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

Will Landau

Department of Statistics
Iowa State University

April 19, 2014

Contents

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 ($g = 1, \ldots, G$) in sample n ($n = 1, \ldots, N$). 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:

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

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

Then:

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

$$p(\theta_\phi \mid \cdots) = \mathrm{N}\left(\theta_\phi \;\middle|\; \frac{c_\phi^2 \sum_{g=1}^G \phi_g}{Gc_\phi^2 + \sigma_\phi^2},\ \frac{c_\phi^2 \sigma_\phi^2}{Gc_\phi^2 + \sigma_\phi^2}\right)$$

$$p(\theta_\alpha \mid \cdots) = N\left(\theta_\alpha \;\middle|\; \frac{c_\alpha^2 \sum_{g=1}^G \alpha_g}{G_\alpha c_\alpha^2 + \sigma_\alpha^2},\ \frac{c_\alpha^2 \sigma_\alpha^2}{G_\alpha c_\alpha^2 + \sigma_\alpha^2}\right)$$

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

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

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

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

$$p\left(\frac{1}{\sigma_\phi^2} \;\middle|\; \cdots\right) = \mathrm{Gamma}\left(\text{shape} = \frac{G-1}{2},\ \text{rate} = \frac{1}{2}\sum_{g=1}^G (\phi_g - \theta_\phi)^2\right) \mathrm{I}\left(\frac{1}{\sigma_\phi^2} > \frac{1}{s_\phi^2}\right)$$

$$p\left(\frac{1}{\sigma_\alpha^2} \;\middle|\; \cdots\right) = \mathrm{Gamma}\left(\text{shape} = \frac{G-1}{2},\ \text{rate} = \frac{1}{2}\sum_{g=1}^G (\alpha_g - \theta_\alpha)^2\right) \mathrm{I}\left(\frac{1}{\sigma_\alpha^2} > \frac{1}{s_\alpha^2}\right)$$

$$p\left(\frac{1}{\sigma_\delta^2} \;\middle|\; \cdots\right) = \mathrm{Gamma}\left(\text{shape} = \frac{G-1}{2},\ \text{rate} = \frac{1}{2}\sum_{g=1}^G (\delta_g - \theta_\delta)^2\right) \mathrm{I}\left(\frac{1}{\sigma_\delta^2} > \frac{1}{s_\delta^2}\right)$$

4 A Metropolis-Hastings Algorithm for Sampling ν

The full conditional distribution of ν is of the form,

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

where

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

Using Stirling's approximation to the gamma function,

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

Hence,

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

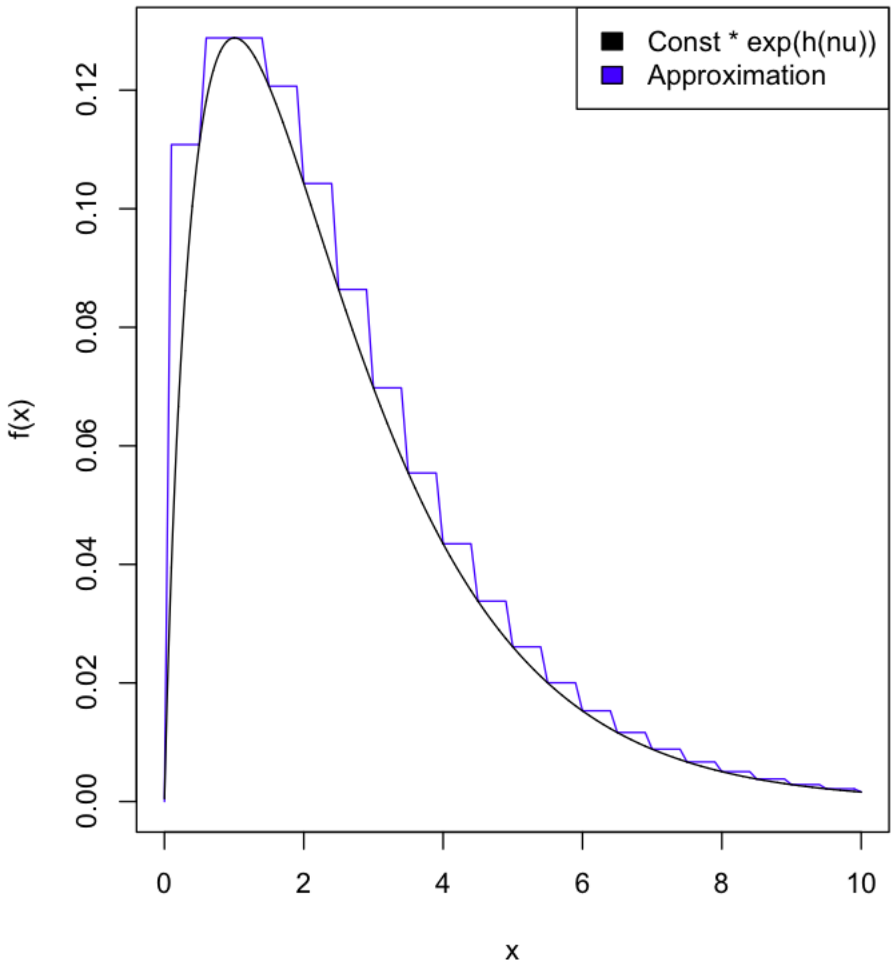
Thus, $\frac{h(\nu)}{G}$ is approximately the log kernel of

