SDS 383D: Exercises 4 – Hierarchical Models

March 27, 2017

Professor Scott

Spencer Woody

Problem 1

Math Tests

We have a model where y_{ij} is the test score of the jth student in school i, with indices $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, N_i$, so N_i is the sample size for school i and there are $N = \sum_{i=1}^{I}$ total test scores. Let $\lambda = 1/\sigma^2$ and $\gamma = 1/\tau^2$ be the precision parameters. Further, let $y_i = [y_{i1}, y_{i2}, \ldots, y_{iN_i}]^T$ and $y = [y_1^T, y_2^T, \ldots, y_I^T]^T$ and $\theta = [\theta_1, \theta_2, \ldots, \theta_I]^T$. As we can see in Figure 1, schools with smaller sample sizes tend to have more extreme average test scores.

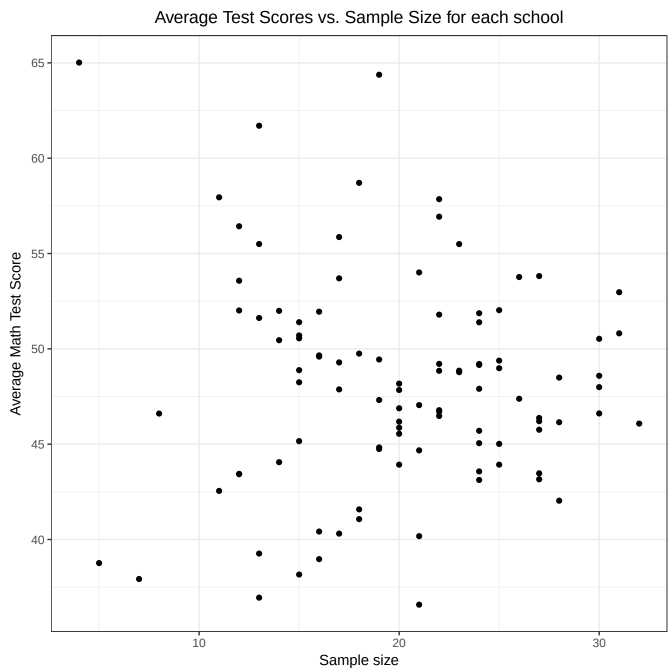

Figure 1: Scatter plot of sample size and average test scores

The hierarchical model for these data is

$$
\begin{aligned}
(y_{ij} | \theta_i, \lambda) &\sim \mathcal{N}\left(\theta_i, \lambda^{-1}\right) \\
(\theta_i | \mu, \lambda, \gamma) &\sim \mathcal{N}\left(\mu, (\lambda\gamma)^{-1}\right).
\end{aligned}
$$

We set the priors

$$
\begin{aligned}
&\pi(\mu) \propto 1, \ -\infty < \mu < \infty \\
&\pi(\lambda) \propto \lambda^{-1}, \ \lambda > 0 \\
&\pi(\gamma) \propto 1, \ \gamma > 0,
\end{aligned}
$$

that is to say, ... In order to implement the Gibbs sampler, we need the posterior full conditionals for each θ_i, μ, λ, and γ.

- For each θ_i,

$$
\begin{aligned}
f(\theta_i | y_i, \mu, \lambda, \gamma) &\propto f(y_i | \theta_i, \lambda) \cdot f(\theta_i | \mu, \lambda, \gamma) \\
&\sim \mathcal{N}\left((N_i\lambda + \lambda\gamma)^{-1} \cdot (N_i\lambda\bar{y}_i + \lambda\gamma\mu), (N_i\lambda + \lambda\gamma)^{-1}\right),
\end{aligned}
$$

  which we know from the normal-normal conjugacy derived in Exercises 1.

- For μ,

$$
\begin{aligned}
\pi(\mu|y,\theta,\lambda,\gamma) &\propto f(\theta|\lambda,\gamma,\mu)\cdot\pi(\mu) \\
&\propto \left(\prod_{i=1}^{I}\exp\left[-\frac{1}{2}\lambda\gamma(\theta_i-\mu)^2\right]\right)\cdot 1 \\
&= \exp\left[-\frac{1}{2}\lambda\gamma\sum_{i=1}^{I}(\theta_i-\mu)^2\right] \\
&= \exp\left[-\frac{1}{2}\lambda\gamma\sum_{i=1}^{I}\left(\theta_i^2-2\theta_i\mu+\mu^2\right)\right] \\
&\propto \exp\left[-\frac{1}{2}\lambda\gamma\left(I\mu^2-2I\bar{\theta}\right)\right] \\
&\sim \mathcal{N}\left(\bar{\theta},(I\lambda\gamma)^{-1}\right).
\end{aligned}
$$

