration for estimation of the variance components is:

$$\boldsymbol{\alpha}^{(r+1)} = \boldsymbol{\alpha}^{(r)} + \mathcal{A}^{-1}(\boldsymbol{\alpha}^{(r)})\mathcal{U}_{\alpha}(\boldsymbol{\alpha}^{(r)}),$$

where $\boldsymbol{\alpha}^{(r)}$ is the current parameter estimate, $\mathcal{A}$ is the average information for $\boldsymbol{\alpha}$ from(1.2.11), and $\mathcal{U}_{\alpha}$ is the restricted score for $\boldsymbol{\alpha}$ from (1.2.9). ♦

Conditional Model Approach

3.1 Mixed Model Equations

Remark 1.3.1. In the marginal model approach, $\boldsymbol{\gamma}$ was treated as unobserved data and integrated away. In the conditional model approach, $\boldsymbol{\gamma}$ is treated as a parameter that requires estimation. ♦

Proposition 1.3.1. The conditional model log likelihood is:

$$\begin{aligned}\ell_C(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\gamma}) &= \ln f(\boldsymbol{y}|\mathcal{D};\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\gamma}) + \ln f(\boldsymbol{\gamma};\boldsymbol{\alpha}) \\ &\propto -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Z\gamma})\mathcal{R}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Z\gamma}) - \frac{1}{2}\gamma'\mathcal{G}^{-1}\gamma.\end{aligned}$$

♦

Result 1.3.1 (Mixed Model Equations). For fixed $\boldsymbol{\alpha}$, the best linear unbiased estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, and the best linear unbiased predictor $\hat{\gamma}$ of $\boldsymbol{\gamma}$ satisfy:

$$\begin{pmatrix} \boldsymbol{X}'\mathcal{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\mathcal{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z}+\mathcal{G}^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\mathcal{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{y} \end{pmatrix}. \tag{1.3.12}$$

♣

Proof. The score equations for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are:

$$\begin{aligned} \mathcal{U}_{\beta} &= \boldsymbol{X}'\mathcal{R}^{-1}\big(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{\gamma}\big) \overset{\text{Set}}{=} 0, \\ \mathcal{U}_{\gamma} &= \boldsymbol{Z}'\mathcal{R}^{-1}\big(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{\gamma}\big)+\mathcal{G}^{-1}\boldsymbol{\gamma} \overset{\text{Set}}{=} 0. \end{aligned}$$

Re-arranging:

$$\begin{aligned} \boldsymbol{X}'\mathcal{R}^{-1}\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{X}'\mathcal{R}^{-1}\boldsymbol{Z}\gamma &= \boldsymbol{X}'\mathcal{R}^{-1}\boldsymbol{y}, \\ \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{X}\boldsymbol{\beta}+(\boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z}+\mathcal{G}^{-1})\gamma &= \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{y}. \end{aligned}$$

In matrix format:

$$\begin{pmatrix} \boldsymbol{X}'\mathcal{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\mathcal{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z}+\mathcal{G}^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\mathcal{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{y} \end{pmatrix}.$$

Define $\boldsymbol{V}$ and $\boldsymbol{W}$ as:

$$\boldsymbol{V} = (\boldsymbol{X},\boldsymbol{Z}), \quad \boldsymbol{W} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \mathcal{G} \end{pmatrix}.$$

Now the normal equations are expressible as:

$$(\boldsymbol{V}'\mathcal{R}^{-1}\boldsymbol{V}+\boldsymbol{W}^{-1})\begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix} = \boldsymbol{V}'\mathcal{R}^{-1}\boldsymbol{y}.$$

Hence, the best linear estimates of $\boldsymbol{\beta},\boldsymbol{\gamma}$ are given by:

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\gamma} \end{pmatrix} = (\boldsymbol{V}'\mathcal{R}^{-1}\boldsymbol{V}+\boldsymbol{W}^{-1})^{-1}\boldsymbol{V}'\mathcal{R}^{-1}\boldsymbol{y}.$$

