Hierarchical Normal Model
Technical Notes

Irving Gómez Méndez

Assume that there are J groups, that we count with a sample Y_{ij}, $i = 1, \dots, n_j$ for the j-th group, where each observation is independent from other observations within the same group and from observations of other groups, and that $Y_{ij}|\theta_j, \sigma_j^2 \sim \mathcal{N}(\theta_j, \sigma_j^2)$ for $i = 1, \dots, n_j$ and $j = 1, \dots, J$.

1 Nonhierarchical models with σ^2 known

Before moving to the hierarchical model, we first consider two simple nonhierarchical models– estimating the mean of each group independently, and complete pooling. Assume for this section that the variance within each group σ_j^2 is known.

1.1 Separate estimates

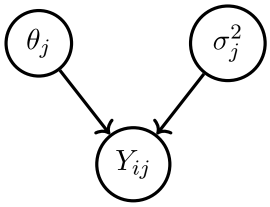

Figure 1: Separate estimates.

Denote by

$$\bar{Y}_{\cdot j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}$$

the sample mean of each group j with variance $\bar{\sigma}_j^2 = \sigma_j^2 / n_j$, then

$$\bar{Y}_{\cdot j}|\theta_j, \sigma_j^2 \sim \mathcal{N}(\theta_j, \bar{\sigma}_j^2).$$

Because we are considering all the within-group variances σ_j's to be known, the likelihood of our model is determined just by the likelihood of each $\bar{Y}_{\cdot j}$, assume that all the θ_j's are independent and $p(\theta_j|\boldsymbol{\sigma}^2) \propto \mathbb{1}_{(-\infty,\infty)}(\theta_j)$, then the posterior distribution for each θ_j is

$$\theta_j|\boldsymbol{\sigma}^2, \mathbf{Y} \sim \mathcal{N}(\bar{Y}_{\cdot j}, \bar{\sigma}_j^2).$$

And

$$Y_j|\boldsymbol{\sigma}^2, \mathbf{Y} \sim \mathcal{N}(\bar{Y}_{\cdot j}, \sigma_j^2 + \bar{\sigma}_j^2),$$

where Y_j represents an observation within group j.

1.2 Pooled estimate

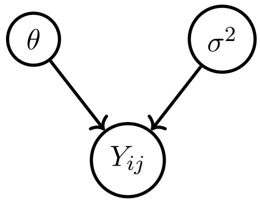

Figure 2: Pooled estimate.

Consider now that we restrict the values θ_j to be equal to a common mean θ and all the values σ_j^2 to be equal to some common known σ^2. Thus

$$\bar{Y}_{\cdot j}|\theta, \sigma^2 \sim \mathcal{N}(\theta, \bar{\sigma}_j^2),$$

where $\bar{\sigma}^2 = \sigma^2/n_j$.

Assume $p(\theta|\sigma^2) \propto \mathbb{1}_{(-\infty,\infty)}(\theta)$, then

$$\begin{aligned} p(\theta|\sigma^2, \mathbf{Y}) &\propto \exp\left\{-\frac{1}{2}\sum_{j=1}^{J}\frac{(\bar{Y}_{\cdot j}-\theta)^2}{\bar{\sigma}_j^2}\right\}\mathbb{1}_{(-\infty,\infty)}(\theta) \\ &= \exp\left\{-\frac{1}{2}\sum_{j=1}^{J}\frac{\bar{Y}_{\cdot j}^2 - 2\bar{Y}_{\cdot j}\theta + \theta^2}{\bar{\sigma}_j^2}\right\}\mathbb{1}_{(-\infty,\infty)}(\theta) \\ &\propto \exp\left\{-\frac{1}{2}\left(\theta^2\sum_{j=1}^{J}\frac{1}{\bar{\sigma}_j^2} - 2\theta\sum_{j=1}^{J}\frac{\bar{Y}_{\cdot j}}{\bar{\sigma}_j^2}\right)\right\}\mathbb{1}_{(-\infty,\infty)}(\theta). \end{aligned}$$

On the other hand, since the logarithm of the posterior is a quadratic function on θ, then $\theta|\sigma^2, \mathbf{Y} \sim \mathcal{N}(\mu, \varphi^2)$, thus

$$\begin{aligned} p(\theta|\sigma^2, \mathbf{Y}) &\propto \exp\left\{-\frac{(\theta-\mu)^2}{2\varphi^2}\right\}\mathbb{1}_{(-\infty,\infty)}(\theta) \\ &= \exp\left\{-\frac{1}{2}\left(\theta^2\frac{1}{\varphi^2} - 2\theta\frac{\mu}{\varphi^2} + \frac{\mu^2}{\varphi^2}\right)\right\}\mathbb{1}_{(-\infty,\infty)}(\theta) \\ &\propto \exp\left\{-\frac{1}{2}\left(\theta^2\frac{1}{\varphi^2} - 2\theta\frac{\mu}{\varphi^2}\right)\right\}\mathbb{1}_{(-\infty,\infty)}(\theta). \end{aligned}$$

From the last expression, we recognize

$$\varphi^2 = \frac{1}{\sum_{j=1}^{J}\frac{1}{\bar{\sigma}_j^2}}$$

that is, the precision is the sum of all the precisions, and

$$\frac{\mu}{\varphi^2} = \sum_{j=1}^{J}\frac{\bar{Y}_{\cdot j}}{\bar{\sigma}_j^2} \Rightarrow \mu = \frac{\sum_{j=1}^{J}\frac{\bar{Y}_{\cdot j}}{\bar{\sigma}_j^2}}{\sum_{j=1}^{J}\frac{1}{\bar{\sigma}_j^2}} \equiv \bar{Y}_{\cdot\cdot}.$$

That is,

$$\theta|\sigma^2, \mathbf{Y} \sim \mathcal{N}(\bar{Y}_{\cdot\cdot}, \varphi^2),$$

and

$$Y_j|\sigma^2, \mathbf{Y} \sim \mathcal{N}(\bar{Y}_{\cdot\cdot}, \sigma^2 + \varphi^2).$$

2 Nonhierarchical models with σ^2 unknown

For this section we consider the more realistic approach where the within-group variances σ_j^2's are unknown. The objective is to determine the full conditional posterior distributions for the parameters. Once we have deduced the full conditional posterior for all our parameters, it is straightforward to implement the Gibbs sampler algorithm, and obtain a sample from the joint posterior distribution.

2.1 Separate estimates

Consider the model

$$Y_{ij}|\theta_j, \sigma_j^2 \sim \mathcal{N}(\theta_j, \sigma_j^2),$$

with the noninformative prior

$$p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2) \propto \prod_{j=1}^{J} \frac{1}{\sigma_j^2} \mathbb{1}_{(-\infty,\infty)}(\theta_j) \mathbb{1}_{(0,\infty)}(\sigma_j^2).$$

We already deduced that the conditional posterior for each θ_j is

$$\theta_j|\boldsymbol{\sigma}^2, \mathbf{Y} \sim \mathcal{N}(\bar{Y}_{\cdot j}, \bar{\sigma}_j^2).$$

On the other hand, note that

$$\begin{aligned} p(\sigma_j^2|\boldsymbol{\theta}, \mathbf{Y}) &\propto \frac{1}{\sigma_j^2} \mathbb{1}_{(0,\infty)}(\sigma_j^2) \prod_{i=1}^{n_j} \mathcal{N}(Y_{ij}|\theta_j, \sigma_j^2) \\ &\propto \frac{1}{\sigma_j^2} \mathbb{1}_{(0,\infty)}(\sigma_j^2) \frac{1}{(\sigma_j^2)^{n_j/2}} \exp\left\{ -\frac{1}{2\sigma_j^2} \sum_{i=1}^{n_j} (Y_{ij} - \theta_j)^2 \right\} \\ &= (\sigma_j^2)^{-(n_j/2+1)} \exp\left\{ -\frac{n_j \hat{\sigma}_j^2}{2\sigma_j^2} \right\} \mathbb{1}_{(0,\infty)}(\sigma^2), \end{aligned}$$

from this expression we observe that

$$\sigma_j^2|\boldsymbol{\theta}, \mathbf{Y} \sim \textsf{Inverse-}\chi^2(n_j, \hat{\sigma}_j^2),$$

where

$$\hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (Y_{ij} - \theta_j)^2.$$

Algorithm 1 Gibbs sampler for separate models.

