
A Fully Bayesian Model for Gene Expression Heterosis in RNA-seq Data

Will Landau

Department of Statistics
Iowa State University

April 2, 2014

Contents

1	Introduction	1
2	The Model	2
3	The Full Conditional Distributions	2
4	A Metropolis-Hastings Algorithm for Sampling ν	4
5	A Metropolis-Hastings Algorithm for Sampling $\rho_n, \phi_g, \alpha_g, \delta_g,$ and $\varepsilon_{g,n}$	6
6	The Full Metropolis-Within-Gibbs Sampler	9
7	Diagnostics	10
7.1	Gelman Factors	10
7.2	Deviance Information Criterion	10
A	Derivations of the Full Conditionals	12
A.1	Transformations of Standard Deviations	13
A.2	$p(\nu \mid \cdots)$: Metropolis	14
A.3	$p(\rho_n \mid \cdots)$: Metropolis	15
A.4	$p(\varepsilon_{g,n} \mid \cdots)$: Metropolis	15
A.5	$p(\phi_g \mid \cdots)$: Metropolis	16
A.6	$p(\alpha_g \mid \cdots)$: Metropolis	17
A.7	$p(\delta_g \mid \cdots)$: Metropolis	18
A.8	$p(\theta_\phi \mid \cdots)$: Normal	20
A.9	$p(\theta_\alpha \mid \cdots)$: Normal	20
A.10	$p(\theta_\delta \mid \cdots)$: Normal	21
A.11	$p(\tau^2 \mid \cdots)$: Gamma	21
A.12	$p\left(\frac{1}{\gamma_g^2} \mid \cdots\right)$: Gamma	22
A.13	$p\left(\frac{1}{\sigma_\rho^2} \mid \cdots\right)$: Truncated Gamma	22
A.14	$p\left(\frac{1}{\sigma_\phi^2} \mid \cdots\right)$: Truncated Gamma	23
A.15	$p\left(\frac{1}{\sigma_\alpha^2} \mid \cdots\right)$: Truncated Gamma	23
A.16	$p\left(\frac{1}{\sigma_\delta^2} \mid \cdots\right)$: Truncated Gamma	23

1 Introduction

This writeup explains a fully Bayesian Markov chain Monte Carlo method for modeling RNA-seq data. The hierarchical model featured focuses on heterosis, or hybrid vigor, a phenomenon that concerns two parental genetic lines and an hybrid line. For each gene in an RNA-seq dataset, we consider three types of heterosis at the level of gene expression:

1. High parent heterosis: the gene is significantly more expressed in the hybrid than in either of the parent lines.
2. Low parent heterosis: the gene is significantly less expressed in the hybrid than in either of the parent lines.
3. Mid parent heterosis: the expression level of the gene in the hybrid is significantly different from the average of the parental expression levels.

Let $y_{g,n}$ be the expression level of gene g ($g = 1, \dots, G$) in sample n ($n = 1, \dots, N$). The samples come from one of three groups: group 1, the first parent, group 2, the hybrid, and group 3, the second parent. Hence, we define:

- μ_{g1} : mean expression level of gene g in the first parent
- μ_{g2} : mean expression level of gene g in the hybrid
- μ_{g3} : mean expression level of gene g in the second parent

In the model below, there are three quantities of primary interest:

- $\phi_g = \frac{\mu_{g1} + \mu_{g3}}{2}$, the parental mean expression level of gene g .
- $\alpha_g = \frac{\mu_{g1} - \mu_{g3}}{2}$, half the parental difference in expression levels of gene g .
- $\delta_g = \mu_{g2} - \phi_g$, the overexpression of gene g in the hybrid relative to the parental mean.

With MCMC samples of these quantities, for some threshold $\varepsilon > 0$, we can calculate empirical estimates of the following probabilities of interest:

- $P(|\alpha_g| \geq \varepsilon \mid \mathbf{y})$, the probability of differential expression.
- $P(\delta_g > |\alpha_g| \mid \mathbf{y})$, the probability of high parent heterosis.
- $P(\delta_g < -|\alpha_g| \mid \mathbf{y})$, the probability of low parent heterosis.
- $P(|\delta_g| \geq \varepsilon \mid \mathbf{y})$, the probability of mid parent heterosis.

2 The Model

$$\begin{aligned}
y_{g,n} &\overset{\text{ind}}{\sim} \text{Poisson}(\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))) \\
\rho_n &\overset{\text{ind}}{\sim} \text{N}(0, \sigma_\rho^2) \\
\sigma_\rho &\sim \text{U}(0, s_\rho) \\
\varepsilon_{g,n} &\overset{\text{ind}}{\sim} \text{N}(0, \gamma_g^2) \\
\gamma_g^2 &\overset{\text{ind}}{\sim} \text{Inv-Gamma}\left(\text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu\tau^2}{2}\right) \\
\nu &\sim \text{U}(0, d) \\
\tau^2 &\sim \text{Gamma}(\text{shape} = a, \text{rate} = b) \\
\phi_g &\overset{\text{ind}}{\sim} \text{N}(\theta_\phi, \sigma_\phi^2) \\
\theta_\phi &\sim \text{N}(0, c_\phi^2) \\
\sigma_\phi &\sim \text{U}(0, s_\phi) \\
\alpha_g &\overset{\text{ind}}{\sim} \text{N}(\theta_\alpha, \sigma_\alpha^2) \\
\theta_\alpha &\sim \text{N}(0, c_\alpha^2) \\
\sigma_\alpha &\sim \text{U}(0, s_\alpha) \\
\delta_g &\overset{\text{ind}}{\sim} \text{N}(\theta_\delta, \sigma_\delta^2) \\
\theta_\delta &\sim \text{N}(0, c_\delta^2) \\
\sigma_\delta &\sim \text{U}(0, s_\delta)
\end{aligned}$$

where:

- Conditional independence is implied unless otherwise specified.
- The parameters to the left of the “ \sim ” are implicitly conditioned on the parameters to the right.
- $\eta(g, n)$ is the function given by:

$$\eta(g, n) = \begin{cases} \phi_g - \alpha_g & \text{library } n \text{ is in treatment group 1 (parent 1)} \\ \phi_g + \delta_g & \text{library } n \text{ is in treatment group 2 (hybrid)} \\ \phi_g + \alpha_g & \text{library } n \text{ is in treatment group 3 (parent 2)} \end{cases}$$

3 The Full Conditional Distributions

Define:

- $k(n)$ = treatment group of library n .

- $\lambda_{g,n} = \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))$
- $G_\alpha = \text{number of genes for which } \alpha_g \neq 0$
- $G_\delta = \text{number of genes for which } \delta_g \neq 0$

