STAT 578: Advanced Bayesian Modeling

Week 4 – Lesson 3

Hierarchical Models: More Details

Fall 2019

Generalizing the Normal Hierarchical Model

Grouped Normal Samples

Let

$$y_{ij} \quad = \quad i\text{th observation in group } j \qquad\qquad i = 1, \ldots, n_j \qquad j = 1, \ldots, J$$

so data are

$$y \quad = \quad (y_{11}, \ldots, y_{n_1 1}, \quad y_{12}, \ldots, y_{n_2 2}, \quad \ldots, \quad y_{1J}, \ldots, y_{n_J J})$$

Assume all independent, and assume identically distributed within groups:

$$y_{ij} \mid \theta_j \quad \sim \quad \mathrm{N}(\theta_j, \sigma^2)$$

Note: We assume σ^2 is known and does not depend on group.

Remark: Typical for a completely randomized experiment with J treatments

Likelihood turns out to depend only on group averages:

$$\bar{y}_{\cdot j} \quad = \quad \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \qquad\qquad j = 1, \ldots, J$$

Their sampling variances are

$$\sigma_j^2 \quad \equiv \quad \mathrm{var}(\bar{y}_{\cdot j} \mid \theta_j) \quad = \quad \sigma^2 / n_j$$

(assumed known)

Thus, sampling model reduces to

$$\bar{y}_{\cdot j} \mid \theta_j \quad \sim \quad \text{indep. N}(\theta_j, \sigma_j^2)$$

(Normality of averages might be justified by central limit theorem, even if y_{ij}s are not exactly normally distributed.)

See BDA3, Sec. 5.4, for classical analysis.

Hierarchical Prior

As in 2016 polls example, let

$$\theta_1, \ldots, \theta_J \mid \mu, \tau \quad \sim \quad \text{iid N}(\mu, \tau^2)$$

(as in a *one-way random effects* model – later).

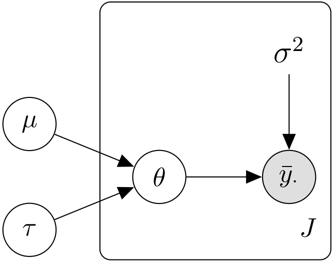
Consider product hyperprior for μ and τ:

$$p(\mu, \tau) \quad = \quad p(\mu)\, p(\tau)$$

Usually feasible to give μ an improper flat prior:

$$p(\mu) \quad \propto \quad 1$$

DAG Model



Note: Because it is on the plate, the σ^2 node is the constant *vector* $(\sigma_1^2, \ldots, \sigma_J^2)$, and not the scalar σ^2.

Posterior

See BDA3, Sec. 5.4, for derivation of posterior marginal density of τ:

$$p(\tau \mid y) \;\; \propto \;\; p(\tau)\, V_\mu^{1/2} \prod_{j=1}^{J} (\sigma_j^2 + \tau^2)^{-1/2} \, \exp\!\left(-\frac{(\bar{y}_{\cdot j} - \hat{\mu})^2}{2(\sigma_j^2 + \tau^2)} \right) \qquad \tau > 0$$

where

$$V_\mu^{-1} \;\; = \;\; \sum_{j=1}^{J} \frac{1}{\sigma_j^2 + \tau^2} \qquad\qquad \hat{\mu} \;\; = \;\; \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^2 + \tau^2}\, \bar{y}_{\cdot j}}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2 + \tau^2}}$$

BDA3 recommends (for $J > 2$)

$$p(\tau) \;\; \propto \;\; 1 \qquad \tau > 0$$

BDA3, Sec. 5.4, also shows

$$\mu \mid \tau, y \quad \sim \quad \mathrm{N}(\hat{\mu}, V_\mu)$$

and

$$\theta_1, \ldots, \theta_J \mid \mu, \tau, y \quad \sim \quad \text{indep. } \mathrm{N}(\hat{\theta}_j, V_j)$$

where

$$\hat{\theta}_j \quad = \quad \frac{\frac{1}{\sigma_j^2}\bar{y}_{\cdot j} + \frac{1}{\tau^2}\mu}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}} \qquad\qquad V_j \quad = \quad \frac{1}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}}$$

Suggests how to simulate from posterior:

1. Draw τ_{sim} from $p(\tau \mid y)$ using an approximate numerical method.
2. Draw μ_{sim} from $p(\mu \mid \tau_{\text{sim}}, y)$.
3. Draw $\theta_{1\text{sim}}, \ldots, \theta_{J\text{sim}}$ independently from

$$p(\theta_1 \mid \mu_{\text{sim}}, \tau_{\text{sim}}, y), \quad \ldots, \quad p(\theta_J \mid \mu_{\text{sim}}, \tau_{\text{sim}}, y)$$

Extensions:

- For unknown σ^2, provide a prior (such as scaled inverse chi-square).
- Allow sampling variances of original observations to differ by group, perhaps with a prior that makes them exchangeable (BDA3, Sec. 14.7).