■

3.2 Random Effect Prediction

Result 1.3.2 (Empirical Bayes Estimation). The best linear unbiased prediction of $\boldsymbol{\gamma}$ is given by:

$$\tilde{\gamma} = E(\boldsymbol{\gamma}|\boldsymbol{y},\mathcal{D}) = \mathcal{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\big(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}\big) = \mathcal{G}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{y}. \tag{1.3.13}$$

♣

Proof. From model (1.2.2), $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$ are jointly distributed as:

$$\binom{\boldsymbol{\gamma}}{\boldsymbol{\epsilon}} \sim N\left(\binom{\mathbf{0}}{\mathbf{0}}, \begin{pmatrix} \mathcal{G} & \mathbf{0} \\ \mathbf{0} & \mathcal{R} \end{pmatrix}\right).$$

The joint distribution of $\boldsymbol{y}$ and $\boldsymbol{\gamma}$ is a linear transformation of $\text{vec}(\boldsymbol{\gamma}, \boldsymbol{\epsilon})$:

$$\binom{\boldsymbol{y}}{\boldsymbol{\gamma}} = \binom{\boldsymbol{X}\boldsymbol{\beta}}{\mathbf{0}} + \begin{pmatrix} \boldsymbol{Z} & \boldsymbol{I} \\ \boldsymbol{I} & \mathbf{0} \end{pmatrix} \binom{\boldsymbol{\gamma}}{\boldsymbol{\epsilon}}.$$

Thus $\text{vec}(\boldsymbol{y}, \boldsymbol{\gamma})$ is normally distributed with mean $\text{vec}(\boldsymbol{X}\boldsymbol{\beta}, \mathbf{0})$ and variance:

$$\begin{aligned} \text{Var}\binom{\boldsymbol{y}}{\boldsymbol{\gamma}} &= \begin{pmatrix} \boldsymbol{Z} & \boldsymbol{I} \\ \boldsymbol{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathcal{G} & \mathbf{0} \\ \mathbf{0} & \mathcal{R} \end{pmatrix} \begin{pmatrix} \boldsymbol{Z}' & \boldsymbol{I} \\ \boldsymbol{I} & \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}' + \mathcal{R} & \boldsymbol{Z}\mathcal{G} \\ \mathcal{G}\boldsymbol{Z}' & \mathcal{G} \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{Z}\mathcal{G} \\ \mathcal{G}\boldsymbol{Z}' & \mathcal{G} \end{pmatrix}. \end{aligned}$$

The conditional distribution of $\boldsymbol{\gamma}$ given $\boldsymbol{y}$ is again normal with expectation and variance:

$$E(\boldsymbol{\gamma}|\boldsymbol{y}, \mathcal{D}) = \mathcal{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \qquad \text{Var}(\boldsymbol{\gamma}|\boldsymbol{y}, \mathcal{D}) = \mathcal{G} - \mathcal{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\mathcal{G}.$$

From the Gauss-Markov theorem, the best linear unbiased estimator of $\boldsymbol{\beta}$ is the generalized least squares estimator:

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

Substituting $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ into $E(\boldsymbol{\gamma}|\boldsymbol{y})$ gives:

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

■

Corollary 1.3.1. The EB prediction $\tilde{\boldsymbol{\gamma}}$ is a weighted average between the GLS estimator $\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ and zero:

$$\tilde{\boldsymbol{\gamma}} = (\mathcal{G}^{-1} + \boldsymbol{Z}'\mathcal{R}\boldsymbol{Z})^{-1}\big\{(\boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z})\hat{\boldsymbol{\gamma}} + \mathcal{G}^{-1}\mathbf{0}\big\}.$$