Then:

$$\begin{aligned}
p(\nu \mid \cdots) &\propto \exp \left(-\log \Gamma(\nu/2) + \frac{\nu}{2} \log \left(\frac{\nu \tau^2}{2} \right) - \nu \frac{1}{G} \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right)^G \\
&\quad \times I(0 < \nu < d) \\
p(\rho_n \mid \cdots) &\propto \exp \left(\rho_n G \bar{y}_{\cdot n} - \frac{\rho_n^2}{2\sigma_\rho^2} - \exp(\rho_n) \sum_{g=1}^G \exp(\varepsilon_{g,n} + \eta(g, n)) \right) \\
p(\varepsilon_{g,n} \mid \cdots) &\propto \exp \left(y_{g,n} \varepsilon_{g,n} - \frac{\varepsilon_{g,n}^2}{2\gamma_g^2} - \exp(\varepsilon_{g,n}) \exp(\rho_n + \eta(g, n)) \right) \\
p(\phi_g \mid \cdots) &\propto \exp \left(\phi_g N \bar{y}_{\cdot g} - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2} - \exp(\phi_g) \left[\exp(-\alpha_g) \sum_{k(n)=1} \exp(\rho_n + \varepsilon_{g,n}) \right. \right. \\
&\quad \left. \left. + \exp(\delta_g) \sum_{k(n)=2} \exp(\rho_n + \varepsilon_{g,n}) + \exp(\alpha_g) \sum_{k(n)=3} \exp(\rho_n + \varepsilon_{g,n}) \right] \right) \\
p(\alpha_g \mid \cdots) &\propto \exp \left(\alpha_g \left(\sum_{k(n)=3} y_{g,n} - \sum_{k(n)=1} y_{g,n} \right) - \frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} \right. \\
&\quad \left. - \exp(\alpha_g) \exp(\phi_g) \sum_{k(n)=3} \exp(\rho_n + \varepsilon_{g,n}) \right. \\
&\quad \left. - \exp(-\alpha_g) \exp(\phi_g) \sum_{k(n)=1} \exp(\rho_n + \varepsilon_{g,n}) \right) \\
p(\delta_g \mid \cdots) &\propto \exp \left(\delta_g \sum_{k(n)=2} y_{g,n} - \frac{(\delta_g - \theta_\delta)^2}{2\sigma_\delta^2} - \exp(\delta_g) \exp(\phi_g) \sum_{k(n) \neq 2} \exp(\rho_n + \varepsilon_{g,n}) \right)
\end{aligned}$$

$$\begin{aligned}
p(\theta_\phi \mid \dots) &= N \left(\theta_\phi \mid \frac{c_\phi^2 \sum_{g=1}^G \phi_g}{Gc_\phi^2 + \sigma_\phi^2}, \frac{c_\phi^2 \sigma_\phi^2}{Gc_\phi^2 + \sigma_\phi^2} \right) \\
p(\theta_\alpha \mid \dots) &= N \left(\theta_\alpha \mid \frac{c_\alpha^2 \sum_{g=1}^G \alpha_g}{Gc_\alpha^2 + \sigma_\alpha^2}, \frac{c_\alpha^2 \sigma_\alpha^2}{Gc_\alpha^2 + \sigma_\alpha^2} \right) \\
p(\theta_\delta \mid \dots) &= N \left(\theta_\delta \mid \frac{c_\delta^2 \sum_{g=1}^G \delta_g}{Gc_\delta^2 + \sigma_\delta^2}, \frac{c_\delta^2 \sigma_\delta^2}{Gc_\delta^2 + \sigma_\delta^2} \right) \\
p(\tau^2 \mid \dots) &= \text{Gamma} \left(\tau^2 \mid \text{shape} = a + \frac{G\nu}{2}, \text{rate} = b + \frac{\nu}{2} \sum_{g=1}^G \frac{1}{\gamma_g^2} \right) \\
p \left(\frac{1}{\gamma_g^2} \mid \dots \right) &= \text{Gamma} \left(\frac{1}{\gamma_g^2} \mid \text{shape} = \frac{N+\nu}{2}, \text{rate} = \frac{1}{2} \left(\nu \tau^2 + \sum_{n=1}^N \varepsilon_{g,n}^2 \right) \right) \\
p \left(\frac{1}{\sigma_\rho^2} \mid \dots \right) &= \text{Gamma} \left(\frac{1}{\sigma_\rho^2} \mid \text{shape} = \frac{N-1}{2}, \text{rate} = \frac{1}{2} \sum_{n=1}^N \rho_n^2 \right) I \left(\frac{1}{\sigma_\rho^2} > \frac{1}{s_\rho^2} \right) \\
p \left(\frac{1}{\sigma_\phi^2} \mid \dots \right) &= \text{Gamma} \left(\text{shape} = \frac{G-1}{2}, \text{rate} = \frac{1}{2} \sum_{g=1}^G (\phi_g - \theta_\phi)^2 \right) I \left(\frac{1}{\sigma_\phi^2} > \frac{1}{s_\phi^2} \right) \\
p \left(\frac{1}{\sigma_\alpha^2} \mid \dots \right) &= \text{Gamma} \left(\text{shape} = \frac{G-1}{2}, \text{rate} = \frac{1}{2} \sum_{g=1}^G (\alpha_g - \theta_\alpha)^2 \right) I \left(\frac{1}{\sigma_\alpha^2} > \frac{1}{s_\alpha^2} \right) \\
p \left(\frac{1}{\sigma_\delta^2} \mid \dots \right) &= \text{Gamma} \left(\text{shape} = \frac{G-1}{2}, \text{rate} = \frac{1}{2} \sum_{g=1}^G (\delta_g - \theta_\delta)^2 \right) I \left(\frac{1}{\sigma_\delta^2} > \frac{1}{s_\delta^2} \right)
\end{aligned}$$

4 A Metropolis-Hastings Algorithm for Sampling ν

The full conditional distribution of ν is of the form,

$$\begin{aligned}
p(\nu \mid \dots) &\propto \exp \left(-\log \Gamma(\nu/2) + \frac{\nu}{2} \log \left(\frac{\nu \tau^2}{2} \right) - \nu \frac{1}{G} \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right)^G \\
&\quad \times I(0 < \nu < d) \\
&= \exp(h(\nu)) \times I(0 < \nu < d)
\end{aligned}$$

where

$$\begin{aligned}
h(\nu) &= G \cdot \left(-\log \Gamma(\nu/2) + \frac{\nu}{2} \log \left(\frac{\nu \tau^2}{2} \right) - \nu \frac{1}{G} \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right) \\
&= G \cdot \left(-\log \Gamma(\nu/2) + \frac{\nu}{2} \log \left(\frac{\nu}{2} \right) + \frac{\nu}{2} \log(\tau^2) - \nu \frac{1}{G} \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right) \\
&= G \cdot \left(-\log \Gamma(\nu/2) + \frac{\nu}{2} \log \left(\frac{\nu}{2} \right) + \underbrace{\nu \left(\frac{\log(\tau^2)}{2} - \frac{1}{G} \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right)}_K \right) \\
&= G \cdot \left(-\log \Gamma(\nu/2) + \frac{\nu}{2} \log \left(\frac{\nu}{2} \right) + \nu K \right)
\end{aligned}$$

Using Stirling's approximation to the gamma function,

$$\log \Gamma \left(\frac{\nu}{2} \right) \approx \frac{\nu}{2} \log \left(\frac{\nu}{2} \right) - \frac{1}{2} \log \left(\frac{\nu}{2} \right) - \frac{\nu}{2} + \frac{1}{2} \log(2\pi)$$

Hence,

$$\begin{aligned}
\frac{h(\nu)}{G} &\approx -\frac{\nu}{2} \log \left(\frac{\nu}{2} \right) + \frac{1}{2} \log \left(\frac{\nu}{2} \right) + \frac{\nu}{2} - \frac{1}{2} \log(2\pi) + \frac{\nu}{2} \log \left(\frac{\nu}{2} \right) + \nu K \\
&= \frac{1}{2} \log \left(\frac{\nu}{2} \right) + \frac{\nu}{2} - \frac{1}{2} \log(2\pi) + \nu K \\
&= \frac{1}{2} \log \left(\frac{\nu}{2} \right) + \nu \left(\frac{1}{2} + K \right) - \frac{1}{2} \log(2\pi) \\
&= \frac{1}{2} \log(\nu) - \frac{1}{2} \log \left(\frac{1}{2} \right) + \nu \left(\frac{1}{2} + K \right) - \frac{1}{2} \log(2\pi) \\
&= \frac{1}{2} \log(\nu) + \nu \left(\frac{1}{2} + K \right) - \frac{1}{2} \log(4\pi) \\
&= \frac{1}{2} \log(\nu) + \nu \left(\frac{1}{2} + \frac{\log(\tau^2)}{2} - \frac{1}{G} \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right) - \frac{1}{2} \log(4\pi)
\end{aligned}$$

Thus, $\frac{h(\nu)}{G}$ is approximately the log kernel of

$$\text{Gamma} \left(\text{shape} = \frac{3}{2}, \text{rate} = -\frac{1}{2} - \frac{\log(\tau^2)}{2} + \frac{1}{G} \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right)$$

Hence, $h(\nu)$ is approximately the log kernel of

$$\text{Gamma} \left(\text{shape} = \frac{G}{2} + 1, \text{rate} = -\frac{G}{2} - \frac{G \log(\tau^2)}{2} + \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right)$$

Let $q(\nu)$ be the log kernel of this gamma approximation to $p(\nu \mid \dots)$. Then, we have a Metropolis-Hastings algorithm for sampling ν .

