STA 360/602L: Module 5.1

Hierarchical normal models with constant variance: two groups

Dr. Olanrewaju Michael Akande

Motivation

- Sometimes, we may have a natural grouping in our data, for example
  - students within schools,
  - patients within hospitals,
  - voters within counties or states,
  - biology data, where animals are followed within natural populations organized geographically and, in some cases, socially.
- For such grouped data, we may want to do inference across all the groups, for example, comparison of the group means.
- Ideally, we should do so in a way that takes advantage of the relationship between observations in the same group, but we should also look to borrow information across groups when possible.
- Hierarchical modeling provides a principled way to do so.

Bayes estimators and bias

- Recall the normal model:

$$y_i|\mu, \sigma^2 \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right).$$

- The MLE for the population mean μ is just the sample mean $\bar{y}$.
- $\bar{y}$ is unbiased for μ. That is, for any data $y_i \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, $\mathbb{E}[\bar{y}] = \mu$.
- However, recall that in the conjugate normal model with known variance for example, the posterior expectation is a **weighted average** of the prior mean and the sample mean.
- That is, the posterior mean is actually biased.

Shrinkage

- Usually through the weighting of the sample data and prior, Bayes procedures have the tendency to pull the estimate of μ toward the prior mean.
- Of course, the magnitude of the pull depends on the sample size.
- This "pulling" phenomenon is referred to as shrinkage.
- Why would we ever want to do this? Why not just stick with the MLE?
- Well, in part, because shrinkage estimators are often "more accurate" in prediction problems – i.e. they tend to do a better job of predicting a future outcome or of recovering the actual parameter values. Remember variance-bias trade off!
- The fact that a biased estimator would do a better job in many prediction problems can be proven rigorously, and is referred to as Stein's paradox.

Modern relevance

- Stein's result implies, in particular, that the sample mean is an *inadmissible* estimator of the mean of a multivariate normal distribution in more than two dimensions – i.e. there are other estimators that will come closer to the true value in expectation.
- In fact, these are Bayes point estimators (the posterior expectation of the parameter μ).
- Most of what we do now in high-dimensional statistics is develop biased estimators that perform better than unbiased ones.
- Examples: lasso regression, ridge regression, various kinds of hierarchical Bayesian models, etc.
- So, here we will get a very basic introduction to Bayesian hierarchical models, which provide a formal and coherent framework for constructing shrinkage estimators.

Why hierarchical models?

- **Bayesian hierarchical models** is a sort of catch-all phrase for a large class of models that have several levels of conditional distributions making up the prior.

- Like simpler one-level priors, they also accomplish shrinkage. However, they are much more flexible.

- Why use them? Several reasons:

  - We may want to exploit more complex dependence structures.
  - We may have many parameters relative to the amount of data that we have, and want to borrow information in estimating them.
  - We may want to shrink toward something other than a simple prior mean/hyper-parameter.

Comparing two groups

- Suppose we want to do inference on mean body mass index (BMI) for two groups (male or female).
- BMI is known to often follow a normal distribution, so let's assume the same here.
- We should expect some relationship between the mean BMI for the two groups.
- We may also think the shape of the two distributions would be relatively the same (at least as a simplifying assumption for now).
- Thus, a reasonable model might be

$$y_{i,male} \overset{iid}{\sim} \mathcal{N}\left(\theta_m, \sigma^2\right); \;\; i = 1, \ldots, n_m;$$
$$y_{i,female} \overset{iid}{\sim} \mathcal{N}\left(\theta_f, \sigma^2\right); \;\; i = 1, \ldots, n_f.$$

but with some relationship between θ_m and θ_f.

Bayesian inference

- One parameterization that can reflect some relationship between θ_m and θ_f is

$$y_{i,male} \overset{iid}{\sim} \mathcal{N}\left(\mu + \delta, \sigma^2\right); \;\; i = 1, \ldots, n_m;$$
$$y_{i,female} \overset{iid}{\sim} \mathcal{N}\left(\mu - \delta, \sigma^2\right); \;\; i = 1, \ldots, n_f.$$

where

- $\theta_m = \mu + \delta$ and $\theta_f = \mu - \delta$,
- $\mu = \dfrac{\theta_m + \theta_f}{2}$ is the average of the population means, and
- $2\delta = \theta_m - \theta_f$ is the difference in population means.