- For λ,

$$
\begin{aligned}
\pi(\lambda|y,\mu,\gamma,\theta) &\propto f(y|\lambda,\theta)\cdot f(\theta|\lambda,\gamma,\mu)\cdot\pi(\lambda) \\
&\propto \left(\prod_{i=1}^{I}\prod_{j=1}^{N_i}\lambda^{1/2}\exp\left[-\frac{1}{2}(y_{ij}-\theta_i)^2\right]\right)\cdot\left(\prod_{i=1}^{I}\lambda^{1/2}\exp\left[-\frac{1}{2}\lambda\gamma(\theta_i-\mu)^2\right]\right)\cdot\lambda^{-1} \\
&= \lambda^{(N+I)/2-1}\exp\left[-\frac{1}{2}\left(\sum_{i=1}^{I}\sum_{j=1}^{N_i}(y_{ij}-\theta_i)^2+\gamma\sum_{i=1}^{I}(\theta_i-\mu)^2\right)\lambda\right] \\
&\sim \text{Gamma}\left(\frac{N+I}{2},\frac{1}{2}\left[\sum_{i=1}^{I}\sum_{j=1}^{N_i}(y_{ij}-\theta_i)^2+\gamma\sum_{i=1}^{I}(\theta_i-\mu)^2\right]\right).
\end{aligned}
$$

- For γ,

$$
\begin{aligned}
\pi(\gamma|y,\mu,\lambda,\theta) &\propto f(\theta|\lambda,\gamma,\mu)\cdot\pi(\gamma) \\
&\propto \left(\prod_{i+1}^{I}\gamma^{1/2}\exp\left[-\frac{1}{2}\lambda\gamma(\theta_i-\mu)^2\right]\right)\cdot 1 \\
&= \gamma^{I/2}\exp\left[-\frac{1}{2}\lambda\sum_{i=1}^{I}(\theta_i-\mu)^2\cdot\gamma\right] \\
&\sim \text{Gamma}\left(\frac{I}{2}-1,\frac{1}{2}\lambda\sum_{i=1}^{I}(\theta_i-\mu)^2\right).
\end{aligned}
$$

Table 1: 95% posterior credible intervals

	2.5%	50%	97.5%
μ	47.03	48.10	49.18
λ	0.0111	0.0118	0.0126
γ	2.43	3.49	5.03

Given the posterior mean $\hat{\theta}_i$ as an estimate of θ_i, define the shrinkage coefficient

$$\kappa_i = \frac{\bar{y}_i-\hat{\theta}_i}{\bar{y}_i},$$

which is a measure incomplete pooling. Figure 2 shows the absolute shrinkage coefficient for each school as a function of sample size. As sample size increases, the shrinkage decreases because we are gaining precision in estimating the school-level mean θ_i.

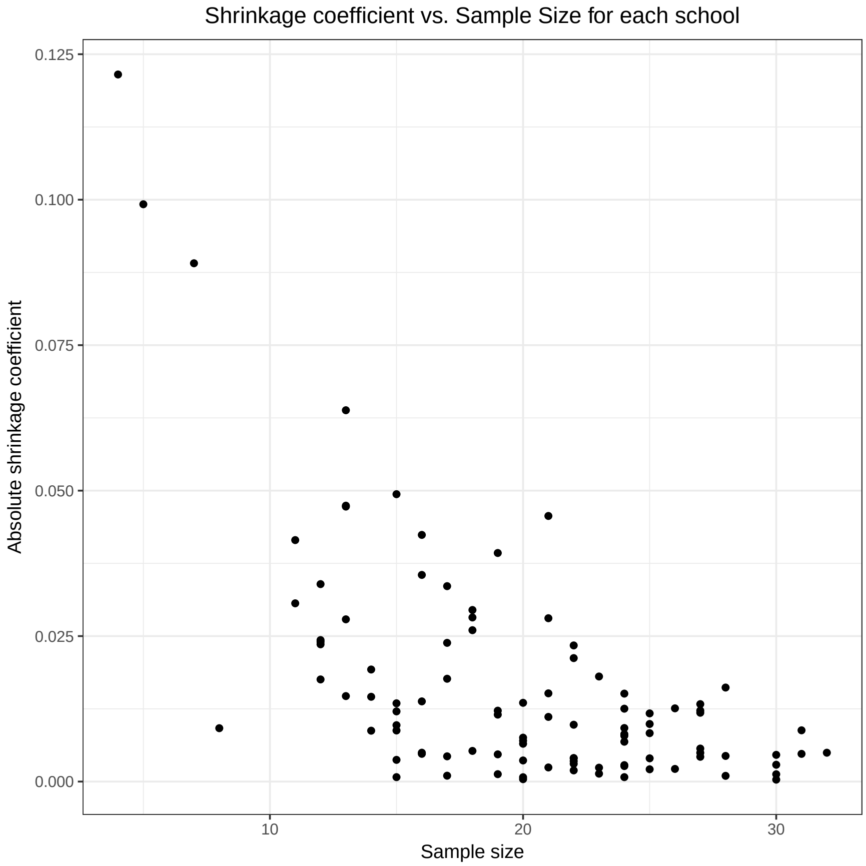

Figure 2: Absolute shrinkage coefficient as a function of sample size