Input: Sample y_{ij} ($i = 1, \ldots, n_j$, $j = 1, \ldots, J$), observed groups' averages $\bar{y}_{\cdot j}$, and posterior sample size S.

Output: Posterior sample for $\theta_j^{(s)}$, $\sigma_j^{2(s)}$ and $Y_j^{(s)}$ ($j = 1, \ldots, J$, $s = 1, \ldots, S$).

Set $\theta_j^{(0)} = \bar{y}_{\cdot j}$
for $s = 1, \ldots, S$ **do**
  for $j = 1, \ldots, J$ **do**
    Compute $\hat{\sigma}_j^{2(s)} = \frac{1}{n_j} \sum_{i=1}^{n_j} \left(y_{ij} - \theta_j^{(s-1)} \right)^2$
    Simulate $\sigma_j^{2(s)} \sim \textsf{Inverse-}\chi^2 \left(n_j, \hat{\sigma}_j^{2(s)} \right)$
    Simulate $\theta_j^{(s)} \sim \mathcal{N} \left(\bar{y}_{\cdot j}, \sigma_j^{2(s)} / n_j \right)$
    Simulate $Y_j^{(s)} \sim \mathcal{N} \left(\theta_j^{(s)}, \sigma_j^{2(s)} \right)$
  end for
end for

From the simulated sample we can estimate posterior intervals. Overlapping intervals for the θ's and the σ's might be considered as evidence against the use of separate models in favor of a pooled estimate.

2.2 Pooled estimate

Consider the model where all the groups share the same mean and variance, that is

$$Y_{ij} | \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2),$$

and the noninformative prior given by

$$p(\theta, \sigma^2) \propto \frac{1}{\sigma^2} \mathbb{1}_{(-\infty,\infty)}(\theta) \mathbb{1}_{(0,\infty)}(\sigma^2).$$

We already deduced that the conditional posterior for θ is

$$\theta | \sigma^2, \mathbf{Y} \sim \mathcal{N}(\bar{Y}_{\cdot\cdot}, \varphi^2),$$

where

$$\bar{Y}_{\cdot\cdot} = \frac{\sum_{j=1}^{J} \frac{\bar{Y}_{\cdot j}}{\bar{\sigma}_j^2}}{\sum_{j=1}^{J} \frac{1}{\bar{\sigma}_j^2}},$$

$$\varphi^2 = \frac{1}{\sum_{j=1}^{J} \frac{1}{\bar{\sigma}_j^2}}$$

and

$$\bar{\sigma}_j^2 = \frac{\sigma^2}{n_j}, \quad j = 1, \ldots, J.$$

On the other hand, note that

$$p(\sigma^2 | \theta, \mathbf{Y}) \propto \frac{1}{\sigma^2} \mathbb{1}_{(0,\infty)}(\sigma^2) \prod_{j=1}^{J} \prod_{i=1}^{n_j} \mathcal{N}(Y_{ij} | \theta, \sigma^2),$$

where we can deduced easily that

$$\sigma^2|\theta, \mathbf{Y} \sim \textsf{Inverse-}\chi^2(n, \hat{\sigma}^2),$$

where $n = \sum_{j=1}^{J} n_j$, and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{J} \sum_{i=1}^{n_j} (Y_{ij} - \theta)^2.$$

Algorithm 2 Gibbs sampler for the pooled model.

Input: Sample y_{ij} $(i = 1, \ldots, n_j,\ j = 1, \ldots, J)$, observed groups' averages $\bar{y}_{\cdot j}$, observed groups' sample variances s_j^2, and posterior sample size S.

Output: Posterior sample for $\theta_j^{(s)}$, $\sigma_j^{2(s)}$ and $Y_j^{(s)}$ $(j = 1, \ldots, J,\ s = 1, \ldots, S)$.

Set

$$\theta^{(0)} = \frac{\sum_{j=1}^{J} \frac{\bar{y}_{\cdot j}}{s_j^2/n_j}}{\sum_{j=1}^{J} \frac{1}{s_j^2/n_j}}$$

for $s = 1, \ldots, S$ **do**

Compute $\hat{\sigma}^{2(s)} = \frac{1}{n} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \left(y_{ij} - \theta^{(s-1)}\right)^2$

Simulate $\sigma_1^{2(s)}, \ldots, \sigma_J^{2(s)} \sim \textsf{Inverse-}\chi^2\left(n, \hat{\sigma}^{2(s)}\right)$

Compute

$$\hat{\theta}^{(s)} = \frac{\sum_{j=1}^{J} \frac{\bar{y}_{\cdot j}}{\sigma_j^{2(s)}/n_j}}{\sum_{j=1}^{J} \frac{1}{\sigma_j^{2(s)}/n_j}}$$

and

$$\varphi^{2(s)} = \frac{1}{\sum_{j=1}^{J} \frac{1}{\sigma_j^{2(s)}/n_j}}$$

Simulate $\theta_1^{(s)}, \ldots, \theta_J^{(s)} \sim \mathcal{N}\left(\hat{\theta}^{(s)}, \varphi^{2(s)}\right)$

Simulate $Y_j^{(s)} \sim \mathcal{N}\left(\theta_j^{(s)}, \sigma_j^{2(s)}\right)$, $j = 1, \ldots, J$

end for

---

3 Hierarchical model with common σ^2 known

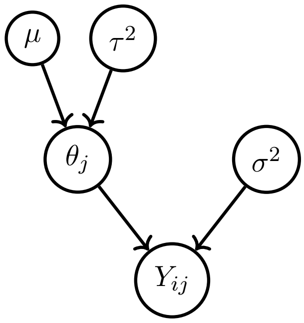

Figure 3: Hierarchical model with common σ^2.

Let be $Y_{ij}|\theta_j, \sigma^2 \sim \mathcal{N}(\theta_j, \sigma^2)$ for $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$, then $\bar{Y}_{\cdot j}|\theta_j, \sigma^2 \sim \mathcal{N}(\theta_j, \bar{\sigma}_j^2)$, where $\bar{\sigma}_j^2 = \sigma^2/n_j$. For the convenience of a conjugate model, we assume

$$\theta_j|\mu, \tau^2 \sim \mathcal{N}(\mu, \tau^2).$$

Due to the conjugacy, we can easily determine the posterior distribution for all the $\theta_j|\mu, \tau^2, \sigma^2, \mathbf{Y}$, which are independent and

$$\theta_j|\mu, \tau^2, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\hat{\theta}_j, V_{\theta_j}),$$

where

$$\hat{\theta}_j = \frac{\frac{1}{\bar{\sigma}_j^2}\bar{Y}_{\cdot j} + \frac{1}{\tau^2}\mu}{\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}}$$

and

$$V_{\theta_j} = \frac{1}{\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}}.$$

It can also been shown easily that the posterior distribution of the observations $Y_j|\mu, \tau^2, \sigma^2, \mathbf{Y}$ are independent and

$$Y_j|\mu, \tau^2, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\hat{\theta}_j, \sigma^2 + V_{\theta_j}).$$

3.1 Empirical Bayes

To assign values for the parameters μ, τ^2 and σ^2, we can take an empirical approach based on the analysis of variance. Let be

$$\bar{n} = \frac{1}{J}\sum_{j=1}^{J} n_j,$$

the mean square within groups MS_W is given by

$$MS_W = \frac{1}{J(\bar{n}-1)}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_{\cdot j})^2.$$

Because $\mathbb{E}(MS_W|\sigma^2) = \sigma^2$ (see page 116 of BDA 3), then an unbiased estimator for σ^2 is given by $\hat{\sigma}^2 = MS_W$. Thus, in the case where σ^2 is not known, we could use $\hat{\sigma}^2$ and act as if it was the real value of σ^2.

