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

Will Landau

Department of Statistics
Iowa State University

July 17, 2013

Contents

1	Introduction	1
2	The Model	2
3	Full Conditional Distributions	3
4	The Gibbs Sampler	4
5	Diagnostics	6
5.1	Gelman Factors	6
5.2	Deviance Information Criterion	7
A	Derivations of the Full Conditionals	8
A.1	Transformations of Standard Deviations	10
A.2	$p(c_n \mid \dots)$: Metropolis	11
A.3	$p(\varepsilon_{g,n} \mid \dots)$ Metropolis	11
A.4	$p\left(\frac{1}{\sigma_c^2} \mid \dots\right)$ Truncated Gamma	12
A.5	$p\left(\frac{1}{\eta_g^2} \mid \dots\right)$ Gamma	12
A.6	$p(d \mid \dots)$: Metropolis	13
A.7	$p(\tau^2 \mid \dots)$: Gamma	13
A.8	$p(\phi_g \mid \dots)$: Metropolis	14
A.9	$p(\alpha_g \mid \dots)$: Metropolis	14
A.10	$p(\delta_g \mid \dots)$: Metropolis	15
A.11	$p(\phi_g, \alpha_g, \delta_g \mid \dots)$: Metropolis	16
A.12	$p(\theta_\phi \mid \dots)$: Normal	17
A.13	$p(\theta_\alpha \mid \dots)$: Normal	17
A.14	$p(\theta_\delta \mid \dots)$: Normal	18
A.15	$p\left(\frac{1}{\sigma_\phi^2} \mid \dots\right)$: Truncated Gamma	18
A.16	$p\left(\frac{1}{\sigma_\alpha^2} \mid \dots\right)$: Truncated Gamma	19
A.17	$p\left(\frac{1}{\sigma_\delta^2} \mid \dots\right)$: Truncated Gamma	20
A.18	$p(\pi_\alpha \mid \dots)$: Beta	20
A.19	$p(\pi_\delta \mid \dots)$: Beta	21
B	Derivations of Metropolis proposals for point mass mixtures	21
B.1	α_g	21
B.2	δ_g	22
	Bibliography	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 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 ($g = 1, \dots, G$) in sample n ($n = 1, \dots, N$). The samples come from one of three groups: group 1, the first parent, group 2, the 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:

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

2 The Model

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

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:

$$\begin{aligned}
p(c_n \mid \cdots) &\propto \exp \left(c_n G \bar{y}_{\cdot 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} \mid \cdots\right) &= \text{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) \\
p\left(\frac{1}{\eta_g^2} \mid \cdots\right) &= \text{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) \\
p(d \mid \cdots) &\propto \Gamma(d/2)^{-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 \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) \\
p(\phi_g \mid \cdots) &\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) \\
p(\alpha_g \mid \cdots) &\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. \\
&\quad \left. - 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) \\
p(\delta_g \mid \cdots) &\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. \\
&\quad \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}$$

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

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

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

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

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

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

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

$$p(\pi_\alpha \mid \dots) = \text{Beta}(\pi_\alpha \mid G - G_\alpha + \alpha_\tau, G_\alpha + b_\tau)$$

$$p(\pi_\delta \mid \dots) = \text{Beta}(\pi_\delta \mid G - G_\delta + \delta_\tau, G_\delta + b_\tau)$$