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

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

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

However, this distribution is stochastically smaller than the true full conditional of ν, so it cannot serve as a Metropolis-Hastings proposal. Instead, I will use a step function approximation to $G * \exp(h(\nu))$ as a proposal distribution.



There will be at least 1000 equally-spaced "bins", unlike in the above demonstration. The support of this step function will be the interval $(0, U)$, where

$$U = \frac{a + 4\sqrt{a}}{b}$$

where

$$a = \frac{G}{2} + 1$$
$$b = -\frac{G}{2} - \frac{G\log(\tau^2)}{2} + \sum_{g=1}^{G}\left[\log\gamma_g + \frac{\tau^2}{2}\frac{1}{\gamma_g^2}\right]$$

i.e., an interval from 0 up to 4 standard deviations above the mean of the crude gamma approximation to the full conditional. There will be an extra interval from U to d in the support as well, where d is the initialization constant used in the prior distribution for ν.

Let $q(\nu)$ be the log of this step function approximation. Then, we have a Metropolis-Hastings algorithm for sampling $\nu^{(i+1)}$, the $(i+1)$'th sample of ν in the MCMC, given $\nu^{(i)}$.

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,

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

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 $p(\theta \mid \cdots)$. Then,

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

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

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

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

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

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

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

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

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

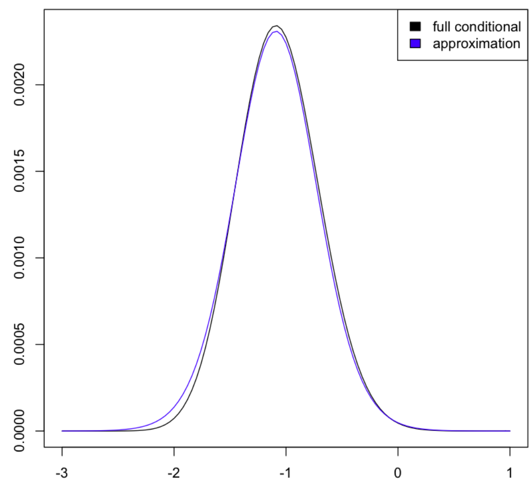
Unfortunately, the tails of this normal approximation are too thin for the proposal to be efficient. However, we can instead sample $\sqrt{|h_2|}(\theta - \widehat{\theta})$ from a t_l distribution, where the degrees of freedom, l, is some large positive number. Let $q(\theta)$ be the log kernel of this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

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

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

$$\lim_{|\theta| \to \infty} \log \frac{\exp(q(\theta))}{p(\theta \mid \cdots)} = \infty$$

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



Hence, the t_l 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 $q(\theta)$.
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 A rejection sampling alternative for sampling ρ_n, ϕ_g, α_g, δ_g, and $\varepsilon_{g,n}$

Instead of sampling ρ_n, ϕ_g, α_g, δ_g, and $\varepsilon_{g,n}$ using a Metropolis-Hastings algorithm as in the previous section, I can use the t_l distribution from before as an envelope in a rejection sampler.

