SDS 383D Ex 04:
Hierarchical Models

March 22, 2016

Jennifer Starling

Math Tests

The data set in "mathtest.csv" shows the scores on a standardized math test from a sample of 10th-grade students at 100 different U.S. urban high schools, all having enrollment of at least 400 10th-grade students. (A lot of educational research involves "survey tests" of this sort, with tests administered to all students being the rare exception.)

Let θ_i be the underlying mean test score for school i, and let y_{ij} be the score for the jth student in school i. Starting with the "mathtest.R" script, you'll notice that the extreme school-level averages $\bar{y}_i$ (both high and low) tend to be at schools where fewer students were sampled.

Part 1

Briefly explain why this would be.

The extreme school-level averages occur in the schools with smaller sample sizes because we do not do a very good job of estimating the mean when sample size is small. These schools do not have min and max observation values that are more extreme than the other schools; they just have fewer observations to balance out the calculation of the mean. The smaller the sample size, the more influential an extreme observation is over the group mean.

Part 2

Fit a normal hierarchical model to these data via Gibbs sampling:

$$\begin{aligned} y_{ij} &\sim \mathrm{N}(\theta_i, \sigma^2) \\ \theta_i &\sim \mathrm{N}(\mu, \tau^2\sigma^2) \end{aligned}$$

Decide upon sensible priors for the unknown model parameters (μ, σ^2, τ^2).
The model is as follows.

$$\begin{aligned} (y_ij|theta_i, \sigma^2) &\sim N(\theta_i, \sigma^2) \\ (\theta_i|\mu, \sigma^2, \tau^2) &\sim N(\mu, \sigma^2\tau^2) \\ \mu &\sim I_{\mathbb{R}}(\mu), \text{ a flat prior on the real line} \\ \tau^2 &\sim I_{\mathbb{R}^+}(\tau^2), \text{ a flat prior on the positive real line} \\ \sigma^2 &\sim \left(\frac{1}{\sigma^2}\right) I_{\mathbb{R}^+}(\sigma^2), \text{ Jeffreys prior} \end{aligned}$$

where

$$\begin{aligned} i &= 1, ..., p \text{ indexes the } p \text{ groups.} \\ n_i &= \text{ sample size in each group.} \\ j &= 1, ..., n_i \text{ indexes observations in a group.} \\ n &= \text{ total number of observations.} \end{aligned}$$

The likelihood is

$$L(y|\theta_1, \dots, \theta_p, \sigma^2) \sim \prod_{i=1}^{p}\prod_{j=1}^{n_i}\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{\sigma^2}\left(y_{ij}-\theta_i\right)^2\right] = \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{ij}-\theta_i\right)^2\right]$$

The full conditionals are as follows.

$(\theta_i|y,\mu,\sigma^2,\tau^2)$

Note that $\bar{y}_i$ is a sufficient statistic for the $y's$, with $\bar{y}_i \sim N\left(\theta_i, \frac{\sigma^2}{n}\right)$.

$$(\theta_i|y,\mu,\sigma^2,\tau^2) \propto \left(\sigma^2\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2/n}(\bar{y}_i-\theta_i)^2\right]\left(\tau^2\sigma^2\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2\tau^2}(\theta_i-\mu)^2\right]$$

This is the normal-normal model, therefore

$$\begin{aligned}
(\theta_i|y,\mu,\sigma^2,\tau^2) &\sim N(m^*,v^*) \\
&\text{with} \\
v^* &= \left[\frac{n_i}{\sigma^2}+\frac{1}{\sigma^2\tau^2}\right]^{-1} = \left[\frac{n_i\tau^2+1}{\sigma^2\tau^2}\right]^{-1} = \sigma^2\left[\frac{\tau^2}{n_i\tau^2+1}\right] \\
m^* &= v^*\left[\left(\frac{n_i}{\sigma^2}\right)\bar{y}_i + \left(\frac{1}{\sigma^2\tau^2}\right)\mu\right] \\
&= \sigma^2\left[\frac{\tau^2}{n_i\tau^2+1}\right]\left[\left(\frac{n_i}{\sigma^2}\right)\bar{y}_i + \left(\frac{1}{\sigma^2\tau^2}\right)\mu\right] \\
&= \left[\frac{n_i\tau^2}{n_i\tau^2+1}\right]\bar{y}_i + \left[\frac{1}{n_i\tau^2+1}\right]\mu \\
&= w\bar{y}_i + (1-w)\mu
\end{aligned}$$

So full conditional is

$$(\theta_i|y,\mu,\sigma^2,\tau^2) \sim N\left(\left[\frac{n_i\tau^2}{n_i\tau^2+1}\right]\bar{y}_i + \left[\frac{1}{n_i\tau^2+1}\right]\mu, \sigma^2\left[\frac{\tau^2}{n_i\tau^2+1}\right]\right) \tag{1}$$

