Bayesian Inference for HLMs

Model: $$y_{ij} = x_{ij}^T \beta + z_{ij}^T a_j + \epsilon_{ij}$$

$$a_1 \ldots a_m \sim \text{IID } N(0, \Psi)$$

Special Simpler Case: $z_{ij} = x_{ij}$

In this case, $y_{ij} = x_{ij}^T (\beta + a_j)$

define $b_j = \beta + a_j$, then $E[b_j] = \beta$
$V[b_j] = \Psi$

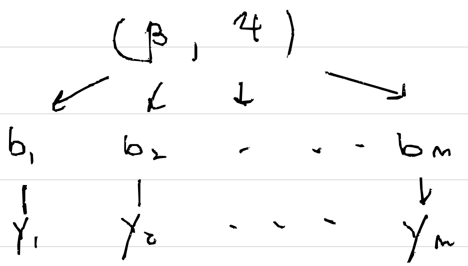
Model becomes

$$y_{ij} = x_{ij}^T b_j + \epsilon_{ij}$$
$$b_1 \ldots b_m \sim \text{IID } N(\beta, \Psi)$$

Vector form: $\underset{n_j \times 1}{\underset{\sim}{y}_j} = \underset{n_j \times p}{X_j}\, b_j + \underset{\sim}{\epsilon}_j$

(within groups) $\epsilon_1 \ldots \epsilon_m \sim \text{IID } N(0, \sigma^2 I)$

(between groups) $b_1 \ldots b_m \sim \text{IID } N(\beta, \Sigma)$



Bayesian Estimation

Perspective 1: Pop. persp.

Model: $Y_j \sim N(X_j\beta,\ X_j \Psi X_j^T + \sigma^2 I)$

Parameters: β, σ^2, Ψ

Bayes Inference: Prior $p(\beta, \sigma^2, \Psi)$

Post $p(\beta, \sigma^2, \Psi \mid Y_1 \dots Y_m)$

Perspective 2: Subpop persp.

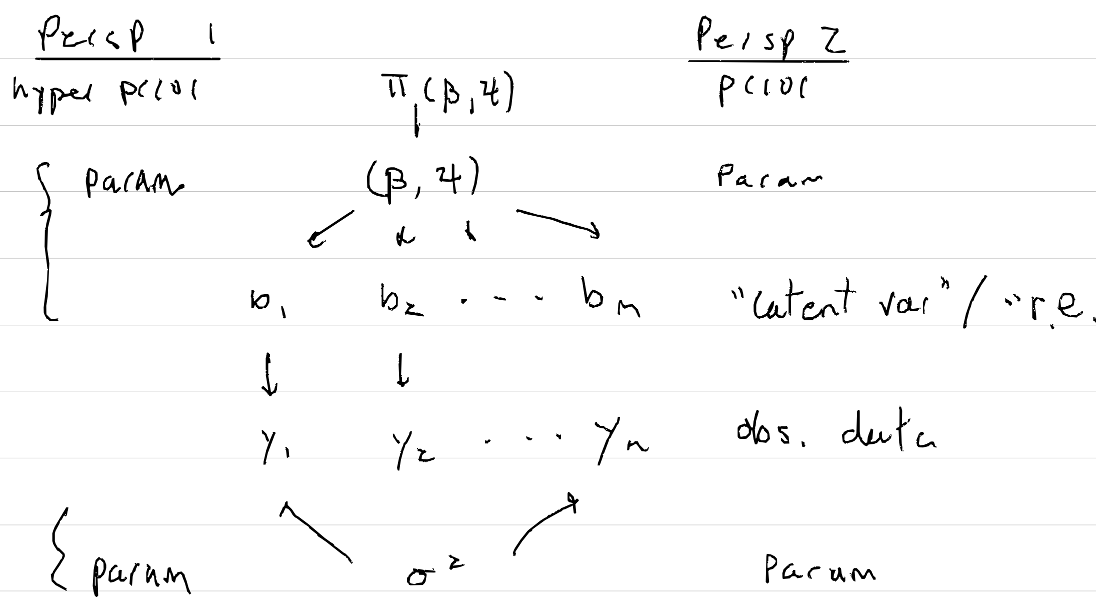
Model $Y_j \sim N(X b_j, \sigma^2 I)$

Parameters: $b_1 \dots b_m, \sigma^2$

Prior: $b_1 \dots b_m \sim_{\text{IID}} p(b_j \mid \beta, \Psi)$, $\sigma^2 \sim p(\sigma^2)$

<u>Hyperprior:</u> $(\beta, \tau) \sim p(\beta, \tau)$

<u>Posterior:</u> $p(\beta, \sigma^2, \tau, b_1 \ldots b_m \mid y_1 \ldots y_m)$



However, the same computational tools can be used for either perspective.

Posterior computation via Gibbs sampling.