1. Sample a proposal θ^* from $q(\theta)$.
2. Sample $u \sim U(0,1)$.
3. If $\log u \le h(\theta^*) - \log M - q(\theta^*)$, set $\theta^{(i+1)} = \theta^*$. Otherwise, return to step 1.

where

$$\log M = \sup_\theta \left(h(\theta) - q(\theta)\right)$$

to find $\log M$, we maximize

$$\begin{aligned} m(\theta) &= h(\theta) - q(\theta) \\ &= A\theta - B(\theta - C)^2 - De^\theta - Ee^{-\theta} + \frac{l+1}{2}\log\left(1 + \frac{\sqrt{|h_2|}(\theta - \widehat{\theta})^2}{l}\right) \end{aligned}$$

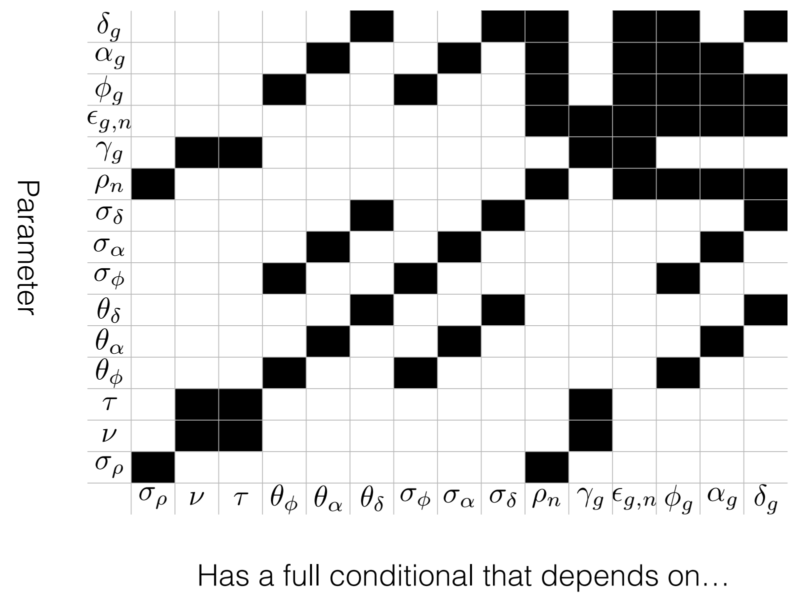
using Newton Raphson with the derivatives,

$$m'(\theta) = A - 2B(\theta - C) - De^\theta + Ee^{-\theta} + \frac{l+1}{2}\frac{2\frac{\sqrt{|h_2|}}{l}(\theta - \widehat{\theta})}{1 + \frac{\sqrt{|h_2|}}{l}(\theta - \widehat{\theta})^2}$$

$$m''(\theta) = -2B - De^\theta - Ee^{-\theta} - \frac{l+1}{2}\left(\frac{2\frac{\sqrt{|h_2|}}{l}(\theta - \widehat{\theta})}{1 + \frac{\sqrt{|h_2|}}{l}(\theta - \widehat{\theta})^2}\right)^2 + \frac{l+1}{2}\frac{2\frac{\sqrt{|h_2|}}{l}}{1 + \frac{\sqrt{|h_2|}}{l}(\theta - \widehat{\theta})^2}$$

7 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.



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

1. ρ_n $(n = 1, \ldots, N)$
2. γ_g $(g = 1, \ldots, G)$
3. $\varepsilon_{g,n}$ $(g = 1, \ldots, G)$, $(n = 1, \ldots, N)$
4. ϕ_g $(g = 1, \ldots, G)$
5. α_g $(g = 1, \ldots, G)$
6. δ_g $(g = 1, \ldots, G)$
7. ν, θ_ϕ, θ_α, θ_δ
8. τ, σ_ρ, σ_ϕ, σ_α, σ_δ

8 Diagnostics

8.1 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, $\widehat{R}$, compares the within-chain variance, W, to the between-chain variance, B. Suppose there are J chains, each with I iterations. Then,

$$\widehat{R} = \sqrt{1 - \frac{1}{I}\left(\frac{B}{W} - 1\right)}$$

$$B = \frac{I}{J-1}\sum_{j=1}^{J}(\overline{\eta}_{.j} - \overline{\eta}_{..})^2, \qquad \overline{\eta}_{.j} = \frac{1}{I}\sum_{i=1}^{I}\eta_{ij}, \quad \overline{\eta}_{..}\sum_{j=1}^{J}\overline{\eta}_{.j}$$

$$W = \frac{1}{J}\sum_{j=1}^{J}s_j^2, \qquad s_j^2 = \frac{1}{I-1}\sum_{i=1}^{I}(\eta_{ij} - \overline{\eta}_{.j})^2$$

$\widehat{R} \to 1$ as $I \to \infty$. An $\widehat{R}$ value far above 1 indicates a lack of convergence, but an $\widehat{R}$ 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:

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

where

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

8.2 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 $\boldsymbol{y}$, the one with the minimum DIC is preferred. DIC is based on the deviance,

