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

Will Landau

Department of Statistics
Iowa State University

February 9, 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 offspring 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 offspring than in either of the parent lines.
2. Low parent heterosis: the gene is significantly less expressed in the offspring than in either of the parent lines.
3. Mid parent heterosis: the expression level of the gene in the offspring 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 offspring, 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 offspring
- 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 offspring relative to the parental mean.

With MCMC samples of these quantities, we can calculate empirical estimates of the following probabilities of interest:

- $P(\alpha_g \neq 0 \mid \boldsymbol{y})$, 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.
- $P(\delta_g \neq 0 \mid \boldsymbol{y})$, the probability of mid parent heterosis.

2 The Model

$$
\begin{aligned}
y_{g,n} &\overset{\text{ind}}{\sim} \text{Poisson}(\exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))) \\
c_n &\overset{\text{ind}}{\sim} \text{N}(0, \sigma_c^2) \\
\sigma_c &\sim \text{U}(0, \sigma_{c0}) \\
\varepsilon_{g,n} &\overset{\text{ind}}{\sim} \text{N}(0, \eta_g^2) \\
\eta_g^2 &\overset{\text{ind}}{\sim} \text{Inv-Gamma}\left(\text{shape} = \frac{d}{2},\ \text{rate} = \frac{d \cdot \tau^2}{2}\right) \\
d &\sim \text{U}(0, d_0) \\
\tau^2 &\sim \text{Gamma}(\text{shape} = a_\tau, \text{rate} = b_\tau) \\
\phi_g &\overset{\text{ind}}{\sim} \text{N}(\theta_\phi, \sigma_\phi^2) \\
\theta_\phi &\sim \text{N}(0, \gamma_\phi^2) \\
\sigma_\phi &\sim \text{U}(0, \sigma_{\phi 0}) \\
\alpha_g &\overset{\text{ind}}{\sim} \pi_\alpha^{1 - I(\alpha_g)} [(1 - \pi_\alpha) \text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)} \\
\theta_\alpha &\sim \text{N}(0, \gamma_\alpha^2) \\
\sigma_\alpha &\sim \text{U}(0, \sigma_{\alpha 0}) \\
\pi_\alpha &\sim \text{Beta}(a_\alpha, b_\alpha) \\
\delta_g &\overset{\text{ind}}{\sim} \pi_\delta^{1 - I(\delta_g)} [(1 - \pi_\delta) \text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)} \\
\theta_\delta &\sim \text{N}(0, \gamma_\delta^2) \\
\sigma_\delta &\sim \text{U}(0, \sigma_{\delta 0}) \\
\pi_\delta &\sim \text{Beta}(a_\delta, b_\delta)
\end{aligned}
$$

where:

- $I(x) = 0$ if $x = 0$ and 1 otherwise.
- Conditional independence is implied unless otherwise specified.
- The parameters to the left of the "$\sim$" are implicitly conditioned on the parameters to the right.
- $\mu(n, \phi_g, \alpha_g, \delta_g)$ is the function given by:

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

3 Full Conditional Distributions

Define:

- $k(n)$ = treatment group of library n.
- $\lambda_{g,n} = \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))\ )$
- G_α = number of genes for which $\alpha_g \neq 0$
- G_δ = number of genes for which $\delta_g \neq 0$
- $I(x) = 0$ if $x = 0$ and 1 otherwise.

Then:

$$p(c_n \mid \cdots) \propto \exp\left(c_n G\overline{y}_{.n} - \exp(c_n)\sum_{g=1}^{G}\exp(\varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)) - \frac{c_n^2}{2\sigma_c^2}\right)$$

$$p(\varepsilon_{g,n} \mid \cdots) \propto \exp\left(y_{g,n}\varepsilon_{g,n} - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)) - \frac{\varepsilon_{g,n}^2}{2\eta_g^2}\right)$$

$$p\left(\frac{1}{\sigma_c^2}\ \middle|\ \cdots\right) = \text{Gamma}\left(\frac{1}{\sigma_c^2}\middle|\text{shape} = \frac{N-1}{2},\ \text{rate} = \frac{1}{2}\sum_{n=1}^{N} c_n^2\right) I\left(\frac{1}{\sigma_c^2} > \frac{1}{\sigma_{c0}^2}\right)$$

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

$$p(d \mid \cdots) \propto \Gamma\left(d/2\right)^{-G}\left(\frac{d\cdot\tau^2}{2}\right)^{Gd/2}\left(\prod_{g=1}^{G}\eta_g^2\right)^{-(d/2+1)}\exp\left(-\frac{d\cdot\tau^2}{2}\sum_{g=1}^{G}\frac{1}{\eta_g^2}\right) I(0 < d < d_0)$$

$$p(\tau^2 \mid \cdots) = \text{Gamma}\left(\tau^2\ \middle|\ \text{shape} = a_\tau + \frac{Gd}{2}\ ,\ \text{rate} = b_\tau + \frac{d}{2}\sum_{g=1}^{G}\frac{1}{\eta_g^2}\right)$$

$$p(\phi_g \mid \cdots) \propto \exp\left(\sum_{n=1}^{N}\left[y_{g,n}\mu(n, \phi_g, \alpha_g, \delta_g) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))\right] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right)$$

$$\begin{aligned}p(\alpha_g \mid \cdots) \propto \exp\Bigg(&\sum_{k(n)\neq 2}\left[y_{g,n}\cdot\mu(n, \phi_g, \alpha_g, \delta_g) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))\right]\\ &-I(\alpha_g)\left(\frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} + \log(1-\pi_\alpha)\right) + (1 - I(\alpha_g))\log\pi_\alpha\Bigg)\end{aligned}$$

$$\begin{aligned}p(\delta_g \mid \cdots) \propto \exp\Bigg(&\sum_{k(n)\neq 2}\left[y_{g,n}\cdot\mu(n, \phi_g, \alpha_g, \delta_g) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))\right]\\ &-I(\delta_g)\left(\frac{(\delta_g - \theta_\delta)^2}{2\sigma_\delta^2} + \log(1-\pi_\delta)\right) + (1 - I(\delta_g))\log\pi_\delta\Bigg)\end{aligned}$$