1. Sample a proposal ν^* from a $q(\nu)$ until $0 < \nu^* < d$.
2. The acceptance probability is

$$p = \min \left\{ 1, \frac{\exp(h(\nu^*))}{\exp(h(\nu^{(i)}))} \frac{\exp(q(\nu^{(i)}))}{\exp(q(\nu^*))} \right\}$$

In practice, calculate p on a log scale.

$$\log p = \min \left\{ 0, h(\nu^*) - h(\nu^{(i)}) + q(\nu^{(i)}) - q(\nu^*) \right\}$$

3. Sample $u \sim U(0, 1)$
4. If $\log u < \log p$, set $\nu^{(i+1)} = \nu^*$ (accept ν^*).. Otherwise, set $\nu^{(i+1)} = \nu^{(i)}$ (do not except ν^*).

5 A Metropolis-Hastings Algorithm for Sampling ρ_n , ϕ_g , α_g , δ_g , and $\varepsilon_{g,n}$

ρ_n , ϕ_g , α_g , δ_g , and $\varepsilon_{g,n}$ all have full conditional distributions of the form,

$$p(\theta \mid \dots) \propto \exp(A\theta - B(\theta - C)^2 - De^\theta - Ee^{-\theta})$$

where θ is the parameter of interest, A, B, C, D , and E are constants, and $B, D, E \geq 0$. Note that E is guaranteed to be 0 except when $\theta = \alpha_g$ for some h . Let $h(\theta)$ be the log kernel of $p(\theta \mid \dots)$. Then,

$$h(\theta) = A\theta - B(\theta - C)^2 - De^\theta - Ee^{-\theta}$$

In addition, let $\hat{\theta}$ be a mode of $h(\theta)$ found via Newton-Raphson. We will need:

$$\begin{aligned}
h_0 &= h(\hat{\theta}) = A\hat{\theta} - B(\hat{\theta} - C)^2 - De^{\hat{\theta}} - Ee^{-\hat{\theta}} \\
h_1 &= h'(\hat{\theta}) = A - 2B(\hat{\theta} - C) - De^{\hat{\theta}} + Ee^{-\hat{\theta}} \\
h_2 &= h''(\hat{\theta}) = -2B - De^{\hat{\theta}} - Ee^{-\hat{\theta}}
\end{aligned}$$

Now, we can approximate $h(\theta)$ with the quadratic Taylor approximation.

$$h(\theta) \approx h_0 + h_1(\theta - \hat{\theta}) + \frac{h_2}{2}(\theta - \hat{\theta})^2$$

$\hat{\theta}$ was found by setting the first derivative equal to 0, so h_1 must be 0. Hence,

$$\begin{aligned}
h(\theta) &\approx h_0 + \frac{h_2}{2}(\theta - \hat{\theta})^2 \\
&= \left(h_0 + \frac{h_2}{2}\hat{\theta}^2 \right) - h_2\hat{\theta}\theta + \frac{h_2}{2}\theta^2 \\
&= -\frac{1}{2|h_2|^{-1}} (\theta - \hat{\theta})^2 + \text{constant}
\end{aligned}$$

Thus, one proposal distribution for θ in the Metropolis might be

$$N\left(\hat{\theta}, |h_2|^{-1}\right)$$

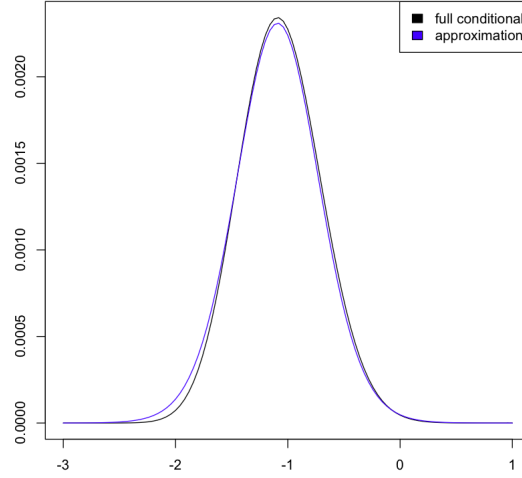
Unfortunately, the tails of this normal approximation are too thin for the proposal to be efficient. However, we can instead sample $\sqrt{|h_2|}(\theta - \hat{\theta})$ from a t_d distribution, where the degrees of freedom, d , is some large positive number. Let $q(\theta)$ be the log density of θ derived from this t_d approximation. Then, for a proportionality constant, k , the ratio of the approximate density to the true full conditional density is

$$\begin{aligned}
\log \frac{\exp(q(\theta))}{p(\theta | \dots)} &= \log q(\theta) - \log p(\theta | \dots) \\
&= \log \left[k \left(1 + \frac{\sqrt{|h_2|}(\theta - \hat{\theta})^2}{d} \right)^{-\frac{d+1}{2}} \right] - [A\theta - B(\theta - C)^2 - De^{\theta} - Ee^{-\theta}] \\
&= \log(k) - \frac{d+1}{2} \log \left(1 + \frac{\sqrt{|h_2|}(\theta - \hat{\theta})^2}{d} \right) - A\theta + B(\theta - C)^2 + De^{\theta} + Ee^{-\theta}
\end{aligned}$$

Since $d, \sqrt{|h_2|}, B, D > 0$ and $E \geq 0$, we see that

$$\lim_{|\theta| \rightarrow \infty} \log \frac{\exp(q(\theta))}{p(\theta | \dots)} = \infty$$

Thus, the proposal has thicker tails than the full conditional. Moreover, for a large d , the t_d density well approximates the full conditional, as seen in an example plot below.



Hence, the t_d proposal should be efficient.

For the full sampler, let $\theta^{(i)}$ be the current value of θ at iteration i of the algorithm. To get $\theta^{(i+1)}$,

1. Sample a proposal θ^* from a $q(\theta)$.
2. The acceptance probability is

$$p = \min \left\{ 1, \frac{\exp(h(\theta^*))}{\exp(h(\theta^{(i)}))} \frac{\exp(q(\theta^{(i)}))}{\exp(q(\theta^*))} \right\}$$

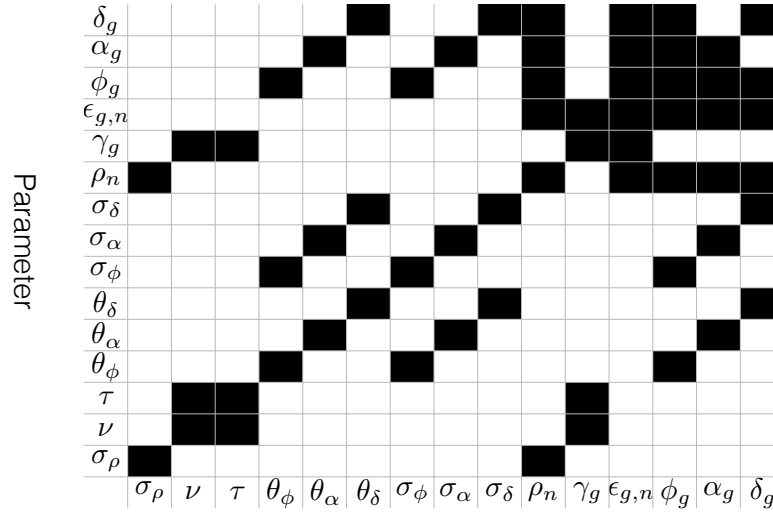
In practice, calculate p on a log scale.

$$\log p = \min \left\{ 0, h(\theta^*) - h(\theta^{(i)}) + q(\theta^{(i)}) - q(\theta^*) \right\}$$

3. Sample $u \sim U(0, 1)$
4. If $\log u < \log p$, set $\theta^{(i+1)} = \theta^*$ (accept θ^*). Otherwise, set $\theta^{(i+1)} = \theta^{(i)}$ (do not except θ^*).

