Hierarchical Linear Models

The classical linear model is

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}.$$

$$\boldsymbol{Y} \text{ is } n \times 1, \qquad \boldsymbol{X} \text{ is } n \times J,$$

$$\boldsymbol{\beta} \text{ is } J \times 1 \quad \text{and} \quad \boldsymbol{\epsilon} \text{ is } n \times 1.$$

Hierarchical linear model:

$$\text{likelihood:} \quad \boldsymbol{Y}|(\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \sim N(\boldsymbol{X\beta}, \boldsymbol{\Sigma})$$

$$\text{first-level prior:} \quad \boldsymbol{\beta}|(\boldsymbol{X}_{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \sim N(\boldsymbol{X}_{\boldsymbol{\beta}}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$$

$$\text{hyperprior:} \quad \boldsymbol{\alpha}|(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}) \sim N(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$$

Here it's assumed that $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_0$ are known. Of course, this could be generalized.

Special cases

Simple random effects model:

(1) Given $\beta_1, \ldots, \beta_n$ and σ^2, $Y_1, \ldots, Y_n$ are independent with $Y_j \sim N(\beta_j, \sigma^2)$.

(2) $\beta_1, \ldots, \beta_n$ are i.i.d. $N(\alpha, \sigma^2_\beta)$

(3) $\alpha \sim N(\alpha_0, \sigma^2_\alpha)$

In this case, $J = n$ and $\boldsymbol{X} = \boldsymbol{I}_n$.

Intragroup correlation:

In a setting as just described, suppose the observations naturally fall into J groups, and we wish to allow for correlation among the observations *within* a group, but assume that observations in *different* groups are independent.

Data in group j:

$$Y_{1j}, Y_{2j}, \ldots, Y_{n_j j}$$

Given $\beta_1, \ldots, \beta_J$ and σ^2, assume that

$$Y_{ij}, \quad i = 1, \ldots, n_j,\ j = 1, \ldots, J,$$

are mutually independent with $Y_{ij} \sim N(\beta_j, \sigma^2)$, $i = 1, \ldots, n_j$, $j = 1, \ldots, J$.

Then suppose that $\beta_1, \ldots, \beta_J$ are i.i.d. $N(\alpha, \sigma_\beta^2)$ and $\alpha \sim N(\alpha_0, \sigma_\alpha^2)$.

Consider the unconditional covariance between Y_{ij} and Y_{rs}.

$$\text{Cov}(Y_{ij}, Y_{rs}) = E(Y_{ij}Y_{rs}) - E(Y_{ij})E(Y_{rs}).$$

We have

$$\begin{aligned} E(Y_{ij}) &= E[E(Y_{ij}|\beta_j)] \\ &= E(\beta_j) \\ &= E[E(\beta_j|\alpha)] \\ &= E(\alpha) = \alpha_0, \end{aligned}$$

and

$$\begin{aligned} E(Y_{ij}Y_{rs}) &= E[E(Y_{ij}Y_{rs}|\boldsymbol{\beta})] \\ &= E(\beta_j\beta_s) \\ &= E[E(\beta_j\beta_s|\alpha)]. \end{aligned}$$

Now, if $j \neq s$, then

$$E[E(\beta_j\beta_s|\alpha)] = E(\alpha^2) = \alpha_0^2 + \sigma_\alpha^2,$$

and if $j = s$

$$E[E(\beta_j\beta_s|\alpha)] = E(\alpha^2 + \sigma_\beta^2) = \alpha_0^2 + \sigma_\alpha^2 + \sigma_\beta^2.$$

So, we have

$$\text{Cov}(Y_{ij}, Y_{rs}) = \begin{cases} \sigma_\alpha^2, & j \neq s \\ \sigma_\alpha^2 + \sigma_\beta^2, & j = s. \end{cases}$$

To find the correlation between Y_{ij} and Y_{rs}, we need the variance of each response.