$$
\begin{aligned}
p(\phi_g, \alpha_g, \delta_g \mid \cdots) \propto \exp \Bigg(& \sum_{n=1}^{N} [y_{g,n}\mu(n, \phi_g, \alpha_g, \delta_g) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2} \\
& - I(\alpha_g) \left(\frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} + \log(1 - \pi_\alpha) \right) + (1 - I(\alpha_g)) \log \pi_\alpha \\
& - I(\delta_g) \left(\frac{(\delta_g - \theta_\delta)^2}{2\sigma_\delta^2} + \log(1 - \pi_\delta) \right) + (1 - I(\delta_g)) \log \pi_\delta \Bigg)
\end{aligned}
$$

$$
\begin{aligned}
p(\theta_\phi \mid \cdots) &= \mathrm{N}\left(\theta_\phi \;\middle|\; \frac{\gamma_\phi^2 \sum_{g=1}^{G} \phi_g}{G\gamma_\phi^2 + \sigma_\phi^2} \;,\; \frac{\gamma_\phi^2 \sigma_\phi^2}{G\gamma_\phi^2 + \sigma_\phi^2} \right) \\
p(\theta_\alpha \mid \cdots) &= N\left(\theta_\alpha \;\middle|\; \frac{\gamma_\alpha^2 \sum_{\alpha_g \neq 0} \alpha_g}{G_\alpha\gamma_\alpha^2 + \sigma_\alpha^2} \;,\; \frac{\gamma_\alpha^2 \sigma_\alpha^2}{G_\alpha\gamma_\alpha^2 + \sigma_\alpha^2} \right) \\
p(\theta_\delta \mid \cdots) &= N\left(\theta_\delta \;\middle|\; \frac{\gamma_\delta^2 \sum_{\delta_g \neq 0} \delta_g}{G_\delta\gamma_\delta^2 + \sigma_\delta^2} \;,\; \frac{\gamma_\delta^2 \sigma_\delta^2}{G_\delta\gamma_\delta^2 + \sigma_\delta^2} \right) \\
p\left(\frac{1}{\sigma_\phi^2} \;\middle|\; \cdots \right) &= \mathrm{Gamma}\left(\frac{1}{\sigma_\phi^2} \;\middle|\; \text{shape} = \frac{G-1}{2},\ \text{rate} = \frac{1}{2}\sum_{g=1}^{G} (\phi_g - \theta_\phi)^2 \right) \mathrm{I}\left(\frac{1}{\sigma_\phi^2} > \frac{1}{\sigma_{\phi 0}^2} \right) \\
p\left(\frac{1}{\sigma_\alpha^2} \;\middle|\; \cdots \right) &= \mathrm{Gamma}\left(\frac{1}{\sigma_\alpha^2} \;\middle|\; \text{shape} = \frac{G_\alpha-1}{2},\ \text{rate} = \frac{1}{2}\sum_{\alpha_g \neq 0} (\alpha_g - \theta_\alpha)^2 \right) \mathrm{I}\left(\frac{1}{\sigma_\alpha^2} > \frac{1}{\sigma_{\alpha 0}^2} \right) \\
p\left(\frac{1}{\sigma_\delta^2} \;\middle|\; \cdots \right) &= \mathrm{Gamma}\left(\frac{1}{\sigma_\delta^2} \;\middle|\; \text{shape} = \frac{G_\delta-1}{2},\ \text{rate} = \frac{1}{2}\sum_{\delta_g \neq 0} (\delta_g - \theta_\delta)^2 \right) \mathrm{I}\left(\frac{1}{\sigma_\delta^2} > \frac{1}{\sigma_{\delta 0}^2} \right) \\
p(\pi_\alpha \mid \cdots) &= \mathrm{Beta}(\pi_\alpha \mid G - G_\alpha + \alpha_\tau,\ G_\alpha + b_\tau) \\
p(\pi_\delta \mid \cdots) &= \mathrm{Beta}(\pi_\delta \mid G - G_\delta + \delta_\tau,\ G_\delta + b_\tau)
\end{aligned}
$$