6 The Full Metropolis-Within-Gibbs Sampler

By inspecting the full conditional distributions, one can see which parameters are conditionally independent. The plot below summarizes this conditional dependence.



Has a full conditional that depends on...

Using this information, I can construct Gibbs steps within each of which the sampled parameters are conditionally independent.

1. ρ_n ($n = 1, \dots, N$)
2. γ_g ($g = 1, \dots, G$)
3. $\varepsilon_{g,n}$ ($g = 1, \dots, G$), ($n = 1, \dots, N$)
4. ϕ_g ($g = 1, \dots, G$)
5. α_g ($g = 1, \dots, G$)
6. δ_g ($g = 1, \dots, G$)
7. $\nu, \theta_\phi, \theta_\alpha, \theta_\delta$
8. $\tau, \sigma_\rho, \sigma_\phi, \sigma_\alpha, \sigma_\delta$

7 Diagnostics

7.1 Gelman Factors

The potential scale reduction factor introduced in the textbook by Gelman ? monitors the lack of convergence of a single variable in an MCMC. Let η_{ij} be the i 'th MCMC draw of a single variable in chain j . Then, the potential scale reduction factor, \hat{R} , compares the within-chain variance, W , to the between-chain variance, B . Suppose there are J chains, each with I iterations. Then,

$$\begin{aligned}\hat{R} &= \sqrt{1 - \frac{1}{I} \left(\frac{B}{W} - 1 \right)} \\ B &= \frac{I}{J-1} \sum_{j=1}^J (\bar{\eta}_{\cdot j} - \bar{\eta}_{\cdot\cdot})^2, & \bar{\eta}_{\cdot j} &= \frac{1}{I} \sum_{i=1}^I \eta_{ij}, & \bar{\eta}_{\cdot\cdot} &= \frac{1}{J} \sum_{j=1}^J \bar{\eta}_{\cdot j} \\ W &= \frac{1}{J} \sum_{j=1}^J s_j^2, & s_j^2 &= \frac{1}{I-1} \sum_{i=1}^I (\eta_{ij} - \bar{\eta}_{\cdot j})^2\end{aligned}$$

$\hat{R} \rightarrow 1$ as $I \rightarrow \infty$. An \hat{R} value far above 1 indicates a lack of convergence, but an \hat{R} value near 1 does not imply convergence.

The Gelman factor used in this analysis is not actually the one given above, but a degrees-of-freedom-adjusted version implemented in the `gleman.diag()` function in the `coda` package in R:

$$\hat{R} = \sqrt{\frac{d+3}{d+1} \frac{\hat{V}}{W}}$$

where

$$d = 2 \frac{\hat{V}^2}{\text{Var}(\hat{V})}, \quad \hat{V} = \hat{\sigma}^2 + \frac{B}{IJ}, \quad \hat{\sigma}^2 = \left(1 - \frac{1}{I}\right) W + \frac{B}{I}$$

7.2 Deviance Information Criterion

The deviance information criterion (DIC) is a model selection heuristic for hierarchical models much like the Akaike information criterion, AIC, and the Bayesian information criterion, BIC. As with AIC and BIC, given a set of models for \mathbf{y} , the one with the minimum DIC is preferred. DIC is based on the deviance,

$$D(\mathbf{y}, \boldsymbol{\eta}) = -2 \log p(\mathbf{y} \mid \boldsymbol{\eta})$$

where \mathbf{y} is the data and $\boldsymbol{\eta}$ is the collection of model parameters. DIC itself is

$$\text{DIC} = 2E(D(\mathbf{y}, \boldsymbol{\eta}) \mid \mathbf{y}) - D(\mathbf{y}, \hat{\boldsymbol{\eta}})$$

where $\hat{\boldsymbol{\eta}}$ is a suitable point estimate of $\boldsymbol{\eta}$. If $\boldsymbol{\eta}_i$ is the collection of parameter estimates of iteration i of the chain and $\bar{\boldsymbol{\eta}}$ is the collection of within-chain parameter means, then we can estimate DIC by

$$\begin{aligned} \widehat{\text{DIC}} &= \sum_{i=1}^I [2D(\mathbf{y} \mid \boldsymbol{\eta}_i)] - D(\mathbf{y}, \hat{\boldsymbol{\eta}}) \\ &= -4 \sum_{i=1}^I \log p(\mathbf{y} \mid \boldsymbol{\eta}_i) + 2 \log p(\mathbf{y} \mid \bar{\boldsymbol{\eta}}) \end{aligned}$$

All that remains is to find $\log p(\mathbf{y} \mid \boldsymbol{\eta})$ for a given set of parameters, $\boldsymbol{\eta}$. Let $\lambda_{g,n} = \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))$, where

$$\eta(g, n) = \begin{cases} \phi_g - \alpha_g & \text{library } n \text{ is in treatment group 1} \\ \phi_g + \delta_g & \text{library } n \text{ is in treatment group 2} \\ \phi_g + \alpha_g & \text{library } n \text{ is in treatment group 3} \end{cases}$$

$$\begin{aligned} \log p(\mathbf{y} \mid \boldsymbol{\eta}) &= \log \prod_{n=1}^N \prod_{g=1}^G \text{Poisson}(y_{g,n} \mid \lambda_{g,n}) \\ &= \sum_{n,g} \log \text{Poisson}(y_{g,n} \mid \lambda_{g,n}) \\ &= \sum_{n,g} \log \left(\frac{\exp(-\lambda_{g,n}) \lambda_{g,n}^{y_{g,n}}}{y_{g,n}!} \right) \\ &= \sum_{n,g} (-\lambda_{g,n} + y_{g,n} \log \lambda_{g,n} - \log(y_{g,n}!)) \end{aligned}$$

Given the size of the data, calculating $\sum_{n,g} -\log(y_{g,n}!)$ is intractable. Hence, in practice, we use

$$\text{DIC} = -4 \sum_{i=1}^I L(\mathbf{y} \mid \boldsymbol{\eta}_i) + 2L(\mathbf{y} \mid \bar{\boldsymbol{\eta}})$$

where

$$L(\mathbf{y}, \boldsymbol{\eta}) = \sum_{n,g} (-\lambda_{g,n} + y_{g,n} \log \lambda_{g,n}).$$

This approach is reasonable because removing the $-\log(y_{g,n}!)$ term inside the sum merely offsets the DIC values of all the models under comparison by the same constant.

