STA 610L: Module 2.9

Random effects ANOVA (Bayesian estimation I)

Dr. Olanrewaju Michael Akande

STA 610L

Introduction

Bayesian estimation is often the approach of choice for fitting hierarchical models.

Two major advantages include

- estimation and computation, particularly in complex, highly structured, or generalized linear models; and
- straightforward uncertainty quantification.

Hierarchical normal model

Recall our data model:

$$y_{ij} = \mu_j + \varepsilon_{ij}$$

where

- $\mu_j = \mu + \alpha_j$, and
- $\alpha_j \overset{iid}{\sim} N\left(0, \tau^2\right) \perp \varepsilon_{ij} \overset{iid}{\sim} N\left(0, \sigma^2\right)$,

  so that $\mu_j \overset{iid}{\sim} N\left(\mu, \tau^2\right)$.

In addition to this data model, we will also need to specify a prior distribution for (μ, τ^2, σ^2), which we will write as $p(\theta) = p(\mu, \tau^2, \sigma^2)$.

Note: this module should be a recap of the derivations you should have covered in STA 360/601/602. Some of the notations might be different so pay attention to those.

Bayesian specification of the model

We will start with a default semi-conjugate prior specification given by

$$p(\mu, \tau^2, \sigma^2) = p(\mu)p(\tau^2)p(\sigma^2),$$

where

$$\begin{aligned}
\pi(\mu) &= \mathcal{N}\left(\mu_0, \gamma_0^2\right) \\
\pi(\tau^2) &= \mathcal{IG}\left(\frac{\eta_0}{2}, \frac{\eta_0\tau_0^2}{2}\right) \\
\pi(\sigma^2) &= \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right).
\end{aligned}$$

STA 610L

Bayesian specification of the model

With this default prior specification, we have nice interpretations of the prior parameters.

- For μ,
  - μ_0: best guess of average of group averages
  - γ_0^2: set based on plausible ranges of values of μ
- For τ^2,
  - τ_0^2: best guess of variance of group averages
  - η_0: set based on how tight prior for τ^2 is around τ_0^2
- For σ^2,
  - σ_0^2: best guess of variance of individual responses around respective group means
  - ν_0: set based on how tight prior for σ^2 is around σ_0^2.

Quick review: inverse-gamma distribution

If $\theta \sim \mathcal{IG}(a,b)$, then the pdf is

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{-(a+1)} e^{-\frac{b}{\theta}} \quad \text{for} \quad a, b > 0,$$

with

- $\mathbb{E}[\theta] = \frac{b}{a-1}$;
- $\mathbb{V}[\theta] = \frac{b^2}{(a-1)^2(a-2)} \;\; \text{for} \;\; a \geq 2.$;
- $\text{Mode}[\theta] = \frac{b}{a+1}$.

STA 610L

Implications on priors

Using an $\mathcal{IG}\left(\dfrac{\eta_0}{2}, \dfrac{\eta_0\tau_0^2}{2}\right)$ distribution for τ^2, we can now see that τ_0^2 is somewhere in the "center" of the distribution (between the mode $\frac{\eta_0\tau_0^2}{\eta_0+2}$ and the mean $\frac{\eta_0\tau_0^2}{\eta_0-2}$).

As the "prior sample size" η_0 increases, the difference between these quantities goes to 0.

We have similar implications on the prior $\pi(\sigma^2) = \mathcal{IG}\left(\dfrac{\nu_0}{2}, \dfrac{\nu_0\sigma_0^2}{2}\right)$.

Fully-specified model

We have now fully-specified our model with the following components.

1. Unknown parameters $(\mu_0, \tau_0^2, \sigma_0^2, \mu_1, \cdots, \mu_J)$
2. Prior distributions, specified in terms of prior guesses $(\mu_0, \tau_0^2, \sigma_0^2)$ and certainty/prior sample sizes $(\gamma_0^2, \eta_0, \nu_0)$
3. Data from our groups.

We can then interrogate the posterior distribution of the parameters using Gibbs sampling, as the full conditional distributions have closed forms.

Full conditionals

- For the full conditionals we will derive here, we will take advantage of results from the regular univariate normal model (from STA 360/601/602). For a refresher, see here.
- Recall that if we assume

$$y_i \sim \mathcal{N}(\mu, \sigma^2), \;\; i = 1, \ldots, n,$$

and set our priors to be

$$\pi(\mu) = \mathcal{N}\left(\mu_0, \gamma_0^2\right).$$
$$\pi(\sigma^2) = \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right),$$

then we have

$$\pi(\mu, \sigma^2 | Y) \propto \left\{ \prod_{i=1}^{n} p(y_i | \mu, \sigma^2) \right\} \cdot \pi(\mu) \cdot \pi(\sigma^2)$$

STA 610L

Full conditionals

- We have

$$\pi(\mu|\sigma^2, Y) = \mathcal{N}\left(\mu_n, \gamma_n^2\right).$$

where

$$\gamma_n^2 = \frac{1}{\dfrac{n}{\sigma^2} + \dfrac{1}{\gamma_0^2}}; \qquad \mu_n = \gamma_n^2 \left[\frac{n}{\sigma^2}\bar{y} + \frac{1}{\gamma_0^2}\mu_0\right],$$

- and

$$\pi(\sigma^2|\mu, Y) = \mathcal{IG}\left(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}\right),$$

where

$$\nu_n = \nu_0 + n; \qquad \sigma_n^2 = \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + \sum_{i=1}^{n}(y_i - \mu)^2\right].$$