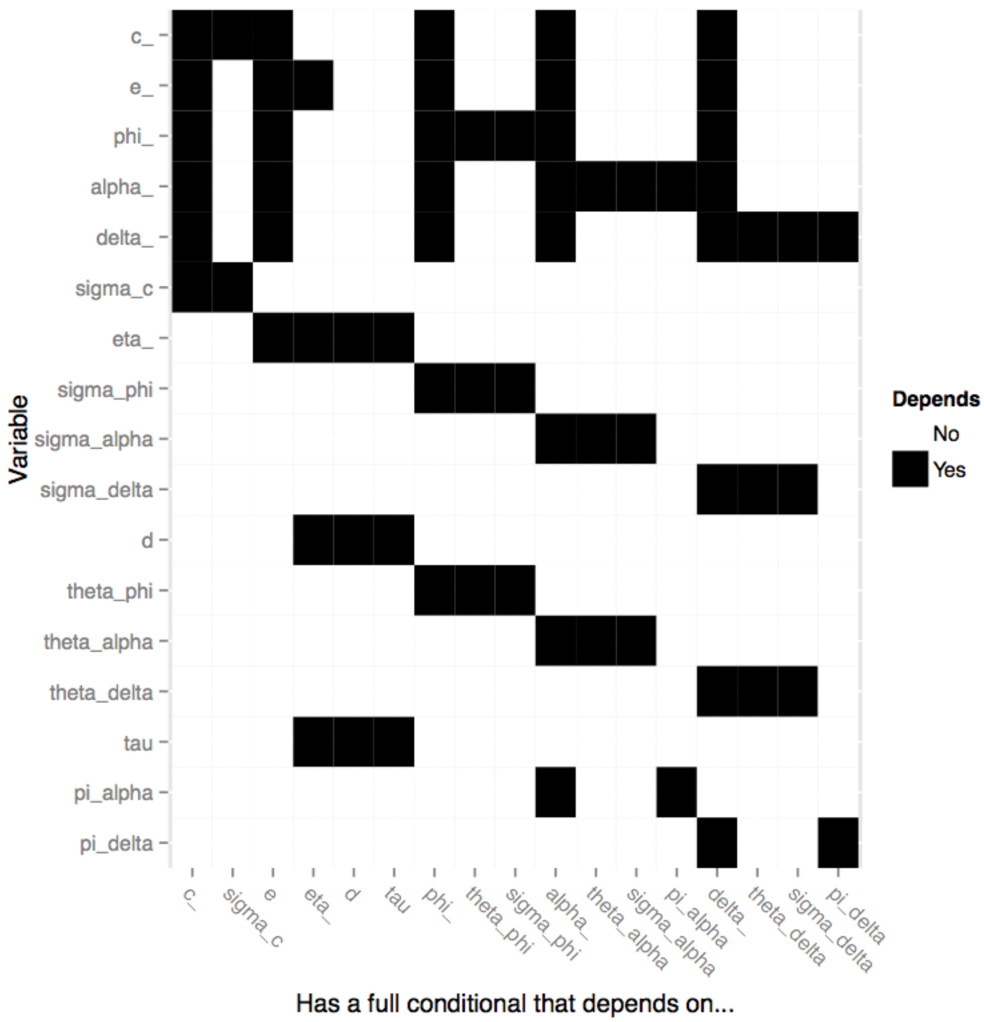
4 The Gibbs Sampler

For certain parameters, the full conditional distribution is independent of other key parameters. For example, the full conditional distribution of c_1 does not contain c_2. Hence, c_1 and c_2 can be sampled in parallel in a single Gibbs step. Obvious sets of parameters that can be jointly sampled are:

- $c_1, \ldots, c_N$
- $\varepsilon_{1,1}, \varepsilon_{1,2}, \ldots, \varepsilon_{1,N}, \varepsilon_{2,N}, \ldots, \varepsilon_{G,N}$
- $\eta_1^2, \ldots, \eta_G^2$
- $\phi_1, \ldots, \phi_G$
- $\alpha_1, \ldots, \alpha_G$

- $\delta_1, \ldots, \delta_G$

The following raster plot gives us a more complete idea of which parameters can be jointly sampled:



Hence, each of the following sets of parameters can be jointly sampled:

1. $c_1, \ \ldots, \ c_N$
2. $\tau, \ \pi_\alpha, \ \pi_\delta$
3. $d, \ \theta_\phi, \ \theta_\alpha, \ \theta_\delta$

4. $\sigma_c,\ \sigma_\phi,\ \sigma_\alpha,\ \sigma_\delta,\ \eta_1^2,\ \ldots,\ \eta_G^2$
5. $\varepsilon_{1,1},\ \varepsilon_{1,2},\ \ldots,\ \varepsilon_{1,N},\ \varepsilon_{2,N},\ \ldots,\ \varepsilon_{G,N}$
6. $\phi_1,\ \ldots,\ \phi_G$
7. $\alpha_1,\ \ldots,\ \alpha_G$
8. $\delta_1,\ \ldots,\ \delta_G$

In order, these are the 8 steps of the Gibbs sampler. Alternatively, one could sample each triplet $(\phi_g, \alpha_g, \delta_g)$ jointly in a single Metropolis step using $p(\phi_g, \alpha_g, \delta_g \cdots)$.

5 Diagnostics

5.1 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 ψ_{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,

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

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

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

$\widehat{R} \to 1$ 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

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

5.2 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,

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

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

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

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

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

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

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

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

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

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

where

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

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.
- $\lambda_{g,n} = \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))$
- G_α = number of genes for which $\alpha_g \neq 0$
- G_δ = number of genes for which $\delta_g \neq 0$
- $I(x) = 0$ if $x = 0$ and 1 otherwise.

Then from the model in Section **??**, we get:

$$
\begin{aligned}
p(c_n \mid \cdots) &\propto \left[\prod_{g=1}^{G} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right] \cdot \text{N}(c_n \mid 0, \sigma_c^2) \\
p(\varepsilon_{g,n} \mid \cdots) &\propto \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))) \cdot \text{N}(\varepsilon_{g,n} \mid 0, \eta_g^2) \\
p\left(\sigma_c \mid \cdots\right) &= \left[\prod_{n=1}^{N} \text{N}(c_n \mid 0, \sigma_c^2)\right] \cdot \text{U}(\sigma_c \mid 0, \sigma_{c0}) \\
p(\eta_g^2 \mid \cdots) &\propto \left[\prod_{n=1}^{N} \text{N}(\varepsilon_{g,n} \mid 0, \eta_g^2)\right] \cdot \text{Inv-Gamma}\left(\eta_g^2 \mid \text{shape} = \frac{d}{2}, \text{rate} = \frac{d \cdot \tau^2}{2}\right) \\
p(d \mid \cdots) &\propto \left[\prod_{g=1}^{G} \text{Inv-Gamma}\left(\eta_g^2 \mid \text{shape} = \frac{d}{2}, \text{rate} = \frac{d \cdot \tau^2}{2}\right)\right] \cdot \text{U}(d \mid 0, d_0) \\
p(\tau^2 \mid \cdots) &\propto \left[\prod_{g=1}^{G} \text{Inv-Gamma}\left(\eta_g^2 \mid \text{shape} = \frac{d}{2}, \text{rate} = \frac{d \cdot \tau^2}{2}\right)\right] \cdot \text{Gamma}(\tau^2 \mid \text{shape} = a_\tau, \text{rate} = b_\tau) \\
p(\phi_g \mid \cdots) &\propto \left[\prod_{n=1}^{N} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right] \cdot \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2) \\
p(\alpha_g \mid \cdots) &\propto \left[\prod_{k(n) \neq 2} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right] \\
&\quad \times \pi_\alpha^{1 - I(\alpha_g)} [(1 - \pi_\alpha) \text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)}
\end{aligned}
$$