A Derivations of the Full Conditionals

Recall:

- $k(n)$ = treatment group of library n .
- $\lambda_{g,n} = \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))$
- G_α = number of genes for which $\alpha_g \neq 0$
- G_δ = number of genes for which $\delta_g \neq 0$

Then from the model in Section 2, we get:

$$\begin{aligned} p(\nu \mid \cdots) &\propto \left[\prod_{g=1}^G \text{Inv-Gamma} \left(\gamma_g^2 \mid \text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu \tau^2}{2} \right) \right] \cdot \text{U}(\nu \mid 0, d) \\ p(\rho_n \mid \cdots) &\propto \left[\prod_{g=1}^G \text{Poisson}(y_{g,n} \mid \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))) \right] \cdot \text{N}(\rho_n \mid 0, \sigma_\rho^2) \\ p(\phi_g \mid \cdots) &\propto \left[\prod_{n=1}^N \text{Poisson}(y_{g,n} \mid \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))) \right] \cdot \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2) \\ p(\alpha_g \mid \cdots) &\propto \left[\prod_{k(n) \neq 2} \text{Poisson}(y_{g,n} \mid \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))) \right] \cdot \text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2) \\ p(\delta_g \mid \cdots) &\propto \left[\prod_{k(n)=2} \text{Poisson}(y_{g,n} \mid \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))) \right] \cdot \text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2) \\ p(\varepsilon_{g,n} \mid \cdots) &\propto \text{Poisson}(y_{g,n} \mid \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))) \cdot \text{N}(\varepsilon_{g,n} \mid 0, \gamma_g^2) \end{aligned}$$

$$\begin{aligned}
p(\sigma_\rho \mid \cdots) &= \left[\prod_{n=1}^N \mathcal{N}(\rho_n \mid 0, \sigma_\rho^2) \right] \cdot \mathcal{U}(\sigma_\rho \mid 0, s_\rho) \\
p(\gamma_g^2 \mid \cdots) &\propto \left[\prod_{n=1}^N \mathcal{N}(\varepsilon_{g,n} \mid 0, \gamma_g^2) \right] \cdot \text{Inv-Gamma} \left(\gamma_g^2 \mid \text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu \tau^2}{2} \right) \\
p(\tau^2 \mid \cdots) &\propto \left[\prod_{g=1}^G \text{Inv-Gamma} \left(\gamma_g^2 \mid \text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu \tau^2}{2} \right) \right] \cdot \text{Gamma}(\tau^2 \mid \text{shape} = a, \text{rate} = b) \\
p(\theta_\phi \mid \cdots) &\propto \left[\prod_{g=1}^G \mathcal{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2) \right] \cdot \mathcal{N}(\theta_\phi \mid 0, c_\phi^2) \\
p(\theta_\alpha \mid \cdots) &\propto \left[\prod_{g=1}^G \mathcal{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2) \right] \cdot \mathcal{N}(\theta_\alpha \mid 0, c_\alpha^2) \\
p(\theta_\delta \mid \cdots) &\propto \left[\prod_{g=1}^G \mathcal{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2) \right] \cdot \mathcal{N}(\theta_\delta \mid 0, c_\delta^2) \\
p(\sigma_\phi \mid \cdots) &\propto \left[\prod_{g=1}^G \mathcal{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2) \right] \cdot \mathcal{U}(\sigma_\phi \mid 0, s_\phi) \\
p(\sigma_\alpha \mid \cdots) &\propto \left[\prod_{g=1}^G \mathcal{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2) \right] \cdot \mathcal{U}(\sigma_\alpha \mid 0, s_\alpha) \\
p(\sigma_\delta \mid \cdots) &\propto \left[\prod_{g=1}^G \mathcal{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2) \right] \cdot \mathcal{U}(\sigma_\delta \mid 0, s_\delta)
\end{aligned}$$

A.1 Transformations of Standard Deviations

Let σ be a standard deviation parameter and let $p(\sigma \mid \cdots)$ be its full conditional distribution. Then, by a transformation of variables,

$$\begin{aligned}
p(\sigma^2 \mid \cdots) &= p(\sqrt{\sigma^2} \mid \cdots) \cdot \left| \frac{d}{d\sigma^2} \sqrt{\sigma^2} \right| \\
&= p(\sigma \mid \cdots) \frac{1}{2} (\sigma^2)^{-1/2}
\end{aligned}$$

I use this transformation several times in the next sections.

A.2 $p(\nu \mid \dots)$: Metropolis

$$\begin{aligned}
p(\nu \mid \dots) &= \left[\prod_{g=1}^G \text{Inv-Gamma} \left(\gamma_g^2 \mid \text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu \tau^2}{2} \right) \right] \cdot \text{U}(\nu \mid 0, d) \\
&= \prod_{g=1}^G \left[\Gamma(\nu/2)^{-1} \left(\frac{\nu \tau^2}{2} \right)^{\nu/2} (\gamma_g^2)^{-(\nu/2+1)} \exp \left(-\frac{1}{\gamma_g^2} \frac{\nu \tau^2}{2} \right) \right] I(0 < \nu < d) \\
&= \Gamma(\nu/2)^{-G} \left(\frac{\nu \tau^2}{2} \right)^{G\nu/2} \left(\prod_{g=1}^G \gamma_g^2 \right)^{-(\nu/2+1)} \exp \left(-\frac{\nu \tau^2}{2} \sum_{g=1}^G \frac{1}{\gamma_g^2} \right) I(0 < \nu < d) \\
&\propto \Gamma(\nu/2)^{-G} \left(\frac{\nu \tau^2}{2} \right)^{G\nu/2} \left(\prod_{g=1}^G \gamma_g^2 \right)^{-\nu/2} \exp \left(-\frac{\nu \tau^2}{2} \sum_{g=1}^G \frac{1}{\gamma_g^2} \right) I(0 < \nu < d) \\
&= \exp \left(-G \log \Gamma(\nu/2) + \frac{G\nu}{2} \log \left(\frac{\nu \tau^2}{2} \right) - \nu \sum_{g=1}^G \log \gamma_g - \frac{\nu \tau^2}{2} \sum_{g=1}^G \frac{1}{\gamma_g^2} \right) \\
&\quad \times I(0 < \nu < d) \\
&= \exp \left(-G \log \Gamma(\nu/2) + \frac{G\nu}{2} \log \left(\frac{\nu \tau^2}{2} \right) - \nu \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right) \\
&\quad \times I(0 < \nu < d) \\
&= \exp \left(-G \log \Gamma(\nu/2) + \frac{G\nu}{2} \log \left(\frac{\nu \tau^2}{2} \right) - G\nu \frac{1}{G} \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right) \\
&\quad \times I(0 < \nu < d) \\
&= \exp \left(-\log \Gamma(\nu/2) + \frac{\nu}{2} \log \left(\frac{\nu \tau^2}{2} \right) - \nu \frac{1}{G} \sum_{g=1}^G \left[\log \gamma_g + \frac{\tau^2}{2} \frac{1}{\gamma_g^2} \right] \right)^G \\
&\quad \times I(0 < \nu < d)
\end{aligned}$$

