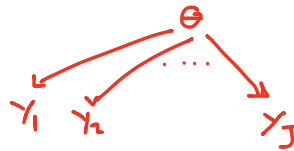
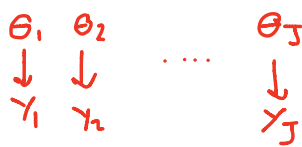
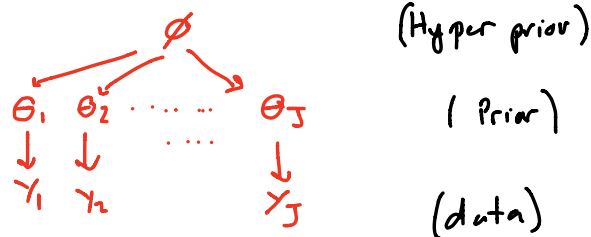


We've been considering the settings where



Is there an inbetween? Hierarchy



Two main advantages

1. Borrowing of strength  $\theta_i | y$  depends on full dataset
2. More flexibility in modeling uncertainty - inherited from  $\phi$

$$P(\theta, \phi | y) \propto P(y | \theta) P(\theta | \phi) P(\phi)$$

This representation makes it simpler to sample from posterior.

- (i) Sample  $\phi^{(s)} \sim P(\phi | y)$
- (ii) Sample  $\theta^{(s)} \sim P(\theta | \phi = \phi^{(s)}, y)$
- (iii) [For post. pred.]  $\tilde{y} \sim P(y | \theta = \theta^{(s)})$

Ex:  $y_{ij}$  = Grade of student  $i$  in school  $j$ .

$$y_{ij} | \theta_j \stackrel{\text{ind}}{\sim} N(\theta_j, \sigma_j^2) \quad \underbrace{\sigma_j^2}_{\text{known}}$$

$$\bar{y}_{.j} | \theta_j \sim N(\theta_j, \sigma_j^2 = \sigma^2/n_j)$$

Frequentist:  $\hat{\theta}_j^{(1)} = \bar{y}_{.j} \quad \hat{\theta}_j^{(2)} = \frac{1}{J} \sum_{j=1}^J \bar{y}_{.j} = \frac{\sum \bar{y}_{.j} / \sigma_j^2}{\sum 1/\sigma_j^2}$

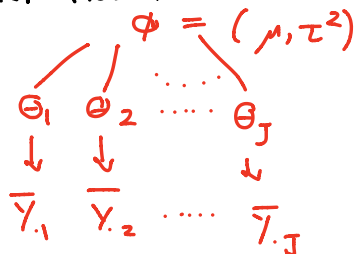
Common approach is to then test

$$H_0: \theta_1 = \theta_2 = \dots = \theta_J \begin{cases} \rightarrow \text{Use } \hat{\theta}_j^{(1)} \text{ if reject} \\ \rightarrow \text{Use } \hat{\theta}_j^{(2)} \text{ if fail to reject} \end{cases}$$

However we could want weighted means

$$\hat{\theta}_j = \lambda_j \hat{\theta}_j^{(1)} + (1 - \lambda_j) \hat{\theta}_j^{(2)}$$

Consider the hierarchical model



$$\bar{y}_{.j} | \theta_j \stackrel{\text{ind}}{\sim} N(\theta_j, \sigma_j^2)$$

$$\theta_j | \mu, \tau^2 \stackrel{\text{ind}}{\sim} N(\mu, \tau^2)$$

From this the full posterior is then

$$P(\theta, \mu | Y) \propto P(Y | \theta) P(\theta | \mu, \tau^2) P(\mu, \tau^2)$$

$$\propto \prod_{j=1}^J N(\bar{y}_j | \theta_j, \sigma_j^2) N(\theta_j | \mu, \tau^2) P(\mu | \tau^2) P(\tau^2)$$

