Math Test Problem.

1. The Hierarchical Model

1.1 The basic model.

- $y_{ij} \sim N(\theta_i, \sigma^2)$, where i – the index of school, $i = 1, 2, \cdots, n$;
  j – the index of sample of i-th school. There are N samples (y_{ij}) in total.
- $\theta_i \sim N(\mu, \tau^2\sigma^2)$.
- μ, σ^2 follow non-informative priors, i.e., $p(\mu), p(1/\sigma^2) \propto 1$.
- τ follows a half-cauchy distribution, i.e., $p(\tau) \propto (\tau^2+1)^{-1}, \tau > 0$.

1.2 Redundant reparameterization.

It could be inviable to directly sample τ from its full conditional – something not well studied. Here I use the reparameterization trick introduced in the paper [1] that:

$$\theta_i = \mu + \eta \cdot \sigma \cdot \varphi_i, \quad \eta \sim N(0,1), \quad \varphi_i \sim N(0, g^2) \text{ where } \underset{(\xi)}{1/g^2} \sim Ga(1/2, 1/2).$$
$$\tau = |\eta| \cdot g. \qquad (1)$$

2. The Gibbs Sampler.

2.1 The full conditionals.

- Full Conds 1 (without reparameterization)

$$f(\underline{\theta}, \mu, \lambda, \tau \mid \underline{y})$$
$$\propto \lambda^{\frac{N}{2}} \exp\left\{-\frac{\lambda}{2}\sum_{i,j}(y_{ij}-\theta_i)^2\right\} \cdot (\lambda/\tau^2)^{n/2} \exp\left\{-\frac{\lambda}{2\tau^2}\sum_i(\theta_i-\mu)^2\right\} \cdot (\tau^2+1)^{-1}$$

- Full Conds 2 (with reparameterization (1))

$$f(\underline{\varphi}, \eta, \mu, \lambda, g \mid \underline{y})$$
$$\propto \lambda^{\frac{N}{2}} \exp\left\{-\frac{\lambda}{2}\sum_{i,j}(y_{ij}-\mu-\eta\cdot\sigma\cdot\varphi_i)^2\right\} \cdot \exp\left\{-\frac{\eta^2}{2}\right\} \cdot \xi^{\frac{n}{2}} \exp\left\{-\frac{\xi}{2}\sum_i \varphi_i^2\right\}$$
$$\cdot\, \xi^{-1/2} \exp\left\{-\frac{1}{2}\xi\right\}.$$

- $f(\mu \mid \cdots) = \mathcal{N}(\mu \mid \mu_m, \mu_v)$ where

$$\mu_v = (\lambda \cdot N)^{-1} \; ; \quad \mu_m = \frac{1}{N} \Sigma_{i,j} (y_{ij} - \eta \cdot \sigma \cdot \varphi_i) = \bar{y} - \eta \cdot \sigma \cdot \bar{\varphi}. \qquad (2)$$

- $f(\eta \mid \cdots) = \mathcal{N}(\eta \mid \eta_m, \eta_v)$ where

$$\eta_v = (1 + \Sigma_i n_i \varphi_i^2)^{-1} \; ; \quad \eta_m = \left(\Sigma_i n_i \varphi_i \cdot \lambda^{\frac{1}{2}} \cdot (\bar{y}_i - \mu) \right) \cdot \eta_v. \qquad (3)$$

- $f(\varphi_i \mid \cdots) = \mathcal{N}(\varphi_i \mid \varphi_{im}, \varphi_{iv})$ where

$$\varphi_{iv} = (n_i \cdot \eta^2 + \xi)^{-1} \; ; \quad \varphi_{im} = \left(n_i \cdot (\eta \sqrt{\lambda}) (\bar{y}_i - \mu) \right) \cdot \varphi_{iv}. \qquad (4)$$

- $f(\xi \mid \cdots) = Ga(\xi \mid \xi_a, \xi_b)$ where

$$\xi_a = \frac{n}{2} + \frac{1}{2} \; ; \quad \xi_b = \frac{1}{2} + \frac{1}{2} \Sigma_i \varphi_i^2. \qquad (5)$$

- $f(\lambda \mid \cdots) = Ga(\lambda \mid \lambda_a, \lambda_b)$ where

$$\lambda_a = \frac{N}{2} + \frac{n}{2} + 1 \; ; \quad \lambda_b = \frac{1}{2} \Sigma_{ij} (y_{ij} - \theta_i)^2 + \frac{1}{2\tau^2} \Sigma_i (\theta_i - \mu)^2$$

$$= \frac{1}{2} \Sigma_i \left[(n_i - 1) s_{y_i}^2 + n_i \cdot (\bar{y}_i - \theta_i)^2 \right] + \frac{1}{2\tau^2} \Sigma_i (\theta_i - \mu)^2 \qquad (6)$$

2.2 Running the chain

1. Start ;
2. initialize $\mu, \eta, \varphi, \xi (1/g^2)$ and $\lambda (1/\sigma^2)$;
3. While (chain does not converge) do :
4. sample μ from $f(\mu \mid \cdots)$ (Eq (2)) ;
5. Sample η from $f(\eta \mid \cdots)$ (Eq (3)) ;
6. Sample φ from $f(\varphi_i \mid \cdots)$ $i = 1, \cdots, n$ (Eq (4)) ;
7. sample ξ from $f(\xi \mid \cdots)$ (Eq (5)) ;
8. sample λ from $f(\lambda \mid \cdots)$ (Eq (6)) ;
9. End While.
10. collect the samples.

Reference

[1] On the Half-Cauchy Prior for a Global Scale Parameter, by N.G. Polson and J.G. Scott. In Bayesian Analysis, Vol 7, Num 4 (2012), 887-902.