A.3 $p(\rho_n \mid \dots)$: Metropolis

$$\begin{aligned}
p(\rho_n \mid \dots) &\propto \left[\prod_{g=1}^G \text{Poisson}(y_{g,n} \mid \lambda_{g,n}) \right] \cdot \text{N}(\rho_n \mid 0, \sigma_\rho^2) \\
&\propto \left[\prod_{g=1}^G \lambda_{g,n}^{y_{g,n}} \exp(-\lambda_{g,n}) \right] \exp\left(-\frac{\rho_n^2}{2\sigma_\rho^2}\right) \\
&= \exp\left(\sum_{g=1}^G [y_{g,n} \log \lambda_{g,n} - \lambda_{g,n}] - \frac{\rho_n^2}{2\sigma_\rho^2}\right) \\
&= \exp\left(\sum_{g=1}^G [y_{g,n}(\rho_n + \varepsilon_{g,n} + \eta(g, n)) - \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] - \frac{\rho_n^2}{2\sigma_\rho^2}\right) \\
&= \exp\left(\rho_n G \bar{y}_{\cdot, n} + \sum_{g=1}^G [y_{g,n}(\varepsilon_{g,n} + \eta(g, n))] - \sum_{g=1}^G \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n)) - \frac{\rho_n^2}{2\sigma_\rho^2}\right) \\
&\propto \exp\left(\rho_n G \bar{y}_{\cdot, n} - \exp(\rho_n) \sum_{g=1}^G \exp(\varepsilon_{g,n} + \eta(g, n)) - \frac{\rho_n^2}{2\sigma_\rho^2}\right) \\
&\propto \exp\left(\rho_n G \bar{y}_{\cdot, n} - \frac{\rho_n^2}{2\sigma_\rho^2} - \exp(\rho_n) \sum_{g=1}^G \exp(\varepsilon_{g,n} + \eta(g, n))\right)
\end{aligned}$$

A.4 $p(\varepsilon_{g,n} \mid \dots)$ Metropolis

$$\begin{aligned}
p(\varepsilon_{g,n} \mid \dots) &= \text{Poisson}(y_{g,n} \mid \lambda_{g,n}) \cdot \text{N}(\varepsilon_{g,n} \mid 0, \gamma_g^2) \\
&\propto \lambda_{g,n}^{y_{g,n}} \exp(-\lambda_{g,n}) \exp\left(-\frac{\varepsilon_{g,n}^2}{2\gamma_g^2}\right) \\
&= \exp\left(y_{g,n} \log \lambda_{g,n} - \lambda_{g,n} - \frac{\varepsilon_{g,n}^2}{2\gamma_g^2}\right) \\
&= \exp\left(y_{g,n}(\rho_n + \varepsilon_{g,n} + \eta(g, n)) - \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n)) - \frac{\varepsilon_{g,n}^2}{2\gamma_g^2}\right) \\
&= \exp\left(y_{g,n} \varepsilon_{g,n} - \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n)) - \frac{\varepsilon_{g,n}^2}{2\gamma_g^2}\right) \\
&= \exp\left(y_{g,n} \varepsilon_{g,n} - \frac{\varepsilon_{g,n}^2}{2\gamma_g^2} - \exp(\varepsilon_{g,n}) \exp(\rho_n + \eta(g, n))\right)
\end{aligned}$$

A.5 $p(\phi_g \mid \cdots)$: Metropolis

$$\begin{aligned}
p(\phi_g \mid \cdots) &= \left[\prod_{n=1}^N \text{Poisson}(y_{g,n} \mid \lambda_{g,n}) \right] \cdot \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2) \\
&\propto \left[\prod_{n=1}^N \lambda_{g,n}^{y_{g,n}} \exp(-\lambda_{g,n}) \right] \cdot \exp\left(-\frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right) \\
&= \exp\left(\sum_{n=1}^N [y_{g,n} \log \lambda_{g,n} - \lambda_{g,n}] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right) \\
&= \exp\left(\sum_{n=1}^N [y_{g,n}(\rho_n + \varepsilon_{g,n} + \eta(g, n)) - \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right) \\
&\propto \exp\left(\sum_{n=1}^N [y_{g,n}\eta(g, n) - \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right) \\
&\propto \exp\left(\sum_{n=1}^N [y_{g,n}\eta(g, n)] - \sum_{n=1}^N [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right) \\
&= \exp\left(\sum_{n=1}^N [y_{g,n}\eta(g, n)] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2} - \sum_{n=1}^N [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))]\right)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{n=1}^N [y_{g,n}\eta(g, n)] &= \sum_{k(n)=1} [y_{g,n}\eta(g, n)] + \sum_{k(n)=2} [y_{g,n}\eta(g, n)] + \sum_{k(n)=3} [y_{g,n}\eta(g, n)] \\
&= \sum_{k(n)=1} [y_{g,n}(\phi_g - \alpha_g)] + \sum_{k(n)=2} [y_{g,n}(\phi_g + \delta_g)] + \sum_{k(n)=3} [y_{g,n}(\phi_g + \alpha_g)] \\
&= \phi_g N \bar{y}_g + \text{constant}
\end{aligned}$$

and

$$\begin{aligned}
\sum_{n=1}^N [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] &= \sum_{k(n)=1} [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] + \sum_{k(n)=2} [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] \\
&\quad + \sum_{k(n)=3} [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] \\
&= \sum_{k(n)=1} [\exp(\rho_n + \varepsilon_{g,n} + \phi_g - \alpha_g)] + \sum_{k(n)=2} [\exp(\rho_n + \varepsilon_{g,n} + \phi_g + \delta_g)] \\
&\quad + \sum_{k(n)=3} [\exp(\rho_n + \varepsilon_{g,n} + \phi_g + \alpha_g)] \\
&= \exp(\phi_g) \left[\sum_{k(n)=1} [\exp(\rho_n + \varepsilon_{g,n} - \alpha_g)] + \sum_{k(n)=2} [\exp(\rho_n + \varepsilon_{g,n} + \delta_g)] \right. \\
&\quad \left. + \sum_{k(n)=3} [\exp(\rho_n + \varepsilon_{g,n} + \alpha_g)] \right] \\
&= \exp(\phi_g) \left[\exp(-\alpha_g) \sum_{k(n)=1} [\exp(\rho_n + \varepsilon_{g,n})] + \exp(\delta_g) \sum_{k(n)=2} [\exp(\rho_n + \varepsilon_{g,n})] \right. \\
&\quad \left. + \exp(\alpha_g) \sum_{k(n)=3} [\exp(\rho_n + \varepsilon_{g,n})] \right]
\end{aligned}$$

so

$$\begin{aligned}
p(\phi_g \mid \dots) &\propto \exp \left(\phi_g N \bar{y}_g - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2} - \exp(\phi_g) \left[\exp(-\alpha_g) \sum_{k(n)=1} [\exp(\rho_n + \varepsilon_{g,n})] \right. \right. \\
&\quad \left. \left. \exp(\delta_g) \sum_{k(n)=2} [\exp(\rho_n + \varepsilon_{g,n})] + \exp(\alpha_g) \sum_{k(n)=3} [\exp(\rho_n + \varepsilon_{g,n})] \right] \right)
\end{aligned}$$

A.6 $p(\alpha_g \mid \dots)$: Metropolis

Similar to ϕ_g ,

$$p(\alpha_g \mid \dots) \propto \exp \left(\sum_{k(n) \neq 2} [y_{g,n} \eta(g, n)] - \frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} - \sum_{k(n) \neq 2} [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] \right)$$

and

$$\begin{aligned}
\sum_{k(n) \neq 2} [y_{g,n} \eta(g, n)] &= \sum_{k(n)=1} [y_{g,n} \eta(g, n)] + \sum_{k(n)=3} [y_{g,n} \eta(g, n)] \\
&= \sum_{k(n)=1} [y_{g,n} (\phi_g - \alpha_g)] + \sum_{k(n)=3} [y_{g,n} (\phi_g + \alpha_g)] \\
&= \alpha_g \left(\sum_{k(n)=3} y_{g,n} - \sum_{k(n)=1} y_{g,n} \right) + \text{constant}
\end{aligned}$$

and

$$\begin{aligned}
\sum_{k(n) \neq 2} \exp(\rho_n + \varepsilon_{g,n} + \eta(g, n)) &= \sum_{k(n)=1} [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] + \sum_{k(n)=3} [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] \\
&= \sum_{k(n)=1} [\exp(\rho_n + \varepsilon_{g,n} + \phi_g - \alpha_g)] + \sum_{k(n)=3} [\exp(\rho_n + \varepsilon_{g,n} + \phi_g + \alpha_g)] \\
&= \exp(-\alpha_g) \exp(\phi_g) \sum_{k(n)=1} [\exp(\rho_n + \varepsilon_{g,n})] + \exp(\alpha_g) \exp(\phi_g) \sum_{k(n)=3} [\exp(\rho_n + \varepsilon_{g,n})]
\end{aligned}$$

so

$$\begin{aligned}
p(\alpha_g \mid \dots) &\propto \exp \left(\alpha_g \left(\sum_{k(n)=3} y_{g,n} - \sum_{k(n)=1} y_{g,n} \right) - \frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} - \exp(-\alpha_g) \exp(\phi_g) \sum_{k(n)=1} [\exp(\rho_n + \varepsilon_{g,n})] \right. \\
&\quad \left. - \exp(\alpha_g) \exp(\phi_g) \sum_{k(n)=3} [\exp(\rho_n + \varepsilon_{g,n})] \right)
\end{aligned}$$