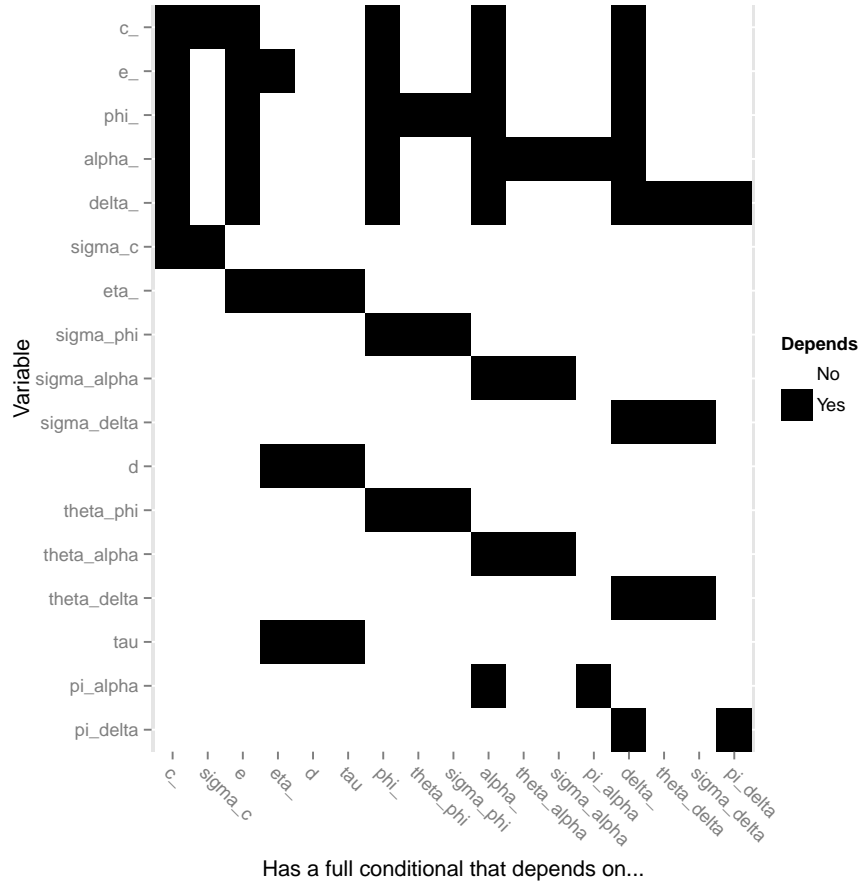
4 The Gibbs Sampler

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, \dots, c_N
- $\varepsilon_{1,1}, \varepsilon_{1,2}, \dots, \varepsilon_{1,N}, \varepsilon_{2,N}, \dots, \varepsilon_{G,N}$
- $\eta_1^2, \dots, \eta_G^2$
- ϕ_1, \dots, ϕ_G
- $\alpha_1, \dots, \alpha_G$

- $\delta_1, \dots, \delta_G$

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, \dots, c_N
2. $\tau, \pi_\alpha, \pi_\delta$
3. $d, \theta_\phi, \theta_\alpha, \theta_\delta$
4. $\sigma_c, \sigma_\phi, \sigma_\alpha, \sigma_\delta, \eta_1^2, \dots, \eta_G^2$
5. $\varepsilon_{1,1}, \varepsilon_{1,2}, \dots, \varepsilon_{1,N}, \varepsilon_{2,N}, \dots, \varepsilon_{G,N}$
6. ϕ_1, \dots, ϕ_G
7. $\alpha_1, \dots, \alpha_G$

8. $\delta_1, \dots, \delta_G$

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 $p(\phi_g, \alpha_g, \delta_g \dots)$.

5 Diagnostics

5.1 Gelman Factors

The potential scale reduction factor introduced in the textbook by Gelman Gelman et al. [2004] 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, \hat{R} , compares the within-chain variance, W , to the between-chain variance, B . Suppose there are J chains, each with I iterations. Then,

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

$\hat{R} \rightarrow 1$ 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

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

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

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

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

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

where $\hat{\boldsymbol{\psi}}$ 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 $\bar{\boldsymbol{\psi}}$ is the collection of within-chain parameter means, then we can estimate DIC by

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

All that remains is to find $\log p(\mathbf{y} \mid \boldsymbol{\psi})$ 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

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

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

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

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

where

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

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 2, we get:

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

$$\begin{aligned}
p(\delta_g \mid \dots) &\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 \dots) &\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 \dots) &\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 \dots) &\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 \dots) &\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 \dots) &\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 \dots) &\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 \dots) &\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 \dots) &\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 \dots) &\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 $p(\sigma \mid \dots)$ be its full conditional distribution. Then, by a transformation of variables,

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

I use this transformation several times in the next sections.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence:

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

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

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

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

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

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

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

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

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

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

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

Hence:

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

A.13 $p(\theta_\alpha \mid \dots)$: Normal

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence:

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

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

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

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

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

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, $\underline{y} = (y_{1,1}, \dots, y_{G,N})$ (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

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

Hence, we let

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

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:

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

Bibliography

A. Gelman, J. B. Carlin, H. S. Stern, and D. S. Rubin. *Bayesian Data Analysis*.
Chapman & Hall/CRC, 2 edition, 2004.