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

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

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

where $\widehat{\boldsymbol{\eta}}$ 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 $\overline{\boldsymbol{\eta}}$ is the collection of within-chain parameter means, then we can estimate DIC by

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

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

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

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

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

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

where

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

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:

$$p(\nu \mid \cdots) \propto \left[\prod_{g=1}^{G} \text{Inv-Gamma}\left(\gamma_g^2 \;\middle|\; \text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu\tau^2}{2} \right) \right] \cdot \text{U}(\nu \mid 0, d)$$

$$p(\rho_n \mid \cdots) \propto \left[\prod_{g=1}^{G} \text{Poisson}(y_{g,n} \mid \exp(\rho_n + \varepsilon_{g,n} + \eta(g,n))) \right] \cdot \text{N}(\rho_n \mid 0, \sigma_\rho^2)$$

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

$$p(\alpha_g \mid \cdots) \propto \left[\prod_{k(n) \neq 2} \text{Poisson}(y_{g,n} \mid \exp(\rho_n + \varepsilon_{g,n} + \eta(g,n))) \right] \cdot \text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)$$

$$p(\delta_g \mid \cdots) \propto \left[\prod_{k(n) = 2} \text{Poisson}(y_{g,n} \mid \exp(\rho_n + \varepsilon_{g,n} + \eta(g,n))) \right] \cdot \text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)$$

$$p(\varepsilon_{g,n} \mid \cdots) \propto \text{Poisson}(y_{g,n} \mid \exp(\rho_n + \varepsilon_{g,n} + \eta(g,n))) \cdot \text{N}(\varepsilon_{g,n} \mid 0, \gamma_g^2)$$

$$p\left(\sigma_\rho \mid \cdots\right) = \left[\prod_{n=1}^{N} \mathrm{N}(\rho_n \mid 0, \sigma_\rho^2)\right] \cdot \mathrm{U}(\sigma_\rho \mid 0, s_\rho)$$

$$p(\gamma_g^2 \mid \cdots) \propto \left[\prod_{n=1}^{N} \mathrm{N}(\varepsilon_{g,n} \mid 0, \gamma_g^2)\right] \cdot \text{Inv-Gamma}\left(\gamma_g^2 \;\middle|\; \text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu\tau^2}{2}\right)$$

$$p(\tau^2 \mid \cdots) \propto \left[\prod_{g=1}^{G} \text{Inv-Gamma}\left(\gamma_g^2 \;\middle|\; \text{shape} = \frac{\nu}{2}, \text{scale} = \frac{\nu\tau^2}{2}\right)\right] \cdot \text{Gamma}(\tau^2 \mid \text{shape} = a, \text{rate} = b)$$

$$p(\theta_\phi \mid \cdots) \propto \left[\prod_{g=1}^{G} \mathrm{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2)\right] \cdot \mathrm{N}(\theta_\phi \mid 0, c_\phi^2)$$

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

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

$$p(\sigma_\phi \mid \cdots) \propto \left[\prod_{g=1}^{G} \mathrm{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2)\right] \cdot \mathrm{U}(\sigma_\phi \mid 0, s_\phi)$$

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

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

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 $p(\nu \mid \cdots)$: Metropolis

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

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

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

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

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

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

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

and

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

so

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

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

Similar to ϕ_g,

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

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

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

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

Similar to ϕ_g,

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

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 $p(\theta_\phi \mid \cdots)$: Normal

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

Hence:

$$
p(\theta_\phi \mid \cdots) = \mathrm{N}\left(\theta_\phi \;\middle|\; \frac{c_\phi^2 \sum_{g=1}^{G} \phi_g}{Gc_\phi^2 + \sigma_\phi^2}, \; \frac{c_\phi^2 \sigma_\phi^2}{Gc_\phi^2 + \sigma_\phi^2}\right)
$$

A.9 $p\left(\theta_\alpha \mid \cdots\right)$: Normal

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

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

$$
p(\theta_\alpha \mid \cdots) = N\left(\theta_\alpha \;\middle|\; \frac{c_\alpha^2 \sum_{g=1}^{G} \alpha_g}{G_\alpha c_\alpha^2 + \sigma_\alpha^2}, \; \frac{c_\alpha^2 \sigma_\alpha^2}{G_\alpha c_\alpha^2 + \sigma_\alpha^2}\right)
$$

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

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

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

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

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

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

Hence:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Analogously to σ_ϕ,

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

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

Analogously to σ_ϕ,

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