On the other hand, let be

$$\bar{Y}_{\cdot\cdot} = \frac{\sum_{j=1}^{J} \frac{n_j}{\sigma^2}\bar{Y}_{\cdot j}}{\sum_{j=1}^{J} \frac{n_j}{\sigma^2}},$$

the mean square between groups MS_B is given by

$$MS_B = \frac{1}{J-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2.$$

Because $\mathbb{E}(MS_B|\sigma^2, \tau^2) = \bar{n}\tau^2 + \sigma^2$ (see page 116 of BDA 3), then unbiased estimators for μ and τ^2 are given by $\hat{\mu} = \bar{Y}_{\cdot\cdot}$ and

$$\hat{\tau}^2 = \frac{MS_B - MS_W}{\bar{n}}.$$

3.2 Full Bayesian approach

Instead of using an empirical approach to assign values for μ, τ^2 and σ^2 based on point estimates, from a fully Bayesian approach we should assign a joint prior distribution $p(\mu, \tau|\sigma^2)$. Because we are considering σ^2 known, then the likelihood of the model can be represented by the sufficient statistics $\bar{Y}_{\cdot j}$, which satisfy that $\bar{Y}_{\cdot j}|\theta_j, \sigma^2 \sim \mathcal{N}(\theta_j, \bar{\sigma}_j^2)$. Considering $\theta_j|\mu, \tau^2 \sim \mathcal{N}(\mu, \tau^2)$, then

$$\bar{Y}_{\cdot j}|\mu, \tau^2, \sigma^2 \sim \mathcal{N}(\mu, \bar{\sigma}_j^2 + \tau^2).$$

The posterior distribution of μ and τ could be calculated from

$$\begin{aligned} p(\mu, \tau|\sigma^2, \mathbf{Y}) &\propto p(\mu, \tau|\sigma^2)\prod_{j=1}^{J} p(\bar{Y}_{\cdot j}|\mu, \tau, \sigma^2) \\ &\propto p(\mu, \tau|\sigma^2)\exp\left\{-\frac{1}{2}\sum_{j=1}^{J}\frac{(\bar{Y}_{\cdot j} - \mu)^2}{\bar{\sigma}_j^2 + \tau^2}\right\}. \end{aligned}$$

3.2.1 Conditional posterior distribution of μ, $p(\mu|\tau, \sigma^2, \mathbf{Y})$

Consider $p(\mu, \tau|\sigma^2) = p(\mu|\tau, \sigma^2)p(\tau|\sigma^2)$ and take $p(\mu|\tau, \sigma^2) \propto \mathbb{1}_{(-\infty,\infty)}(\mu)$, thus

$$p(\mu, \tau|\sigma^2, \mathbf{Y}) \propto p(\tau|\sigma^2)\exp\left\{-\frac{1}{2}\sum_{j=1}^{J}\frac{(\bar{Y}_{\cdot j} - \mu)^2}{\bar{\sigma}_j^2 + \tau^2}\right\}\mathbb{1}_{(-\infty,\infty)}(\mu),$$

then

$$p(\mu|\tau, \sigma^2, \mathbf{Y}) \propto \exp\left\{-\frac{1}{2}\sum_{j=1}^{J}\frac{(\bar{Y}_{\cdot j} - \mu)^2}{\bar{\sigma}_j^2 + \tau^2}\right\}\mathbb{1}_{(-\infty,\infty)}(\mu).$$

From our analysis on the pooled estimate, we recognize immediately that

$$\mu|\tau, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\hat{\mu}, V_\mu),$$

where

$$\hat{\mu} = \frac{\sum_{j=1}^{J}\frac{\bar{Y}_{\cdot j}}{\bar{\sigma}_j^2 + \tau^2}}{\sum_{j=1}^{J}\frac{1}{\bar{\sigma}_j^2 + \tau^2}}, \text{ and } V_\mu = \frac{1}{\sum_{j=1}^{J}\frac{1}{\bar{\sigma}_j^2 + \tau^2}}.$$

3.2.2 $\mathbb{E}(\theta_j|\tau, \sigma^2, \mathbf{Y})$ and $\mathbb{V}(\theta_j|\tau, \sigma^2, \mathbf{Y})$

Using these expressions, we can compute

$$\begin{aligned}
\mathbb{E}(\theta_j|\tau, \sigma^2, \mathbf{Y}) &= \mathbb{E}_{\mu|\tau,\sigma^2,\mathbf{Y}}[\mathbb{E}(\theta_j|\mu, \tau^2, \sigma^2, \mathbf{Y})] \\
&= \mathbb{E}_{\mu|\tau,\sigma^2,\mathbf{Y}}[\hat{\theta}_j] \\
&= \mathbb{E}_{\mu|\tau,\sigma,^2\mathbf{Y}}\left[\frac{\frac{1}{\bar{\sigma}_j^2}\bar{Y}_{\cdot j} + \frac{1}{\tau^2}\mu}{\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}}\right] \\
&= \frac{\frac{1}{\bar{\sigma}_j^2}\bar{Y}_{\cdot j} + \frac{1}{\tau^2}\hat{\mu}}{\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}(\theta_j|\tau, \sigma^2, \mathbf{Y}) &= \mathbb{E}_{\mu|\tau,\sigma^2,\mathbf{Y}}[\mathbb{V}(\theta_j|\mu, \tau^2, \sigma^2, \mathbf{Y})] + \mathbb{V}_{\mu|\tau,\sigma^2,\mathbf{Y}}[\mathbb{E}(\theta_j|\mu, \tau^2, \sigma^2, \mathbf{Y})] \\
&= \mathbb{E}_{\mu|\tau,\sigma^2,\mathbf{Y}}[V_{\theta_j}] + \mathbb{V}_{\mu|\tau,\sigma^2,\mathbf{Y}}[\hat{\theta}_j] \\
&= \mathbb{E}_{\mu|\tau,\sigma^2,\mathbf{Y}}\left[\frac{1}{\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}}\right] + \mathbb{V}_{\mu|\tau,\sigma^2,\mathbf{Y}}\left[\frac{\frac{1}{\bar{\sigma}_j^2}\bar{Y}_{\cdot j} + \frac{1}{\tau^2}\mu}{\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}}\right] \\
&= \frac{1}{\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}} + \frac{\left(\frac{1}{\tau^2}\right)^2 V_\mu}{\left(\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}\right)^2}
\end{aligned}$$

Because $\theta_j|\tau^2, \sigma^2, \mathbf{Y}$ follows a normal distribution, we conclude that

$$\theta_j|\tau^2, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\mathbb{E}(\theta_j|\tau, \sigma^2, \mathbf{Y}), \mathbb{V}(\theta_j|\tau, \sigma^2, \mathbf{Y}))$$

and

$$Y_j|\tau^2, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\mathbb{E}(\theta_j|\tau, \sigma^2, \mathbf{Y}), \sigma^2 + \mathbb{V}(\theta_j|\tau, \sigma^2, \mathbf{Y})).$$

Note that $\mathbb{E}(\theta_j|\tau, \sigma^2, \mathbf{Y}) \xrightarrow[\tau\to 0]{} \hat{\mu}$ and $\hat{\mu} \xrightarrow[\tau\to 0]{} \frac{\sum_{j=1}^J \frac{\bar{Y}_{\cdot j}}{\bar{\sigma}_j^2}}{\sum_{j=1}^J \frac{1}{\bar{\sigma}_j^2}} \equiv \bar{Y}_{\cdot\cdot}$, while $\mathbb{V}(\theta_j|\tau, \sigma^2, \mathbf{Y}) \xrightarrow[\tau\to 0]{} V_\mu$ and $V_\mu \xrightarrow[\tau\to 0]{} \frac{1}{\sum_{j=1}^J \frac{1}{\bar{\sigma}_j^2}} \equiv \varphi^2$. On the other hand, $\mathbb{E}(\theta_j|\tau, \sigma^2, \mathbf{Y}) \xrightarrow[\tau\to\infty]{} \bar{Y}_{\cdot j}$, while $\mathbb{V}(\theta_j|\tau, \sigma^2, \mathbf{Y}) \xrightarrow[\tau\to\infty]{} \bar{\sigma}_j^2$.

