A Fully Bayesian Model for RNA-seq Data

Will Landau

Department of Statistics
Iowa State University

May 23, 2013

Contents

1 The Model

Let $y_{g,n}$ be the expression level of gene g ($g = 1, \ldots, G$) in library n ($n = 1, \ldots, N$). Let $\mu(n, \phi_g, \alpha_g, \delta_g)$ be the function given by:

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

Then:

$$\begin{aligned}
y_{g,n} &\sim \text{Poisson}(\log(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))) \\
c_n &\sim \text{N}(0, \sigma_c^2) \\
\sigma_c &\sim \text{U}(0, 1) \\
\varepsilon_{g,n} &\sim \text{N}(0, \sigma_g^2) \\
\sigma_g &\sim \text{Gamma}\left(\frac{d \cdot \sigma_0^2}{2}, \frac{d}{2}\right) \\
d &\sim \text{U}(0, 1000) \\
\sigma_0^2 &\sim \text{Exp}(1) \\
\phi_g &\sim \text{N}(\theta_\phi, \sigma_\phi^2) \\
\theta_\phi &\sim \text{N}(0, 1000) \\
\sigma_\phi &\sim \text{U}(0, 1000) \\
\alpha_g &\sim \pi_\alpha \text{I}(\alpha_g = 0) + (1 - \pi_\alpha)\text{I}(\alpha_g \neq 0)\text{N}(\theta_\alpha, \sigma_\alpha^2) \\
\theta_\alpha &\sim \text{N}(0, 1000) \\
\sigma_\alpha &\sim \text{U}(0, 1000) \\
\pi_\alpha &\sim \text{U}(0, 1) \\
\delta_g &\sim \pi_\delta \text{I}(\delta_g = 0) + (1 - \pi_\delta)\text{I}(\delta_g \neq 0)\text{N}(\theta_\delta, \sigma_\delta^2) \\
\theta_\delta &\sim \text{N}(0, 1000) \\
\sigma_\delta &\sim \text{U}(0, 1000) \\
\pi_\delta &\sim \text{U}(0, 1)
\end{aligned}$$

where:

- Independence is implied unless otherwise specified.
- The parameters to the left of the "$\sim$" are implicitly conditioned on the parameters to the right.

This gives us the full conditional posteriors up to a normalizing constant. Below, $k(n)$ is the treatment group of library n.

$$
\begin{aligned}
p(c_n \mid \cdots) &\propto \left[\prod_{g=1}^{G} \text{Poisson}(y_{g,n} \mid \log(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right] \text{N}(c_n \mid 0, \sigma_c^2)^G \\
p(\sigma_c \mid \cdots) &\propto \left[\prod_{n=1}^{N} \text{N}(c_n \mid 0, \sigma_c^2)\right]^G \text{U}(\sigma_c \mid 0, 1)^{G \cdot N} \\
p(\varepsilon_{g,n} \mid \cdots) &\propto \text{Poisson}(y_{g,n} \mid \log(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))) \cdot \text{N}(\varepsilon_{g,n} \mid 0, \sigma_g^2) \\
p(\sigma_g \mid \cdots) &\propto \left[\prod_{n=1}^{N} \text{N}(\varepsilon_{g,n} \mid 0, \sigma_g^2)\right] \text{Gamma}\left(\sigma_g \mid \frac{d \cdot \sigma_0^2}{2}, \frac{d}{2}\right)^N \\
p(d \mid \cdots) &\propto \left[\prod_{g=1}^{G} \text{Gamma}\left(\sigma_g \mid \frac{d \cdot \sigma_0^2}{2}, \frac{d}{2}\right)\right]^N \text{U}(d \mid 0, 1000)^{G \cdot N} \\
p(\sigma_0^2 \mid \cdots) &\propto \left[\prod_{g=1}^{G} \text{Gamma}\left(\sigma_g \mid \frac{d \cdot \sigma_0^2}{2}, \frac{d}{2}\right)\right]^N \text{Exp}(\sigma^2 \mid 1)^{G \cdot N} \\
p(\phi_g \mid \cdots) &\propto \left[\prod_{n=1}^{N} \text{Poisson}(y_{g,n} \mid \log(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right] \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2)^N \\
p(\theta_\phi \mid \cdots) &\propto \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2)^{G \cdot N} \cdot \text{N}(\theta_\phi \mid 0, 1000)^{G \cdot N} \\
p(\sigma_\phi \mid \cdots) &\propto \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2)^{G \cdot N} \cdot \text{U}(\sigma_\phi \mid 0, 1000)^{G \cdot N} \\
p(\alpha_g \mid \cdots) &\propto \left[\prod_{k(n) \neq 2} \text{Poisson}(y_{g,n} \mid \log(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))\right. \\
&\qquad \left. \times \left[\pi_\alpha \text{I}(\alpha_g = 0) + (1 - \pi_\alpha)\text{I}(\alpha_g \neq 0)\text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)\right]\right] \\
p(\theta_\alpha \mid \cdots) &\propto \left[\prod_{k(n) \neq 2} \left[\text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2) \cdot \text{N}(\theta_\alpha \mid 0, 1000)\right]\right]^G \\
p(\sigma_\alpha \mid \cdots) &\propto \left[\prod_{k(n) \neq 2} \left[\text{N}(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2) \cdot \text{U}(\sigma_\alpha \mid 0, 1000)\right]\right]^G \\
p(\delta_g \mid \cdots) &\propto \left[\prod_{k(n) = 2} \text{Poisson}(y_{g,n} \mid \log(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \delta_g, \delta_g)))\right. \\
&\qquad \left. \times \left[\pi_\delta \text{I}(\delta_g = 0) + (1 - \pi_\delta)\text{I}(\delta_g = 0)\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2)\right]\right] \\
p(\theta_\delta \mid \cdots) &\propto \left[\prod_{k(n) = 2} \left[\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2) \cdot \text{N}(\theta_\delta \mid 0, 1000)\right]\right]^G \\
p(\sigma_\delta \mid \cdots) &\propto \left[\prod_{k(n) = 2} \left[\text{N}(\delta_g \mid \theta_\delta, \sigma_\delta^2) \cdot \text{U}(\sigma_\delta \mid 0, 1000)\right]\right]^G
\end{aligned}
$$