Hierarchical models

Dr. Jarad Niemi

Iowa State University

August 29, 2017

Normal hierarchical model

Let

$$Y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2)$$

for $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $\sum_{g=1}^{G} n_g = n$.

Normal hierarchical model

Let

$$Y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2)$$

for $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $\sum_{g=1}^{G} n_g = n$. Now consider the following model assumptions:

- $\theta_i \stackrel{ind}{\sim} N(\mu, \tau^2)$

Normal hierarchical model

Let

$$Y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2)$$

for $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $\sum_{g=1}^{G} n_g = n$. Now consider the following model assumptions:

- $\theta_i \stackrel{ind}{\sim} N(\mu, \tau^2)$
- $\theta_i \stackrel{ind}{\sim} La(\mu, \tau)$

Normal hierarchical model

Let

$$Y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2)$$

for $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $\sum_{g=1}^{G} n_g = n$. Now consider the following model assumptions:

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

Normal hierarchical model

Let

$$Y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2)$$

for $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $\sum_{g=1}^{G} n_g = n$. Now consider the following model assumptions:

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

Normal hierarchical model

Let

$$Y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2)$$

for $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $\sum_{g=1}^{G} n_g = n$. Now consider the following model assumptions:

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

Normal hierarchical model

Let

$$Y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2)$$

for $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $\sum_{g=1}^{G} n_g = n$. Now consider the following model assumptions:

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

Normal hierarchical model

Let

$$Y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2)$$

for $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $\sum_{g=1}^{G} n_g = n$. 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 (in the case of the discrete mixture) π.

Normal hierarchical model

Consider the model

$$Y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2)$$

Normal hierarchical model

Consider the model

$$\begin{aligned} Y_{ig} &\stackrel{ind}{\sim} N(\theta_g, \sigma^2) \\ \theta_g &\stackrel{ind}{\sim} N(\mu, \tau^2) \end{aligned}$$

Normal hierarchical model

Consider the model

$$\begin{aligned} Y_{ig} &\stackrel{ind}{\sim} N(\theta_g, \sigma^2) \\ \theta_g &\stackrel{ind}{\sim} N(\mu, \tau^2) \end{aligned}$$

where $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $n = \sum_{g=1}^{G} n_g$

Normal hierarchical model

Consider the model

$$\begin{aligned} Y_{ig} &\stackrel{ind}{\sim} N(\theta_g, \sigma^2) \\ \theta_g &\stackrel{ind}{\sim} N(\mu, \tau^2) \end{aligned}$$

where $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $n = \sum_{g=1}^{G} n_g$ with prior distribution

$$p(\mu, \sigma^2, \tau^2) = p(\mu)p(\sigma^2)p(\tau^2) \propto \frac{1}{\sigma^2} Ca^+(\tau; 0, C).$$

Normal hierarchical model

Consider the model

$$\begin{aligned} Y_{ig} &\stackrel{ind}{\sim} N(\theta_g, \sigma^2) \\ \theta_g &\stackrel{ind}{\sim} N(\mu, \tau^2) \end{aligned}$$

where $i = 1, \ldots, n_g$, $g = 1, \ldots, G$, and $n = \sum_{g=1}^{G} n_g$ with prior distribution

$$p(\mu, \sigma^2, \tau^2) = p(\mu)p(\sigma^2)p(\tau^2) \propto \frac{1}{\sigma^2} Ca^+(\tau; 0, C).$$

For background on why we are using these priors for the variances, see Gelman (2006) `https://projecteuclid.org/euclid.ba/1340371048`: "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)".

Gibbs sampler for normal hierarchical model

Here is a possible Gibbs sampler for this model:

Gibbs sampler for normal hierarchical model

Here is a possible Gibbs sampler for this model:

- For $g = 1, \ldots, G$, sample $\theta_g \sim p(\theta_g | \cdots)$.

Gibbs sampler for normal hierarchical model

Here is a possible Gibbs sampler for this model:

- For $g = 1, \ldots, G$, sample $\theta_g \sim p(\theta_g | \cdots)$.
- Sample $\sigma^2 \sim p(\sigma^2 | \cdots)$.

