Sampling Scheme for CyTOF Model3

Arthur Lui

April 4, 2018

1 Notation

Samples are taken from I subjects, $i = 1, 2, ..., I$. Sample i consists of N_i cells, and for each cell, expression levels of J markers are measured. Let a J-dimensional vector $\boldsymbol{y}_{in} = (y_{in1}, \ldots, y_{inJ})'$, $i = 1, \ldots, I$ and $n = 1, \ldots, N_i$ denote the measurements of expression levels of J markers for cell n in sample i where element $\tilde{y}_{inj} \in \mathbb{R}^+$ represents the raw measurement of an expression level of marker j of cell n in sample i. Let c_{ij} denote the "cutoff" values for sample i, marker j. The cutoff values are obtained as part of the CyTOF analysis. Each marker and sample comes with a cutoff value. Beyond the cutoff, "positive" expression for a marker starts. In practice, some noise is detected from the environment when the samples are measured. Hence, a set of cutoffs are recorded for each marker for each sample as well. But this variation should be small. We consider the logarithm transformation after scaling $\tilde{y}_{inj}$ by c_{ij},

$$y_{inj} = \log\left(\frac{\tilde{y}_{inj}}{c_{ij}}\right) \in \mathbb{R}.$$

For some (i, n, j), $\tilde{y}_{inj}$ is missing. Expression levels are recorded as missing when the markers are not expressed and background signals for the marker are minimal. To account for missing data, we introduce a binary indicator,

$$m_{inj} = \begin{cases} 0, & \text{if } \tilde{y}_{inj} \text{ is observed,} \\ 1, & \text{if } \tilde{y}_{inj} \text{ is missing.} \end{cases}$$

That is, $m_{inj} = 1$ indicates that the expression level of marker j of cell n in sample i is not observed.

2 Sampling Model

We first assume K latent cell phenotypes and a cell takes one of the phenotypes. Each phenotype is defined based on a combination of expression or no expression of J markers in a cell. We let a $J \times K$ binary matrix $\boldsymbol{Z}$ characterize the latent phenotypes, a column for a phenotype, where values 0 and 1 of z_{jk} represent no expression and expression of marker j in cell phenotype k. We introduce latent indicators for latent cell phenotypes of cell n in sample i, λ_{in}, $i = 1, \ldots, I$ and $n = 1, \ldots, N_i$. $\lambda_{in} \in \{1, \ldots, K\}$ denotes the phenotype of cell n. Given $z_{j,\lambda_{i,n}}$, we assume a mixture of normals for y_{inj},

$$y_{inj} \mid \mu^*, \sigma_i^{2*} \overset{ind}{\sim} \begin{cases} \sum_{\ell=1}^{L^0} \eta_{ij\ell}^0 \text{Normal}(\mu_{0\ell}^\star, \sigma_{0i\ell}^{2\star}), & \text{if } z_{j,k} = 0, \\ \sum_{\ell=1}^{L^1} \eta_{ij\ell}^1 \text{Normal}(\mu_{1\ell}^\star, \sigma_{1i\ell}^{2\star}), & \text{if } z_{j,k} = 1, \end{cases}$$

where L^0 and L^1 are fixed, $0 < \eta_{ij\ell}^0 < 1$ with $\sum_\ell \eta_{ij\ell}^0 = 1$ and $0 < \eta_{ij\ell}^1 < 1$ with $\sum_\ell \eta_{ij\ell}^1 = 1$. Given $\lambda_{in} = k$ we define γ_{inj}; for $i = 1, \ldots, I$, $n = 1, \ldots, N_i$ and $j = 1, \ldots, J$,

$$p(\gamma_{inj} = \ell) = \eta_{in\ell}^{z_{jk}}, \text{ where } \ell \in \{1, \ldots, L^{z_{jk}}\}.$$

Given $\lambda_{in} = k$ and $\gamma_{inj} = \ell$, we assume a normal distribution for y_{inj}; for $i = 1, \ldots, I$, $n = 1, \ldots, N_i$ and $j = 1, \ldots, J$,

$$y_{inj} \mid \mu_{inj}, \sigma_{inj}^2 \overset{ind}{\sim} \text{Normal}(\mu_{inj}, \sigma_{inj}^2),$$

where $\mu_{inj} = \mu_{z_{j,k},\ell}^\star$ and $\sigma_{inj}^2 = {\sigma^2}_{iz_{j,k}\ell}^\star$. Given y_{inj}, we assume a Bernoulli distribution for m_{inj}; for $i = 1, \ldots, I$, $n = 1, \ldots, N_i$ and $j = 1, \ldots, J$,

$$\begin{aligned} m_{inj} \mid p_{inj} &\overset{ind}{\sim} \text{Bernoulli}(p_{inj}) \\ \text{logit}(p_{inj}) &:= \begin{cases} \beta_{0i} - \beta_{1i}(y_{inj} - c_0)^2, & \text{if } y_{inj} < c_0, \\ \beta_{0i} - \beta_{1i} c_1 \sqrt{y_{inj} - c_0}, & \text{otherwise}, \end{cases} \end{aligned}$$

where c_0 and c_1 are fixed.