Recall, a MCMC approx to $p(\theta_A, \theta_B, \theta_C | y)$ can be simulated as follows:

given $\theta_B^{(0)}, \theta_C^{(0)}, y$, iterate for $s = 0 \dots S$

1) simulate $\theta_A^{(s+1)} \sim p(\theta_A | \theta_B^{(s)}, \theta_C^{(s)}, y)$

2) simulate $\theta_B^{(s+1)} \sim p(\theta_B | \theta_A^{(s+1)}, \theta_C^{(s)}, y)$

3) simulate $\theta_C^{(s+1)} \sim p(\theta_C | \theta_A^{(s+1)} \theta_B^{(s+1)}, y)$

Then the empirical dist of

$$(\theta_A^{(1)}, \theta_B^{(1)}, \theta_C^{(1)}), (\theta_A^{(2)} \theta_B^{(2)}, \theta_C^{(2)}) \dots (\theta_A^{(S)}, \theta_B^{(S)}, \theta_C^{(S)})$$

is approximately $p(\theta_A, \theta_B, \theta_C | y)$

From the Gibbs sampler, we can obtain

- Posterior point estimates

$$\hat{\theta}_A = \frac{1}{S} \sum_{s=1}^{S} \theta_A^{(s)} \approx E[\hat{\theta}_A \mid y]$$

- Posterior CIs

quantile $(\theta_A^{(1)} \ldots \theta_A^{(S)}, .025, .975)$ is an approx 95% post. CI

and more: $P(\theta_A > \theta_B \mid y) \approx \dfrac{\#(\theta_A^{(s)} > \theta_B^{(s)})}{S}$

Gibbs sampling for the HLM

given starting vals:

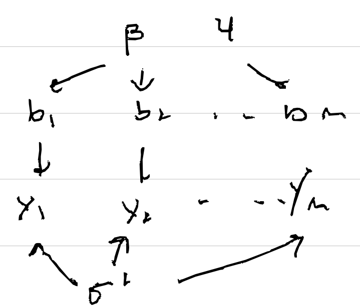
1) sim $\beta \sim p(\beta \mid y_1 \ldots y_m, b_1 \ldots b_m, \sigma^2, \tau)$
2) sim $\sigma^2 \sim p(\sigma^2 \mid y_1 \ldots y_m, b_1 \ldots b_m, \beta, \tau)$
3) sim $\tau \sim p(\tau \mid y_1 \ldots y_m, b_1 \ldots b_m, \sigma^2, \tau)$
4) for each $j = 1 \ldots m$

   sim $b_j \sim p(b_j \mid \{\text{everything but } b_j\})$

under particular priors for (β, σ^2, Ψ)
these full cond are easy to calc. + s.m.

let $\beta \sim N(\beta_0, V_0)$
$\sigma^2 \sim invgam(\nu_0/2, \nu_0\sigma_0^2/2)$
$\Psi \sim$ inv Wishart (γ_0, S_0^{-1})

<u>Full cond calculation</u>



$p(\text{params} \mid y \ldots y_m) \propto p(y \ldots y_m, \text{params})$

$p(y_1 \ldots y_m, b_1 \ldots b_m, \beta, \Psi, \sigma^2)$

$=$

$p(y \ldots y_m \mid b_1 \ldots b_m, \beta, \Psi, \sigma^2)\, p(b_1 \ldots b_m \mid \beta, \Psi, \sigma^2)$
$\times\, p(\beta, \Psi, \sigma^2)$

$$= \left\{ \prod_{j=1}^{m} p(\underset{\sim}{y_j} \mid b_j, \sigma^2) \right\} \left\{ \prod_{j=1}^{m} p(b_j \mid \beta, \Psi) \right\} p(\beta) p(\Psi) p(\sigma^2)$$

Recall: $p(\theta_A \mid \theta_B, \theta_C, y) = \dfrac{p(\theta_A, \theta_B, \theta_C, y)}{p(\theta_B, \theta_C, y)}$

$$\underset{\theta_A}{\propto} \; p(\theta_A, \theta_B, \theta_C, y)$$

So to find full cond for θ_A, just need the parts of joint that dep on θ_A.

<u>FC of $\underset{\sim}{b}_1 \ldots \underset{\sim}{b}_m$:</u>

$$p(b_j \mid \cdots) \propto p(b_j \mid \beta, \Psi) \times p(y_j \mid b_j, \sigma^2)$$

$N(\beta, \Psi)$ $\qquad$ $N_n(X_j b_j, \sigma^2 I_{n_j})$

Recall Bayes regression from 602:

If $y \mid b \sim N(Xb, \sigma^2 I)$
$\quad b \sim N(b_0, 4)$

Then $p(b \mid y) = p(b)\, p(y \mid b) / p(y)$

$\quad\quad \propto p(b)\, p(y \mid b)$

$\quad\quad\quad\quad p(b)$: $N(b_0, 4)$ $\quad\quad$ $p(y \mid b)$: $N(Xb, \sigma^2 I)$