This result is consistent from a classical analysis of variance, if the ratio of between to within mean squares is significantly greater than 1, then the analysis of variance suggests separate estimates, $\hat{\theta}_j = \bar{Y}_{\cdot j}$, at the same time this would also mean that τ is large. If the ratio of mean squares is not 'statistical significant' different from 1, then pooling is reasonable and $\hat{\theta}_j = \bar{Y}_{\cdot\cdot}$, for all j, which would also mean that the F test cannot reject the hypothesis $\tau = 0$. From these observations we can conclude that the Bayesian analysis under the hierarchical model provides a compromise that combines information from all the groups without assuming all the θ_j's to be equal.

3.2.3 Conditional posterior distribution of τ, $p(\tau|\sigma^2, \mathbf{Y})$

Because $p(\mu, \tau|\sigma^2, \mathbf{Y}) = p(\mu|\tau, \sigma^2, \mathbf{Y})p(\tau|\sigma^2, \mathbf{Y})$, then

$$\begin{aligned}
p(\tau|\sigma^2, \mathbf{Y}) &= \frac{p(\mu, \tau|\sigma^2, \mathbf{Y})}{p(\mu|\tau, \sigma^2, \mathbf{Y})} \\
&\propto \frac{p(\mu, \tau|\sigma^2) \prod_{j=1}^{J} p(\bar{Y}_{\cdot j}|\mu, \tau^2, \sigma^2)}{p(\mu|\tau, \sigma^2, \mathbf{Y})} \\
&= \frac{p(\mu|\tau, \sigma^2)p(\tau|\sigma^2) \prod_{j=1}^{J} \mathcal{N}(\bar{Y}_{\cdot j}|\mu, \bar{\sigma}_j^2 + \tau^2)}{\mathcal{N}(\mu|\hat{\mu}, V_\mu)},
\end{aligned}$$

because we assume $p(\mu|\tau, \sigma^2) \propto \mathbb{1}_{(-\infty,\infty)}(\mu)$, then

$$p(\tau|\sigma^2, \mathbf{Y}) \propto \frac{p(\tau|\sigma^2) \prod_{j=1}^{J} \mathcal{N}(\bar{Y}_{\cdot j}|\mu, \bar{\sigma}_j^2 + \tau^2)}{\mathcal{N}(\mu|\hat{\mu}, V_\mu)}.$$

Remember that all the factors of μ must cancel when the expression is simplified. This means that the identity must hold for any value of μ, in particular it holds if we set μ to $\hat{\mu}$, which makes evaluation of the expression simple,

$$\begin{aligned}
p(\tau|\sigma^2, \mathbf{Y}) &\propto \frac{p(\tau|\sigma^2) \prod_{j=1}^{J} \mathcal{N}(\bar{Y}_{\cdot j}|\hat{\mu}, \bar{\sigma}_j^2 + \tau^2)}{\mathcal{N}(\hat{\mu}|\hat{\mu}, V_\mu)} \\
&\propto p(\tau|\sigma^2) V_\mu^{1/2} \prod_{j=1}^{J} (\bar{\sigma}_j^2 + \tau^2)^{-1/2} \exp\left\{ -\frac{(\bar{Y}_{\cdot j} - \hat{\mu})}{2(\bar{\sigma}_j^2 + \tau^2)} \right\}.
\end{aligned}$$

To complete the analysis we must assign a prior distribution for τ. However, we must examine the posterior and ensure it has a finite integral for the chosen prior. Remember that $\hat{\mu} \xrightarrow[\tau \to 0]{} \bar{Y}_{\cdot\cdot}$ and $V_\mu \xrightarrow[\tau \to 0]{} \varphi^2$. Then, everything multiplying $p(\tau|\sigma^2)$ approaches a nonzero constant limit as τ tends to zero. Thus, the behavior of the posterior density near $\tau = 0$ is determined by the prior density. The usual 'noninformative' function $p(\tau|\sigma^2) \propto \frac{1}{\tau} \mathbb{1}_{(0,\infty)}(\tau)$ is not integrable for any small interval including $\tau = 0$ and yields a nonintegrable posterior density. Meanwhile, the uniform prior distribution $p(\tau|\sigma^2) \propto \mathbb{1}_{(0,\infty)}(\tau)$ yields a proper posterior density.

Thus, making $p(\tau|\sigma^2) \propto \mathbb{1}_{(0,\infty)}(\tau)$, the conditional posterior distribution of τ is given by

$$p(\tau|\sigma^2, \mathbf{Y}) \propto V_\mu^{1/2} \prod_{j=1}^{J} (\bar{\sigma}_j^2 + \tau^2)^{-1/2} \exp\left\{ -\frac{(\bar{Y}_{\cdot j} - \hat{\mu})}{2(\bar{\sigma}_j^2 + \tau^2)} \right\} \mathbb{1}_{(0,\infty)}(\tau).$$

Furthermore, let be $\hat{\tau}$ the MAP estimate of τ then, under regularity conditions, an approximate interval of probability $(1-\alpha) \times 100\%$ for τ is given by

$$\left\{ \tau : \frac{p(\tau|\sigma^2, \mathbf{Y})}{p(\hat{\tau}|\sigma^2, \mathbf{Y})} \geq \exp\left\{ -\frac{q^{1-\alpha}_{\chi^2_1}}{2} \right\} \right\}.$$

4 Hierarchical model with common σ^2 unknown

Unfortunately, for most of the practical problems σ^2 is unknown. Because σ is a scalar parameter of our observations, we can take a uniform prior distribution for $\log \sigma$. Then, using the priors previously defined for the other parameters, we get the following joint prior

$$p(\boldsymbol{\theta}, \mu, \tau, \log\sigma) \propto \mathbb{1}_{(-\infty,\infty)}(\mu)\mathbb{1}_{(0,\infty)}(\tau)\mathbb{1}_{(-\infty,\infty)}(\log\sigma)\prod_{j=1}^{J}\mathcal{N}(\theta_j|\mu,\tau^2)$$

or, equivalently

$$p(\boldsymbol{\theta}, \mu, \tau^2, \sigma^2) \propto \frac{1}{\tau}\frac{1}{\sigma^2}\mathbb{1}_{(-\infty,\infty)}(\mu)\mathbb{1}_{(0,\infty)}(\tau^2)\mathbb{1}_{(0,\infty)}(\sigma^2)\prod_{j=1}^{J}\mathcal{N}(\theta_j|\mu,\tau^2).$$

Thus, the joint posterior is given by

$$p(\boldsymbol{\theta}, \mu, \tau^2, \sigma^2|\mathbf{Y}) \propto \frac{1}{\tau}\frac{1}{\sigma^2}\mathbb{1}_{(-\infty,\infty)}(\mu)\mathbb{1}_{(0,\infty)}(\tau^2)\mathbb{1}_{(0,\infty)}(\sigma^2)\prod_{j=1}^{J}\mathcal{N}(\theta_j|\mu,\tau^2)\prod_{j=1}^{J}\prod_{i=1}^{n_j}\mathcal{N}(Y_{ij}|\theta_j,\sigma^2).$$