Let $\boldsymbol{\theta}$ represent all parameters (discussed in the next section). Let $\boldsymbol{y}$ represent all y_{inj} $\forall (i, n, j)$. Let $\boldsymbol{m}$ represent all m_{inj} $\forall (i, n, j)$. The resulting **likelihood** is as follows:

$$\begin{aligned} \mathcal{L} = p(\boldsymbol{y}, \boldsymbol{m} \mid \boldsymbol{\theta}) &= p(\boldsymbol{m} \mid \boldsymbol{y}) p(\boldsymbol{y} \mid \boldsymbol{\theta}) \\ &= \prod_{i,n,j} p(m_{inj} \mid y_{inj}, \boldsymbol{\theta}) p(y_{inj} \mid \boldsymbol{\theta}) \end{aligned}$$

$$= \prod_{i,n,j} \left\{ p_{inj}^{m_{inj}} (1 - p_{inj})^{1-m_{inj}} \times \frac{1}{\sqrt{2\pi\sigma^2_{inj}}} \exp\left\{ -\frac{(y_{inj} - \mu_{inj})^2}{2\sigma^2_{inj}} \right\} \right\}. \quad (1)$$

The model is fully specified after priors are placed on all unknown parameters. The marginal density for $y_{i,n,j}$ after integrating out λ and γ is

$$p(y_{inj} \mid \theta) = \sum_{k=1}^{K} W_{ik} \sum_{\ell=1}^{L^{Z_{jk}}} \eta_{ij\ell}^{Z_{jk}} \cdot \text{Normal}(y_{inj} \mid \mu^*_{Z_{jk},\ell}, {\sigma^2}^*_{i,Z_{jk},\ell}). \quad (2)$$

3 Priors

Let $\boldsymbol{\theta}$ be ... **TODO**

Latent Phenotype for $J \times K$ binary matrix $\boldsymbol{Z}$,

$$\begin{aligned} v_k \mid \alpha &\overset{iid}{\sim} \text{Beta}(\alpha/K, 1),\ k = 1, \ldots, K, \\ \alpha &\sim \text{Gamma}(a_\alpha, b_\alpha) \\ \boldsymbol{h}_k &\overset{iid}{\sim} \text{Normal}_J(\boldsymbol{0}, \boldsymbol{G}) \\ Z_{jk} \mid h_{jk}, v_k &:= \mathbb{1}\left\{\Phi(h_{jk} \mid 0, \boldsymbol{G}_{jj}) < v_k\right\} \end{aligned}$$

Our main inferential objective is to obtain a binary matrix $\boldsymbol{Z}$ which contains the latent phenotypes present in the samples. The number of phenotypes is not known beforehand. A suitable prior for the latent phenotypes is an Indian buffet process (IBP), a distribution over binary matrices of infinite dimensions. When K is taken to the limit and $\boldsymbol{G}$ is the identity matrix, the representation yields the traditional IBP. Our model additionally accounts for the correlation between markers through $\boldsymbol{G}$.

We represent a sample as an admixture of heterogeneous cells having K different latent cell phenotypes characterized by $\boldsymbol{Z}$. A $K-$dimensional vector $\boldsymbol{W}_i$ denotes the proportion of a mixture for sample i and we assume

$$\begin{aligned} &\boldsymbol{W}_i \overset{iid}{\sim} \text{Dirichlet}_K(d/K, \ldots, d/K), \\ &p(\lambda_{in} = k \mid \boldsymbol{W}_i) = W_{ik}. \end{aligned}$$

Recall that the likelihood for y_{inj} is a mixture of normals with means $\eta_{0\ell}$, $\ell = 1, \ldots, L^0$

and $\eta_{1\ell}$, $\ell = 1, \ldots, L^1$ conditional on $z_{j\lambda_{in}} = 0$ and 1, respectively. Given that maker j is expressed in phenotype k, $z_{jk} = 1$, cells taking phenotype k may have a positive mean for the observed expression level of the marker. Similarly, for $z_{jk} = 0$ the mean of y_{inj} for cells taking phenotype k may take a negative value. We thus assume

$$\begin{aligned}
\mu^*_{0\ell} \mid \psi_0, \tau_0^2 &\overset{iid}{\sim} \text{Normal}_-(\psi_0, \tau_0^2), \quad \ell \in \left\{1, ..., L^0\right\}, \\
\mu^*_{1\ell} \mid \psi_1, \tau_1^2 &\overset{iid}{\sim} \text{Normal}_+(\psi_1, \tau_1^2), \quad \ell \in \left\{1, ..., L^1\right\}.
\end{aligned}$$

Note that X Normal$_-(m, s^2)$ denotes that X is distributed Normally with mean m and variance s^2, truncated to take on only **negative** values. Similarly, X Normal$_+(m, s^2)$ denotes that X is distributed Normally with mean m and variance s^2, truncated to take on only **positive** values.