For partial conjugacy let's set  $\mu | \tau^2 \sim N(\mu_0, \tau^2 / k_0)$

Still need to understand

(1)  $P(\theta_j | \mu, \tau^2, y)$

(2)  $P(\mu, \tau^2 | y)$

(3)  $P(\tau^2 | y)$

Rmk: Useful fact:  $X | \theta \sim N(\theta, \sigma^2)$   $\theta \sim N(\theta_0, \tau^2)$

(i)  $X \sim N(\theta_0, \sigma^2 + \tau^2)$

(ii)  $\theta | X \sim N\left(\frac{X/\sigma^2 + \theta_0/\tau^2}{1/\sigma^2 + 1/\tau^2}, \frac{1}{1/\sigma^2 + 1/\tau^2}\right)$

(1)  $P(\theta_j | \mu, \tau^2, y) \propto N(\bar{y}_j | \theta_j, \sigma_j^2) N(\theta_j | \mu, \tau^2)$

$$\propto N\left(\frac{\bar{y}_j / \sigma_j^2 + \mu / \tau^2}{1/\sigma_j^2 + 1/\tau^2}, \frac{1}{1/\sigma_j^2 + 1/\tau^2}\right)$$

$$\equiv N(\hat{\theta}_j, v_j)$$

Rmk:  $\tau^2 \rightarrow 0 : \hat{\theta}_j \rightarrow \mu$  (pooled across all)  
 $\tau^2 \rightarrow \infty : \hat{\theta}_j \rightarrow \bar{y}_{.j}$  (no pooling)

$$(2) P(\mu | \tau^2, y) = \int P(\theta, \mu | \tau^2, y) d\theta$$

$$\propto \prod_{j=1}^J \underbrace{\int N(\bar{y}_{.j} | \theta_j, \sigma_j^2) P(\theta_j | \mu, \tau^2) d\theta_j}_{\bar{y}_{.j} | \mu, \tau^2 \sim N(\mu, \sigma_j^2 + \tau^2)} P(\mu | \tau^2)$$

$$\propto \prod_{j=1}^J N(\bar{y}_{.j} | \mu, \sigma_j^2 + \tau^2) N(\mu | \mu_0, \tau^2/k_0)$$

note

$$\cdot \bar{y}_{.1} | \mu, \tau^2 \sim N(\mu, \sigma_1^2 + \tau^2)$$

$$\cdot \mu | \tau^2 \sim N(\mu_0, \tau^2/k_0)$$

$$\mu | \bar{y}_{.1}, \tau^2 \sim N\left( \frac{\bar{y}_{.1} / \sigma_1^2 + \tau^2 + \mu_0 / \tau^2 / k_0}{\frac{1}{\sigma_1^2 + \tau^2} + \frac{k_0}{\tau^2}}, \frac{1}{\frac{1}{\sigma_1^2 + \tau^2} + \frac{k_0}{\tau^2}} \right)$$

Iteratively updating

$$\mu | \tau^2, y \sim N \left( \hat{\mu} = \frac{\sum_j (\bar{y}_j / \sigma_j^2 + \tau^2) + \frac{\mu_0 k_0}{\tau^2}}{\sum_j \frac{1}{\sigma_j^2 + \tau^2} + \frac{k_0}{\tau^2}}, \hat{\tau}^2 = \frac{1}{\sum_j \frac{1}{\sigma_j^2 + \tau^2} + \frac{k_0}{\tau^2}} \right)$$

$$(3) P(\tau^2 | y) = \int P(\mu, \tau^2 | y) d\mu$$

$$\propto \int_j \pi N(\bar{y}_j | \mu, \sigma_j^2 + \tau^2) N(\mu | \mu_0, \tau^2 / k_0) P(\tau^2) d\mu$$

$$\propto \hat{\tau}^{1/2} \prod_{j=1}^J (\sigma_j^2 + \tau^2)^{-1/2} \exp \left\{ -\frac{1}{2} \left[ \frac{\sum_j (\bar{y}_j - \hat{\mu})^2}{\sigma_j^2 + \tau^2} + \frac{(\hat{\mu} - \mu_0)^2}{\tau^2 / k_0} \right] \right\} P(\tau^2)$$

→ Need numerical methods here.