$$
\begin{aligned}
p(\delta_g \mid \cdots) &\propto \left[\prod_{k(n)=2} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \delta_g, \delta_g)))\right] \\
p(\phi_g, \alpha_g, \delta_g \mid \cdots) &\propto \left[\prod_{n=1}^{N} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right] \cdot \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2) \\
&\quad \times \pi_\alpha^{1-I(\alpha_g)}[(1-\pi_\alpha)\text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)} \times \pi_\delta^{1-I(\delta_g)}[(1-\pi_\delta)\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)} \\
&\quad \times \pi_\delta^{1-I(\delta_g)}[(1-\pi_\delta)\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)} \\
p(\theta_\phi \mid \cdots) &\propto \left[\prod_{g=1}^{G} \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2)\right] \cdot \text{N}(\theta_\phi \mid 0, \gamma_\phi^2) \\
p(\theta_\alpha \mid \cdots) &\propto \left[\prod_{g=1}^{G} \pi_\alpha^{1-I(\alpha_g)}[(1-\pi_\alpha)\text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)}\right] \cdot \text{N}(\theta_\alpha \mid 0, \gamma_\alpha^2) \\
p(\theta_\delta \mid \cdots) &\propto \left[\prod_{g=1}^{G} \pi_\delta^{1-I(\delta_g)}[(1-\pi_\delta)\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)}\right] \cdot \text{N}(\theta_\delta \mid 0, \gamma_\delta^2) \\
p(\sigma_\phi \mid \cdots) &\propto \left[\prod_{g=1}^{G} \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2)\right] \cdot \text{U}(\sigma_\phi \mid 0, \sigma_{\phi 0}) \\
p(\sigma_\alpha \mid \cdots) &\propto \left[\prod_{g=1}^{G} \pi_\alpha^{1-I(\alpha_g)}[(1-\pi_\alpha)\text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)}\right] \cdot \text{U}(\sigma_\alpha \mid 0, \sigma_{\alpha 0}) \\
p(\sigma_\delta \mid \cdots) &\propto \left[\prod_{g=1}^{G} \pi_\delta^{1-I(\delta_g)}[(1-\pi_\delta)\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)}\right] \cdot \text{U}(\sigma_\delta \mid 0, \sigma_{\delta 0}) \\
p(\pi_\alpha \mid \cdots) &\propto \left[\prod_{g=1}^{G} \pi_\alpha^{1-I(\alpha_g)}[(1-\pi_\alpha)\text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)}\right] \cdot \text{Beta}(\pi_\alpha \mid a_\alpha, b_\alpha) \\
p(\pi_\delta \mid \cdots) &\propto \left[\prod_{g=1}^{G} \pi_\delta^{1-I(\delta_g)}[(1-\pi_\delta)\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)}\right] \cdot \text{Beta}(\pi_\delta \mid a_\delta, b_\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 $p(c_n \mid \cdots)$: Metropolis

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

A.3 $p(\varepsilon_{g,n} \mid \cdots)$ Metropolis

$$
\begin{aligned}
p(\varepsilon_{g,n} \mid \cdots) &= \text{Poisson}(y_{g,n} \mid \lambda_{g,n}) \cdot \text{N}(\varepsilon_{g,n} \mid 0, \eta_g^2) \\
&\propto \lambda_{g,n}^{y_{g,n}} \exp(-\lambda_{g,n}) \exp\left(-\frac{\varepsilon_{g,n}^2}{2\eta_g^2}\right) \\
&= \exp\left(y_{g,n} \log \lambda_{g,n} - \lambda_{g,n} - \frac{\varepsilon_{g,n}^2}{2\eta_g^2}\right) \\
&= \exp\left(y_{g,n}(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)) - \frac{\varepsilon_{g,n}^2}{2\eta_g^2}\right) \\
&= \exp\left(y_{g,n}\varepsilon_{g,n} - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)) - \frac{\varepsilon_{g,n}^2}{2\eta_g^2}\right)
\end{aligned}
$$

A.4 $p\left(\frac{1}{\sigma_c^2} \mid \cdots\right)$ Truncated Gamma

$$
\begin{aligned}
p(\sigma_c^2 \mid \cdots) &= p(\sigma_c \mid \cdots)\frac{1}{2}(\sigma_c^2)^{-1/2} \qquad \text{(transformation in Section ??)} \\
&\propto \left[\prod_{n=1}^{N} \mathrm{N}(c_n \mid 0, \sigma_c^2)\right] \cdot \mathrm{U}(\sigma_c \mid 0, \sigma_{c0})\frac{1}{2}(\sigma_c^2)^{-1/2} \\
&\propto \prod_{n=1}^{N}\left[\frac{1}{\sqrt{\sigma_c^2}}\exp\left(-\frac{c_n^2}{2\sigma_c^2}\right)\right] \cdot \mathrm{I}(0 < \sigma_c < \sigma_{c0})(\sigma_c^2)^{-1/2} \\
&= (\sigma_c^2)^{-N/2}\exp\left(-\frac{1}{\sigma_c^2}\frac{1}{2}\sum_{n=1}^{N} c_n^2\right) \cdot \mathrm{I}(0 < \sigma_c < \sigma_{c0})(\sigma_c^2)^{-1/2} \\
&= (\sigma_c^2)^{-(N/2-1/2+1)}\exp\left(-\frac{1}{\sigma_c^2}\frac{1}{2}\sum_{n=1}^{N} c_n^2\right) \cdot \mathrm{I}(0 < \sigma_c < \sigma_{c0})
\end{aligned}
$$

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

$$
p\left(\frac{1}{\sigma_c^2} \mid \cdots\right) = \mathrm{Gamma}\left(\frac{1}{\sigma_c^2} \mid \text{shape} = \frac{N-1}{2},\ \text{rate} = \frac{1}{2}\sum_{n=1}^{N} c_n^2\right) I\left(\frac{1}{\sigma_c^2} > \frac{1}{\sigma_{c0}^2}\right)
$$