For $i = 1, \ldots, I$,

$$\begin{aligned}
\sigma^2_{0i\ell} \mid s_i &\overset{ind}{\sim} \text{Inverse-Gamma}(a_\sigma, s_i), \quad \ell \in \left\{1, ..., L^0\right\}, \\
\sigma^2_{1i\ell} \mid s_i &\overset{ind}{\sim} \text{Inverse-Gamma}(a_\sigma, s_i), \quad \ell \in \left\{1, ..., L^1\right\}, \\
s_i &\overset{iid}{\sim} \text{Gamma}(a_s, b_s), \quad i \in \{1, ..., I\}.
\end{aligned}$$

For $i = 1, \ldots, I$, $n = 1, \ldots, N_i$ and $j = 1, \ldots, J$,

$$\begin{aligned}
p(\gamma_{inj} = \ell \mid \boldsymbol{\eta}_{i,j}^{Z_{j\lambda_{in}}}) = \eta_{ijl}^{Z_{j\lambda_{in}}}, \quad \ell \in \left\{1, ..., L^{Z_{j\lambda_{in}}}\right\}, \\
\boldsymbol{\eta}^0_{ij} \overset{iid}{\sim} \text{Dirichlet}_{L^0}(a_{\eta^0}/L^0, \ldots, a_{\eta^0}/L^0), \\
\boldsymbol{\eta}^1_{ij} \overset{iid}{\sim} \text{Dirichlet}_{L^1}(a_{\eta^1}/L^1, \ldots, a_{\eta^1}/L^1).
\end{aligned}$$

For $i = 1, \ldots, I$,

$$\begin{aligned}
\beta_{0i} &\overset{iid}{\sim} \text{Normal}(m_{\beta_0}, s^2_{\beta_0}) \quad \text{(hyper-parameters determined empirically)}, \\
\beta_{1i} &\overset{iid}{\sim} \text{Gamma}(a_{\beta_1}, b_{\beta_1}) \quad \text{(with mean } a_{\beta_1}/b_{\beta_1}\text{, determined empirically)}.
\end{aligned}$$

The Gamma distribution with parameters (a, b) has mean a/b. The inverse-Gamma distribution with parameters (a, b) has mean $b/(a-1)$. For the mixture locations of the likelihood,

4 Sampling via MCMC

Sampling can be done via Gibbs sampling by repeatedly updating each parameter one at a time until convergence. Parameter updates are made by sampling from it full conditional distribution. Where this cannot be done conveniently, a metropolis step will be necessary.

To sample from a distribution which is otherwise difficult to sample from, the Metropolis-Hastings algorithm can be used. This is particularly useful when sampling from a full conditional distribution of one of many parameters in an MCMC based sampling scheme (such as a Gibbs sampler). Say B samples from a distribution with density $p(\theta)$ is desired, one can do the following:

1. Provide an initial value for the sampler, e.g. $\theta^{(0)}$.
2. Repeat the following steps for $i = 1, ..., B$.
3. Sample a new value $\tilde{\theta}$ for $\theta^{(i)}$ from a proposal distribution $Q(\cdot \mid \theta^{(i-1)})$.
    - Let $q(\tilde{\theta} \mid \theta)$ be the density of the proposal distribution.
4. Compute the "acceptance ratio" to be

$$\rho = \min\left\{1, \frac{p(\tilde{\theta})}{p(\theta^{(i-1)})} \bigg/ \frac{q(\tilde{\theta} \mid \theta^{(i-1)})}{q(\theta^{(i-1)} \mid \tilde{\theta})}\right\}$$

5. Set

$$\theta^{(i)} := \begin{cases} \tilde{\theta} & \text{with probability } \rho \\ \theta^{(i-1)} & \text{with probability } 1-\rho. \end{cases}$$

Note that in the case of a symmetric proposal distribution, the acceptance ratio simplifies further to be $\frac{p(\tilde{\theta})}{p(\theta^{(i-1)})}$.

The proposal distribution should be chosen to have the same support as the parameter. Transforming parameters to have infinite support can, therefore, be convenient as a Normal proposal distribution can be used. Moreover, as previously mentioned, the use of symmetric proposal distributions (such as the Normal distribution) can simplify the computation of the acceptance ratio.