Bayesian inference

- Convenient prior:
  - $\pi(\mu, \delta, \sigma^2) = \pi(\mu) \cdot \pi(\delta) \cdot \pi(\sigma^2)$, where
    - $\pi(\mu) = \mathcal{N}(\mu_0, \gamma_0^2)$,
    - $\pi(\delta) = \mathcal{N}(\delta_0, \tau_0^2)$, and
    - $\pi(\sigma^2) = \mathcal{IG}(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2})$.

Bayesian inference

- Note that we can rewrite

$$y_{i,male} \overset{iid}{\sim} \mathcal{N}\left(\mu + \delta, \sigma^2\right); \;\; i = 1, \ldots, n_m;$$
$$y_{i,female} \overset{iid}{\sim} \mathcal{N}\left(\mu - \delta, \sigma^2\right); \;\; i = 1, \ldots, n_f$$

as

$$(y_{i,male} - \delta) \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right); \;\; i = 1, \ldots, n_m;$$
$$(y_{i,female} + \delta) \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right); \;\; i = 1, \ldots, n_f$$

or

$$(y_{i,male} - \mu) \overset{iid}{\sim} \mathcal{N}\left(\delta, \sigma^2\right); \;\; i = 1, \ldots, n_m;$$
$$(-1)(y_{i,female} - \mu) \overset{iid}{\sim} \mathcal{N}\left(\delta, \sigma^2\right); \;\; i = 1, \ldots, n_f.$$

as needed, so we can leverage past results for the full conditionals.

Full conditionals

- For the full conditionals we will derive here, we will take advantage of previous results from the regular univariate normal model.

- 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

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

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

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

Full conditionals

- With $\pi(\mu) = \mathcal{N}(\mu_0, \gamma_0^2)$, and

$$(y_{i,male} - \delta) \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right); \quad i = 1, \ldots, n_m;$$
$$(y_{i,female} + \delta) \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right); \quad i = 1, \ldots, n_f,$$

we have

$$\mu | Y, \delta, \sigma^2 \sim \mathcal{N}(\mu_n, \gamma_n^2), \quad \text{where}$$

$$\gamma_n^2 = \frac{1}{\dfrac{1}{\gamma_0^2} + \dfrac{n_m + n_f}{\sigma^2}}$$

$$\mu_n = \gamma_n^2 \left[\frac{\mu_0}{\gamma_0^2} + \frac{\sum_{i=1}^{n_m} (y_{i,male} - \delta) + \sum_{i=1}^{n_f} (y_{i,female} + \delta)}{\sigma^2} \right].$$

Full conditionals

- With $\pi(\delta) = \mathcal{N}(\delta_0, \tau_0^2)$, and

$$(y_{i,male} - \mu) \overset{iid}{\sim} \mathcal{N}\left(\delta, \sigma^2\right); \quad i = 1, \ldots, n_m;$$
$$(-1)(y_{i,female} - \mu) \overset{iid}{\sim} \mathcal{N}\left(\delta, \sigma^2\right); \quad i = 1, \ldots, n_f,$$

we have

$$\delta | Y, \mu, \sigma^2 \sim \mathcal{N}(\delta_n, \tau_n^2), \quad \text{where}$$

$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n_m + n_f}{\sigma^2}}$$

$$\delta_n = \tau_n^2 \left[\frac{\delta_0}{\tau_0^2} + \frac{\sum_{i=1}^{n_m}(y_{i,male} - \mu) + (-1)\sum_{i=1}^{n_f}(y_{i,female} - \mu)}{\sigma^2} \right].$$

Full conditionals

- With $\pi(\sigma^2) = \mathcal{IG}(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2})$, and

$$y_{i,male} \overset{iid}{\sim} \mathcal{N}\left(\mu + \delta, \sigma^2\right); \ \ i = 1, \ldots, n_m;$$
$$y_{i,female} \overset{iid}{\sim} \mathcal{N}\left(\mu - \delta, \sigma^2\right); \ \ i = 1, \ldots, n_f$$

we have

$$\sigma^2 | Y, \mu, \delta \sim \mathcal{IG}(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}), \quad \text{where}$$

$$\nu_n = \nu_0 + n_m + n_f$$

$$\sigma_n^2 = \frac{1}{\nu_n}\left[\nu_0 \sigma_0^2 + \sum_{i=1}^{n_m}(y_{i,male} - [\mu + \delta])^2 + \sum_{i=1}^{n_f}(y_{i,female} - [\mu - \delta])^2\right].$$

- We will use write a Gibbs sampler for this model and fit the model to real data in the next module.

What's next?

Move on to the readings for the next module!