Conditional posterior for θ_j. We already deduced that the posterior distribution of $\theta_j|\mu,\tau^2,\sigma^2,\mathbf{Y}$ are independent and

$$\theta_j|\mu,\tau^2,\sigma^2,\mathbf{Y} \sim \mathcal{N}(\hat{\theta}_j, V_{\theta_j}),$$

where

$$\hat{\theta}_j = \frac{\frac{1}{\bar{\sigma}_j^2}\bar{Y}_{\cdot j} + \frac{1}{\tau^2}\mu}{\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}}$$

and

$$V_{\theta_j} = \frac{1}{\frac{1}{\bar{\sigma}_j^2} + \frac{1}{\tau^2}}.$$

Conditional posterior for μ.

$$\begin{aligned} p(\mu|\boldsymbol{\theta}, \tau^2, \sigma^2, \mathbf{Y}) &\propto \mathbb{1}_{(-\infty,\infty)}(\mu)\prod_{j=1}^{J}\mathcal{N}(\theta_j|\mu,\tau^2) \\ &\propto \exp\left\{-\frac{1}{2}\sum_{j=1}^{J}\frac{(\theta_j-\mu)^2}{\tau^2}\right\}\mathbb{1}_{(-\infty,\infty)}(\mu). \end{aligned}$$

From our analysis on the pooled model, it is immediate that

$$\mu|\boldsymbol{\theta}, \tau^2, \sigma^2, \mathbf{Y} \sim \mathcal{N}(\hat{\mu}, \tau^2/J),$$

where $\hat{\mu} = \frac{1}{J}\sum_{j=1}^{J}\theta_j$.

Conditional posterior for σ^2.

$$\begin{aligned} p(\sigma^2|\boldsymbol{\theta}, \mu, \tau^2, \mathbf{Y}) &\propto \frac{1}{\sigma^2}\mathbb{1}_{(0,\infty)}(\sigma^2)\prod_{j=1}^{J}\prod_{i=1}^{n_j}\mathcal{N}(Y_{ij}|\theta_j,\sigma^2) \\ &\propto \frac{1}{\sigma^2}\mathbb{1}_{(0,\infty)}(\sigma^2)\frac{1}{(\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij}-\theta_j)^2\right\} \\ &= (\sigma^2)^{-(n/2+1)}\exp\left\{-\frac{n\hat{\sigma}^2}{2\sigma^2}\right\}\mathbb{1}_{(0,\infty)}(\sigma^2). \end{aligned}$$

Thus,

$$\sigma^2|\boldsymbol{\theta}, \mu, \tau^2, \mathbf{Y} \sim \textsf{Inverse-}\chi^2(n, \hat{\sigma}^2),$$

where $n = \sum_{j=1}^J n_j$, and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - \theta_j)^2.$$

Conditional posterior for τ^2.

$$\begin{aligned}
p(\tau^2|\boldsymbol{\theta}, \mu, \sigma^2, \mathbf{Y}) &\propto \frac{1}{(\tau^2)^{1/2}}\mathbb{1}_{(0,\infty)}(\tau^2)\prod_{j=1}^{J}\mathcal{N}(\theta_j|\mu, \tau^2) \\
&\propto \frac{1}{(\tau^2)^{1/2}}\frac{1}{(\tau^2)^{J/2}}\exp\left\{-\frac{1}{2\tau^2}\sum_{j=1}^{J}(\theta_j - \mu)^2\right\}\mathbb{1}_{(0,\infty)}(\tau^2) \\
&= (\tau^2)^{-(\frac{J-1}{2}+1)}\exp\left\{-\frac{(J-1)\hat{\tau}^2}{2\tau^2}\right\}\mathbb{1}_{(0,\infty)}(\tau^2).
\end{aligned}$$

Thus,

$$\tau^2|\theta, \mu, \sigma^2, \mathbf{Y} \sim \textsf{Inverse-}\chi^2(J-1, \hat{\tau}^2),$$

where

$$\hat{\tau}^2 = \frac{1}{J-1}\sum_{j=1}^{J}(\theta_j - \mu)^2.$$

Algorithm 3 Gibbs sampler for the hierarchical model with common σ^2 unknown.

Input: Sample y_{ij} ($i = 1, \dots, n_j$, $j = 1, \dots, J$), observed groups' averages $\bar{y}_{\cdot j}$, punctual estimates for θ_j, e.g. $\hat{\theta}_j = \bar{y}_{\cdot j}$, punctual estimate for μ, e.g.

$$\hat{\mu} = \frac{\sum_{j=1}^{J} \bar{y}_{\cdot j} n_j / s_j^2}{\sum_{j=1}^{J} n_j / s_j^2},$$

and posterior sample size S.

Output: Posterior sample for $\theta_j^{(s)}$, $\mu^{(s)}$, $\tau^{(s)}$, $\sigma^{(s)}$ and $Y_j^{(s)}$ ($j = 1, \dots, J$, $s = 1, \dots, S$).

Set $\theta_j^{(0)} = \hat{\theta}_j$ and $\mu^{(0)} = \hat{\mu}$

for $s = 1, \dots, S$ **do**

  Compute $\hat{\tau}^{2(s)} = \frac{1}{J-1} \sum_{j=1}^{J} \left(\theta_j^{(s-1)} - \mu^{(s-1)} \right)^2$

  Simulate $\tau^{2(s)} \sim \mathsf{Inverse}\text{-}\chi^2 \left(J - 1, \hat{\tau}^{2(s)} \right)$

  Compute $\hat{\mu}^{(s)} = \frac{1}{J} \sum_{j=1}^{J} \theta_j^{(s-1)}$

  Simulate $\mu^{(s)} \sim \mathcal{N} \left(\hat{\mu}^{(s)}, \tau^{2(s)} / J \right)$

  Compute $\hat{\sigma}^{2(s)} = \frac{1}{n} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \left(y_{ij} - \theta_j^{(s-1)} \right)^2$

  Simulate $\sigma^{2(s)} \sim \mathsf{Inverse}\text{-}\chi^2 \left(n, \hat{\sigma}^{2(s)} \right)$

  for $j = 1, \dots, J$ **do**

    Compute

$$\hat{\theta}_j^{(s)} = \frac{\frac{n_j}{\sigma^{2(s)}} \bar{y}_{\cdot j} + \frac{1}{\tau^{2(s)}} \mu^{(s)}}{\frac{n_j}{\sigma^{2(s)}} + \frac{1}{\tau^{2(s)}}}$$

and

$$V_{\theta_j}^{(s)} = \frac{1}{\frac{n_j}{\sigma^{2(s)}} + \frac{1}{\tau^{2(s)}}}$$

    Simulate $\theta_j^{(s)} \sim \mathcal{N} \left(\hat{\theta}_j^{(s)}, V_{\theta_j}^{(s)} \right)$

    Simulate $\mathbf{Y}_j^{(s)} \sim \mathcal{N} \left(\theta_j^{(s)}, \sigma^{2(s)} \right)$

  end for

end for

5 Hierarchical model with non-common σ^2

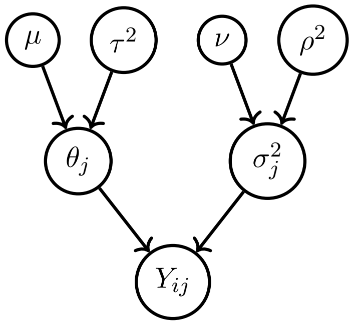

Figure 4: Hierarchical model with non-common σ^2.

Up to now, we have consider for the hierarchical model that all the groups have the same within-variance σ^2. In this section we consider the case where these variances are not the same, but they are not completely independent either. As we did with the groups' means, we assume that these variances share a common structure, yielding a hierarchical model for the within-variances.