Some common parameter transformations are therefore presented here:

1. For parameters bounded between $(0, 1)$, a logit-transformation may be used. Specifically, if a random variable X with density $f_X(x)$ has support in the unit interval, then $Y = \text{logit}(X) = \log\left(\frac{p}{1-p}\right)$ will have density $f_Y(y) = f_X\left(\frac{1}{1+\exp(-y)}\right)\frac{e^{-y}}{(1+e^{-y})^2}$.
2. For parameters with support $(0, \infty)$, a log-transformation may be used. Specifically, if

a random variable X with density $f_X(x)$ has positive support, then $Y = \log(X)$ has pdf $f_Y(y) = f_X(e^y)e^y$.

5 Full Conditionals

5.1 Full Conditional for $\boldsymbol{\beta}$

Define f_{inj} to be

$$
\begin{aligned}
f_{inj} &:= P(m_{inj} \mid p_{inj}, y_{inj}) \\
&= p_{inj}^{m_{inj}}(1 - p_{inj})^{1-m_{inj}} \\
&= \left(\frac{1}{1 + e^{-x_{inj}}}\right)^{m_{inj}} \left(\frac{1}{1 + e^{x_{inj}}}\right)^{1-m_{inj}},
\end{aligned}
$$

where

$$
x_{inj} := \begin{cases} \beta_{0i} - \beta_{1i}(y_{inj} - c_0)^2, & \text{if } y_{inj} < c_0 \\ \beta_{0i} - \beta_{1i}c_1\sqrt{y_{inj} - c_0}, & \text{otherwise,} \end{cases}
$$

where c_0 and c_1 are fixed.

Write a little about the following...

Predetermine c_{low}, c_{high}, p_{low}, p_{high}, and p_0, which will determine the shape of the missing mechanism. Then

$$
\begin{aligned}
\beta_0 &:= \text{logit}(p_0) \\
\beta_1 &:= \frac{\beta_0 - \text{logit}(p_{\text{low}})}{(y_{\text{low}} - c_0)^2} \\
c_1 &:= \frac{\beta_0 - \text{logit}(p_{\text{high}})}{\beta_1 \sqrt{y_{\text{high}} - c_0}}.
\end{aligned}
$$

5.1.1 Full Conditional for β_{0i}

Recall that $\beta_{0i} \overset{iid}{\sim} \text{Normal}(m_{\beta_0}, s^2_{\beta_0})$.

$$
\begin{aligned}
p(\beta_{0i} \mid \boldsymbol{y}, \text{rest}) &\propto p(\beta_{0i}) \times \prod_{n=1}^{N_i} \prod_{j=1}^{J} f_{inj} \\
&\propto \exp\left\{ \frac{-(\beta_{0i} - m_{\beta_0})^2}{2s^2_{\beta_0}} \right\} \prod_{n=1}^{N_i} \prod_{j=1}^{J} f_{inj}
\end{aligned}
$$

Since the full conditional distribution cannot be directly sampled from, it may be sampled from by a Metropolis step with a Normal proposal distribution. The proposed state is accepted with probability

$$
\min\left\{ 1, \frac{p(\tilde{\beta}_{0i} \mid \boldsymbol{y}, \text{rest})}{p(\beta_{0i} \mid \boldsymbol{y}, \text{rest})} \right\}.
$$

5.1.2 Full Conditional for β_{1i}

Recall that $\beta_{1i} \overset{ind}{\sim} \text{Gamma}(a_{\beta_1}, b_{\beta_1})$.

$$
\begin{aligned}
p(\beta_{1i} \mid \boldsymbol{y}, \text{rest}) &\propto p(\beta_{1i}) \times \prod_{n=1}^{N_i} \prod_{j=1}^{J} f_{inj} \\
&\propto \beta_{1i}^{a_{\beta_1}-1} \exp\left\{ -b_{\beta_1}\beta_{1i} \right\} \prod_{n=1}^{N_i} \prod_{j=1}^{J} f_{inj}
\end{aligned}
$$

Since the full conditional distribution for β_{1i} cannot be directly sampled from, it may be sampled from by a Metropolis step with a Normal proposal distribution. The parameter first needs to be log-transformed. Let the full conditional of the transformed parameter be $p(\xi \mid \boldsymbol{y}, \text{rest}) = p_{\beta_{1i}}(\exp(\xi) \mid \boldsymbol{y}, \text{rest}) \exp(\xi)$. Then, the proposed state of the transformed parameter (ξ) is accepted with probability

$$
\min\left\{ 1, \frac{p(\tilde{\xi} \mid \boldsymbol{y}, \text{rest})}{p(\xi \mid \boldsymbol{y}, \text{rest})} \right\}.
$$

Exponentiating the updated value for ξ returns the updated value for β_{1i}.

