

STA 360/602L: MODULE 5.4

HIERARCHICAL NORMAL MODELING OF MEANS AND VARIANCES

DR. OLANREWAJU MICHAEL AKANDE

HIERARCHICAL MODELING OF MEANS RECAP

- We've looked at the hierarchical normal model of the form

$$\begin{aligned} y_{ij} | \theta_j, \sigma^2 &\sim \mathcal{N}(\theta_j, \sigma^2); \quad i = 1, \dots, n_j \\ \theta_j | \mu, \tau^2 &\sim \mathcal{N}(\mu, \tau^2); \quad j = 1, \dots, J. \end{aligned}$$

- The model gives us an extra hierarchy through the prior on the means, leading to sharing of information across the groups, when estimating the group-specific means.
- We set the variance, σ^2 , as the same for all groups, to simplify posterior inference.
- We will relax that assumption in this module.

HIERARCHICAL MODELING OF MEANS AND VARIANCES

- Often researchers emphasize differences in means. However, variances can be very important.
- If we think means vary across groups, why shouldn't we worry about variances also varying across groups?
- In that case, we have the model

$$\begin{aligned} y_{ij} | \theta_j, \sigma_j^2 &\sim \mathcal{N}(\theta_j, \sigma_j^2); \quad i = 1, \dots, n_j \\ \theta_j | \mu, \tau^2 &\sim \mathcal{N}(\mu, \tau^2); \quad j = 1, \dots, J, \end{aligned}$$

- However, now we also need a model on all the σ_j^2 's that lets us borrow information about across groups.

POSTERIOR INFERENCE

- Now we need to find a semi-conjugate distribution for the σ_j^2 's. Before, with one σ^2 , we had

$$\pi(\sigma^2) = \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right),$$

which was nicely semi-conjugate.

- That suggests that maybe we should start with.

$$\sigma_1^2, \dots, \sigma_J^2 | \nu_0, \sigma_0^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

- However, if we just fix the hyperparameters ν_0 and σ_0^2 in advance, the prior on the σ_j^2 's does not allow borrowing of information across other values of σ_j^2 , to aid in estimation.
- Thus, we actually need to treat ν_0 and σ_0^2 as parameters in a hierarchical model for both means and variances.

POSTERIOR INFERENCE

- Therefore, the full posterior is now:

$$\begin{aligned}\pi(\theta_1, \dots, \theta_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, \sigma_0^2 | Y) &\propto p(y | \theta_1, \dots, \theta_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, \sigma_0^2) \\ &\quad \times p(\theta_1, \dots, \theta_J | \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, \sigma_0^2) \\ &\quad \times p(\sigma_1^2, \dots, \sigma_J^2 | \mu, \tau^2, \nu_0, \sigma_0^2) \\ &\quad \times \pi(\mu, \tau^2, \nu_0, \sigma_0^2) \\ &= p(y | \theta_1, \dots, \theta_J, \sigma^2) \\ &\quad \times p(\theta_1, \dots, \theta_J | \mu, \tau^2) \\ &\quad \times p(\sigma_1^2, \dots, \sigma_J^2 | \nu_0, \sigma_0^2) \\ &\quad \times \pi(\mu) \cdot \pi(\tau^2) \cdot \pi(\nu_0) \cdot \pi(\sigma_0^2) \\ &= \left\{ \prod_{j=1}^J \prod_{i=1}^{n_j} p(y_{ij} | \theta_j, \sigma_j^2) \right\} \\ &\quad \times \left\{ \prod_{j=1}^J p(\theta_j | \mu, \tau^2) \right\} \\ &\quad \times \left\{ \prod_{j=1}^J p(\sigma_j^2 | \nu_0, \sigma_0^2) \right\} \\ &\quad \times \pi(\mu) \cdot \pi(\tau^2) \cdot \pi(\nu_0) \cdot \pi(\sigma_0^2)\end{aligned}$$