A.5 $p\left(\frac{1}{\eta_g^2} \mid \cdots\right)$ Gamma

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

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

$$
p\left(\frac{1}{\eta_g^2} \mid \cdots\right) = \mathrm{Gamma}\left(\frac{1}{\eta_g^2} \mid \text{shape} = \frac{N+d}{2},\ \text{rate} = \frac{1}{2}\left(d \cdot \tau^2 + \sum_{n=1}^{N}\varepsilon_{g,n}^2\right)\right)
$$

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

$$
\begin{aligned}
p(d \mid \cdots) &= \left[\prod_{g=1}^{G} \text{Inv-Gamma}\left(\eta_g^2 \mid \text{shape} = \frac{d}{2}, \text{rate} = \frac{d \cdot \tau^2}{2}\right)\right] \cdot \text{U}(d \mid 0, d_0) \\
&\propto \prod_{g=1}^{G} \left[\Gamma\left(d/2\right)^{-1} \left(\frac{d \cdot \tau^2}{2}\right)^{d/2} \left(\eta_g^2\right)^{-(d/2+1)} \exp\left(-\frac{1}{\eta_g^2}\frac{d \cdot \tau^2}{2}\right)\right] I(2 < d < d_0) \\
&\propto \Gamma\left(d/2\right)^{-G} \left(\frac{d \cdot \tau^2}{2}\right)^{Gd/2} \left(\prod_{g=1}^{G} \eta_g^2\right)^{-(d/2+1)} \exp\left(-\frac{d \cdot \tau^2}{2} \sum_{g=1}^{G} \frac{1}{\eta_g^2}\right) I(0 < d < d_0)
\end{aligned}
$$

A.7 $p(\tau^2 \mid \cdots)$: Gamma

$$
\begin{aligned}
p(\tau^2 \mid \cdots) &= \left[\prod_{g=1}^{G} \text{Inv-Gamma}\left(\eta_g^2 \mid \text{shape} = \frac{d}{2}, \text{rate} = \frac{d \cdot \tau^2}{2}\right)\right] \cdot \text{Gamma}(\tau^2 \mid \text{shape} = a_\tau, \text{rate} = b_\tau) \\
&\propto \left[\Gamma\left(d/2\right)^{-G} \left(\frac{d \cdot \tau^2}{2}\right)^{Gd/2} \left(\prod_{g=1}^{G} \eta_g^2\right)^{-(d/2+1)} \exp\left(-\frac{d \cdot \tau^2}{2} \sum_{g=1}^{G} \frac{1}{\eta_g^2}\right)\right] \cdot (\tau^2)^{a_\tau - 1} \exp\left(-b_\tau \tau^2\right) \\
&\propto \left[\left(\tau^2\right)^{Gd/2} \exp\left(-\tau^2 \cdot \frac{d}{2} \sum_{g=1}^{G} \frac{1}{\eta_g^2}\right)\right] \cdot (\tau^2)^{a_\tau - 1} \exp\left(-b_\tau \tau^2\right) \\
&= (\tau^2)^{Gd/2 + a_\tau - 1} \exp\left(-\tau^2 \left(b_\tau + \frac{d}{2} \sum_{g=1}^{G} \frac{1}{\eta_g^2}\right)\right)
\end{aligned}
$$

Hence:

$$
p(\tau^2 \mid \cdots) = \text{Gamma}\left(\tau^2 \mid \text{shape} = a_\tau + \frac{Gd}{2},\ \text{rate} = b_\tau + \frac{d}{2} \sum_{g=1}^{G} \frac{1}{\eta_g^2}\right)
$$

A.8 $p(\phi_g \mid \cdots)$: Metropolis

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

A.9 $p(\alpha_g \mid \cdots)$: Metropolis

$$
\begin{aligned}
p(\alpha_g \mid \cdots) &= \left[\prod_{k(n) \neq 2} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right] \pi_\alpha^{1 - I(\alpha_g)} [(1 - \pi_\alpha)\text{N}(\theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)} \\
&\propto \left[\prod_{k(n) \neq 2} \lambda_{g,n}^{y_{g,n}} \exp(-\lambda_{g,n})\right] \exp\left(-\frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2}\right)^{I(\alpha_g)} \pi_\alpha^{1 - I(\alpha_g)} (1 - \pi_\alpha)^{I(\alpha_g)} \\
&= \exp\left(\sum_{k(n) \neq 2} [y_{g,n} \log \lambda_{g,n} - \lambda_{g,n}] - I(\alpha_g) \frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} + (1 - I(\alpha_g)) \log \pi_\alpha + I(\alpha_g) \log(1 - \pi_\alpha)\right) \\
&= \exp\left(\sum_{k(n) \neq 2} [y_{g,n} \log \lambda_{g,n} - \lambda_{g,n}] - I(\alpha_g) \left(\frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} + \log(1 - \pi_\alpha)\right) + (1 - I(\alpha_g)) \log \pi_\alpha\right) \\
&= \exp\Bigg(\sum_{k(n) \neq 2} [y_{g,n}(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))] \\
&\qquad - I(\alpha_g) \left(\frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} + \log(1 - \pi_\alpha)\right) + (1 - I(\alpha_g)) \log \pi_\alpha\Bigg) \\
&\propto \exp\Bigg(\sum_{k(n) \neq 2} [y_{g,n} \cdot \mu(n, \phi_g, \alpha_g, \delta_g) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))] \\
&\qquad - I(\alpha_g) \left(\frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} + \log(1 - \pi_\alpha)\right) + (1 - I(\alpha_g)) \log \pi_\alpha\Bigg)
\end{aligned}
$$