5.2 Full Conditional for Missing $\boldsymbol{y}$

$$\begin{aligned} p(y_{inj} \mid m_{inj}, \text{rest}) &\propto p(y_{inj} \mid \text{rest})\; p(m_{inj} \mid y_{inj}, \text{rest}) \\ &\propto f_{inj} \sum_{\ell=1}^{L^{Z_{jk}}} \eta_{ij\ell}^{Z_{jk}} \cdot \text{Normal}(y_{inj} \mid \mu^*_{Z_{jk},\ell}, {\sigma^2}^*_{i,Z_{jk},\ell}). \end{aligned}$$

Since the full conditional distribution cannot be directly sampled from, it may be sampled from by a Metropolis step with a Normal proposal distribution. The proposed state is accepted with probability

$$\min\left\{1, \frac{p(\tilde{y}_{inj} \mid \boldsymbol{y}, \text{rest})}{p(y_{inj} \mid \boldsymbol{y}, \text{rest})}\right\}.$$

Note that f_{inj} is a function of y_{inj} and should be computed accordingly.

5.3 Full Conditional for $\boldsymbol{\mu}^\star$

For $\mu^*_{0\ell}$, let $S_{0i\ell} = \{(i,n,j) : (Z_{j,\lambda_{in}} = 0 \;\cap\; \gamma_{inj} = \ell)\}$g and $|S_{0i\ell}|$ the cardinality of $S_{0i\ell}$.

$$\begin{aligned} p(\mu^*_{0\ell} \mid \boldsymbol{y}, \text{rest}) &\propto p(\mu^*_{0\ell} \mid \psi_0, \tau_0^2) \times p(\boldsymbol{y} \mid \mu^*_{0\ell}, \text{rest}) \\ &\propto \mathbb{1}\{\mu^*_{0\ell} < 0\} \exp\left\{\frac{-(\mu^*_{0\ell} - \psi_0)^2}{2\tau_0^2}\right\} \prod_{i=1}^{I} \prod_{(i,n,j)\in S_{0i\ell}} \exp\left\{\frac{-(y_{inj} - \mu^*_{0\ell})^2}{2{\sigma^2}^\star_{i0\ell}}\right\} \\ &\propto \exp\left\{-\frac{(\mu^*_{0\ell})^2}{2}\left(\frac{1}{\tau_0^2} + \sum_{i=1}^{I} \frac{|S_{0i\ell}|}{{\sigma^2}^\star_{0i\ell}}\right) + \mu^*_{0\ell}\left(\frac{\psi_0}{\tau_0^2} + \sum_{i=1}^{I} \sum_{S_{0i\ell}} \frac{y_{inj}}{{\sigma^2}^*_{0i\ell}}\right)\right\} \\ &\quad \times \mathbb{1}\{\mu^*_{0i\ell} < 0\} \end{aligned}$$

$$\therefore \mu^*_{0l} \mid \boldsymbol{y}, \text{rest} \overset{ind}{\sim} \text{Normal}_-\left(\frac{\psi_0 + \tau_0^2 \sum_{i=1}^{I} \sum_{S_{0i\ell}} (y_{inj}/{\sigma^2}^\star_{0i\ell})}{1 + \tau_0^2 \sum_{i=1}^{I} (|S_{0i\ell}|/{\sigma^2}^\star_{0i\ell})}, \frac{\tau_0^2}{1 + \tau_0^2 \sum_{i=1}^{I} (|S_{0i\ell}|/{\sigma^2}^\star_{0i\ell})}\right)$$

Similarly for $\mu^*_{1\ell}$, let $S_{1i\ell} = \{(i,n,j) : (Z_{j,\lambda_{in}} = 1 \;\cap\; \gamma_{inj} = \ell)\}$g and $|S_{1i\ell}|$ the cardinality of $S_{1i\ell}$.

$$\therefore \mu^*_{1l} \mid \boldsymbol{y}, \text{rest} \overset{ind}{\sim} \text{Normal}_+\left(\frac{\psi_1 + \tau_1^2 \sum_{i=1}^{I} \sum_{S_{1i\ell}} (y_{inj}/{\sigma^2}^\star_{1i\ell})}{1 + \tau_1^2 \sum_{i=1}^{I} (|S_{1i\ell}|/{\sigma^2}^\star_{1i\ell})}, \frac{\tau_1^2}{1 + \tau_1^2 \sum_{i=1}^{I} (|S_{1i\ell}|/{\sigma^2}^\star_{1i\ell})}\right)$$

5.4 Full Conditional for $\boldsymbol{\sigma^{2^*}}$

Let $S_{0i\ell} = \{(i,n,j) : Z_{j,\lambda_{in}} = 0 \ \cap \ \gamma_{inj} = \ell\}$, $i = 1, \ldots, I$.

