Hierarchical models

Dr. Jarad Niemi

Iowa State University

August 24, 2017

Normal hierarchical model

Let

$$Y_{ij} \stackrel{ind}{\sim} N(\theta_i, \sigma^2)$$

for $i = 1, \ldots, \mathrm{I}$, $j = 1, \ldots, n_i$, and $\sum_{i=1}^{\mathrm{I}} n_i = n$. Now consider the following model assumptions:

- $\theta_i \stackrel{ind}{\sim} N(\mu, \tau^2)$
- $\theta_i \stackrel{ind}{\sim} La(\mu, \tau)$
- $\theta_i \stackrel{ind}{\sim} t_v(\mu, \tau^2)$
- $\theta_i \stackrel{ind}{\sim} \pi\delta_0 + (1-\pi)N(\mu, \tau^2)$
- $\theta_i \stackrel{ind}{\sim} \pi\delta_0 + (1-\pi)t_v(\mu, \tau^2)$

To perform a Bayesian analysis, we need a prior on μ, τ^2, and (possibly) π.

Scale mixtures of normals

Recall that if

$$\theta|\phi \sim N(\phi, V) \text{ and } \phi \sim N(m, C)$$

then

$$\theta \sim N(m, V + C).$$

This is called a location mixture.

Now, if

$$\theta|\phi \sim N(m, C\phi)$$

and we assume a mixing distribution for ϕ, we have a scale mixture. Since the top level distributional assumption is normal, we refer to this as a scale mixture of normals.

t distribution

Let

$$\theta|\phi \sim N(m, \phi C) \text{ and } \phi \sim IG(a, b)$$

then

$$\begin{aligned}
p(\theta) &= \int p(\theta|\phi)p(\phi)d\phi \\
&= (2\pi\sqrt{C})^{-1/2}\frac{b^a}{\Gamma(a)}\int \phi^{-1/2}e^{-(\theta-m)^2/2\phi C}\phi^{-(a+1)}e^{-b/\phi}d\phi \\
&= (2\pi C)^{-1/2}\frac{b^a}{\Gamma(a)}\int \phi^{-(a+1/2+1)}e^{-[b+(\theta-m)^2/2C]/\phi}d\phi \\
&= (2\pi C)^{-1/2}\frac{b^a}{\Gamma(a)}\frac{\Gamma(a+1/2)}{[b+(\theta-m)^2/2C]^{a+1/2}} \\
&= \frac{\Gamma([2a+1]/2)}{\Gamma(2a/2)\sqrt{2a\pi bC/a}}\left[1+\frac{1}{2a}\frac{(\theta-m)^2}{bC/a}\right]^{-[2a+1]/2}
\end{aligned}$$

Thus

$$\theta \sim t_{2a}(m, bC/a)$$

i.e. θ has a t distribution with $2a$ degrees of freedom, location m, and scale (variance) bC/a. Thus if $C = 1$, $a = v/2$, and $b = v\tau^2/2$, we have a t distribution with v degrees of freedom, location m, and scale τ^2.

Laplace distribution

Let

$$\theta|\phi \sim N(m, \phi C) \text{ and } \phi \sim Exp(1/2b^2)$$

then, by equation (4) in Park and Casella (2008),

$$\theta \sim La(m, b/\sqrt{C})$$

where $z = (\theta - m)/\sqrt{C}$ and $a = 1/b$ and thus θ has a Laplace distribution with location m and scale $b/\sqrt{C}$.

Normal hierarchical model

Recall our hierarchical model

$$Y_{ij} \stackrel{ind}{\sim} N(\theta_i, \sigma^2)$$

for $i = 1, \ldots, \mathrm{I}$ and $j = 1, \ldots, n_i$. Now consider the following model assumptions:

- $\theta_i \stackrel{ind}{\sim} N(\mu, \phi_i), \phi_i = \tau^2 \implies \theta_i \stackrel{ind}{\sim} N(\mu, \tau^2)$
- $\theta_i|\phi_i \stackrel{ind}{\sim} N(\mu, \phi_i), \phi_i \stackrel{ind}{\sim} Exp(1/2\tau^2) \implies \theta_i \stackrel{ind}{\sim} La(\mu, \tau)$
- $\theta_i|\phi_i \stackrel{ind}{\sim} N(\mu, \phi_i), \phi_i \stackrel{ind}{\sim} IG(v/2, v\tau^2/2) \implies \theta_i \stackrel{ind}{\sim} t_v(\mu, \tau^2)$

For simplicity, let's assume $\sigma^2 \sim IG(a, b)$, $\mu \sim N(m, C)$, and $\tau \sim Ca^+(0, c)$ and that σ^2, μ, and τ are *a priori* independent.

Gibbs sampling

The following Gibbs sampler will converge to the posterior $p(\theta, \sigma, \mu, \phi, \tau|y)$:

1. Sample $\mu \sim p(\mu| \ldots)$.
2. Independently, sample $\theta_i \sim p(\theta_i| \ldots)$.
3. Sample $\sigma \sim p(\sigma| \ldots)$.
4. Independently, sample $\phi_i \sim p(\phi_i| \ldots)$.
5. Sample $\tau \sim p(\tau| \ldots)$.