$\quad\quad \propto \text{dnorm}(E[b \mid y], V[b \mid y])$

where

$$V[b \mid y] = [X^T X / \sigma^2 + 4^{-1}]^{-1}$$

$X^T X / \sigma^2$: data precision $\quad\quad$ 4^{-1}: prior precision

$$E[b \mid y] = [X^T X / \sigma^2 + 4^{-1}]^{-1} [X^T y / \sigma^2 + 4^{-1} b_0]$$

The full conditional dist for $\underset{\sim}{b}_j$ has exactly this form:

$$b_j \mid \cdot\cdot \sim N\left(E[b_j \mid \cdot\cdot],\ V[b_j \mid \cdot\cdot] \right)$$

where
$$V[b_j \mid \cdot\cdot] = \left(X_j^T X_j / \sigma^2 + \Psi^{-1} \right)^{-1}$$

$$E[b_j \mid \cdot\cdot] = \left(\;\;"\;\; \right)^{-1} \left(X_j^t Y_j / \sigma^2 + \Psi^{-1} \beta \right)$$

FC of σ^2:

from the joint dist, we see

$$p(\sigma^2 \mid \ldots) \propto p(\sigma^2) \prod_{j=1}^{h} \prod_{i=1}^{n_j} p(y_{ij} \mid x_{ij}, b_j, \sigma^2)$$

$p(\sigma^2)$ ↖ inv gamma

$p(y_{ij} \mid \ldots)$ ↑ $N(x_{ij}^T b_j, \sigma^2)$

10

$$p(\sigma^2 \mid \cdots) \propto p(\sigma^2) \times (\sigma^2)^{-\frac{nm}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{ij}\sum (y_{ij} - x_{ij}^T b_j)^2\right]$$

$$= p(\sigma^2) \times \left((\sigma^2)^{-\frac{nm}{2}} e^{-\frac{1}{2\sigma^2} SSR} \right)$$

$$= (\sigma^2)^{-(\frac{\nu_0}{2}+1)} e^{-\frac{\nu_0 \sigma_0^2}{2\sigma^2}} \times (\quad)$$

$$= \sigma^{2^{-\left(\frac{\nu_0 + nm}{2}+1\right)}} e^{-\frac{1}{2\sigma^2}[\nu_0 \sigma_0^2 + SSR]}$$

$$\propto \text{inv gamma}\left(\frac{\nu_0 + nm}{2}, \frac{\nu_0 \sigma_0^2 + SSR}{2}\right]$$

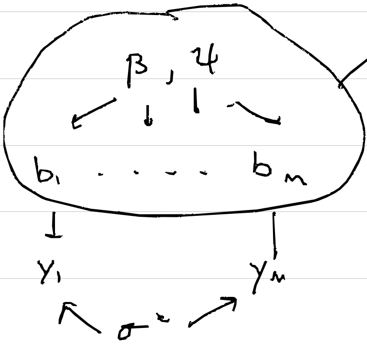
(note post is concentrated around

$$\frac{\nu_0 \sigma_0^2 + SSR}{\nu_0 + nm}$$

$$= \frac{\nu_0}{\nu_0 + nm} \sigma_0^2 + \frac{nm}{\nu_0 + nm} \frac{SSR}{nm}$$

↑ "prior est."

↑ "data" est.



Last part

$p(\beta, \Psi \mid b_1 \dots b_m \dots) \propto$

$p(b_1 \dots b_m \mid \beta, \Psi)\ p(\beta)\ p(\Psi)$

$b_1 \dots b_m \sim \text{iid } N_p(\beta, \Psi)$ ↑ ; $N(\beta_0, V_0)$ ↑ ; inv. Wishart ↑

Recall from 602:

Inference for Multivariate Normal Model

If $b_1 \dots b_m \sim \text{iid } N(\beta, \Psi)$

$\beta \sim N(\beta_0, V_0)$

$\Psi^{-1} \sim \text{Wish}(\nu_0, S_0^{-1})$

$$\Rightarrow \quad \beta \mid b_1 \dots b_m, \Psi \sim N\left([m\Psi^{-1} + V_0^{-1}]^{-1} [m\Psi^{-1}\bar{b} + V_0^{-1}\beta_0],\ [m\Psi^{-1} + V_0^{-1}]^{-1} \right)$$

$\Sigma^{-1} \sim b_1 \ldots b_m, \beta \sim \text{Wish}\left(\nu_0 + M, \left[S_0 + \sum_j (b_j - \beta)(b_j - \beta)'\right]^{-1}\right)$

So finally, we have the Gibbs sampler:
Given starting values $(\mu^{(0)}, \Sigma^{(0)}, b_1^0 \ldots b_M^0, \sigma^{2(0)})$,
iterate for $s = 0 \ldots S$

1) sim $\mu^{(s+1)} \sim N(\ )$

2) sim $\Sigma^{(s+1)} \sim$ inv wishart ( )

3) sim $\sigma^2 \sim$ inv gamma ( )

4) for $j = 1 \ldots M$
   sim $b_j \sim N(\ )$