$$
\begin{aligned}
p({\sigma^2}^*_{0i\ell} \mid \boldsymbol{y}, \text{rest}) &\propto p({\sigma^2}^*_{0i\ell} \mid s_i) \times p(\boldsymbol{y} \mid {\sigma^2}^*_{0i\ell}, \text{rest}) \\
&\propto ({\sigma^2}^*_{0i\ell})^{-a_\sigma - 1} \exp\left\{-\frac{s_i}{{\sigma^2}^*_{0i\ell}}\right\} \prod_{(i,n,j)\in S_{0i\ell}} \left\{\frac{1}{\sqrt{2{\sigma^2}^*_{0i\ell}}} \exp\left\{\frac{-(y_{inj} - \mu^*_{0\ell})^2}{2{\sigma^2}^*_{0i\ell}}\right\}\right\} \\
&\propto ({\sigma^2}^*_{0i\ell})^{-(a_\sigma + \frac{|S_{0i\ell}|}{2}) - 1} \exp\left\{\left(\frac{1}{{\sigma^2}^*_{0i\ell}}\right)\left(s_i + \sum_{(i,n,j)\in S_{0i\ell}} \frac{(y_{inj} - \mu^*_{0\ell})^2}{2}\right)\right\}.
\end{aligned}
$$

$$
\therefore {\sigma^2}^*_{0i\ell} \mid \boldsymbol{y}, \text{rest} \overset{ind}{\sim} \text{Inverse-Gamma}\left(a_\sigma + \frac{|S_{0i\ell}|}{2}, \quad s_i + \sum_{(i,n,j)\in S_{0i\ell}} \frac{(y_{inj} - \mu^*_{0\ell})^2}{2}\right).
$$

Similarly, let $S_{1i\ell} = \{(i,n,j) : Z_{j,\lambda_{in}} = 1 \ \cap \ \gamma_{inj} = \ell\}$. Then, the full conditional for ${\sigma^2}^*_{1i\ell}$ is

$$
\therefore {\sigma^2}^*_{1i\ell} \mid \boldsymbol{y}, \text{rest} \overset{ind}{\sim} \text{Inverse-Gamma}\left(a_\sigma + \frac{|S_{1i\ell}|}{2}, \quad s_i + \sum_{(i,n,j)\in S_{1i\ell}} \frac{(y_{inj} - \mu^*_{1\ell})^2}{2}\right).
$$

5.5 Full Conditional for s_i

$$
\begin{aligned}
p(s_i \mid \boldsymbol{y}, \text{rest}) &\propto p(s_i) \times \prod_{z=0}^{1}\prod_{\ell=1}^{L^z} p({\sigma^2}^*_{zi\ell} \mid s_i) \\
&\propto s_i^{a_s - 1} \exp\{-b_s s_i\} \times \prod_{z=0}^{1}\prod_{\ell=1}^{L^z} s_i^{a_\sigma} \exp\{-s_i/{\sigma^2}^*_{zi\ell}\} \\
&\propto s_i^{a_s + (L^0 + L^1)a_\sigma - 1} \exp\left\{-s_i\left(b_s + \sum_{z=0}^{1}\sum_{\ell=1}^{L^z} 1/{\sigma^2}^*_{zi\ell}\right)\right\}.
\end{aligned}
$$

$$
\therefore s_i \mid \boldsymbol{y}, \text{rest} \sim \text{Gamma}\left(a_s + (L^0 + L^1)a_\sigma, \quad b_s + \sum_{z=0}^{1}\sum_{\ell=1}^{L^z} \frac{1}{{\sigma^2}^*_{zi\ell}}\right).
$$

5.6 Full Conditional for $\boldsymbol{\gamma}$

The prior for γ_{inj} is $p(\gamma_{inj} = \ell \mid Z_{j\lambda_{in}} = z, \eta^z_{ij\ell}) = \eta^z_{ij\ell}$, where $\ell \in \{1, ..., L^z\}$.

$$\begin{aligned}
p(\gamma_{inj} = \ell \mid \boldsymbol{y}, Z_{j\lambda_{in}} = z, \text{rest}) &\propto p(\gamma_{inj} = \ell) \times p(y_{inj} \mid \gamma_{inj} = \ell, \text{rest}) \\
&\propto p(\gamma_{inj} = \ell) \times p(y_{inj} \mid \mu^*_{z\ell}, {\sigma^2}^*_{zi\ell}, \text{rest}) \\
&\propto \eta^z_{ij\ell} \times \text{Normal}(y_{inj} \mid \mu^*_{z\ell}, {\sigma^2}^*_{zi\ell}) \\
&\propto \eta^z_{ij\ell} \times ({\sigma^2}^*_{zi\ell})^{-1/2} \exp\left\{ -\frac{(y_{inj} - \mu^*_{z\ell})^2}{2{\sigma^2}^*_{zi\ell}} \right\}
\end{aligned}$$

The normalizing constant is obtained by summing the last expression over $\ell = 1, ..., L^z$. Moreover, since ℓ is discrete, a Gibbs update can be done on γ_{inj}.