Consider that $Y_{ij}|\theta_j, \sigma_j^2 \sim \mathcal{N}(\theta_j, \sigma_j^2)$. For the simplicity of a conjugate model, we consider

$$\theta_j|\mu, \tau^2 \sim \mathcal{N}(\mu, \tau^2),$$

and

$$\sigma_j^2|\nu, \rho^2 \sim \textsf{Inverse-}\chi^2(\nu, \rho^2).$$

It is easy to see that the conditional posterior distributions for each θ_j, μ and τ^2 are the same as those calculated in the previous section. Thus, we only have to determine the conditional posterior distributions for σ_j^2, ν and ρ^2.

5.1 Conditional posterior for σ_j^2

Due to the conjugacy property of the model,

$$\sigma_j^2|\boldsymbol{\theta}, \nu, \rho^2, \mathbf{Y} \sim \textsf{Inverse-}\chi^2(\nu_j, \hat{\sigma}_j^2),$$

where $\nu_j = \nu + n_j$,

$$\hat{\sigma}_j^2 = \frac{\nu\rho^2 + n_j v_j}{\nu + n_j},$$

and

$$v_j = \frac{1}{n_j}\sum_{i=1}^{n_j}(Y_{ij} - \theta_j)^2.$$

Note that

$$\begin{aligned} \mathbb{E}(\sigma_j^2|\boldsymbol{\theta}, \nu, \rho^2, \mathbf{Y}) &= \frac{\nu_j}{\nu_j - 2}\hat{\sigma}_j^2 \\ &= \frac{\nu\rho^2 + n_j v_j}{\nu + n_j - 2}, \end{aligned}$$

and

$$\mathbb{V}(\sigma_j^2|\boldsymbol{\theta},\nu,\rho^2,\mathbf{Y}) = \frac{2\nu_j^2}{(\nu_j-2)^2(\nu_j-4)}\hat{\sigma}_j^4$$
$$= \frac{2(\nu\rho^2+n_jv_j)^2}{(\nu+n_j-2)^2(\nu+n_j-4)},$$

from this expressions is easy to see that in the limit case when $\nu\to\infty$, $\mathbb{E}(\sigma_j^2|\boldsymbol{\theta},\nu,\rho^2,\mathbf{Y}) \xrightarrow[\nu\to\infty]{} \rho^2$ and $\mathbb{V}(\sigma_j^2|\boldsymbol{\theta},\nu,\rho^2,\mathbf{Y}) \xrightarrow[\nu\to\infty]{} 0$. On the other hand, if $\nu\to 0$, then $\sigma_j^2|\boldsymbol{\theta},\nu,\rho^2,\mathbf{Y}\sim \textsf{Inverse-}\chi^2(n_j,v_j)$, corresponding with the separate estimates.

5.2 Estimating ρ^2 and ν

Before establishing prior distributions for ρ^2 and ν, we present a practical approach to get punctual estimates for these parameters that we have seen in simulations to be a good approximation for their real values.

Since $\sigma^2\sim\textsf{Inverse-}\chi^2(\nu,\rho^2)$, then

$$\mathbb{E}(\sigma^2|\nu,\rho^2) = \frac{\nu}{\nu-2}\rho^2$$

and

$$\mathbb{V}(\sigma^2|\nu,\rho^2) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\rho^4.$$

Let be E_{s^2} the average of the observed sample within-group variances $s_1^2,\dots,s_J^2$ and V_{s^2} their variance, thus using the method of moments we have

$$E_{s^2} = \frac{\hat{\nu}}{\hat{\nu}-2}\hat{\rho}^2 \Rightarrow \hat{\rho}^2 = \frac{\hat{\nu}-2}{\hat{\nu}}E_{s^2},$$

and

$$V_{s^2} = \frac{2\cancel{\hat{\nu}^2}}{\cancel{(\hat{\nu}-2)^2}(\hat{\nu}-4)}\frac{\cancel{(\hat{\nu}-2)^2}}{\cancel{\hat{\nu}^2}}(E_{s^2})^2$$
$$\Rightarrow \hat{\nu} = \frac{2(E_{s^2})^2}{V_{s^2}}+4,$$

thus

$$\hat{\rho}^2 = \left(1-\frac{2V_{s^2}}{2(E_{s^2})^2+4V_{s^2}}\right)E_{s^2}$$
$$= \left(\frac{2(E_{s^2})^2+2V_{s^2}}{2(E_{s^2})^2+4V_{s^2}}\right)E_{s^2}.$$

We could set the value of ρ^2 in this estimate and plot $\mathbb{E}(\sigma_j^2|\boldsymbol{\theta},\nu,\rho^2,\mathbf{Y})$ and $\mathbb{V}(\sigma_j^2|\boldsymbol{\theta},\nu,\rho^2,\mathbf{Y})$ as functions of ν. In case that we don't know the value for $\boldsymbol{\theta}$, we could substitute v_j by the sample variance of the group s_j^2.

5.3 Conditional posterior for ρ^2

The joint distribution is now given by

$$p(\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\sigma}^2, \nu, \rho^2|\mathbf{Y}) \propto \frac{1}{\tau}\mathbb{1}_{(-\infty,\infty)}(\mu)\mathbb{1}_{(0,\infty)}(\tau^2)p(\nu)p(\rho^2) \times \prod_{j=1}^{J}\mathcal{N}(\theta_j|\mu,\tau^2)\text{Inverse-}\chi^2(\sigma_j^2|\nu,\rho^2)\prod_{j=1}^{J}\prod_{i=1}^{n_j}\mathcal{N}(Y_{ij}|\theta_j,\sigma_j^2).$$

To calculate the conditional posterior distribution of ρ^2, note that

$$\begin{aligned}
p(\rho^2|\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\sigma}^2, \nu, \mathbf{Y}) &= p(\rho^2|\boldsymbol{\sigma}^2, \nu) \\
&\propto p(\rho^2|\nu)p(\boldsymbol{\sigma}^2|\nu, \rho^2) \\
&\propto p(\rho^2|\nu)\prod_{j=1}^{J}(\rho^2)^{\nu/2}\exp\left\{-\frac{\nu}{2\sigma_j^2}\rho^2\right\} \\
&= p(\rho^2|\nu)(\rho^2)^{\frac{J\nu}{2}}\exp\left\{-\left(\frac{\nu}{2}\sum_{j=1}^{J}\frac{1}{\sigma_j^2}\right)\rho^2\right\}.
\end{aligned}$$

Setting $p(\rho^2|\nu) = \frac{1}{\rho^2}\mathbb{1}_{(0,\infty)}(\rho^2)$ yields

$$p(\rho^2|\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \nu, \mathbf{Y}) \propto (\rho^2)^{\frac{J\nu}{2}-1}\exp\left\{-\left(\frac{\nu}{2}\sum_{j=1}^{J}\frac{1}{\sigma_j^2}\right)\rho^2\right\}\mathbb{1}_{(0,\infty)}(\rho^2).$$

It is immediate from the previous expression that

$$\rho^2|\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\sigma}^2, \nu, \mathbf{Y} \sim \text{Gamma}\left(\alpha = \frac{J\nu}{2}, \beta = \frac{J\nu}{2\hat{\rho}^2}\right),$$

where

$$\hat{\rho}^2 = \frac{J}{\sum_{j=1}^{J}\frac{1}{\sigma_j^2}}.$$

Note that $\mathbb{E}(\rho^2|\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\sigma}^2, \nu, \mathbf{Y}) = \hat{\rho}^2$, which is the harmonic mean of the within-variances. Thus ρ^2 models to the “common” within-variance from which the within-variance of each group deviates.

At this point, we could set the value of ν to its empirical estimate $\hat{\nu}$ and get a sample posterior for all the other parameters using Gibbs sampler.

Algorithm 4 Gibbs sampler for the extended hierarchical model with fixed ν.