A.10 $p(\delta_g \mid \cdots)$: Metropolis

$$
\begin{aligned}
p(\delta_g \mid \cdots) &= \left[\prod_{k(n)\neq 2} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right] \pi_\delta^{1-I(\delta_g)} [(1-\pi_\delta)\text{N}(\theta_\delta, \sigma_\delta^2)]^{I(\delta_g)} \\
&\propto \left[\prod_{k(n)\neq 2} \lambda_{g,n}^{y_{g,n}} \exp(-\lambda_{g,n})\right] \exp\left(-\frac{(\delta_g - \theta_\delta)^2}{2\sigma_\delta^2}\right)^{I(\delta_g)} \pi_\delta^{1-I(\delta_g)} (1-\pi_\delta)^{I(\delta_g)} \\
&= \exp\left(\sum_{k(n)\neq 2} [y_{g,n}\log\lambda_{g,n} - \lambda_{g,n}] - I(\delta_g)\frac{(\delta_g-\theta_\delta)^2}{2\sigma_\delta^2} + (1 - I(\delta_g))\log\pi_\delta + I(\delta_g)\log(1-\pi_\delta)\right) \\
&= \exp\left(\sum_{k(n)\neq 2} [y_{g,n}\log\lambda_{g,n} - \lambda_{g,n}] - I(\delta_g)\left(\frac{(\delta_g-\theta_\delta)^2}{2\sigma_\delta^2} + \log(1-\pi_\delta)\right) + (1 - I(\delta_g))\log\pi_\delta\right) \\
&= \exp\left(\sum_{k(n)\neq 2} [y_{g,n}(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))]\right. \\
&\qquad \left. - I(\delta_g)\left(\frac{(\delta_g-\theta_\delta)^2}{2\sigma_\delta^2} + \log(1-\pi_\delta)\right) + (1 - I(\delta_g))\log\pi_\delta\right) \\
&\propto \exp\left(\sum_{k(n)\neq 2} [y_{g,n}\cdot\mu(n, \phi_g, \alpha_g, \delta_g) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))]\right. \\
&\qquad \left. - I(\delta_g)\left(\frac{(\delta_g-\theta_\delta)^2}{2\sigma_\delta^2} + \log(1-\pi_\delta)\right) + (1 - I(\delta_g))\log\pi_\delta\right)
\end{aligned}
$$

A.11 $p(\phi_g, \alpha_g, \delta_g \mid \cdots)$: Metropolis

$$
\begin{aligned}
p(\phi_g, \alpha_g, \delta_g \mid \cdots) &\propto \left[\prod_{n=1}^{N} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right] \cdot \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2) \\
&\quad \times \pi_\alpha^{1-I(\alpha_g)}[(1-\pi_\alpha)\text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)} \times \pi_\delta^{1-I(\delta_g)}[(1-\pi_\delta)\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)} \\
&\propto \exp\left(\sum_{n=1}^{N} [y_{g,n}\mu(n, \phi_g, \alpha_g, \delta_g) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right) \\
&\quad \times \pi_\alpha^{1-I(\alpha_g)}[(1-\pi_\alpha)\text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)} \times \pi_\delta^{1-I(\delta_g)}[(1-\pi_\delta)\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)} \\
&\propto \exp\left(\sum_{n=1}^{N} [y_{g,n}\mu(n, \phi_g, \alpha_g, \delta_g) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2}\right. \\
&\qquad - I(\alpha_g)\left(\frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2} + \log(1-\pi_\alpha)\right) + (1 - I(\alpha_g))\log \pi_\alpha \\
&\qquad \left. - I(\delta_g)\left(\frac{(\delta_g - \theta_\delta)^2}{2\sigma_\delta^2} + \log(1-\pi_\delta)\right) + (1 - I(\delta_g))\log \pi_\delta\right)
\end{aligned}
$$

A.12 $p(\theta_\phi \mid \cdots)$: Normal

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

Hence:

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

A.13 $p\left(\theta_\alpha \mid \cdots\right)$: Normal

$$\begin{aligned}
p(\theta_\alpha \mid \cdots) &= \left[\prod_{g=1}^{G} \pi_\alpha^{1-I(\alpha_g)} [(1-\pi_\alpha)\mathrm{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)}\right] \cdot \mathrm{N}(\theta_\alpha \mid 0, \gamma_\alpha^2) \\
&\propto \left[\prod_{\alpha_g \neq 0} \mathrm{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]\right] \cdot \mathrm{N}(\theta_\alpha \mid 0, \gamma_\alpha^2)
\end{aligned}$$

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

$$p(\theta_\alpha \mid \cdots) = N\left(\theta_\alpha \;\middle|\; \frac{\gamma_\alpha^2 \sum_{\alpha_g \neq 0} \alpha_g}{G_\alpha \gamma_\alpha^2 + \sigma_\alpha^2}, \quad \frac{\gamma_\alpha^2 \sigma_\alpha^2}{G_\alpha \gamma_\alpha^2 + \sigma_\alpha^2}\right)$$

A.14 $p(\theta_\delta \mid \cdots)$: Normal

$$\begin{aligned} p(\theta_\delta \mid \cdots) &= \left[\prod_{g=1}^{G} \pi_\delta^{1-I(\delta_g)} [(1-\pi_\delta)\mathrm{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)}\right] \cdot \mathrm{N}(\theta_\delta \mid 0, \gamma_\delta^2) \\ &\propto \left[\prod_{\delta_g \neq 0} \mathrm{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]\right] \cdot \mathrm{N}(\theta_\delta \mid 0, \gamma_\delta^2) \end{aligned}$$

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

$$p(\theta_\delta \mid \cdots) = N\left(\frac{\gamma_\delta^2 \sum_{\delta_g \neq 0} \delta_g}{G_\delta \gamma_\delta^2 + \sigma_\delta^2}, \; \frac{\gamma_\delta^2 \sigma_\delta^2}{G_\delta \gamma_\delta^2 + \sigma_\delta^2}\right)$$

where G_δ is the number of genes for which $\delta_g \neq 0$.