That is, $\tilde{\boldsymbol{\gamma}}$ is a *shrinkage estimator*. ♣

Proof. From the induced marginal model (1.2.3), $\boldsymbol{\Sigma} = \mathcal{R} + \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'$. Multiplying by $\boldsymbol{Z}'\mathcal{R}^{-1}$ on the left and rearranging gives:

$$\begin{gathered} \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{\Sigma} = \boldsymbol{Z}'\mathcal{R}^{-1}\big(\mathcal{R} + \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'\big) = \boldsymbol{Z}' + \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z}\mathcal{G}\boldsymbol{Z}' = (\mathcal{G}^{-1} + \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z})\mathcal{G}\boldsymbol{Z}', \\ (\mathcal{G}^{-1} + \boldsymbol{Z}'\mathcal{R}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\mathcal{R}^{-1} = \mathcal{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}. \end{gathered}$$

Suppose $\boldsymbol{\gamma}$ were treated as a fixed effect and estimated from the model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} \sim N\big(\boldsymbol{0}, \mathcal{R}\big)$. The BLUE of $\boldsymbol{\gamma}$ is:

$$\hat{\gamma} = \big(\boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z}\big)^{-1}\boldsymbol{Z}'\mathcal{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).$$

From (1.3.13), the EB estimator of $\boldsymbol{\gamma}$ is:

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

Using the equivalence $\mathcal{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1} = (\mathcal{G}^{-1} + \boldsymbol{Z}'\mathcal{R}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\mathcal{R}^{-1}$, the EM estimator of $\boldsymbol{\gamma}$ is expressible as:

$$\begin{aligned}
\tilde{\gamma} &= (\mathcal{G}^{-1} + \boldsymbol{Z}'\mathcal{R}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\mathcal{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= (\mathcal{G}^{-1} + \boldsymbol{Z}'\mathcal{R}\boldsymbol{Z})^{-1}\big\{(\boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z})\hat{\gamma} + \mathcal{G}^{-1}\boldsymbol{0}\big\}.
\end{aligned}$$

■

Corollary 1.3.2. The empirical Bayes prediction of $\boldsymbol{y}$ is:

$$\tilde{\boldsymbol{y}} = \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} + \big(\boldsymbol{I} - \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\big)\boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

The EB prediction $\tilde{\boldsymbol{y}}$ is interpretable as a weighted average of the observations $\boldsymbol{y}$ and the fitted values $\boldsymbol{X}\hat{\boldsymbol{\beta}}$. ♣

Proof. The first term is expressible as:

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

The second term is expressible as:

$$\boldsymbol{Z}\hat{\gamma} = \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{Z}\mathcal{G}\boldsymbol{Z}'\big\{\boldsymbol{I} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\big\}\boldsymbol{y}$$

Combining like terms gives the result. ■

EM Algorithm

Proof. Regarding $\boldsymbol{\gamma}$ as missing data, the *complete data* log likelihood is:

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}) &\propto -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\gamma)'\mathcal{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\gamma) - \frac{1}{2}\gamma'\mathcal{G}^{-1}\gamma \\
&= -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\mathcal{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
&\quad - \frac{1}{2}\gamma'\boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z}\gamma + \frac{2}{2}\gamma'\boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{y} - \frac{1}{2}\gamma'\mathcal{G}^{-1}\gamma \\
&= -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\mathcal{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
&\quad + \gamma'\boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{y} - \frac{1}{2}\gamma'(\mathcal{G}^{-1} + \boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z})\gamma.
\end{aligned}$$

From the derivation of the EB estimator for $\boldsymbol{\gamma}$, the conditional distribution of $\boldsymbol{\gamma}$ given the observed data is normal with mean and covariance:

$$E(\boldsymbol{\gamma}|\boldsymbol{y},\mathcal{D}) = \boldsymbol{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}), \qquad \mathrm{Var}(\boldsymbol{\gamma}|\boldsymbol{y},\mathcal{D}) = \boldsymbol{G}-\boldsymbol{G}\boldsymbol{Z}'\Sigma^{-1}\boldsymbol{Z}\boldsymbol{G}.$$