FULL CONDITIONALS

- Notice that this new factorization won't affect the full conditionals for μ and τ^2 from before, since those have nothing to do with all the new σ_j^2 's.
- That is,

$$\pi(\mu | \dots) = \mathcal{N}(\mu_n, \gamma_n^2) \quad \text{where}$$

$$\gamma_n^2 = \frac{1}{\frac{J}{\tau^2} + \frac{1}{\gamma_0^2}}; \quad \mu_n = \gamma_n^2 \left[\frac{J}{\tau^2} \bar{\theta} + \frac{1}{\gamma_0^2} \mu_0 \right],$$

and

$$\pi(\tau^2 | \dots) = \mathcal{IG} \left(\frac{\eta_n}{2}, \frac{\eta_n \tau_n^2}{2} \right) \quad \text{where}$$

$$\eta_n = \eta_0 + J; \quad \tau_n^2 = \frac{1}{\eta_n} \left[\eta_0 \tau_0^2 + \sum_{j=1}^J (\theta_j - \mu)^2 \right].$$

FULL CONDITIONALS

- The full conditional for each θ_j , we have

$$\pi(\theta_j | \theta_{-j}, \mu, \sigma_1^2, \dots, \sigma_J^2, \tau^2, Y) \propto \left\{ \prod_{i=1}^{n_j} p(y_{ij} | \theta_j, \sigma_j^2) \right\} \cdot p(\theta_j | \mu, \tau^2)$$

with the only change from before being σ_j^2 .

- That is, those terms still include a normal density for θ_j multiplied by a product of normals in which θ_j is the mean, again mirroring the previous case, so you can show that

$$\pi(\theta_j | \theta_{-j}, \mu, \sigma_1^2, \dots, \sigma_J^2, \tau^2, Y) = \mathcal{N}(\mu_j^*, \tau_j^*) \quad \text{where}$$

$$\tau_j^* = \frac{1}{\frac{n_j}{\sigma_j^2} + \frac{1}{\tau^2}}; \quad \mu_j^* = \tau_j^* \left[\frac{n_j}{\sigma_j^2} \bar{y}_j + \frac{1}{\tau^2} \mu \right]$$

HOW ABOUT WITHIN-GROUP VARIANCES?

- Before we get to the choice of the priors for ν_0 and σ_0^2 , we have enough to derive the full conditional for each σ_j^2 . This actually takes a similar form to what we had before we indexed by j , that is,

$$\pi(\sigma_j^2 | \sigma_{-j}^2, \theta_1, \dots, \theta_J, \mu, \tau^2, \nu_0, \sigma_0^2, Y) \propto \left\{ \prod_{i=1}^{n_j} p(y_{ij} | \theta_j, \sigma_j^2) \right\} \cdot \pi(\sigma_j^2 | \nu_0, \sigma_0^2)$$

- This still looks like what we had before, that is, products of normals and one inverse-gamma, so that

$$\pi(\sigma_j^2 | \sigma_{-j}^2, \theta_1, \dots, \theta_J, \mu, \tau^2, \nu_0, \sigma_0^2, Y) = \mathcal{IG} \left(\frac{\nu_j^*}{2}, \frac{\nu_j^* \sigma_j^{2(*)}}{2} \right) \quad \text{where}$$

$$\nu_j^* = \nu_0 + n_j; \quad \sigma_j^{2(*)} = \frac{1}{\nu_j^*} \left[\nu_0 \sigma_0^2 + \sum_{i=1}^{n_j} (y_{ij} - \theta_j)^2 \right].$$

REMAINING HYPER-PRIORS

- Now we can get back to priors for ν_0 and σ_0^2 . Turns out that a semi-conjugate prior for σ_0^2 (you have seen this on the homework) is a gamma distribution. That is, if we set

$$\pi(\sigma_0^2) = \mathcal{Ga}(a, b),$$

then,

$$\begin{aligned}\pi(\sigma_0^2 | \theta_1, \dots, \theta_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, Y) &\propto \left\{ \prod_{j=1}^J p(\sigma_j^2 | \nu_0, \sigma_0^2) \right\} \cdot \pi(\sigma_0^2) \\ &\propto \mathcal{IG}\left(\sigma_j^2; \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right) \cdot \mathcal{Ga}(\sigma_0^2; a, b)\end{aligned}$$

- Recall that

- $\mathcal{Ga}(y; a, b) \equiv \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}, \text{ and}$