A.7 $p(\delta_g \mid \dots)$: Metropolis

Similar to ϕ_g ,

$$p(\delta_g \mid \dots) \propto \exp \left(\sum_{k(n) \neq 2} [y_{g,n} \eta(g, n)] - \frac{(\delta_g - \theta_\delta)^2}{2\sigma_\delta^2} - \sum_{k(n) \neq 2} [\exp(\rho_n + \varepsilon_{g,n} + \eta(g, n))] \right)$$

and

$$\begin{aligned}
\sum_{k(n)=2} [y_{g,n}\eta(g,n)] &= \sum_{k(n)=2} [y_{g,n}(\phi_g + \delta_g)] \\
&= \delta_g \sum_{k(n)=2} y_{g,n} + \text{constant}
\end{aligned}$$

and

$$\begin{aligned}
\sum_{k(n) \neq 2} \exp(\rho_n + \varepsilon_{g,n} + \eta(g,n)) &= \sum_{k(n) \neq 2} \exp(\rho_n + \varepsilon_{g,n} + \phi_g + \delta_g) \\
&= \exp(\delta_g) \exp(\phi_g) \sum_{k(n) \neq 2} \exp(\rho_n + \varepsilon_{g,n})
\end{aligned}$$

so

$$p(\delta_g \mid \cdots) \propto \exp \left(\delta_g \sum_{k(n)=2} y_{g,n} - \frac{(\delta_g - \theta_\delta)^2}{2\sigma_\delta^2} - \exp(\delta_g) \exp(\phi_g) \sum_{k(n) \neq 2} \exp(\rho_n + \varepsilon_{g,n}) \right)$$

A.8 $p(\theta_\phi \mid \dots)$: Normal

$$\begin{aligned}
p(\theta_\phi \mid \dots) &= \left[\prod_{g=1}^G \mathcal{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2) \right] \cdot \mathcal{N}(\theta_\phi \mid 0, c_\phi^2) \\
&\propto \left[\prod_{g=1}^G \exp \left(-\frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2} \right) \right] \exp \left(-\frac{\theta_\phi^2}{2c_\phi^2} \right) \\
&= \exp \left(-\sum_{g=1}^G \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2} \right) \exp \left(-\frac{\theta_\phi^2}{2c_\phi^2} \right) \\
&= \exp \left(-\frac{\sum_{g=1}^G \phi_g^2 - 2\theta_\phi \sum_{g=1}^G \phi_g + G\theta_\phi^2}{2\sigma_\phi^2} \right) \exp \left(-\frac{\theta_\phi^2}{2c_\phi^2} \right) \\
&= \exp \left(-\frac{\sum_{g=1}^G \phi_g^2 - 2\theta_\phi \sum_{g=1}^G \phi_g + G\theta_\phi^2}{2\sigma_\phi^2} - \frac{\theta_\phi^2}{2c_\phi^2} \right) \\
&= \exp \left(-\frac{c_\phi^2 \sum_{g=1}^G \phi_g^2 - 2c_\phi^2 (\sum_{g=1}^G \phi_g) \theta_\phi + Gc_\phi^2 \theta_\phi^2}{2\sigma_\phi^2 c_\phi^2} - \frac{\sigma_\phi^2 \theta_\phi^2}{2\sigma_\phi^2 c_\phi^2} \right) \\
&= \exp \left(-\frac{c_\phi^2 \sum_{g=1}^G \phi_g^2 - 2c_\phi^2 (\sum_{g=1}^G \phi_g) \theta_\phi + (Gc_\phi^2 + \sigma_\phi^2) \theta_\phi^2}{2\sigma_\phi^2 c_\phi^2} \right) \\
&\propto \exp \left(-\frac{(Gc_\phi^2 + \sigma_\phi^2) \left(\theta_\phi - \frac{c_\phi^2 (\sum_{g=1}^G \phi_g)}{Gc_\phi^2 + \sigma_\phi^2} \right)^2}{2\sigma_\phi^2 c_\phi^2} \right)
\end{aligned}$$

Hence:

$$p(\theta_\phi \mid \dots) = \mathcal{N} \left(\theta_\phi \mid \frac{c_\phi^2 \sum_{g=1}^G \phi_g}{Gc_\phi^2 + \sigma_\phi^2}, \frac{c_\phi^2 \sigma_\phi^2}{Gc_\phi^2 + \sigma_\phi^2} \right)$$

A.9 $p(\theta_\alpha \mid \dots)$: Normal

$$p(\theta_\alpha \mid \dots) \propto \left[\prod_{g=1}^G \mathcal{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2) \right] \cdot \mathcal{N}(\theta_\alpha \mid 0, c_\alpha^2)$$

From algebra similar to the derivation of $p(\theta_\phi \mid \dots)$,

$$p(\theta_\alpha \mid \dots) = \mathcal{N} \left(\theta_\alpha \mid \frac{c_\alpha^2 \sum_{g=1}^G \alpha_g}{Gc_\alpha^2 + \sigma_\alpha^2}, \frac{c_\alpha^2 \sigma_\alpha^2}{Gc_\alpha^2 + \sigma_\alpha^2} \right)$$

A.10 $p(\theta_\delta \mid \dots)$: Normal

$$p(\theta_\delta \mid \dots) \propto \left[\prod_{g=1}^G \text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2) \right] \cdot \text{N}(\theta_\delta \mid 0, c_\delta^2)$$

From algebra similar to the derivation of $p(\theta_\phi \mid \dots)$,

$$p(\theta_\delta \mid \dots) = N \left(\theta_\delta \mid \frac{c_\delta^2 \sum_{g=1}^G \delta_g}{G_\delta c_\delta^2 + \sigma_\delta^2}, \frac{c_\delta^2 \sigma_\delta^2}{G_\delta c_\delta^2 + \sigma_\delta^2} \right)$$

A.11 $p(\tau^2 \mid \dots)$: Gamma

$$\begin{aligned} p(\tau^2 \mid \dots) &= \left[\prod_{g=1}^G \text{Inv-Gamma} \left(\gamma_g^2 \mid \text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu \tau^2}{2} \right) \right] \cdot \text{Gamma}(\tau^2 \mid \text{shape} = a, \text{rate} = b) \\ &\propto \left[\Gamma(\nu/2)^{-G} \left(\frac{\nu \tau^2}{2} \right)^{G\nu/2} \left(\prod_{g=1}^G \gamma_g^2 \right)^{-(\nu/2+1)} \exp \left(-\frac{\nu \tau^2}{2} \sum_{g=1}^G \frac{1}{\gamma_g^2} \right) \right] \cdot (\tau^2)^{a-1} \exp(-b\tau^2) \\ &\propto \left[(\tau^2)^{G\nu/2} \exp \left(-\tau^2 \cdot \frac{\nu}{2} \sum_{g=1}^G \frac{1}{\gamma_g^2} \right) \right] \cdot (\tau^2)^{a-1} \exp(-b\tau^2) \\ &= (\tau^2)^{G\nu/2+a-1} \exp \left(-\tau^2 \left(b + \frac{\nu}{2} \sum_{g=1}^G \frac{1}{\gamma_g^2} \right) \right) \end{aligned}$$