Input: Sample y_{ij} ($i = 1, \ldots, n_j$, $j = 1, \ldots, J$), observed groups' averages $\bar{y}_{\cdot j}$, punctual estimates for θ_j, e.g. $\hat{\theta}_j = \bar{y}_{\cdot j}$, punctual estimate for μ, e.g.

$$\hat{\mu} = \frac{\sum_{j=1}^{J} \bar{y}_{\cdot j} n_j / s_j^2}{\sum_{j=1}^{J} n_j / s_j^2},$$

punctual estimates for σ_j^2, e.g. $\hat{\sigma}_j^2 = s_j^2$, and posterior sample size S.

Output: Posterior sample for $\theta_j^{(s)}$, $\mu^{(s)}$, $\tau^{(s)}$, $\sigma_j^{(s)}$, $\rho^{(s)}$ and $Y_j^{(s)}$ ($j = 1, \ldots, J$, $s = 1, \ldots, S$).

Set $\theta_j^{(0)} = \hat{\theta}_j$, $\mu^{(0)} = \hat{\mu}$, $\sigma_j^{(0)} = \hat{\sigma}_j$

for $s = 1, \ldots, S$ **do**

Compute $\hat{\tau}^{2(s)} = \frac{1}{J-1} \sum_{j=1}^{J} \left(\theta_j^{(s-1)} - \mu^{(s-1)} \right)^2$

Simulate $\tau^{2(s)} \sim \textsf{Inverse-}\chi^2 \left(J - 1, \hat{\tau}^{2(s)} \right)$

Compute $\hat{\mu}^{(s)} = \frac{1}{J} \sum_{j=1}^{J} \theta_j^{(s-1)}$

Simulate $\mu^{(s)} \sim \mathcal{N} \left(\hat{\mu}^{(s)}, \tau^{2(s)} / J \right)$

Compute $\hat{\rho}^{2(s)} = J / \sum_{j=1}^{J} \frac{1}{\sigma_j^{2(s-1)}}$

Simulate $\rho^{2(s)} \sim \textsf{Gamma} \left(\alpha = \frac{J\nu}{2}, \beta = \frac{J\nu}{2\hat{\rho}^{2(s)}} \right)$

for $j = 1, \ldots, J$ **do**

Compute

$$\hat{\theta}_j^{(s)} = \frac{\frac{n_j}{\sigma^{2(s)}} \bar{y}_{\cdot j} + \frac{1}{\tau^{2(s)}} \mu^{(s)}}{\frac{n_j}{\sigma^{2(s)}} + \frac{1}{\tau^{2(s)}}}$$

and

$$V_{\theta_j}^{(s)} = \frac{1}{\frac{n_j}{\sigma^{2(s)}} + \frac{1}{\tau^{2(s)}}}$$

Simulate $\theta_j^{(s)} \sim \mathcal{N} \left(\hat{\theta}_j^{(s)}, V_{\theta_j}^{(s)} \right)$

Compute

$$v_j^{(s)} = \frac{1}{n_j} \sum_{i=1}^{n_j} \left(y_{ij} - \theta_j^{(s)} \right)^2$$

and

$$\hat{\sigma}_j^{2(s)} = \frac{\nu \rho^{2(s)} + n_j v_j^{(s)}}{\nu + n_j}$$

Simulate $\sigma_j^{2(s)} \sim \textsf{Inverse-}\chi^2 (\nu + n_j, \hat{\sigma}_j^{2(s)})$

Simulate $\mathbf{Y}_j^{(s)} \sim \mathcal{N} \left(\theta_j^{(s)}, \sigma^{2(s)} \right)$

end for

end for

5.4 Conditional posterior for ν

To calculate the conditional posterior distribution of ν, note that

$$\begin{aligned}
p(\nu|\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\sigma}^2, \rho^2, \mathbf{Y}) &= p(\nu|\boldsymbol{\sigma}^2, \rho^2) \\
&\propto p(\nu|\rho^2)p(\boldsymbol{\sigma}^2|\nu, \rho^2) \\
&\propto p(\nu|\rho^2)\prod_{j=1}^{J}\frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}(\rho^2)^{\nu/2}(\sigma_j^2)^{-\nu/2}\exp\left\{-\frac{\nu\rho^2}{2\sigma_j^2}\right\} \\
&= p(\nu|\rho^2)\frac{(\nu/2)^{J\nu/2}}{\Gamma^J(\nu/2)}(\rho^2)^{J\nu/2}\exp\left\{-\frac{\nu\rho^2}{2}\sum_{j=1}^{J}\frac{1}{\sigma_j^2}\right\}\prod_{j=1}^{J}(\sigma_j^2)^{-\nu/2},
\end{aligned}$$

let be $\omega = \nu/2$, then

$$p(\nu|\boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{\sigma}^2, \rho^2, \mathbf{Y}) \propto p(\nu|\rho^2)\frac{\omega^{J\omega}}{\Gamma^J(\omega)}(\rho^2)^{J\omega}\exp\left\{-J\omega\frac{\rho^2}{\hat{\rho}^2}\right\}\prod_{j=1}^{J}(\sigma_j^2)^{-\omega}$$

This is an intricate expression, which gives little guide for the selection of a prior distribution for ν yielding a known distribution. We have considered different prior distributions for ν of the form $p(\nu|\rho^2) \propto \nu^{-h}\mathbb{1}_{(0,\infty)}(\nu)$ for different values of $h > 0$, such as $h = 3, 2, 1.5$ or 1.

The impact of h in the posterior distribution. We have seen in simulations that larger values for h tend to generate a posterior distribution for ν concentrated in lower values and, thus, being models closer to the case where the within-variances share no-common structure. Therefore, larger values for h would make each within-variance σ_j^2 to concentrate in the observed sample variance s_j^2 at the cost of increasing the uncertainty in their estimates, since the observations have less impact on other groups beyond the one they belong to. Meanwhile, smaller values for h have the opposite effect, generating models that are closer to the case where a single common within-variance σ^2 is considered for all the groups. The limit case $h \to 0$, corresponding with the prior $p(\nu) = \mathbb{1}_{(0,\infty)}(\nu)$, generates an improper monotonically increasing posterior for ν, which makes all the within-variances to concentrate in the common-variance ρ^2. Thus, lower values for h would introduce a bias from the observed sample variance, but would reduce the uncertainty in the common estimate. Therefore the value for h models the well-known bias-variance compromise for the within-variances.

Algorithm 5 Gibbs sampler for the extended hierarchical model.

Input: Sample y_{ij} $(i = 1, \dots, n_j,\ j = 1, \dots, J)$, observed groups' averages $\bar{y}_{\cdot j}$, punctual estimates for θ_j, e.g. $\hat{\theta}_j = \bar{y}_{\cdot j}$, punctual estimate for μ, e.g.

$$\hat{\mu} = \frac{\sum_{j=1}^{J} \bar{y}_{\cdot j} n_j / s_j^2}{\sum_{j=1}^{J} n_j / s_j^2},$$

punctual estimates for σ_j^2, e.g. $\hat{\sigma}_j^2 = s_j^2$, punctual estimate for ρ^2, e.g.

$$\hat{\rho}^2 = \frac{J}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2}}$$

and posterior sample size S.

Output: Posterior sample for $\theta_j^{(s)}$, $\mu^{(s)}$, $\tau^{(s)}$, $\sigma_j^{(s)}$, $\nu^{(s)}$, $\rho^{(s)}$ and $Y_j^{(s)}$ $(j = 1, \dots, J,\ s = 1, \dots, S)$.