$$\begin{aligned} \text{Var}(Y_{ij}) &= \text{Var}\left(E(Y_{ij}|\beta_j)\right) + E\left(\text{Var}(Y_{ij}|\beta_j)\right) \\ &= \text{Var}(\beta_j) + \sigma^2 \\ &= \text{Var}(\alpha) + \sigma_\beta^2 + \sigma^2 \\ &= \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2. \end{aligned}$$

Therefore,

$$\text{Corr}(Y_{ij}, Y_{rs}) = \begin{cases} \frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma^2}, & j \neq s \\ \frac{\sigma_\alpha^2+\sigma_\beta^2}{\sigma_\alpha^2+\sigma_\beta^2+\sigma^2}, & j = s. \end{cases}$$

Take $\sigma_\alpha = 0$ and $\sigma_\beta > 0$ to get the desired model.

Random effects with multiple populations:

Suppose the components of $\boldsymbol{\beta}$ fall into K groups. The β_js in group k are a random sample from $N(\alpha_k, \sigma^2_{\beta k})$.

Many regressions:

To illustrate the idea, consider a case where we have three different regression models.

$$\boldsymbol{Y}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1, \quad \boldsymbol{Y}_2 = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2,$$

$$\boldsymbol{Y}_3 = \boldsymbol{X}_3\boldsymbol{\beta}_3 + \boldsymbol{\epsilon}_3$$

$$\boldsymbol{Y}_i|(\boldsymbol{X}_i, \boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i) \sim N(\boldsymbol{X}_i\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i), \quad i = 1, 2, 3,$$

with Y_1, Y_2, Y_3 independent conditional on $(\boldsymbol{X}_i, \boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2, 3$.

We can see that this model falls within the framework of HLM since $\boldsymbol{Y}^T = (Y_1^T, Y_2^T, Y_3^T)$ and $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{X}_3 \end{bmatrix},$$

$\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T)$ and $\boldsymbol{\epsilon}^T = (\boldsymbol{\epsilon}_1^T, \boldsymbol{\epsilon}_2^T, \boldsymbol{\epsilon}_3^T)$.

Gibbs sampler applied to the "many regressions model"

For convenience, let's say the $\boldsymbol{\beta}_i$s have the same dimension, p. Let $\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{I}$, $i = 1, 2, 3$.

Given γ, $\boldsymbol{\alpha}$ and $\boldsymbol{D}$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, σ_1, σ_2, σ_3 are mutually independent with

$$\sigma_i \sim g(\,\cdot\,; \gamma), \quad i = 1, 2, 3,$$

and

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\alpha}, \boldsymbol{D}),$$

where $\boldsymbol{D}$ is diagonal and γ, $\boldsymbol{\alpha}$ and $\boldsymbol{D}$ are independent.

Let

$$\boldsymbol{\alpha} \sim N(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0)$$

and

$$\boldsymbol{\gamma} \sim \pi_1.$$

The diagonal elements of $\boldsymbol{D}$ are i.i.d. as inverse gamma.

On each full iteration of the Gibbs sampler, we need to generate

$$\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_1, \sigma_2, \sigma_3, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{D}).$$

The conditional posterior distribution of $\boldsymbol{\beta}_i$ given all other parameters is normal with mean vector

$$(\boldsymbol{X}_i^T \boldsymbol{X}_i + \sigma_i^2 \boldsymbol{D}^{-1})^{-1} (\boldsymbol{X}_i^T \boldsymbol{y}_i + \sigma_i^2 \boldsymbol{D}^{-1} \boldsymbol{\alpha})$$

and covariance matrix

$$\sigma_i^2 (\boldsymbol{X}_i^T \boldsymbol{X}_i + \sigma_i^2 \boldsymbol{D}^{-1})^{-1}.$$

To apply the Gibbs sampler, one needs also to derive all the other full conditional posterior distributions. In doing so, one may treat as constants all parameters besides the one whose conditional distribution is being considered.

For example, suppose we want the conditional posterior of σ_1. This distribution is proportional to

$$g(\sigma_1; \boldsymbol{\gamma})\sigma_1^{-n_1} \exp\left[-\frac{1}{2\sigma_1^2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i)^T(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i)\right]$$

since none of the other multiplicative factors in the posterior depend on σ_1.