5.7 Full Conditional for $\boldsymbol{\eta}$

The prior for $\boldsymbol{\eta}^z_{ij}$ is $\boldsymbol{\eta}^z_{ij} \sim \text{Dirichlet}_{L^z}(a_{\eta^z}/L^z)$, for $z \in \{0, 1\}$. So the full conditional for $\boldsymbol{\eta}^z_{ij}$ is:

$$\begin{aligned}
p(\boldsymbol{\eta}^z_{ij} \mid \text{rest}) &\propto \; p(\boldsymbol{\eta}^z_{ij}) \times \prod_{n=1}^{N_i} p(\gamma_{inj} \mid \boldsymbol{\eta}^z_{ij}) \\
&\propto \; p(\boldsymbol{\eta}^z_{ij}) \times \prod_{n=1}^{N_i} \prod_{\ell=1}^{L^z} \left(\eta^z_{ij\ell}\right)^{\mathbb{1}\left\{\gamma_{inj} = \ell \,\cap\, Z_{j\lambda_{in}} = z\right\}} \\
&\propto \; \prod_{\ell=1}^{L^z} \left(\eta^z_{ij\ell}\right)^{a_{\eta^z}/L^z - 1} \times \prod_{n=1}^{N_i} \prod_{\ell=1}^{L^z} \left(\eta^z_{ij\ell}\right)^{\mathbb{1}\left\{\gamma_{inj} = \ell \,\cap\, Z_{j\lambda_{in}} = z\right\}} \\
&\propto \; \prod_{\ell=1}^{L^z} \left(\eta^z_{ij\ell}\right)^{\left(a_{\eta^z}/L^z + \sum_{n=1}^{N_i} \mathbb{1}\left\{\gamma_{inj} = \ell \,\cap\, Z_{j\lambda_{in}} = z\right\}\right) - 1}
\end{aligned}$$

Therefore,

$$\boldsymbol{\eta}^z_{ij} \mid \boldsymbol{y}, \text{rest} \;\sim\; \text{Dirichlet}_{L^z}\left(a^*_1, ..., a^*_{L^z}\right)$$

where $a^*_\ell = a_{\eta^z}/L^z + \sum_{n=1}^{N_i} \mathbb{1}\left\{\gamma_{inj} = \ell \,\cap\, Z_{j\lambda_{in}} = z\right\}$. Consequently, the full conditional for $\boldsymbol{\eta}^z_{ij}$ can be sampled from directly from a Dirichlet distribution of the form above.

5.8 Full Conditional for $\boldsymbol{v}$

The prior distribution for v_k are $v_k \mid \alpha \overset{ind}{\sim} \text{Beta}(\alpha/K, 1)$, for $k = 1, ..., K$. So, $p(v_k \mid \alpha) = \alpha v_k^{\alpha/K-1}$. Let $S = \{(i,n) \colon \lambda_{in} = k\}$.

$$\begin{aligned} p(v_k \mid \boldsymbol{y}, \text{rest}) &\propto p(v_k) \prod_{j=1}^{J} \prod_{(i,n)\in S} p(\boldsymbol{y} \mid v_k, \text{rest}) \\ &\propto (v_k)^{\alpha/K-1} \prod_{j=1}^{J} \prod_{(i,n)\in S} \sum_{\ell=1}^{L^{Z_{jk}}} \eta_{ij\ell}^{Z_{jk}} \cdot \text{Normal}(y_{inj} \mid \mu^*_{Z_{jk}\ell}, {\sigma^2}^*_{Z_{jk}i\ell}) \end{aligned}$$

Since the full conditional distribution for v_k cannot be directly sampled from, it may be sampled from by a Metropolis step with a Normal proposal distribution. The parameter first needs to be logit-transformed. Let the full conditional of the transformed parameter be $p(\xi \mid \boldsymbol{y}, \text{rest}) = p_{v_k}\left(\frac{1}{1+\exp(-\xi)} \mid \boldsymbol{y}, \text{rest}\right) \frac{\exp(-\xi)}{(1+\exp(-\xi))^2}$. The proposed state of the transformed parameter (ξ) is accepted with probability

$$\min\left\{1, \frac{p(\tilde{\xi} \mid \boldsymbol{y}, \text{rest})}{p(\xi \mid \boldsymbol{y}, \text{rest})}\right\}$$

Taking the inverse-logit of the updated value for ξ returns the updated value for v_k. Note also that $\mu^*_{Z_{jk}\ell}$ and ${\sigma^2}^*_{Z_{jk}i\ell}$ are functions of v_k, and should be computed accordingly. Moreover, we will only recompute the likelihood (in the metropolis acceptance ratio) when Z_{jk} becomes different.