- $\mathcal{IG}(y; a, b) \equiv \frac{b^a}{\Gamma(a)} y^{-(a+1)} e^{-\frac{b}{y}}.$

REMAINING HYPER-PRIORS

- So $\pi(\sigma_0^2 | \theta_1, \dots, \theta_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \nu_0, Y)$

$$\begin{aligned}
 & \propto \left\{ \prod_{j=1}^J p(\sigma_j^2 | \nu_0, \sigma_0^2) \right\} \cdot \pi(\sigma_0^2) \\
 & \propto \mathcal{IG} \left(\sigma_j^2; \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2} \right) \cdot \mathcal{Ga}(\sigma_0^2; a, b) \\
 & = \prod_{j=1}^J \frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\left(\frac{\nu_0}{2} \right)}}{\Gamma \left(\frac{\nu_0}{2} \right)} (\sigma_j^2)^{-\left(\frac{\nu_0}{2} + 1 \right)} e^{-\frac{\nu_0 \sigma_0^2}{2(\sigma_j^2)}} \cdot \left[\frac{b^a}{\Gamma(a)} (\sigma_0^2)^{a-1} e^{-b\sigma_0^2} \right] \\
 & \propto \prod_{j=1}^J (\sigma_0^2)^{\left(\frac{\nu_0}{2} \right)} e^{-\frac{\nu_0 \sigma_0^2}{2(\sigma_j^2)}} \cdot \left[(\sigma_0^2)^{a-1} e^{-b\sigma_0^2} \right] \\
 & \propto (\sigma_0^2)^{\left(\frac{J\nu_0}{2} \right)} e^{-\sigma_0^2 \left[\frac{\nu_0}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right]} \cdot \left[(\sigma_0^2)^{a-1} e^{-b\sigma_0^2} \right]
 \end{aligned}$$

REMAINING HYPER-PRIORS

- That is, the full conditional is

$$\begin{aligned}\pi(\sigma_0^2 | \dots) &\propto \left[(\sigma_0^2)^{\left(\frac{J\nu_0}{2}\right)} e^{-\sigma_0^2 \left[\frac{\nu_0}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2}\right]} \right] \cdot \left[(\sigma_0^2)^{a-1} e^{-b\sigma_0^2} \right] \\ &\propto \left[(\sigma_0^2)^{\left(a + \frac{J\nu_0}{2} - 1\right)} e^{-\sigma_0^2 \left[b + \frac{\nu_0}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2}\right]} \right] \\ &\equiv \mathcal{Ga}(\sigma_0^2; a_n, b_n),\end{aligned}$$

where

$$a_n = a + \frac{J\nu_0}{2}; \quad b_n = b + \frac{\nu_0}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2}.$$

REMAINING HYPER-PRIORS

- Ok that leaves us with one parameter to go, i.e., ν_0 . Turns out there is no simple conjugate/semi-conjugate prior for ν_0 .
- Common practice is to restrict ν_0 to be an integer (which makes sense when we think of it as being degrees of freedom, which also means it cannot be zero). With the restriction, we need a discrete distribution as the prior with support on $\nu_0 = 1, 2, 3, \dots$
- **Question: Can we use either a binomial or a Poisson prior on for ν_0 ?**
- A popular choice is the geometric distribution with pmf $p(\nu_0) = (1 - p)^{\nu_0 - 1} p$.
- However, we will rewrite the kernel as $\pi(\nu_0) \propto e^{-\alpha \nu_0}$. How did we get here from the geometric pmf and what is α ?