The EM objective function is defined as the expectation of the complete data log likelihood given the observed data and the current parameter state:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) \equiv E\big\{\ell(\boldsymbol{\beta},\boldsymbol{\alpha})|\boldsymbol{y},\mathcal{D},\boldsymbol{\theta}^{(r)}\big\}.$$

Define the following working expectations:

$$\begin{aligned}\hat{\gamma}^{(r)} &\equiv E(\boldsymbol{\gamma}|\boldsymbol{y},\mathcal{D},\boldsymbol{\theta}^{(r)}) = \boldsymbol{G}^{(r)}\boldsymbol{Z}'\boldsymbol{\Sigma}^{(r),-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^{(r)}),\\ \hat{\boldsymbol{V}}^{(r)} &\equiv \mathrm{Var}(\boldsymbol{\gamma}|\boldsymbol{y},\mathcal{D},\boldsymbol{\theta}^{(r)}) = \boldsymbol{G}^{(r)}-\boldsymbol{G}^{(r)}\boldsymbol{Z}'\boldsymbol{\Sigma}^{(r),-1}\boldsymbol{Z}\boldsymbol{G}^{(r)}.\end{aligned}$$

Let $\boldsymbol{A} \equiv \mathcal{G}^{-1}+\boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{Z}$. Using the working expectations:

$$\begin{aligned}E\big(\boldsymbol{\gamma}'\boldsymbol{Z}'\mathcal{R}^{-1}\boldsymbol{y}|\boldsymbol{y},\mathcal{D}\big) &= (\hat{\gamma}^{(r)})'\boldsymbol{Z}\mathcal{R}^{-1}\boldsymbol{y},\\ E\big(\boldsymbol{\gamma}'\boldsymbol{A}\boldsymbol{\gamma}|\boldsymbol{y},\mathcal{D}\big) &= \mathrm{tr}\big\{\hat{\boldsymbol{V}}^{(r)}\boldsymbol{A}\big\}+(\hat{\gamma}^{(r)})'\boldsymbol{A}(\hat{\gamma}^{(r)}).\end{aligned}$$

The EM objective function is now expressible as:

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = &-\frac{1}{2}\ln\det(\boldsymbol{\Sigma})-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\mathcal{R}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})\\ &+(\hat{\gamma}^{(r)})'\boldsymbol{Z}\mathcal{R}^{-1}\boldsymbol{y}-\frac{1}{2}\mathrm{tr}\big\{\hat{\boldsymbol{V}}^{(r)}\boldsymbol{A}\big\}-\frac{1}{2}(\hat{\gamma}^{(r)})'\boldsymbol{A}(\hat{\gamma}^{(r)}).\end{aligned}$$

Consider a conditional maximization approach. Recall that the MLE of $\boldsymbol{\beta}$ is available in closed form. Let $\boldsymbol{\beta}^{(r+1)} \leftarrow \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}^{(r)})$ denote the GLS estimate of $\boldsymbol{\beta}$ given the current variance components $\boldsymbol{\alpha}^{(r)}$. Score equations for the variance components are obtained by differentiating the EM objective function:

$$\begin{aligned}\mathcal{U}_{\alpha_p}(\boldsymbol{\alpha}|\boldsymbol{\beta}^{(r+1)},\boldsymbol{\theta}^{(r)}) = &-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\alpha_p}\right)+\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^{(r+1)})\mathcal{R}^{-1}\frac{\partial\mathcal{R}}{\partial\alpha_p}\mathcal{R}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}^{(r+1)})\\ &-(\hat{\gamma}^{(r)})'\boldsymbol{Z}\mathcal{R}^{-1}\frac{\partial\mathcal{R}}{\partial\alpha_p}\mathcal{R}^{-1}\boldsymbol{y}-\frac{1}{2}\mathrm{tr}\left(\hat{\boldsymbol{V}}^{(r)}\frac{\partial\boldsymbol{A}}{\partial\alpha_p}\right)-\frac{1}{2}(\hat{\gamma}^{(r)})'\frac{\partial\boldsymbol{A}}{\partial\alpha_p}(\hat{\gamma}^{(r)}),\end{aligned}$$