STA 610L

Posterior inference

- Our hierarchical model can be written as

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

- Under our prior specification, we can factor the posterior as follows:

$$\begin{aligned} \pi(\mu_1,\ldots,\mu_J,\mu,\sigma^2,\tau^2|Y) &\propto p(y|\mu_1,\ldots,\mu_J,\mu,\sigma^2,\tau^2) \\ &\quad \times p(\mu_1,\ldots,\mu_J|\mu,\sigma^2,\tau^2) \\ &\quad \times \pi(\mu,\sigma^2,\tau^2) \\ \\ &= p(y|\mu_1,\ldots,\mu_J,\sigma^2) \\ &\quad \times p(\mu_1,\ldots,\mu_J|\mu,\tau^2) \\ &\quad \times \pi(\mu)\cdot\pi(\sigma^2)\cdot\pi(\tau^2) \\ \\ &= \left\{\prod_{j=1}^{J}\prod_{i=1}^{n_j} p(y_{ij}|\mu_j,\sigma^2)\right\} \\ &\quad \times \left\{\prod_{j=1}^{J} p(\mu_j|\mu,\tau^2)\right\} \\ &\quad \times \pi(\mu)\cdot\pi(\sigma^2)\cdot\pi(\tau^2) \end{aligned}$$

STA 610L

Full conditional for grand mean

- The full conditional distribution of μ is proportional to the part of the joint posterior $\pi(\mu_1, \ldots, \mu_J, \mu, \sigma^2, \tau^2 | Y)$ that involves μ.

- That is,

$$\pi(\mu|\mu_1, \ldots, \mu_J, \sigma^2, \tau^2, Y) \propto \left\{ \prod_{j=1}^{J} p(\mu_j|\mu, \tau^2) \right\} \cdot \pi(\mu).$$

- This looks like the full conditional distribution from the one-sample normal case, so you can show that

$$\pi(\mu|\mu_1, \ldots, \mu_J, \sigma^2, \tau^2, Y) = \mathcal{N}\left(\mu_n, \gamma_n^2\right) \quad \text{where}$$

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

and $\bar{\theta} = \frac{1}{J} \sum_{j=1}^{J} \mu_j$.

STA 610L

Full conditionals for group means

- Similarly, the full conditional distribution of each μ_j is proportional to the part of the joint posterior $\pi(\mu_1, \ldots, \mu_J, \mu, \sigma^2, \tau^2 | Y)$ that involves μ_j.
- That is,

$$\pi(\mu_j | \mu, \sigma^2, \tau^2, Y) \propto \left\{ \prod_{i=1}^{n_j} p(y_{ij} | \mu_j, \sigma^2) \right\} \cdot p(\mu_j | \mu, \tau^2)$$

- Those terms include a normal for μ_j multiplied by a product of normals in which μ_j is the mean, again mirroring the one-sample case, so you can show that

$$\pi(\mu_j | \mu, \sigma^2, \tau^2, Y) = \mathcal{N}\left(\mu_j^\star, \nu_j^\star\right) \quad \text{where}$$

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

STA 610L

Full conditionals for group means

- Our estimate for each μ_j is a weighted average of $\bar{y}_j$ and μ, ensuring that we are borrowing information across all levels through μ and τ^2.
- The weights for the weighted average is determined by relative precisions from the data and from the second level model.
- The groups with smaller n_j have estimated $\mu_j^\star$ closer to μ than schools with larger n_j.
- Thus, degree of shrinkage of μ_j depends on ratio of within-group to between-group variances.

Full conditionals for across-group variance

- The full conditional distribution of τ^2 is proportional to the part of the joint posterior $\pi(\mu_1, \ldots, \mu_J, \mu, \sigma^2, \tau^2 | Y)$ that involves τ^2.
- That is,

$$\pi(\tau^2 | \mu_1, \ldots, \mu_J, \mu, \sigma^2, Y) \propto \left\{ \prod_{j=1}^{J} p(\mu_j | \mu, \tau^2) \right\} \cdot \pi(\tau^2)$$

- As in the case for μ, this looks like the one-sample normal problem, and our full conditional posterior is

$$\pi(\tau^2 | \mu_1, \ldots, \mu_J, \mu, \sigma^2, Y) = \mathcal{IG}\left(\frac{\eta_n}{2}, \frac{\eta_n \tau_n^2}{2} \right) \quad \text{where}$$

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

STA 610L

Full conditionals for within-group variance

- Finally, the full conditional distribution of σ^2 is proportional to the part of the joint posterior $\pi(\mu_1, \ldots, \mu_J, \mu, \sigma^2, \tau^2 | Y)$ that involves σ^2.

- That is,

$$\pi(\sigma^2 | \mu_1, \ldots, \mu_J, \mu, \tau^2, Y) \propto \left\{ \prod_{j=1}^{J} \prod_{i=1}^{n_j} p(y_{ij} | \mu_j, \sigma^2) \right\} \cdot \pi(\sigma^2)$$

- We can again take advantage of the one-sample normal problem, so that our full conditional posterior (homework) is

$$\pi(\sigma^2 | \mu_1, \ldots, \mu_J, \mu, \tau^2, Y) = \mathcal{IG}\left(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2} \right) \quad \text{where}$$

$$\nu_n = \nu_0 + \sum_{j=1}^{J} n_j; \qquad \sigma_n^2 = \frac{1}{\nu_n} \left[\nu_0 \sigma_0^2 + \sum_{j=1}^{J} \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2 \right].$$

STA 610L

What's next?

Move on to the readings for the next module!