The first three steps will be common to all models while the last two steps will be unique to each model (without a point mass).

Sample μ

$$\theta_i \stackrel{ind}{\sim} N(\mu, \phi_i) \text{ and } \mu \sim N(m, C)$$

Immediately, we should know that

$$\mu | \ldots \sim N(m', C')$$

with

$$\begin{array}{rl} C' &= \left(\frac{1}{C} + \sum_{i=1}^{\mathrm{I}} \frac{1}{\phi_i} \right)^{-1} \\ m' &= C' \left(\frac{m}{C} + \sum_{i=1}^{\mathrm{I}} \frac{\theta_i}{\phi_i} \right) \end{array}$$

Sample θ

$$Y_{ij} \stackrel{ind}{\sim} N(\theta_i, \sigma^2) \text{ and } \theta_i \sim N(\mu, \phi_i)$$

$$\begin{aligned} p(\theta|\ldots) &\propto \left[\prod_{i=1}^{\mathrm{I}} \prod_{j=1}^{n_i} e^{-(y_{ij}-\theta_i)^2/2\sigma^2}\right] \left[\prod_{i=1}^{\mathrm{I}} e^{-(\theta_i-\mu)^2/2\phi_i}\right] \\ &\propto \prod_{i=1}^{\mathrm{I}} \left[\prod_{j=1}^{n_i} e^{-(y_{ij}-\theta_i)^2/2\sigma^2} e^{-(\theta_i-\mu)^2/2\phi_i}\right] \end{aligned}$$

Thus θ_i are conditionally independent given everything else. It should be obvious that

$$\theta_i|\ldots \sim N\left(\left[\frac{\mu}{\phi_i} + \frac{n_i}{\sigma^2}\overline{y}_i\right], \left[\frac{1}{\phi_i} + \frac{n_i}{\sigma^2}\right]^{-1}\right)$$

where $\overline{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$.

Sample σ^2

$$Y_{ij} \stackrel{ind}{\sim} N(\theta_i, \sigma^2) \text{ and } \sigma^2 \sim IG(a, b)$$

This is just a normal data model with an unknown variance that has the conjugate prior. The only difficulty is that we have several groups here. But very quickly you should be able to determine that

$$\sigma^2 | \ldots \sim IG(a', b')$$

where

$$\begin{aligned} a' &= a + \textstyle\sum_{i=1}^{\mathrm{I}} n_i/2 = a + n/2 \\ b' &= b + \textstyle\sum_{i=1}^{\mathrm{I}} \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2/2. \end{aligned}$$

Distributional assumption for θ_i

$$Y_{ij} \stackrel{ind}{\sim} N(\theta_i, \sigma^2) \text{ and } \theta_i \stackrel{ind}{\sim} N(\mu, \phi_i)$$

$$\begin{array}{ll} \phi_i & = \tau \\ \phi_i & \sim Exp(1/2\tau^2) \\ \phi_i & \sim IG(v/2, v\tau^2/2) \end{array}$$

The steps that are left are 1) sample ϕ and 2) sample τ^2,

Sample ϕ for normal model

For normal model, $\phi_i = \tau$, so we will adress this when we sample τ.

Sample ϕ for Laplace model

For Laplace model,

$$\theta_i \stackrel{ind}{\sim} N(\mu, \phi_i) \text{ and } \phi_i \stackrel{ind}{\sim} Exp(1/2\tau^2),$$

so the full conditional is

$$p(\phi | \ldots) \propto \left[\prod_{i=1}^{\mathrm{I}} N(\theta_i; \mu, \phi_i) Exp(\phi_i; 1/2\tau^2) \right].$$

So the individual ϕ_i are conditionally independent with

$$p(\phi_i | \ldots) \propto N(\theta_i; \mu, \phi_i) Exp(\phi_i; 1/2\tau^2) \propto \phi_i^{-1/2} e^{-(\theta_i - \mu)^2/2\phi_i} e^{-\phi_i/2\tau^2}$$

If we perform the transformation $\eta_i = 1/\phi_i$, we have

$$p(\eta_i | \ldots) \propto \eta_i^{-3/2} e^{-\frac{(\theta_i - \mu)^2}{2}\eta_i - \frac{1}{2\tau^2\eta_i}}$$

which is the kernel of an inverse Gaussian distribution with mean $\sqrt{1/\tau^2(\theta_i - \mu)^2}$ and scale $1/\tau^2$ where the parameterization is such that the variance is μ^3/λ (different from the `mgcv::rig` parameterization).

Sample ϕ for t model

For the t model,

$$\theta_i \stackrel{ind}{\sim} N(\mu, \phi_i) \text{ and } \phi_i \stackrel{ind}{\sim} IG(v/2, v\tau^2/2),$$

so we have

$$\phi_i|\ldots \stackrel{ind}{\sim} IG([v+1]/2, [v\tau^2 + (\theta_i - \mu)^2]/2).$$

Since this is just I independent normal data models with a known mean and independent conjugate inverse gamma priors on the variance.