where:

$$\frac{\partial\boldsymbol{A}}{\partial\alpha_p} = -\mathcal{G}^{-1}\frac{\partial\mathcal{G}}{\partial\alpha_p}\mathcal{G}^{-1}-\boldsymbol{Z}'\mathcal{R}^{-1}\frac{\partial\mathcal{R}}{\partial\alpha_p}\mathcal{R}^{-1}\boldsymbol{Z}.$$

The score equations for the variance components are solved numerically to obtain $\boldsymbol{\alpha}^{(r+1)}$. The algorithm iterates between updating $\boldsymbol{\beta}^{(r)}$ and updating $\boldsymbol{\alpha}^{(r)}$ until the improvement $\ell(\boldsymbol{\beta}^{(r+1)},\boldsymbol{\alpha}^{(r+1)})-\ell(\boldsymbol{\beta}^{(r)},\boldsymbol{\alpha}^{(r)})$ in the marginal log likelihood (1.2.4) falls below the tolerance. ■

Inference

5.1 Fixed Effects

Remark 1.5.1. Throughout, consider the marginalized LMM:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} \sim N\big(\mathbf{0}, \boldsymbol{\Sigma}\big)$. Partition the regression parameter $\boldsymbol{\beta} = (\boldsymbol{\beta}_A, \boldsymbol{\beta}_B)$. Suppose that $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_A^*$ is of interest. Let $\boldsymbol{X} = (\boldsymbol{X}_A, \boldsymbol{X}_B)$ denote the corresponding partition of the fixed effect design matrix. ♦

Definition 1.5.1. From (1.2.6), the information for $\boldsymbol{\beta}$ is $\mathcal{I}_{\beta\beta'} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$. Partition the information as:

$$\mathcal{I}_{\beta\beta'} = \begin{pmatrix} \mathcal{I}_{\beta_A\beta_A'} & \mathcal{I}_{\beta_A\beta_B'} \\ \mathcal{I}_{\beta_B\beta_A'} & \mathcal{I}_{\beta_B\beta_B'} \end{pmatrix}.$$

The **efficient information** for $\boldsymbol{\beta}_A$ is:

$$\mathcal{I}_{\beta_A\beta_A'|\beta_B} = \mathcal{I}_{\beta_A\beta_A'} - \mathcal{I}_{\beta_A\beta_A'}\mathcal{I}_{\beta_B\beta_B'}^{-1}\mathcal{I}_{\beta_B\beta_A'}.$$

■

Proposition 1.5.1. The Wald test of $H_0 : H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_A^*$ is:

$$T_W = (\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^*)'\mathcal{I}_{\beta_A\beta_A'|\beta_B}(\hat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^*) \dot{\sim} \chi^2_{\dim(\boldsymbol{\beta}_A^*)}.$$

♦

Proposition 1.5.2. Let $\tilde{\boldsymbol{\beta}}_B$ denote a consistent estimate of $\boldsymbol{\beta}_B$ under $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_A^*$, such as a solution to the score equation:

$$\mathcal{U}_{\beta_B}(\boldsymbol{\beta}_A = \boldsymbol{\beta}_A^*, \boldsymbol{\beta}_B) \overset{\text{Set}}{=} \mathbf{0}.$$

Let $\tilde{\mathcal{U}}_{\beta_A}$ denote $\mathcal{U}_{\beta_A}(\boldsymbol{\beta}_A = \boldsymbol{\beta}_A^*, \boldsymbol{\beta}_B = \tilde{\boldsymbol{\beta}}_B)$. The score test of $H_0 : \boldsymbol{\beta}_A = \boldsymbol{\beta}_A^*$ is:

$$T_S = \tilde{\mathcal{U}}_{\beta_A}'\mathcal{I}_{\beta_A\beta_A'|\beta_B}^{-1}\tilde{\mathcal{U}}_{\beta_A} \dot{\sim} \chi^2_{\dim(\boldsymbol{\beta}_A^*)}.$$