Gibbs sampler for normal hierarchical model

Here is a possible Gibbs sampler for this model:

- For $g = 1, \ldots, G$, sample $\theta_g \sim p(\theta_g | \cdots)$.
- Sample $\sigma^2 \sim p(\sigma^2 | \cdots)$.
- Sample $\mu \sim p(\mu | \cdots)$.

Gibbs sampler for normal hierarchical model

Here is a possible Gibbs sampler for this model:

- For $g = 1, \ldots, G$, sample $\theta_g \sim p(\theta_g | \cdots)$.
- Sample $\sigma^2 \sim p(\sigma^2 | \cdots)$.
- Sample $\mu \sim p(\mu | \cdots)$.
- Sample $\tau^2 \sim p(\tau^2 | \cdots)$.

Gibbs sampler for normal hierarchical model

Here is a possible Gibbs sampler for this model:

- For $g = 1, \ldots, G$, sample $\theta_g \sim p(\theta_g|\cdots)$.
- Sample $\sigma^2 \sim p(\sigma^2|\cdots)$.
- Sample $\mu \sim p(\mu|\cdots)$.
- Sample $\tau^2 \sim p(\tau^2|\cdots)$.

How many steps exist in this Gibbs sampler?

Gibbs sampler for normal hierarchical model

Here is a possible Gibbs sampler for this model:

- For $g = 1, \ldots, G$, sample $\theta_g \sim p(\theta_g | \cdots)$.
- Sample $\sigma^2 \sim p(\sigma^2 | \cdots)$.
- Sample $\mu \sim p(\mu | \cdots)$.
- Sample $\tau^2 \sim p(\tau^2 | \cdots)$.

How many steps exist in this Gibbs sampler? G+3? 4?

2-Step Gibbs sampler for normal hierarchical model

Here is a 2-step Gibbs sampler:

2-Step Gibbs sampler for normal hierarchical model

Here is a 2-step Gibbs sampler:

1. Sample $\theta = (\theta_1, \ldots, G) \sim p(\theta|\cdots)$.

2-Step Gibbs sampler for normal hierarchical model

Here is a 2-step Gibbs sampler:

1. Sample $\theta = (\theta_1, \ldots, G) \sim p(\theta|\cdots)$.
2. Sample $\mu, \sigma^2, \tau^2 \sim p(\mu, \sigma^2, \tau^2|\cdots)$.

2-Step Gibbs sampler for normal hierarchical model

Here is a 2-step Gibbs sampler:

1. Sample $\theta = (\theta_1, \ldots, G) \sim p(\theta | \cdots)$.
2. Sample $\mu, \sigma^2, \tau^2 \sim p(\mu, \sigma^2, \tau^2 | \cdots)$.

There is stronger theoretical support for 2-step Gibbs sampler, thus, if we can, it is prudent to construct a 2-step Gibbs sampler.

Sampling θ

The full conditional for θ is

$$p(\theta|\cdots) \quad \propto p(\theta, \mu, \sigma^2, \tau^2|y)$$

Sampling θ

The full conditional for θ is

$$\begin{aligned} p(\theta|\cdots) &\propto p(\theta,\mu,\sigma^2,\tau^2|y) \\ &\propto p(y|\theta,\sigma^2)p(\theta|\mu,\tau^2)p(\mu,\sigma^2,\tau^2) \end{aligned}$$

Sampling θ

The full conditional for θ is

$$
\begin{aligned}
p(\theta|\cdots) &\propto p(\theta,\mu,\sigma^2,\tau^2|y) \\
&\propto p(y|\theta,\sigma^2)p(\theta|\mu,\tau^2)p(\mu,\sigma^2,\tau^2) \\
&\propto p(y|\theta,\sigma^2)p(\theta|\mu,\tau^2)
\end{aligned}
$$

Sampling θ

The full conditional for θ is

$$\begin{aligned}
p(\theta|\cdots) &\propto p(\theta,\mu,\sigma^2,\tau^2|y) \\
&\propto p(y|\theta,\sigma^2)p(\theta|\mu,\tau^2)p(\mu,\sigma^2,\tau^2) \\
&\propto p(y|\theta,\sigma^2)p(\theta|\mu,\tau^2) \\
&= \prod_{g=1}^{G} p(y_g|\theta_g,\sigma^2)p(\theta_g|\mu,\tau^2)
\end{aligned}$$

where $y_g = (y_{1g},\ldots,y_{n_g g})$.