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

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

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

$$p\left(\frac{1}{\sigma_\phi^2} \;\middle|\; \cdots\right) = \text{Gamma}\left(\text{shape} = \frac{G-1}{2}, \; \text{rate} = \frac{1}{2}\sum_{g=1}^{G} (\phi_g - \theta_\phi)^2\right) \mathrm{I}\left(\frac{1}{\sigma_\phi^2} > \frac{1}{\sigma_{\phi 0}^2}\right)$$

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

$$\begin{aligned}
p(\sigma_\alpha^2 \mid \cdots) &= p(\sigma_\alpha \mid \cdots)\frac{1}{2}(\sigma_\alpha^2)^{-1/2} \qquad \text{(transformation in Section \textbf{??})} \\
&\propto \left[\prod_{g=1}^{G} \pi_\alpha^{1-I(\alpha_g)}[(1-\pi_\alpha)\mathrm{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)}\right] \cdot \mathrm{U}(\sigma_\alpha \mid 0, \sigma_{\alpha 0})(\sigma_\alpha^2)^{-1/2} \\
&\propto \prod_{\alpha_g \neq 0} N(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2) \cdot I(0 < \sigma_\alpha^2 < \sigma_{\alpha 0}^2)(\sigma_\alpha^2)^{-1/2} \\
&\propto \prod_{\alpha_g \neq 0} (\sigma_\alpha^2)^{-1/2} \exp\left(-\frac{(\alpha_g - \theta_\alpha)^2}{2\sigma_\alpha^2}\right) \cdot I(0 < \sigma_\alpha^2 < \sigma_{\alpha 0}^2)(\sigma_\alpha^2)^{-1/2} \\
&= (\sigma_\alpha^2)^{-G_\alpha/2} \exp\left(-\frac{1}{\theta_\alpha^2}\frac{1}{2}\sum_{\alpha_g \neq 0}(\alpha_g - \theta_\alpha)^2\right) \cdot I(0 < \sigma_\alpha^2 < \sigma_{\alpha 0}^2)(\sigma_\alpha^2)^{-1/2} \\
&= (\sigma_\alpha^2)^{-(G_\alpha/2 - 1/2 + 1)} \exp\left(-\frac{1}{\theta_\alpha^2}\frac{1}{2}\sum_{\alpha_g \neq 0}(\alpha_g - \theta_\alpha)^2\right) \cdot I(0 < \sigma_\alpha^2 < \sigma_{\alpha 0}^2)
\end{aligned}$$

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

$$p\left(\frac{1}{\sigma_\alpha^2} \mid \cdots\right) = \mathrm{Gamma}\left(\frac{1}{\sigma_\alpha^2} \mid \text{shape} = \frac{G_\alpha - 1}{2},\ \text{rate} = \frac{1}{2}\sum_{\alpha_g \neq 0}(\alpha_g - \theta_\alpha)^2\right) \mathrm{I}\left(\frac{1}{\sigma_\alpha^2} > \frac{1}{\sigma_{\alpha 0}^2}\right)$$

A.17 $p\left(\frac{1}{\sigma_\delta^2} \mid \cdots\right)$: Truncated Gamma

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

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

$$
p\left(\frac{1}{\sigma_\delta^2} \mid \cdots\right) = \text{Gamma}\left(\frac{1}{\sigma_\delta^2} \mid \text{shape} = \frac{G_\delta - 1}{2},\ \text{rate} = \frac{1}{2}\sum_{\delta_g \neq 0}(\delta_g - \theta_\delta)^2\right) \mathrm{I}\left(\frac{1}{\sigma_\delta^2} > \frac{1}{\sigma_{\delta 0}^2}\right)
$$

A.18 $p(\pi_\alpha \mid \cdots)$: Beta

$$
\begin{aligned}
p(\pi_\alpha \mid \cdots) &= \left[\prod_{g=1}^{G} \pi_\alpha^{1-I(\alpha_g)}[(1-\pi_\alpha)\mathrm{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)}\right] \cdot \mathrm{Beta}(\pi_\alpha \mid a_\alpha, b_\alpha) \\
&\propto [\pi_\alpha^{G-G_\alpha}(1-\pi_\alpha)^{G_\alpha}]\pi_\alpha^{a_\tau - 1}(1-\pi_\alpha)^{b_\tau - 1} \\
&= \pi_\alpha^{G - G_\alpha + a_\tau - 1}(1-\pi_\alpha)^{G_\alpha + b_\tau - 1}
\end{aligned}
$$

Hence:

$$
p(\pi_\alpha \mid \cdots) = \mathrm{Beta}(G - G_\alpha + a_\tau,\ G_\alpha + b_\tau)
$$

A.19 $p(\pi_\delta \mid \cdots)$: Beta

$$
\begin{aligned}
p(\pi_\delta \mid \cdots) &= \left[\prod_{g=1}^{G} \pi_\delta^{1-I(\delta_g)}[(1-\pi_\delta)\mathrm{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)}\right] \cdot \mathrm{Beta}(\pi_\delta \mid a_\delta, b_\delta) \\
&\propto [\pi_\delta^{G-G_\delta}(1-\pi_\delta)^{G_\delta}]\pi_\delta^{a_\tau - 1}(1-\pi_\delta)^{b_\tau - 1} \\
&= \pi_\delta^{G-G_\delta+a_\tau-1}(1-\pi_\delta)^{G_\delta+b_\tau-1}
\end{aligned}
$$

where G_δ is the number of genes for which $\delta_g \neq 0$. Hence:

$$
p(\pi_\delta \mid \cdots) = \mathrm{Beta}(G - G_\delta + a_\tau, \ G_\delta + b_\tau)
$$