FINAL FULL CONDITIONAL

- With this prior, $\pi(\nu_0 | \theta_1, \dots, \theta_J, \sigma_1^2, \dots, \sigma_J^2, \mu, \tau^2, \sigma_0^2, Y)$

$$\begin{aligned}
 &\propto \left\{ \prod_{j=1}^J p(\sigma_j^2 | \nu_0, \sigma_0^2) \right\} \cdot \pi(\nu_0) \\
 &\propto \mathcal{IG} \left(\sigma_j^2; \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2} \right) \cdot e^{-\alpha \nu_0} \\
 &= \left[\prod_{j=1}^J \frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\left(\frac{\nu_0}{2} \right)}}{\Gamma \left(\frac{\nu_0}{2} \right)} (\sigma_j^2)^{-\left(\frac{\nu_0}{2} + 1 \right)} e^{-\frac{\nu_0 \sigma_0^2}{2(\sigma_j^2)}} \right] \cdot e^{-\alpha \nu_0} \\
 &\propto \left[\left(\frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\left(\frac{\nu_0}{2} \right)}}{\Gamma \left(\frac{\nu_0}{2} \right)} \right)^J \cdot \left(\prod_{j=1}^J \frac{1}{\sigma_j^2} \right)^{\left(\frac{\nu_0}{2} + 1 \right)} \cdot e^{-\nu_0 \left[\frac{\sigma_0^2}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right]} \right] \cdot e^{-\alpha \nu_0}
 \end{aligned}$$

FINAL FULL CONDITIONAL

- That is, the full conditional is

$$\pi(\nu_0 | \dots) \propto \left[\left(\frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right) \left(\frac{\nu_0}{2} \right)}{\Gamma\left(\frac{\nu_0}{2}\right)} \right)^J \cdot \left(\prod_{j=1}^J \frac{1}{\sigma_j^2} \right) \left(\frac{\nu_0}{2} + 1 \right) \cdot e^{-\nu_0 \left[\alpha + \frac{\sigma_0^2}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right]} \right],$$

which is not a known kernel and is thus unnormalized (i.e., does not integrate to 1 in its current form).

- This sure looks like a lot, but it will be relatively easy to compute in R.
- Now, technically, the support is $\nu_0 = 1, 2, 3, \dots$, however, we can compute this to compute the unnormalized distribution across a grid of ν_0 values, say, $\nu_0 = 1, 2, 3, \dots, K$ for some large K , and then sample.

FINAL FULL CONDITIONAL

- One more thing, computing these probabilities on the raw scale can be problematic particularly because of the product inside. Good idea to transform to the log scale instead.
- That is,

$$\pi(\nu_0 | \dots) \propto \left[\left(\frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\left(\frac{\nu_0}{2} \right)^J}}{\Gamma\left(\frac{\nu_0}{2}\right)} \right) \cdot \left(\prod_{j=1}^J \frac{1}{\sigma_j^2} \right)^{\left(\frac{\nu_0}{2} - 1 \right)} \cdot e^{-\nu_0 \left[\alpha + \frac{\sigma_0^2}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right]} \right]$$

$$\begin{aligned} \Rightarrow \ln \pi(\nu_0 | \dots) &\propto \left(\frac{J\nu_0}{2} \right) \ln \left(\frac{\nu_0 \sigma_0^2}{2} \right) - J \ln \left[\Gamma\left(\frac{\nu_0}{2}\right) \right] \\ &+ \left(\frac{\nu_0}{2} + 1 \right) \left(\sum_{j=1}^J \ln \left[\frac{1}{\sigma_j^2} \right] \right) \\ &- \nu_0 \left[\alpha + \frac{\sigma_0^2}{2} \sum_{j=1}^J \frac{1}{\sigma_j^2} \right] \end{aligned}$$

FULL MODEL

As a recap, the final model is therefore:

$$y_{ij}|\theta_j, \sigma^2 \sim \mathcal{N}(\theta_j, \sigma_j^2); \quad i = 1, \dots, n_j; \quad j = 1, \dots, J$$

$$\theta_j|\mu, \tau^2 \sim \mathcal{N}(\mu, \tau^2); \quad j = 1, \dots, J$$

$$\sigma_1^2, \dots, \sigma_J^2|\nu_0, \sigma_0^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right); \quad j = 1, \dots, J$$

$$\mu \sim \mathcal{N}(\mu_0, \gamma_0^2)$$

$$\tau^2 \sim \mathcal{IG}\left(\frac{\eta_0}{2}, \frac{\eta_0 \tau_0^2}{2}\right).$$

$$\pi(\nu_0) \propto e^{-\alpha \nu_0}$$

$$\sigma_0^2 \sim \mathcal{Ga}(a, b).$$

WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!