A Hierarchical Normal Model for Data from Several Groups

- From the above, we see the full conditionals for ϕ and τ^2 satisfy:

$$p(\phi|\mu_1,\ldots,\mu_m,\tau^2,\sigma^2,\mathbf{y}_1,\ldots,\mathbf{y}_m) \propto p(\phi)\prod_{j=1}^{m} p(\mu_j|\phi,\tau^2)$$

$$p(\tau^2|\mu_1,\ldots,\mu_m,\phi,\sigma^2,\mathbf{y}_1,\ldots,\mathbf{y}_m) \propto p(\tau^2)\prod_{j=1}^{m} p(\mu_j|\phi,\tau^2)$$

A Hierarchical Normal Model for Data from Several Groups

- It can be shown that the full conditional for ϕ is normal and the full conditional for τ^2 is inverse gamma. Specifically:

$$\phi|\mu_1, \ldots, \mu_m, \tau^2 \sim N\left(\frac{\frac{m\bar{\mu}}{\tau^2} + \frac{\phi_0}{\gamma^2}}{\frac{m}{\tau^2} + \frac{1}{\gamma^2}}, \frac{1}{\frac{m}{\tau^2} + \frac{1}{\gamma^2}}\right)$$

and

$$\frac{1}{\tau^2}|\mu_1, \ldots, \mu_m, \phi \sim \text{ gamma}\left(\frac{\eta_1 + m}{2}, \frac{\eta_1\eta_2 + \sum_j (\mu_j - \phi)^2}{2}\right)$$

- Similarly, the full conditional for any μ_j satisfies:

$$p(\mu_j|\phi, \tau^2, \sigma^2, \mathbf{y}_1, \ldots, \mathbf{y}_m) \propto p(\mu_j|\phi, \tau^2) \prod_{i=1}^{n_j} p(y_{ij}|\mu_j, \sigma^2)$$

- **Conditional** on ϕ, τ^2, σ^2, μ_j is independent of the other μ's **and** of the data in the **other** groups.

A Hierarchical Normal Model for Data from Several Groups

- Then it can be shown:

$$\mu_j | \mathbf{y}_j, \sigma^2, \tau^2, \phi \sim N\left(\frac{\frac{n_j \bar{y}_j}{\sigma^2} + \frac{\phi}{\tau^2}}{\frac{n_j}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n_j}{\sigma^2} + \frac{1}{\tau^2}} \right)$$

- Similarly, the full conditional for σ^2 is conditionally independent of $\{\phi, \tau^2\}$, given $\{\mathbf{y}_1, \dots, \mathbf{y}_m, \mu_1, \dots, \mu_m\}$:

$$p(\sigma^2 | \mu_1, \dots, \mu_m, \mathbf{y}_1, \dots, \mathbf{y}_m) \propto p(\sigma^2) \prod_{j=1}^{m} \prod_{i=1}^{n_j} p(y_{ij} | \mu_j, \sigma^2)$$

$$\propto (\sigma^2)^{-\nu_1/2+1} e^{-\frac{\nu_1 \nu_2}{2\sigma^2}} (\sigma^2)^{-\frac{\sum n_j}{2}} e^{-\frac{1}{2\sigma^2} \sum_j \sum_i (y_{ij} - \mu_j)^2}$$

Collecting terms, this is an **inverse gamma**, and:

$$\frac{1}{\sigma^2} | \boldsymbol{\mu}, \mathbf{y}_1, \dots, \mathbf{y}_m \sim \text{ gamma}\left(\frac{1}{2}\left(\nu_1 + \sum_{j=1}^{m} n_j\right), \frac{1}{2}\left[\nu_1 \nu_2 + \sum_j \sum_i (y_{ij} - \mu_j)^2\right] \right)$$

Example: Data from Several Groups

- **Example 3** (Math scores): The data are math scores for 10th-grade students from $m = 100$ different urban high schools.
- The sample sizes $n_1, \ldots, n_m$ are quite different across schools.
- The nationwide **total** (between plus within) variance for this test is 100, and the **nationwide** mean is 50.
- We choose the priors

$$\begin{aligned} 1/\sigma^2 &\sim \text{ gamma}(1/2, 100/2) \\ 1/\tau^2 &\sim \text{ gamma}(1/2, 100/2) \\ \phi &\sim N(50, 25) \end{aligned}$$

- We can then repeatedly cycle through $\phi^{[s]}, \tau^{2[s]}, \sigma^{2[s]}, \mu_1^{[s]}, \ldots, \mu_m^{[s]}$ (for $s = 1, \ldots, S$) using their full conditionals and the Gibbs sampler.
- See `R` example with real schools data.

Bayesian Estimation and Shrinkage

- The posterior mean of μ_j (given ϕ, τ^2, σ^2 and $\mathbf{y}_j$) is

$$E[\mu_j|\mathbf{y}_j, \phi, \tau^2, \sigma^2] = \frac{\frac{n_j \bar{y}_j}{\sigma^2} + \frac{\phi}{\tau^2}}{\frac{n_j}{\sigma^2} + \frac{1}{\tau^2}}$$
$$= \left(\frac{n_j/\sigma^2}{n_j/\sigma^2 + 1/\tau^2}\right)\bar{y}_j + \left(\frac{1/\tau^2}{n_j/\sigma^2 + 1/\tau^2}\right)\phi$$

- So the posterior mean of μ_j is pulled **away from** $\bar{y}_j$ and **toward** ϕ, the **mean** of the distribution of **all** the μ_j's.
- This is called **shrinkage**.
- How much is each μ_j shrunk? It depends on n_j.
- For schools with a large sample size (large n_j), shrinkage is minimal.
- For schools with a few students (small n_j), shrinkage is substantial.

Bayesian Estimation and Shrinkage

- **Example 1**: (Schools 82 vs. 46)

$$\text{Data: } \bar{y}_{82} = 38.76, \;\; n_{82} = 5, \;\; \hat{\mu}_{82} = 42.53$$
$$\bar{y}_{46} = 40.18, \;\; n_{46} = 21, \;\; \hat{\mu}_{46} = 41.31$$

- Note $\hat{\phi} = 48.12$.
- For school 82, we have substantial shrinkage toward $\hat{\phi}$.
- For school 46, we have less shrinkage toward $\hat{\phi}$.
- We might then rank school 82 ahead of school 46, because we doubt that $\bar{y}_{82}$ is a good estimate of school 82's true mean, being based on only 5 students.

Bayesian Estimation and Shrinkage

- **Example 2**: (Schools 67 and 51)

$$\text{Data: } \bar{y}_{67} = 65.02, \quad n_{67} = 4, \quad \hat{\mu}_{67} = 57.14$$
$$\bar{y}_{51} = 64.37, \quad n_{51} = 19, \quad \hat{\mu}_{51} = 61.84$$

- School 67 is shrunk down more toward $\hat{\phi}$.
- We expect school 51 to have a higher true mean even though its sample mean was lower.
- **Intuition**: Whom would you trust more to make a free throw, someone who has made 4 out of 4, or someone who has made 96 out of 100?