B Derivations of Metropolis proposals for point mass mixtures

B.1 α_g

I choose a proposal for α_g with the form,

$$
q(\alpha_g \mid \theta'_\alpha, \sigma'_\alpha, \pi'_\alpha) = I(\alpha_g = 0)\pi'_\alpha + I(\alpha_g \neq 0)(1-\pi'_\alpha)N(\alpha_g \mid \theta'_\alpha, (\sigma'_\alpha)^2),
$$

which resembles the prior for α_g except that the parameters are updated to reflect the data, $\underline{y} = (y_{1,1}, \ldots, y_{G,N})$ (except for π'_α, for which we simply use π_α). To find θ'_α and σ'_α, we pretend that α_g has a $N(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)$ conditional likelihood, θ_α has a $N(\theta_\alpha \mid 0, \gamma_\alpha^2)$ prior, and σ_α is fixed. From the rule on pages 46 and 47 of Gelman's book, the conditional posterior distribution of θ_α is

$$
N\left(\theta_\alpha \,\middle|\, \frac{\sigma_\alpha^{-2}\alpha_g}{\gamma_\alpha^{-2} + \sigma_\alpha^{-2}} \ , \ (\gamma_\alpha^{-2} + \sigma_\alpha^{-2})^{-1}\right)
$$

Hence, we let

$$
\begin{aligned}
\theta'_\alpha &= \frac{\sigma_\alpha^{-2}\alpha_g}{\gamma_\alpha^{-2} + \sigma_\alpha^{-2}} \\
(\sigma_\alpha^2)' &= \mathrm{Var}(\alpha_g) \\
&= \mathrm{Var}(E(\alpha_g \mid \theta_\alpha)) + E(\mathrm{Var}(\alpha_g \mid \theta_\alpha)) \\
&= \underbrace{\mathrm{Var}(\theta_\alpha)}_{\text{Use prior variance.}} + E(\sigma_\alpha^2) \\
&= \gamma_\alpha^2 + \sigma_\alpha^2
\end{aligned}
$$

B.2 δ_g

The proposal for δ_g is analogous to that of α_g:

$$q(\delta_g \mid \theta'_\delta, \sigma'_\delta, \pi'_\delta) = I(\delta_g = 0)\pi'_\delta + I(\delta_g \neq 0)(1 - \pi'_\delta)N(\delta_g \mid \theta'_\delta, (\sigma'_\delta)^2),$$

where:

$$\begin{aligned} \theta'_\delta &= \frac{\sigma_\delta^{-2}\delta_g}{\gamma_\delta^{-2} + \sigma_\delta^{-2}} \\ (\sigma'_\delta)^2 &= \gamma_\delta^2 + \sigma_\delta^2 \\ \pi'_\delta &= \pi_\delta \end{aligned}$$

C Using normal approximations to improve Metropolis proposals

We can turn the Metropolis steps into Metropolis-Hastings steps using proposal distributions that are normal approximations of the full conditionals. Let $g(\theta) = \log(p(\theta|\cdots))$ be the full conditional distribution of some parameter, θ. Let $\widehat{\theta}$ be some point estimate of θ (for example, the MLE). Then,

$$\begin{aligned} g(\theta) &\approx g(\widehat{\theta}) + g'(\widehat{\theta})(\theta - \widehat{\theta}) + \frac{g''(\widehat{\theta})}{2}(\theta - \widehat{\theta})^2 \\ &= g(\widehat{\theta}) + g'(\widehat{\theta})\theta - g'(\widehat{\theta})\widehat{\theta} + \frac{g''(\widehat{\theta})}{2}\theta^2 - g''(\widehat{\theta})\widehat{\theta}\theta + \frac{g''(\widehat{\theta})}{2}\widehat{\theta}^2 \\ &= \underbrace{\left[g(\widehat{\theta}) - g'(\widehat{\theta})\widehat{\theta} + \frac{g''(\widehat{\theta})}{2}\widehat{\theta}^2\right]}_{A} + \underbrace{\left[g'(\widehat{\theta}) - g''(\widehat{\theta})\widehat{\theta}\right]}_{B}\theta + \underbrace{\frac{g''(\widehat{\theta})}{2}}_{C}\theta^2 \\ &= A\left(\theta + \frac{B}{2A}\right)^2 + C - \frac{B^2}{4A} \end{aligned}$$

That means

$$
\begin{aligned}
\exp(g(\theta)) &\approx \exp\left[A\left(\theta + \frac{B}{2A}\right)^2 + C - \frac{B^2}{4A}\right] \\
&\propto \exp\left[A\left(\theta + \frac{B}{2A}\right)^2\right] \\
&= \exp\left[\frac{\left(\theta + \frac{B}{2A}\right)^2}{2(2A)^{-1}}\right] \\
&\propto N\left(-\frac{B}{2A},\ \frac{1}{2A}\right)
\end{aligned}
$$

Now,

$$
\begin{aligned}
-\frac{B}{2A} &= \frac{-g'(\widehat{\theta}) + g''(\widehat{\theta})\widehat{\theta}}{2g(\widehat{\theta}) - 2g'(\widehat{\theta})\widehat{\theta} + g''(\widehat{\theta})\widehat{\theta}^2} \\
\frac{1}{2A} &= \frac{1}{2g(\widehat{\theta}) - 2g'(\widehat{\theta})\widehat{\theta} + g''(\widehat{\theta})\widehat{\theta}^2}
\end{aligned}
$$

so that

$$
\exp(g(\theta)) \approx N\left(\frac{-g'(\widehat{\theta}) + g''(\widehat{\theta})\widehat{\theta}}{2g(\widehat{\theta}) - 2g'(\widehat{\theta})\widehat{\theta} + g''(\widehat{\theta})\widehat{\theta}^2},\ \frac{1}{2g(\widehat{\theta}) - 2g'(\widehat{\theta})\widehat{\theta} + g''(\widehat{\theta})\widehat{\theta}^2}\right)
$$

Now, let $q(\theta)$ be the above normal distribution and $\theta^{(i)}$ be the current value of θ at iteration i of the chain. To get $\theta^{(i+1)}$, we first sample a proposal θ^* from q. Then, we compute the probability,

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

We set $\theta^{(i+1)} = \theta^*$ with probability p and $\theta^{(i+1)} = \theta^{(i)}$ with probability $1 - p$.