Hierarchial Linear Models

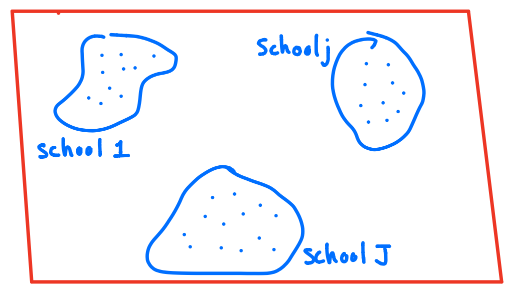
Oftentimes data y_i are observed in groups.

Motivates a covariance structure.

Ex: Longitudinal Studies

$$
\begin{array}{lcccc}
\text{Sub. } 1 & Y_{11} & Y_{21} & \cdots & Y_{n_1 1} \\
\vdots & & & & \\
\text{sub } j & Y_{1j} & Y_{2j} & \cdots & Y_{n_j j} \\
\vdots & & & & \\
\text{sub } J & Y_{1J} & Y_{2J} & \cdots & Y_{n J}
\end{array}
$$

Cluster Data structures



In these cases, $Y_j = \begin{bmatrix} y_{1j} \\ \vdots \\ y_{n_j J} \end{bmatrix}$ is such that $Var[Y_j] = \Sigma_j$

- Frequently assume some structure on the Σ_j

Pooling $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_J \end{bmatrix}$ then gives $Var(Y) = diag(\Sigma_j)_{j=1}^{J} = \bigoplus_{j=1}^{J} \Sigma_j$

Writting a model for this structure is given by

focus

$$Y_{ij} = \mu + \delta_j + e_{ij} \quad \delta_j \overset{iid}{\sim} N(0, \tau^2), \; e_{ij} \overset{iid}{\sim} \underbrace{N(0, \sigma^2)}$$

population effect — between group — with group

will change with more complex data types.

In frequentist lingo: - μ is the fixed effect

- δ, e random effects (i.e. to be marginalized later)

Usually assume $\mathbb{P}(\mu) \propto 1$ or $\mu \sim N(\mu_0, \Sigma_0)$ $\Sigma_0^{-1} = 0$ which gives a mixed effects model.

Suppose $Y_{ij} \mid \theta_j, \sigma^2 \sim N(\theta_j, \sigma^2)$ s.t. $\theta_j = \mu + \delta_j$

then $Y_j \mid \theta_j, \sigma^2 \sim N(\theta_j, \sigma^2 I_{n_j})$

and $$Y \mid \theta, \sigma^2 \sim N\left(\underbrace{\begin{bmatrix} 1_{n_1} & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ 0 & & \ddots & 1_{n_J} \end{bmatrix}}_{=X} \theta, \sigma^2 I_n\right)$$

also we have

$$\theta \mid \mu, \tau^2 \sim N(\mu 1_J, \tau^2 I_J)$$

Remember: $Y \mid \theta \sim N(X\theta, \Sigma)$, $\theta \sim N(\theta_0, \Omega)$

$$Y = \theta X + e, \quad e \sim N(0, \Sigma) \qquad \theta = \theta_0 + f \quad f \sim N(0, \Omega)$$

$$Y = X\theta_0 + Xf + e \sim N(X\theta_0, X\Omega X^T + \Sigma)$$

Thus, $$Y \mid \mu, \sigma^2, \tau^2 \sim N(X 1_J \mu, \sigma^2 I_n + \tau^2 X I_J X^T)$$

$$\stackrel{D}{=} N\left(1_n \mu, \sigma^2 I_n + \tau^2 \bigoplus_{j=1}^{J} 1_{n_j} 1_{n_j}^T\right)$$

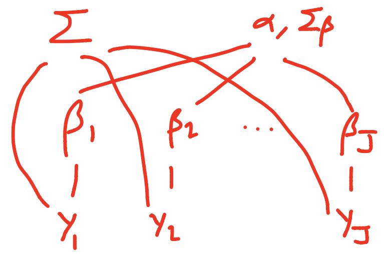
$$= \begin{bmatrix} \boxed{\Sigma} & & \\ & \ddots & \\ & & \boxed{} \end{bmatrix} \qquad \begin{bmatrix} \sigma^2+\tau^2 & \cdots & \tau^2 \\ \vdots & \ddots & \vdots \\ \tau^2 & \cdots & \sigma^2+\tau^2 \end{bmatrix}$$

<u>Rmk</u>: τ^2 captures exchangeable correlation structure.

In more generality,

$Y_j | \beta_j\, \Sigma_j \overset{ind}{\sim} N(X_j \beta_j, \Sigma_j)$ group-wise linear models.

$\beta_j | \alpha, \Sigma_\beta \overset{iid}{\sim} N((X_\beta)_j \alpha, (\Sigma_\beta)_j)$



$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_J \end{bmatrix} \Bigg| \beta, \Sigma \sim N\left(\underbrace{\begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & X_J \end{bmatrix}}_{=X} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_J \end{bmatrix}}_{=\beta}, \underbrace{\bigoplus_{j=1}^{J} \Sigma_j}_{=\Sigma_y} \right)$$

$$\beta | \alpha, \Sigma_\beta \sim N\left(\underbrace{\begin{bmatrix} X_{\beta_1} \\ \vdots \\ X_{\beta_J} \end{bmatrix}}_{=X_\beta} \alpha, \underbrace{\bigoplus_{j=1}^{J} \Sigma_{\beta_j}}_{=\Sigma_\beta} \right)$$

Mixed Effect Formulation: $\beta = X_\beta \alpha + \delta \,;\ \delta \sim N(0, \Sigma_\beta)$

$$Y = X\beta + e = \underbrace{X X_\beta}_{=Z} \alpha + X\delta + e$$

$$Y | \alpha, \delta, \Sigma_y \sim N(Z \underbrace{\alpha}_{\text{fixed}} + X \underbrace{\delta}_{\text{Random}}, \Sigma_y)$$

From here

$$Y | \alpha, \Sigma_Y, \Sigma_\beta \sim N(Z\alpha, X \Sigma_\beta X^T + \Sigma_y)$$

Still need priors for all of these

$p(\alpha) \propto 1$, $\Sigma_\beta, \Sigma_y \sim \text{Inv-Wishart}$ for each block.

Simpler models also exist (e.g. $\Sigma_y = \sigma^2 I_n$, $\Sigma_\beta = \tau^2 I$)

and have priors on $\sigma^2 \sim \text{Inv-}\chi^2(\nu_\sigma, S_\sigma)$, $\tau^2 \sim \text{Inv-}\chi^2(\nu_\tau, S_\tau)$

Model Fitting

Gibbs sampler: $(\alpha, \sigma^2, \tau^2, \beta)$

1. $\sigma^2 \mid \alpha, \beta, \tau^2, Y \sim \text{Inv-}\chi^2\left(\nu_\sigma + n, \dfrac{\nu_0 S_\sigma + RSS(\beta)}{n + \nu_0}\right)$

2. $\tau^2 \mid \alpha, \beta, \sigma^2, Y \sim \text{Inv-}\chi^2\left(\nu_\tau + \underbrace{pJ}_{\text{length of } \beta_j}, \dfrac{\nu_\tau S_\tau + (\beta - X_p\alpha)^T(\beta - X_p\alpha)}{pJ + \nu_\tau}\right)$

3. $\beta \mid \alpha, \sigma^2, \tau^2, Y$

4. $\alpha \mid \beta, \sigma^2, \tau^2, Y$