Sampling θ

The full conditional for θ is

$$
\begin{array}{rl}
p(\theta|\cdots) & \propto p(\theta, \mu, \sigma^2, \tau^2|y) \\
& \propto p(y|\theta, \sigma^2)p(\theta|\mu, \tau^2)p(\mu, \sigma^2, \tau^2) \\
& \propto p(y|\theta, \sigma^2)p(\theta|\mu, \tau^2) \\
& = \prod_{g=1}^{G} p(y_g|\theta_g, \sigma^2)p(\theta_g|\mu, \tau^2)
\end{array}
$$

where $y_g = (y_{1g}, \ldots, y_{n_g g})$. We now know that the θ_g are conditionally independent of each other.

Sampling θ_g

The full conditional for θ_g is

Sampling θ_g

The full conditional for θ_g is

$$p(\theta_g|\cdots) \propto p(y_g|\theta_g,\sigma^2)p(\theta_g|\mu,\tau^2)$$

Sampling θ_g

The full conditional for θ_g is

$$\begin{aligned} p(\theta_g|\cdots) &\propto p(y_g|\theta_g,\sigma^2)p(\theta_g|\mu,\tau^2) \\ &= \prod_{i=1}^{n_g} N(y_{ig};\theta_g,\sigma^2)N(\theta_g;\mu,\tau^2) \end{aligned}$$

Sampling θ_g

The full conditional for θ_g is

$$\begin{aligned} p(\theta_g|\cdots) &\propto p(y_g|\theta_g,\sigma^2)p(\theta_g|\mu,\tau^2) \\ &= \prod_{i=1}^{n_g} N(y_{ig};\theta_g,\sigma^2)N(\theta_g;\mu,\tau^2) \end{aligned}$$

Notice that this does not include $\theta_{g'}$ for any $g' \neq g$.

Sampling θ_g

The full conditional for θ_g is

$$\begin{array}{rl} p(\theta_g|\cdots) & \propto p(y_g|\theta_g,\sigma^2)p(\theta_g|\mu,\tau^2) \\ & = \prod_{i=1}^{n_g} N(y_{ig};\theta_g,\sigma^2)N(\theta_g;\mu,\tau^2) \end{array}$$

Notice that this does not include $\theta_{g'}$ for any $g' \neq g$. This is an alternative way to conclude that the θ_g are conditionally independent of each other.

Sampling θ_g

The full conditional for θ_g is

$$\begin{aligned} p(\theta_g|\cdots) &\propto p(y_g|\theta_g,\sigma^2)p(\theta_g|\mu,\tau^2) \\ &= \prod_{i=1}^{n_g} N(y_{ig};\theta_g,\sigma^2)N(\theta_g;\mu,\tau^2) \end{aligned}$$

Notice that this does not include $\theta_{g'}$ for any $g' \neq g$. This is an alternative way to conclude that the θ_g are conditionally independent of each other.

Thus

$$\theta_g|\cdots \stackrel{ind}{\sim} N(\mu_g,\tau_g^2)$$

where

$$\begin{aligned} \tau_g^2 &= [\tau^{-2}+n_g\sigma^{-2}]^{-1} \\ \mu_g &= \tau_g^2[\mu\tau^{-2}+\overline{y}_g n_g\sigma^{-2}] \\ \overline{y}_g &= \frac{1}{n_g}\sum_{i=1}^{n_g} y_{ig}. \end{aligned}$$

Sampling μ, σ^2, τ^2

The full conditional for μ, σ^2, τ^2 is

$$p(\mu, \sigma^2, \tau^2 | \cdots) \quad \propto p(y|\theta, \sigma^2) p(\theta|\mu, \tau^2) p(\mu) p(\sigma^2) p(\tau^2)$$

