Recap $\underline{y} = X\underline{\beta} + Z\underline{b} + \underline{\varepsilon}$

Then we will estimate the fixed effects by maximum likelihood, $\hat{\beta} = \arg\max L(\beta, \theta, \sigma^2 | y)$

$$= (X^T V^{-1} X)^{-1} X^T V^{-1} y$$

& estimate parameters in the distⁿ of the random effects, θ, σ^2, by either maximum likelihood or restricted maximum likelihood.

$$\hat{\theta}, \hat{\sigma}^2 = \arg\max L_R(\theta, \sigma^2 | y) = \arg\max \int L(\beta, \theta, \sigma^2 | y) \, d\beta$$

$\hat{\theta}, \hat{\sigma}^2$ are then unbiased for fixed effect models, but are still biased for ~~not~~ random effect models.

Results on multivariate Gaussians

Lemma 1 Any linear combination of Gaussians is Gaussian.

$$X \sim MVN_n(\mu, \Sigma) \quad \text{then} \quad AX \sim MVN_p(A\mu, A\Sigma A^T)$$

where A is $p \times n$ constant matrix

Lemma 2 If U & V are Gaussian & $\mathrm{Cov}(U,V) = 0$ then U & V are independent.

Proposition Suppose $\underline{S} = \begin{pmatrix} \underline{S}_1 \\ \underline{S}_2 \end{pmatrix}$ where

$$\begin{pmatrix} \underline{S}_1 \\ \underline{S}_2 \end{pmatrix} \sim MVN\left(\begin{pmatrix} \underline{m}_1 \\ \underset{\mathbb{E}\,\underline{S}_2}{\underline{m}_2} \end{pmatrix}, \begin{pmatrix} \overset{\mathrm{Var}(S_1)}{V_{11}} & \overset{\mathrm{Cov}(\underline{S}_1, \underline{S}_2)}{V_{12}} \\ V_{21} & V_{22} \end{pmatrix} \right)$$

Then

$$\underline{S}_2 \mid \underline{S}_1 \sim MVN\Big(\underbrace{\underset{\uparrow \text{ mean of } S_2}{\underline{m}_2} + V_{21} V_{11}^{-1} (\underline{S}_1 - \underline{m}_1)}_{\text{mean of } \underline{S}_2 | S_1}, \overbrace{\underset{\text{variance of } S_2}{V_{22}} - V_{21} V_{11}^{-1} V_{12}}^{\text{variance of } S_2 | S_1} \Big)$$

Proof Let $Z = S_2 + AS_1$ where $A = -V_{21}V_{11}^{-1}$

Then
$$\begin{aligned}
\mathrm{cov}(Z, S_1) &= \mathrm{Cov}(S_2, S_1) + \mathrm{Cov}(AS_1, S_1) \\
&= V_{21} + A\,\mathrm{Var}(S_1) \\
&= V_{21} + (-V_{21}V_{11}^{-1}) V_{11} \\
&= V_{21} - V_{21} = 0
\end{aligned}$$

$\therefore$ Z & S_1 are uncorrelated, & since they are both Gaussian they are independent.

Clearly $\mathbb{E}Z = \underline{m}_2 + A\underline{m}_1$

Thus $\mathbb{E}(S_2|S_1) = \mathbb{E}(Z - AS_1 | S_1)$

$= \mathbb{E}(Z|S_1) - A\,\mathbb{E}(S_1|S_1)$

$= \underbrace{m_2 + Am_1}_{\mathbb{E}Z} - AS_1$ as $Z \perp\!\!\!\perp S_1$

$= m_2 + V_{21}V_{11}^{-1}(S_1 - m_1)$

For the covariance matrix

$Var(S_2|S_1) = Var(Z - AS_1|S_1)$

$= Var(Z|S_1)$ as $Var(S_1|S_1) = 0$

$= Var(Z)$

$= Var(S_2 + AS_1)$

$= Var(S_2) + Var(AS_1) + 2\,Cov(S_2, AS_1)$

$= V_{22} + AV_{11}A^T + 2AV_{12}$

$= V_{22} + V_{21}V_{11}^{-1}V_{11}V_{11}^{-1}V_{12} - 2V_{21}V_{11}^{-1}V_{12}$

$= V_{22} - V_{21}V_{11}^{-1}V_{12}$

as $V_{21}^T = V_{12}$ as required.

Predicting random effects

$$\underline{Y} = X\underline{\beta} + Z\underline{b} + \underline{\varepsilon}$$

$$\underline{b} \sim N(0, V_b) \qquad \underline{\varepsilon} \sim N(0, \sigma^2 I)$$

We don't think of the b_i as parameters

Instead we predict them (i.e. prediction not estimation).

$\underline{Y} \sim N(X\underline{\beta}, V_Y)$ & $\underline{b}$ are both multivariate Gaussian

& so is $\begin{pmatrix} \underline{Y} \\ \underline{b} \end{pmatrix}$. We predict $\underline{b}$ by $\mathbb{E}(\underline{b} | \underline{Y})$

This is the Best Linear Unbiased Predictor (BLUP) for $\underline{b}$.

Write $\begin{pmatrix} \underline{Y} \\ \underline{b} \end{pmatrix} \sim MVN\left(\begin{pmatrix} X\underline{\beta} \\ \underline{0} \end{pmatrix}, \begin{pmatrix} V_Y & C_{Y,b} \\ C_{b,Y} & V_b \end{pmatrix} \right)$ Cov$(\underline{Y}, \underline{b})$

$$C_{Y,b} = \text{Cov}(\underline{Y}, \underline{b}) = \text{Cov}(X\underline{\beta} + Z\underline{b} + \underline{\varepsilon}, \underline{b}) = \text{Cov}(Z\underline{b}, \underline{b})$$
$$= Z V_b$$

So by the proposition

$$\underline{b} | \underline{Y} \sim MVN\left(\underline{0} + C_{b,Y} V_Y^{-1}(Y - X\beta), V_b - C_{b,Y} V_Y^{-1} C_{Y,b}\right)$$

So the BLUP of $\underline{b}$ is

$$\hat{b} = \mathbb{E}(\underline{b} | \underline{Y}) = C_{b,Y} V_Y^{-1}(Y - X\beta)$$

but typically $\hat{\beta}$ & $C_{b,Y}, V_Y$ are unknown so we use their estimates

$$\hat{b} = \hat{C}_{b,Y} \hat{V}_Y^{-1}(Y - X\hat{\beta}).$$