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

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

- 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

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

- 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.

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

- 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:

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

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,

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

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

and

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

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

Full conditionals

- The full conditional for each θ_j, we have

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

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

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

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

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,

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

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

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

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

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

$$\pi(\sigma_0^2) = \mathcal{G}a\left(a, b\right),$$

then,

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

- Recall that
  - $\mathcal{G}a(y;a,b) \equiv \dfrac{b^a}{\Gamma(a)} y^{a-1} e^{-by}$, and
  - $\mathcal{IG}(y;a,b) \equiv \dfrac{b^a}{\Gamma(a)} y^{-(a+1)} e^{-\frac{b}{y}}$.

Remaining hyper-priors

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

$$\propto \left\{\prod_{j=1}^{J} p(\sigma_j^2|\nu_0,\sigma_0^2)\right\}\cdot\pi(\sigma_0^2)$$

$$\propto \prod_{j=1}^{J} \mathcal{IG}\left(\sigma_j^2;\frac{\nu_0}{2},\frac{\nu_0\sigma_0^2}{2}\right)\cdot\mathcal{G}a\left(\sigma_0^2;a,b\right)$$

$$= \left[\prod_{j=1}^{J}\frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\left(\frac{\nu_0}{2}\right)}}{\Gamma\left(\frac{\nu_0}{2}\right)}(\sigma_j^2)^{-\left(\frac{\nu_0}{2}+1\right)}e^{-\frac{\nu_0\sigma_0^2}{2(\sigma_j^2)}}\right]\cdot\left[\frac{b^a}{\Gamma(a)}(\sigma_0^2)^{a-1}e^{-b\sigma_0^2}\right]$$

$$\propto \left[\prod_{j=1}^{J}(\sigma_0^2)^{\left(\frac{\nu_0}{2}\right)}e^{-\frac{\nu_0\sigma_0^2}{2(\sigma_j^2)}}\right]\cdot\left[(\sigma_0^2)^{a-1}e^{-b\sigma_0^2}\right]$$

$$\propto \left[(\sigma_0^2)^{\left(\frac{J\nu_0}{2}\right)}e^{-\sigma_0^2\left[\frac{\nu_0}{2}\sum_{j=1}^{J}\frac{1}{\sigma_j^2}\right]}\right]\cdot\left[(\sigma_0^2)^{a-1}e^{-b\sigma_0^2}\right]$$

Remaining hyper-priors

- That is, the full conditional is

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

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 $\nu_0 = 1, 2, 3, \ldots$.

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

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

Final full conditional

- That is, the full conditional is

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

  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 $\nu_0 = 1, 2, 3, \ldots$, however, we can compute this to compute the unnormalized distribution across a grid of ν_0 values, say, $\nu_0 = 1, 2, 3, \ldots, K$ 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,

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

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

Full model

As a recap, the final model is therefore:

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

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

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

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

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

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

$$\sigma_0^2 \sim \mathcal{G}a\left(a, b\right).$$

What's next?

Move on to the readings for the next module!