Sampling μ, σ^2, τ^2

The full conditional for μ, σ^2, τ^2 is

$$
\begin{aligned}
p(\mu, \sigma^2, \tau^2 | \cdots) &\propto p(y|\theta, \sigma^2) p(\theta|\mu, \tau^2) p(\mu) p(\sigma^2) p(\tau^2) \\
&= p(y|\theta, \sigma^2) p(\sigma^2) p(\theta|\mu, \tau^2) p(\mu) p(\tau^2)
\end{aligned}
$$

Sampling μ, σ^2, τ^2

The full conditional for μ, σ^2, τ^2 is

$$\begin{aligned} p(\mu, \sigma^2, \tau^2 | \cdots) &\propto p(y|\theta, \sigma^2) p(\theta|\mu, \tau^2) p(\mu) p(\sigma^2) p(\tau^2) \\ &= p(y|\theta, \sigma^2) p(\sigma^2) p(\theta|\mu, \tau^2) p(\mu) p(\tau^2) \end{aligned}$$

So we know that σ^2 is independent of μ and τ^2.

Sampling σ^2

Recall that

$$y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2) \text{ and } p(\sigma^2) \propto 1/\sigma^2.$$

Sampling σ^2

Recall that

$$y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2) \text{ and } p(\sigma^2) \propto 1/\sigma^2.$$

Thus, we are in the scenario of normal data with a known mean and unknown variance and the unknown variance has our default prior.

Sampling σ^2

Recall that

$$y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2) \text{ and } p(\sigma^2) \propto 1/\sigma^2.$$

Thus, we are in the scenario of normal data with a known mean and unknown variance and the unknown variance has our default prior. Thus, we should know the full conditional is
$\sigma^2 | \cdots \sim IG\left(\frac{n}{2}, \frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{n_g}(y_{ig}-\theta_g)^2\right)$.

Sampling σ^2

Recall that

$$y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2) \text{ and } p(\sigma^2) \propto 1/\sigma^2.$$

Thus, we are in the scenario of normal data with a known mean and unknown variance and the unknown variance has our default prior. Thus, we should know the full conditional is
$\sigma^2| \cdots \sim IG\left(\frac{n}{2}, \frac{1}{2}\sum_{g=1}^G \sum_{i=1}^{n_g} (y_{ig} - \theta_g)^2\right)$.

To derive the full conditional, use

$$\begin{aligned} p(\sigma^2|\cdots) &\propto \prod_{g=1}^G \prod_{i=1}^{n_g} (\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(y_{ig} - \theta_g)^2\right) \frac{1}{\sigma^2} \\ &= (\sigma^2)^{-n/2-1} \exp\left(-\frac{1}{2}\sum_{g=1}^G \sum_{i=1}^{n_g} (y_{ig} - \theta_g)^2 \Big/ \sigma^2\right) \end{aligned}$$

Sampling σ^2

Recall that

$$y_{ig} \stackrel{ind}{\sim} N(\theta_g, \sigma^2) \text{ and } p(\sigma^2) \propto 1/\sigma^2.$$

Thus, we are in the scenario of normal data with a known mean and unknown variance and the unknown variance has our default prior. Thus, we should know the full conditional is
$\sigma^2|\cdots \sim IG\left(\frac{n}{2}, \frac{1}{2}\sum_{g=1}^G \sum_{i=1}^{n_g}(y_{ig} - \theta_g)^2\right)$.

To derive the full conditional, use

$$\begin{array}{rl} p(\sigma^2|\cdots) & \propto \prod_{g=1}^G \prod_{i=1}^{n_g} (\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(y_{ig} - \theta_g)^2\right) \frac{1}{\sigma^2} \\ & = (\sigma^2)^{-n/2-1} \exp\left(-\frac{1}{2}\sum_{g=1}^G \sum_{i=1}^{n_g}(y_{ig} - \theta_g)^2 \Big/ \sigma^2\right) \end{array}$$

which is the kernel of a $IG\left(\frac{n}{2}, \frac{1}{2}\sum_{g=1}^G \sum_{i=1}^{n_g}(y_{ig} - \theta_g)^2\right)$.