$(\mu|\theta,y,\sigma^2,\tau^2)$

$$\begin{aligned}
\left(\mu|\theta,y,\sigma^2,\tau^2\right) &\propto \exp\left[-\frac{1}{2\sigma^2\tau^2}\sum_{i=1}^{p}(\theta_i-\mu)^2\right]\cdot 1 \\
&= \exp\left[-\frac{1}{2\sigma^2\tau^2}\left\{(\theta_1-\mu)(\theta_1-\mu)+\ldots+(\theta_p-\mu)(\theta_p-\mu)\right\}\right] \\
&= \exp\left[-\frac{1}{2\sigma^2\tau^2}\left\{p\mu^2 - 2\mu\sum_{i=1}^{p}\theta_i + \sum_{i=1}^{p}\theta_i^2\right\}\right] \\
&= \exp\left[-\frac{p}{2\sigma^2\tau^2}\left\{\mu^2 - 2\mu\left(\frac{\sum_{i=1}^{p}\theta_i}{p}\right) + \frac{\sum_{i=1}^{p}\theta_i^2}{p}\right\}\right] \\
&\propto \exp\left[-\frac{p}{2\sigma^2\tau^2}\left\{\mu^2 - 2\mu\bar{\theta}_i\right\}\right]
\end{aligned}$$

We recognize this as a Normal kernel, therefore

$$\left(\mu|\theta,y,\sigma^2,\tau^2\right) \sim N\left(\bar{\theta}_i, \frac{\sigma^2\tau^2}{p}\right) \tag{2}$$

$(\sigma^2|\theta,y,\mu,\tau^2)$

$$(\sigma^2|\theta,y,\mu,\tau^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2\right] \left(\sigma^2\right)^{-\frac{p}{2}} \exp\left[-\frac{1}{2\sigma^2\tau^2}\sum_{i=1}^{p}(\theta_i-\mu)^2\right]\left(\frac{1}{\sigma^2}\right)$$

$$= \left(\sigma^2\right)^{-\frac{(n+p)}{2}-1} \exp\left[-\left(\frac{1}{\sigma^2}\right)\cdot\left\{\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2+\frac{1}{2\tau^2}\sum_{i=1}^{p}(\theta_i-\mu)^2\right\}\right]$$

We recognize this as an Inverse-Gamma kernel, therefore

$$(\sigma^2|\theta,y,\mu,\tau^2) \sim IG\left(\frac{(n+p)}{2},\left\{\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\theta_i)^2+\frac{1}{2\tau^2}\sum_{i=1}^{p}(\theta_i-\mu)^2\right\}\right) \tag{3}$$

$(\tau^2|\theta,y,\mu,\sigma^2)$

$$(\tau^2|\theta,y,\mu,\sigma^2) \propto \left(\tau^2\right)^{-\frac{p}{2}} \exp\left[-\frac{1}{2\sigma^2\tau^2}\sum_{i=1}^{p}(\theta_i-\mu)^2\right]\cdot 1$$

We recognize this as an Inverse Gamma kernel, therefore

$$(\tau^2|\theta,y,\mu,\sigma^2) \sim IG\left(\frac{p}{2}-1,\frac{1}{2\sigma^2}\sum_{i=1}^{p}(\theta_i-\mu)^2\right) \tag{4}$$

Part 3

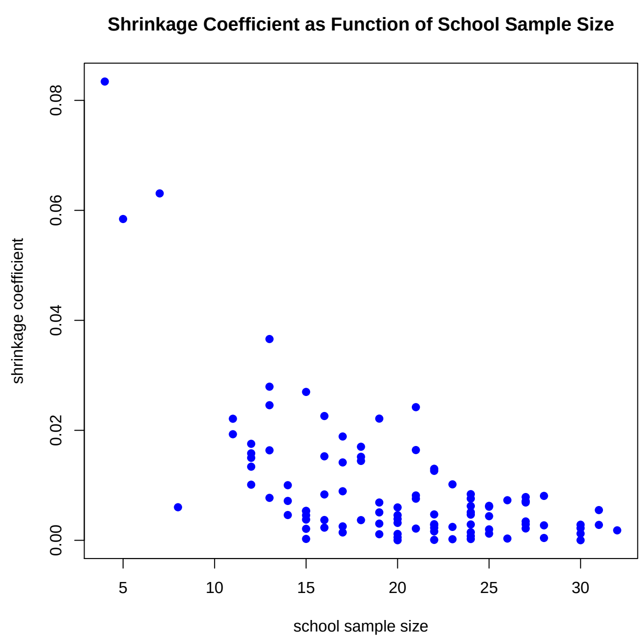

Figure 1: Shrinkage estimator by school sample size

Price Elasticity of Demand

A Hierarchical Probit Model via Data Augmentation

Gene Expression Over Time