Hence:

$$p(\tau^2 \mid \dots) = \text{Gamma} \left(\tau^2 \mid \text{shape} = a + \frac{G\nu}{2}, \text{rate} = b + \frac{\nu}{2} \sum_{g=1}^G \frac{1}{\gamma_g^2} \right)$$

A.12 $p\left(\frac{1}{\gamma_g^2} \mid \dots\right)$ Gamma

$$\begin{aligned}
p(\gamma_g^2 \mid \dots) &= \left[\prod_{n=1}^N N(\varepsilon_{g,n} \mid 0, \gamma_g^2) \right] \cdot \text{Inv-Gamma} \left(\gamma_g^2 \mid \text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu\tau^2}{2} \right) \\
&\propto \left[\prod_{n=1}^N (\gamma_g^2)^{-1/2} \exp \left(-\frac{1}{\gamma_g^2} \frac{\varepsilon_{g,n}^2}{2} \right) \right] \cdot (\gamma_g^2)^{-(\nu/2+1)} \exp \left(-\frac{1}{\gamma_g^2} \frac{\nu\tau^2}{2} \right) \\
&= \left[(\gamma_g^2)^{-N/2} \exp \left(-\frac{1}{\gamma_g^2} \frac{1}{2} \sum_{n=1}^N \varepsilon_{g,n}^2 \right) \right] \cdot (\gamma_g^2)^{-(\nu/2+1)} \exp \left(-\frac{1}{\gamma_g^2} \frac{\nu\tau^2}{2} \right) \\
&= (\gamma_g^2)^{-((N+\nu)/2+1)} \exp \left(-\frac{1}{\gamma_g^2} \frac{1}{2} \left(\nu\tau^2 + \sum_{n=1}^N \varepsilon_{g,n}^2 \right) \right)
\end{aligned}$$

which is the kernel of an inverse gamma distribution. Hence:

$$p\left(\frac{1}{\gamma_g^2} \mid \dots\right) = \text{Gamma} \left(\frac{1}{\gamma_g^2} \mid \text{shape} = \frac{N+\nu}{2}, \text{rate} = \frac{1}{2} \left(\nu\tau^2 + \sum_{n=1}^N \varepsilon_{g,n}^2 \right) \right)$$

A.13 $p\left(\frac{1}{\sigma_\rho^2} \mid \dots\right)$ Truncated Gamma

$$\begin{aligned}
p(\sigma_\rho^2 \mid \dots) &= p(\sigma_\rho \mid \dots) \frac{1}{2} (\sigma_\rho^2)^{-1/2} \quad (\text{transformation in Section A.1}) \\
&\propto \left[\prod_{n=1}^N N(\rho_n \mid 0, \sigma_\rho^2) \right] \cdot U(\sigma_\rho \mid 0, s_\rho) \frac{1}{2} (\sigma_\rho^2)^{-1/2} \\
&\propto \prod_{n=1}^N \left[\frac{1}{\sqrt{\sigma_\rho^2}} \exp \left(-\frac{\rho_n^2}{2\sigma_\rho^2} \right) \right] \cdot I(0 < \sigma_\rho < s_\rho) (\sigma_\rho^2)^{-1/2} \\
&= (\sigma_\rho^2)^{-N/2} \exp \left(-\frac{1}{\sigma_\rho^2} \frac{1}{2} \sum_{n=1}^N \rho_n^2 \right) \cdot I(0 < \sigma_\rho < s_\rho) (\sigma_\rho^2)^{-1/2} \\
&= (\sigma_\rho^2)^{-(N/2-1/2+1)} \exp \left(-\frac{1}{\sigma_\rho^2} \frac{1}{2} \sum_{n=1}^N \rho_n^2 \right) \cdot I(0 < \sigma_\rho < s_\rho)
\end{aligned}$$

which is the kernel of a truncated inverse gamma distribution. Hence:

$$p\left(\frac{1}{\sigma_\rho^2} \mid \dots\right) = \text{Gamma} \left(\frac{1}{\sigma_\rho^2} \mid \text{shape} = \frac{N-1}{2}, \text{rate} = \frac{1}{2} \sum_{n=1}^N \rho_n^2 \right) I \left(\frac{1}{\sigma_\rho^2} > \frac{1}{s_\rho^2} \right)$$

A.14 $p\left(\frac{1}{\sigma_\phi^2} \mid \dots\right)$: Truncated Gamma

$$\begin{aligned}
p(\sigma_\phi^2 \mid \dots) &= p(\sigma_\phi \mid \dots) \frac{1}{2} (\sigma_\phi^2)^{-1/2} \quad (\text{transformation in Section A.1}) \\
&\propto \left[\prod_{g=1}^G \mathcal{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2) \right] \cdot \mathcal{U}(\sigma_\phi \mid 0, s_\phi) (\sigma_\phi^2)^{-1/2} \\
&\propto \left[\prod_{g=1}^G (\sigma_\phi^2)^{-1/2} \exp\left(-\frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right) \right] \mathcal{I}(0 < \sigma_\phi^2 < s_\phi^2) (\sigma_\phi^2)^{-1/2} \\
&= (\sigma_\phi^2)^{-G/2} \exp\left(-\sum_{g=1}^G \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right) \mathcal{I}(0 < \sigma_\phi^2 < s_\phi^2) (\sigma_\phi^2)^{-1/2} \\
&= (\sigma_\phi^2)^{-(G/2-1/2+1)} \exp\left(-\frac{1}{\sigma_\phi^2} \frac{1}{2} \sum_{g=1}^G (\phi_g - \theta_\phi)^2\right) \mathcal{I}(0 < \sigma_\phi^2 < s_\phi^2)
\end{aligned}$$

which is the kernel of a truncated inverse gamma distribution. Hence:

$$p\left(\frac{1}{\sigma_\phi^2} \mid \dots\right) = \text{Gamma}\left(\text{shape} = \frac{G-1}{2}, \text{rate} = \frac{1}{2} \sum_{g=1}^G (\phi_g - \theta_\phi)^2\right) \mathcal{I}\left(\frac{1}{\sigma_\phi^2} > \frac{1}{s_\phi^2}\right)$$

A.15 $p\left(\frac{1}{\sigma_\alpha^2} \mid \dots\right)$: Truncated Gamma

Analogously to σ_ϕ ,

$$p\left(\frac{1}{\sigma_\alpha^2} \mid \dots\right) = \text{Gamma}\left(\text{shape} = \frac{G-1}{2}, \text{rate} = \frac{1}{2} \sum_{g=1}^G (\alpha_g - \theta_\alpha)^2\right) \mathcal{I}\left(\frac{1}{\sigma_\alpha^2} > \frac{1}{s_\alpha^2}\right)$$

A.16 $p\left(\frac{1}{\sigma_\delta^2} \mid \dots\right)$: Truncated Gamma

Analogously to σ_ϕ ,

$$p\left(\frac{1}{\sigma_\delta^2} \mid \dots\right) = \text{Gamma}\left(\text{shape} = \frac{G-1}{2}, \text{rate} = \frac{1}{2} \sum_{g=1}^G (\delta_g - \theta_\delta)^2\right) \mathcal{I}\left(\frac{1}{\sigma_\delta^2} > \frac{1}{s_\delta^2}\right)$$