Sampling μ, τ^2

Recall that

$$\theta_g \stackrel{ind}{\sim} N(\mu, \tau^2) \text{ and } p(\mu, \tau^2) \propto Ca^+(\tau; 0, C).$$

Sampling μ, τ^2

Recall that

$$\theta_g \stackrel{ind}{\sim} N(\mu, \tau^2) \text{ and } p(\mu, \tau^2) \propto Ca^+(\tau; 0, C).$$

This is a non-standard distribution, but is extremely close a normal model with unknown mean and variance with the standard non-informative prior $p(\mu, \tau^2) \propto 1/\tau^2$ or the conjugate normal-inverse-gamma prior.

Sampling μ, τ^2

Recall that

$$\theta_g \stackrel{ind}{\sim} N(\mu, \tau^2) \text{ and } p(\mu, \tau^2) \propto Ca^+(\tau; 0, C).$$

This is a non-standard distribution, but is extremely close a normal model with unknown mean and variance with the standard non-informative prior $p(\mu, \tau^2) \propto 1/\tau^2$ or the conjugate normal-inverse-gamma prior.

Here are some options for sampling from this distribution:

- random-walk Metropolis (in 2 dimensions),
- independent Metropolis-Hastings using posterior from standard non-informative prior as the proposal, or
- rejection sampling using posterior from standard non-informative prior as the proposal

The posterior under the standard non-informative prior is

$$\tau^2 | \cdots \sim \text{Inv-}\chi^2(G-1, s_\theta^2) \text{ and } \mu|\tau^2, \ldots \sim N(\overline{\theta}, \tau^2/G)$$

where $\overline{\theta} = \frac{1}{G}\sum_{g=1}^G \theta_g$ and $s_\theta^2 = \frac{1}{G-1}(\theta_g - \overline{\theta})^2$.

Sampling μ, τ^2

Recall that

$$\theta_g \stackrel{ind}{\sim} N(\mu, \tau^2) \text{ and } p(\mu, \tau^2) \propto Ca^+(\tau; 0, C).$$

This is a non-standard distribution, but is extremely close a normal model with unknown mean and variance with the standard non-informative prior $p(\mu, \tau^2) \propto 1/\tau^2$ or the conjugate normal-inverse-gamma prior.

Here are some options for sampling from this distribution:

- random-walk Metropolis (in 2 dimensions),
- independent Metropolis-Hastings using posterior from standard non-informative prior as the proposal, or
- rejection sampling using posterior from standard non-informative prior as the proposal

The posterior under the standard non-informative prior is

$$\tau^2|\cdots \sim \text{Inv-}\chi^2(G-1, s_\theta^2) \text{ and } \mu|\tau^2, \ldots \sim N(\overline{\theta}, \tau^2/G)$$

where $\overline{\theta} = \frac{1}{G}\sum_{g=1}^G \theta_g$ and $s_\theta^2 = \frac{1}{G-1}(\theta_g - \overline{\theta})^2$. What is the MH ratio?

Markov chain Monte Carlo for normal hierarchical model

1. Sample $\theta \sim p(\theta|\cdots)$:
   a. For $g = 1, \ldots, G$, sample $\theta_g \sim N(\mu_g, \tau_g^2)$.
2. Sample μ, σ^2, τ^2:
   a. Sample $\sigma^2 \sim IG(n/2, SSE)$.
   b. Sample μ, τ^2 using independent Metropolis-Hastings using posterior from standard non-informative prior as the proposal.

Markov chain Monte Carlo for normal hierarchical model

1. Sample $\theta \sim p(\theta|\cdots)$:
   a. For $g = 1, \ldots, G$, sample $\theta_g \sim N(\mu_g, \tau_g^2)$.
2. Sample μ, σ^2, τ^2:
   a. Sample $\sigma^2 \sim IG(n/2, SSE)$.
   b. Sample μ, τ^2 using independent Metropolis-Hastings using posterior from standard non-informative prior as the proposal.