Set $\theta_j^{(0)} = \hat{\theta}_j$, $\mu^{(0)} = \hat{\mu}$, $\sigma_j^{(0)} = \hat{\sigma}_j$, $\rho^{(0)} = \hat{\rho}$

for $s = 1, \dots, S$ **do**

Compute $\hat{\tau}^{2(s)} = \frac{1}{J-1} \sum_{j=1}^{J} \left(\theta_j^{(s-1)} - \mu^{(s-1)} \right)^2$

Simulate $\tau^{2(s)} \sim \mathsf{Inverse}\text{-}\chi^2 \left(J - 1, \hat{\tau}^{2(s)} \right)$

Compute $\hat{\mu}^{(s)} = \frac{1}{J} \sum_{j=1}^{J} \theta_j^{(s-1)}$

Simulate $\mu^{(s)} \sim \mathcal{N} \left(\hat{\mu}^{(s)}, \tau^{2(s)}/J \right)$

Simulate $\nu^{(s)} \sim p(\nu | \sigma^{2(s-1)}, \rho^{2(s-1)})$

Compute $\hat{\rho}^{2(s)} = J / \sum_{j=1}^{J} \frac{1}{\sigma_j^{2(s-1)}}$

Simulate $\rho^{2(s)} \sim \mathsf{Gamma} \left(\alpha = \frac{J\nu^{(s)}}{2}, \beta = \frac{J\nu^{(s)}}{2\hat{\rho}^{2(s)}} \right)$

for $j = 1, \dots, J$ **do**

Compute

$$\hat{\theta}_j^{(s)} = \frac{\frac{n_j}{\sigma^{2(s-1)}} \bar{y}_{\cdot j} + \frac{1}{\tau^{2(s)}} \mu^{(s)}}{\frac{n_j}{\sigma^{2(s-1)}} + \frac{1}{\tau^{2(s)}}}$$

and

$$V_{\theta_j}^{(s)} = \frac{1}{\frac{n_j}{\sigma^{2(s-1)}} + \frac{1}{\tau^{2(s)}}}$$

Simulate $\theta_j^{(s)} \sim \mathcal{N} \left(\hat{\theta}_j^{(s)}, V_{\theta_j}^{(s)} \right)$

Compute

$$v_j^{(s)} = \frac{1}{n_j} \sum_{i=1}^{n_j} \left(y_{ij} - \theta_j^{(s)} \right)^2$$

and

$$\hat{\sigma}_j^{2(s)} = \frac{\nu^{(s)} \rho^{2(s)} + n_j v_j^{(s)}}{\nu^{(s)} + n_j}$$

Simulate $\sigma_j^{2(s)} \sim \mathsf{Inverse}\text{-}\chi^2 (\nu^{(s)} + n_j, \hat{\sigma}_j^{2(s)})$

Simulate $\mathbf{Y}_j^{(s)} \sim \mathcal{N} \left(\theta_j^{(s)}, \sigma^{2(s)} \right)$

end for

end for

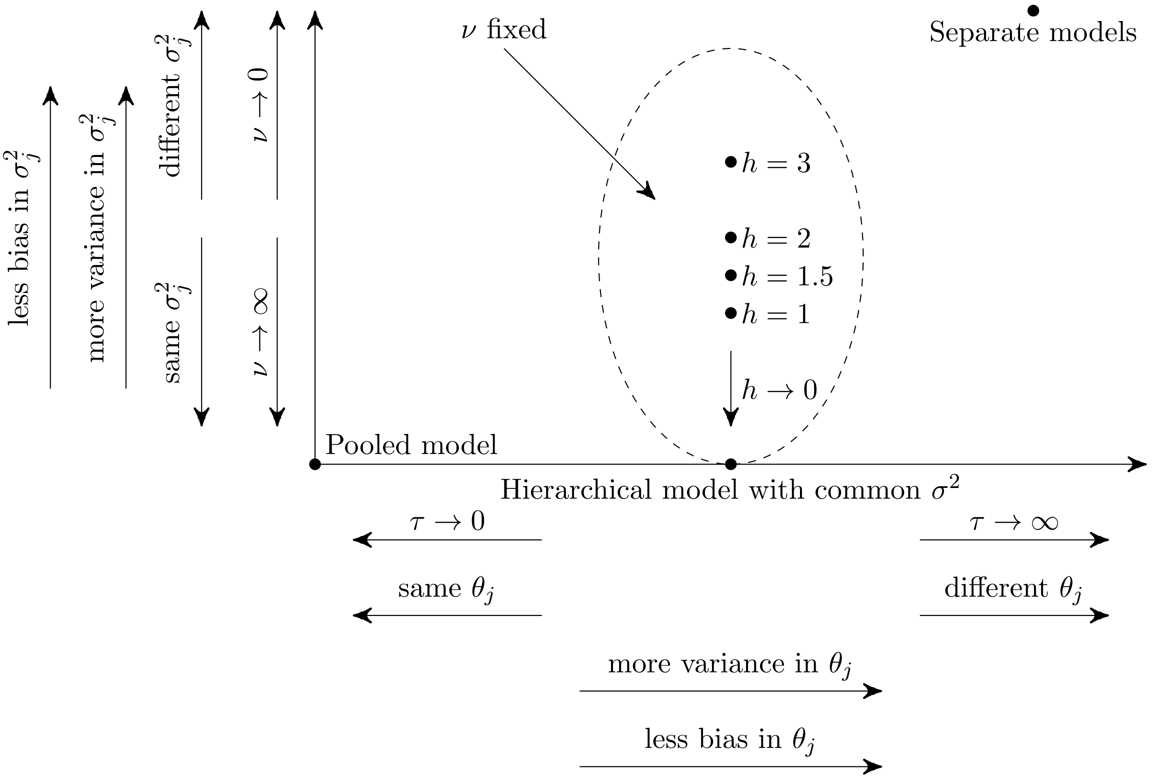

Figure 5: Relation between models.

A Appendix

A.1 Assessing convergence of an MCMC

Since we know the form of the joint posterior density (except for some constant), we can estimate the log-posterior for each simulated sample from the Gibss sampler, and monitor this estimand to infer if we are already simulating from the stationary distribution.

Together with the graph of the log-posterior, we can use the potential scale reduction approach that we explain now (see BDA page 284).

Between- and within-sequences variances

Suppose that we have simulated m chains, each of length n, and let be ψ_{ij} $(i = 1, \dots, n;$ $j = 1, \dots, m)$ any scalar estimand that we wish to monitor. Then, we define the between- and within-sequences variances, B and W, as

$$B = \frac{n}{m-1}\sum_{j=1}^{m}(\bar{\psi}_{\cdot j} - \bar{\psi}_{\cdot\cdot})^2, \text{ where } \bar{\psi}_{\cdot j} = \frac{1}{n}\sum_{i}^{n}\psi_{ij},\ \bar{\psi}_{\cdot\cdot} = \frac{1}{m}\sum_{j=1}^{m}\bar{\psi}_{\cdot j},$$

$$W = \frac{1}{m}\sum_{j=1}^{m}s_j^2, \text{ where } s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\psi_{ij} - \bar{\psi}_{\cdot j})^2.$$

Potential scale reduction

The marginal posterior variance of the estimand, $\mathbb{V}(\psi|\mathbf{Y})$ can be estimated through

$$\widehat{\mathbb{V}}^{+}(\psi|\mathbf{Y}) = \frac{n-1}{n}W + \frac{1}{n}B.$$

The potential scale reduction is then defined as

$$\widehat{R} = \sqrt{\frac{\widehat{\mathbb{V}}^{+}(\psi|\mathbf{Y})}{W}},$$

which declines to 1 as $n \to \infty$. If the potential scale reduction is high, then we have reason to believe that proceeding with further simulations may improve the inference about the target distribution of the associated scalar estimand.

A.2 Technical note about the scaled inverse-χ^2

It happens often that the scaled inverse-χ^2 is not available in some libraries, however, if the inverse-gamma is available in the library, we can use the well known relation

$$\textsf{Inverse-}\chi^2(\nu, s^2) \equiv \textsf{Inverse-Gamma}\left(\alpha = \frac{\nu}{2}, \beta = \frac{\nu}{2}s^2\right),$$

where α is the parameter of shape and β the parameter of scale.