Sample τ for normal model

Let

$$\theta_i \stackrel{ind}{\sim} N(\mu, \tau^2) \text{ and } \tau \sim Ca^+(0, c).$$

so the full conditional is

$$p(\eta | \ldots) \propto \eta^{-\mathrm{I}/2} e^{-\sum_{i=1}^{\mathrm{I}} (\theta_i - \mu)^2 / 2\eta} \left(1 + \eta/c^2\right)^{-1} \eta^{-1/2}$$

where we performed the transformation $\eta = \tau^2$ on the prior.

Let's use Metropolis-Hastings with proposal distribution

$$IG\left(\frac{\mathrm{I}-1}{2}, \sum_{i=1}^{\mathrm{I}} \frac{(\theta_i - \mu)^2}{2}\right)$$

and acceptance probability $\min\{1, \rho\}$ where

$$\rho = \frac{\left(1 + \eta^*/c^2\right)^{-1}}{\left(1 + \eta^{(i)}/c^2\right)^{-1}} = \frac{1 + \eta^{(i)}/c^2}{1 + \eta^*/c^2}$$

where $\eta^{(i)}$ and η^* are the current and proposed value respective.

Sample τ for Laplace model

Let

$$\phi_i \sim Exp(1/2\tau^2) \text{ and } \tau \sim Ca^+(0, c)$$

so the full conditional is

$$p(\eta|\ldots) \propto \eta^{-\mathrm{I}} e^{-\sum_{i=1}^{\mathrm{I}} \phi_i/2\eta} \left(1 + \eta/c^2\right)^{-1} \eta^{-1/2}.$$

Let's use Metropolis-Hastings with proposal distribution

$$IG\left(\mathrm{I} - \frac{1}{2}, \sum_{i=1}^{\mathrm{I}} \frac{\phi_i}{2}\right)$$

and acceptance probability $\min\{1, \rho\}$ where again

$$\rho = \frac{1 + \eta^{(i)}/c^2}{1 + \eta^*/c^2}.$$

Then we calculate $\tau = \sqrt{\eta}$.

Sample τ for t model

Let

$$\phi_i \sim IG(v/2, v\tau^2/2) \text{ and } \tau \sim Ca^+(0, c)$$

so the full conditional is

$$p(\eta|\ldots) \propto \eta^{\mathrm{I}v/2} e^{-\frac{\eta}{2}\sum_{i=1}^{\mathrm{I}} \frac{1}{\phi_i}} \left(1 + \eta/c^2\right)^{-1} \eta^{-1/2}.$$

Let's use Metropolis-Hastings with proposal distribution

$$Ga\left(\frac{\mathrm{I}v + 1}{2}, \frac{1}{2}\sum_{i=1}^{\mathrm{I}} \frac{1}{\phi_i}\right)$$

and acceptance probability $\min\{1, \rho\}$ where again

$$\rho = \frac{1 + \eta^{(i)}/c^2}{1 + \eta^*/c^2}.$$

Then we calculate $\tau = \sqrt{\eta}$.

Dealing with point-mass distributions

We would also like to consider models with

$$\theta_i \stackrel{ind}{\sim} \pi\delta_0 + (1-\pi)N(\mu, \phi_i)$$

where $\phi_i = \tau^2$ corresponds to a normal and

$$\phi_i \stackrel{ind}{\sim} IG(v/2, v\tau^2/2)$$

corresponds to a t distribution for the non-zero θ_i.

Similar to the previous, the θ_i are conditionally independent. To sample θ_i, we calculate

$$\begin{aligned}
\pi' &= \frac{\pi \prod_{j=1}^{n_i} N(y_{ij};0,\sigma^2)}{\pi \prod_{j=1}^{n_i} N(y_{ij};0,\sigma^2) + (1-\pi)\prod_{j=1}^{n_i} N(y_{ij};\mu,\phi_i+\sigma^2)}. \\
\phi_i' &= \left(\frac{1}{\phi_i} + \frac{n_i}{\sigma^2}\right)^{-1} \\
\mu_i' &= \phi_i' \left(\frac{\mu}{\phi_i} + \frac{n_i}{\sigma^2}\overline{y}_i\right)
\end{aligned}$$

Dealing with point-mass distributions (cont.)

Let

$$\theta_i \stackrel{ind}{\sim} \pi\delta_0 + (1-\pi)N(\mu, \phi_i)$$

and independently $\pi \sim Beta(s, f)$, $\mu \sim N(m, C)$, and $\phi_i = \tau^2$ for normal model or $\phi_i \stackrel{ind}{\sim} IG(v/2, v\tau^2/2)$ for the t model.

The full conditional for π is

$$\pi | \ldots \sim Beta\left(s + \sum_{i=1}^{\mathrm{I}} \mathrm{I}(\theta_i = 0), f + \sum_{i=1}^{\mathrm{I}} \mathrm{I}(\theta_i \neq 0)\right)$$

and μ and ϕ_i get updated using only those θ_i that are non-zero.