What happens if $\theta_g \stackrel{ind}{\sim} La(\mu, \tau)$ or $\theta_g \stackrel{ind}{\sim} t_v(\mu, \tau^2)$?

Scale mixtures of normals

Recall that if

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

Scale mixtures of normals

Recall that if

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

then

$$\theta$$

Scale mixtures of normals

Recall that if

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

then

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

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

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

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,

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.

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

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

t distribution

Let

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

then

$$p(\theta) = \int p(\theta|\phi)p(\phi)d\phi$$

t distribution

Let

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

then

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

t distribution

Let

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

then

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

t distribution

Let

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

then

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

t distribution

Let

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

then

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

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

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

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, scale bC/a, and variance $\frac{bC}{a-1}$.

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, scale bC/a, and variance $\frac{bC}{a-1}$.

Hierarchical t distribution

Let $m = \mu$, $C = 1$, $a = \nu/2$, and $b = \nu\tau^2/2$,

Hierarchical t distribution

Let $m = \mu$, $C = 1$, $a = \nu/2$, and $b = \nu\tau^2/2$, i.e.

$$\theta|\phi \sim N(\mu, \phi) \text{ and } \phi \sim IG(\nu/2, \nu\tau^2/2).$$

Hierarchical t distribution

Let $m = \mu$, $C = 1$, $a = \nu/2$, and $b = \nu\tau^2/2$, i.e.

$$\theta|\phi \sim N(\mu, \phi) \text{ and } \phi \sim IG(\nu/2, \nu\tau^2/2).$$

Then, we have

$$\theta \sim t_\nu(\mu, \tau^2),$$

i.e. a t distribution with ν degrees of freedom, location μ, and scale τ^2.

Hierarchical t distribution

Let $m = \mu$, $C = 1$, $a = \nu/2$, and $b = \nu\tau^2/2$, i.e.

$$\theta|\phi \sim N(\mu, \phi) \text{ and } \phi \sim IG(\nu/2, \nu\tau^2/2).$$

Then, we have

$$\theta \sim t_\nu(\mu, \tau^2),$$

i.e. a t distribution with ν degrees of freedom, location μ, and scale τ^2.

Notice that the parameterization has a redundancy between C and a/b, i.e. we could have chosen $C = \tau^2$, $a = \nu/2$, and $b = \nu/2$ and we would have obtained the same marginal distribution for θ.

Laplace distribution

Let

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

where $E[\phi] = 2b^2$ and $Var[\phi] = 4b^4$.

Laplace distribution

Let

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

where $E[\phi] = 2b^2$ and $Var[\phi] = 4b^4$. then, by an extension of equation (4) in Park and Casella (2008), we have

$$p(\theta) = \frac{1}{2Cb} e^{-\frac{|\theta - m|}{Cb}}.$$

Laplace distribution

Let

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

where $E[\phi] = 2b^2$ and $Var[\phi] = 4b^4$. then, by an extension of equation (4) in Park and Casella (2008), we have

$$p(\theta) = \frac{1}{2Cb} e^{-\frac{|\theta - m|}{Cb}}.$$

This is the pdf for a Laplace (double exponential) distribution with location m and scale Cb

Laplace distribution

Let

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

where $E[\phi] = 2b^2$ and $Var[\phi] = 4b^4$. then, by an extension of equation (4) in Park and Casella (2008), we have

$$p(\theta) = \frac{1}{2Cb} e^{-\frac{|\theta - m|}{Cb}}.$$

This is the pdf for a Laplace (double exponential) distribution with location m and scale Cb which we write

$$\theta \sim La(m, Cb).$$

Laplace distribution

Let

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

where $E[\phi] = 2b^2$ and $Var[\phi] = 4b^4$. then, by an extension of equation (4) in Park and Casella (2008), we have

$$p(\theta) = \frac{1}{2Cb} e^{-\frac{|\theta - m|}{Cb}}.$$

This is the pdf for a Laplace (double exponential) distribution with location m and scale Cb which we write