♦

5.2 Variance Components

Example 1.5.1 (Kernel Regression). Consider the kernel regression model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + f(\boldsymbol{z}_i) + \epsilon_i,$$

where $\epsilon_i \sim N\big(0, \sigma^2\big)$, and $f(\cdot)$ is an unknown function belonging to a reproducing kernel Hilbert space $\mathcal{H}$, with reproducing kernel $k(\cdot,\cdot)$. This model is isomorphic to the following random intercept model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \gamma + \boldsymbol{\epsilon},$$
$$\boldsymbol{\gamma} \sim N\big(\boldsymbol{0}, \tau^2\boldsymbol{K}\big) \perp \boldsymbol{\epsilon} \sim N\big(0, \sigma^2\boldsymbol{I}\big),$$

where $K_{ij} = k(\boldsymbol{z}_i, \boldsymbol{z}_j)$. Consider evaluating $H_0 : f(\boldsymbol{z}_i) \equiv 0$, or equivalently $H_0 : \tau^2 = 0$. The covariance of the induced marginal model is:

$$\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I} + \tau^2\boldsymbol{K}.$$

Identify $\boldsymbol{\alpha} = (\sigma^2, \tau^2)$. The restricted log likelihood is:

$$\ell_r(\boldsymbol{\alpha}) \propto -\frac{1}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\ln\det(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}) - \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}.$$

The score equation for τ^2 is:

$$\mathcal{U}_{\tau^2}(\sigma^2, \tau^2) = \frac{\partial \ell_r}{\partial \tau^2} = -\frac{1}{2}\mathrm{tr}\big(\boldsymbol{Q}\boldsymbol{K}\big) + \frac{1}{2}\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{K}\boldsymbol{Q}\boldsymbol{y}.$$

Under $H_0 : \tau^2 = 0$, the error projection reduces to:

$$\boldsymbol{Q} = \sigma^{-2}\big\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\big\} = \sigma^{-2}\boldsymbol{P}_X^{\perp}.$$

The score at $\tau^2 = 0$ is:

$$\mathcal{U}_{\tau^2}(\sigma^2, \tau^2 = 0) = -\frac{1}{2\sigma^2}\mathrm{tr}\big(\boldsymbol{P}_X^{\perp}\boldsymbol{K}\big) - \frac{1}{2\sigma^4}\boldsymbol{y}'\boldsymbol{P}_X^{\perp}\boldsymbol{K}\boldsymbol{P}_X^{\perp}\boldsymbol{y}.$$

Since the trace term does not depend on $\boldsymbol{y}$, consider the test statistic:

$$T_K = \boldsymbol{y}'\boldsymbol{P}_X^{\perp}\boldsymbol{K}\boldsymbol{P}_X^{\perp}\boldsymbol{y}.$$

Let $\boldsymbol{L}\boldsymbol{L}'$ denote the Cholesky decomposition of the kernel matrix $\boldsymbol{K}$. Under $H_0 : \tau^2 = 0$, $\boldsymbol{P}_X^{\perp}\boldsymbol{y} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{P}_X^{\perp})$. Thus T_K follows a mixture of central χ_1^2 distributions:

$$T_K \sim \sum_{i=1}^{n} \lambda_i \chi_1^2,$$

where (λ_i) are eigenvalues of the matrix:

$$\boldsymbol{\Xi} = \boldsymbol{L}'\mathrm{Var}(\boldsymbol{P}_X^{\perp}\boldsymbol{y})\boldsymbol{L} = \sigma^2\boldsymbol{L}'\boldsymbol{P}_X^{\perp}\boldsymbol{L}.$$

♠