$$\theta \sim La(m, Cb).$$

Laplace distribution

Let

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

where $E[\phi] = 2b^2$ and $Var[\phi] = 4b^4$. then, by an extension of equation (4) in Park and Casella (2008), we have

$$p(\theta) = \frac{1}{2Cb} e^{-\frac{|\theta - m|}{Cb}}.$$

This is the pdf for a Laplace (double exponential) distribution with location m and scale Cb which we write

$$\theta \sim La(m, Cb).$$

and say θ has a Laplace distribution with location m and scale Cb

Laplace distribution

Let

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

where $E[\phi] = 2b^2$ and $Var[\phi] = 4b^4$. then, by an extension of equation (4) in Park and Casella (2008), we have

$$p(\theta) = \frac{1}{2Cb} e^{-\frac{|\theta - m|}{Cb}}.$$

This is the pdf for a Laplace (double exponential) distribution with location m and scale Cb which we write

$$\theta \sim La(m, Cb).$$

and say θ has a Laplace distribution with location m and scale Cb and $E[\theta] = m$ and $Var[\theta] = 2[Cb]^2 = 2C^2b^2$.

Hierarchical Laplace distribution

Let $m = \mu$, $C = 1$, and $b = \tau$

Hierarchical Laplace distribution

Let $m = \mu$, $C = 1$, and $b = \tau$ i.e.

$$\theta|\phi \sim N(\mu, \phi) \text{ and } \phi \sim Exp(1/2\tau^2).$$

Hierarchical Laplace distribution

Let $m = \mu$, $C = 1$, and $b = \tau$ i.e.

$$\theta|\phi \sim N(\mu, \phi) \text{ and } \phi \sim Exp(1/2\tau^2).$$

Then, we have

$$\theta \sim La(\mu, \tau),$$

i.e. a Laplace distribution with location μ and scale τ.

Hierarchical Laplace distribution

Let $m = \mu$, $C = 1$, and $b = \tau$ i.e.

$$\theta|\phi \sim N(\mu, \phi) \text{ and } \phi \sim Exp(1/2\tau^2).$$

Then, we have

$$\theta \sim La(\mu, \tau),$$

i.e. a Laplace distribution with location μ and scale τ.

Notice that the parameterization has a redundancy between C and b, i.e. we could have chosen $C = \tau$ and $b = 1$ and we would have obtained the same marginal distribution for θ.

Normal hierarchical model

Recall our hierarchical model

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

for $i = 1, \ldots, \mathrm{I}$ and $j = 1, \ldots, n_i$.

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:

Normal hierarchical model

Recall our hierarchical model

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

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

- $\theta_i \overset{ind}{\sim} N(\mu, \phi_i), \phi_i = \tau^2$

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:

- $\theta_i \stackrel{ind}{\sim} N(\mu, \phi_i), \phi_i = \tau^2$

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:

- $\theta_i \stackrel{ind}{\sim} N(\mu, \phi_i), \phi_i = \tau^2 \implies \theta_i \stackrel{ind}{\sim} N(\mu, \tau^2)$

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:

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

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:

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

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:

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

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:

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

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:

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

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:

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

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

Gibbs sampling

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

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

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

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

Sample μ

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

Immediately, we should know that

$$\mu| \cdots \sim$$

Sample μ

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

Immediately, we should know that

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

with

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

Sample θ

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

Sample θ

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

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

Sample θ

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

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

Thus θ_i are conditionally independent given everything else.

Sample θ

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

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

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

$$\theta_i|\cdots \sim$$

Sample θ

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

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

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

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

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

Sample σ^2

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

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.

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.

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

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

where

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

Distributional assumption for θ_i

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

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

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,

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

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

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

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

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

So the individual ϕ_i are conditionally independent

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

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

So the individual ϕ_i are conditionally independent with

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

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

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

So the individual ϕ_i are conditionally independent with

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

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

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

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

So the individual ϕ_i are conditionally independent with

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

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

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

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

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

So the individual ϕ_i are conditionally independent with

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

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

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

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,

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

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

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

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

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

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

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

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

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

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

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 normal model

Let

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

so the full conditional is

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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.

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

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

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.

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

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

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