5.9 Full Conditional for α

$$\begin{aligned} p(\alpha \mid \boldsymbol{y}, \text{rest}) &\propto p(\alpha) \times \prod_{k=1}^{K} p(v_k \mid \alpha) \\ &\propto \alpha^{a_\alpha - 1} \exp\{-b_\alpha \alpha\} \times \prod_{k=1}^{K} \alpha \; v_k^{\alpha/K} \\ &\propto \alpha^{a_\alpha + K - 1} \exp\left\{-\alpha\left(b_\alpha + \frac{\sum_{k=1}^{K} \log v_k}{K}\right)\right\} \end{aligned}$$

$$\therefore \alpha \mid \boldsymbol{y}, \text{rest} \sim \text{Gamma}\left(a_\alpha + K, \; b_\alpha + \frac{\sum_{k=1}^{K} \log v_k}{K}\right)$$

5.10 Full Conditional for $\boldsymbol{H}$

The prior for $\boldsymbol{h}_k$ is $\boldsymbol{h}_k \sim \text{Normal}_J(0, \boldsymbol{G})$. We can analytically compute the conditional distribution $h_{j,k} \mid \boldsymbol{h}_{-j,k}$, which is

$$h_{jk} \mid \boldsymbol{h}_{-j,k} \sim \text{Normal}(m_j, S_j^2),$$

where

$$\begin{cases} m_j & = \boldsymbol{G}_{j,-j}\boldsymbol{G}^{-1}_{-j,-j}(\boldsymbol{h}_{-j,k}) \\ S_j^2 & = \boldsymbol{G}_{j,j} - \boldsymbol{G}_{j,-j}\boldsymbol{G}^{-1}_{-j,-j}\boldsymbol{G}_{-j,j} \end{cases}$$

and the notation $\boldsymbol{h}_{-j,k}$ refers to the vector h_k excluding the j^{th} element. Likewise, $\boldsymbol{G}_{-j,k}$ refers to the k^{th} column of the matrix $\boldsymbol{G}$ excluding the j^{th} row.

Note that if $\boldsymbol{G} = \mathbf{I}_J$, then $m_j = 0$ and $S_j^2 = 1$. Let $S = \{(i,n) \colon \lambda_{in} = k\}$.

$$\begin{aligned} p(h_{jk} \mid \boldsymbol{y}, \text{rest}) &\propto p(h_{jk}) \prod_{(i,n)\in S} p(y_{inj} \mid h_{jk}, \text{rest}) \\ &\propto \exp\left\{\frac{-(h_{jk} - m_j)^2}{2S_j^2}\right\} \prod_{j=1}^{J} \prod_{(i,n)\in S} \sum_{\ell=1}^{L^{Z_{jk}}} \eta_{ij\ell}^{Z_{jk}} \cdot \text{Normal}(y_{inj} \mid \mu^*_{Z_{jk}\ell}, {\sigma^2}^*_{Z_{jk}i\ell}) \end{aligned}$$

Since the full conditional distribution cannot be directly sampled from, it may be sampled from by a Metropolis step with a Normal proposal distribution. The proposed state is accepted with probability

$$\min\left\{1, \frac{p(\tilde{h}_{jk} \mid \boldsymbol{y}, \text{rest})}{p(h_{jk} \mid \boldsymbol{y}, \text{rest})}\right\}.$$

Note also that $\mu^*_{Z_{jk}\ell}$ and ${\sigma^2}^*_{Z_{jk}i\ell}$ are functions of h_{jk}, and should be computed accordingly. Moreover, we will only recompute the likelihood (in the metropolis acceptance ratio) when Z_{jk} becomes different.

5.11 Full Conditional for $\boldsymbol{\lambda}$

The prior for λ_{in} is $p(\lambda_{in} = k \mid \boldsymbol{W}_i) = W_{ik}$.

$$p(\lambda_{in} = k \mid \boldsymbol{y}, \text{rest}) \propto p(\lambda_{in} = k)\ p(\boldsymbol{y} \mid \lambda_{in} = k, \text{rest})$$

$$\propto W_{ik} \prod_{j=1}^{J} \left(\sum_{\ell=1}^{L^{Z_{jk}}} \eta_{ij\ell}^{Z_{jk}} \cdot \text{Normal}(y_{inj} \mid \mu^*_{Z_{jk}\ell}, \sigma^{2*}_{Z_{jk}i\ell}). \right)$$

The normalizing constant is obtained by summing the last expression over $k = 1, ..., K$. Moreover, since k is discrete, a Gibbs update can be done on λ_{in}.

5.12 Full Conditional for $\boldsymbol{W}$

The prior for $\boldsymbol{W}_i$ is $\boldsymbol{W}_i \sim \text{Dirichlet}(d/K, \cdots, d/K)$. So the full conditional for $\boldsymbol{W}_i$ is:

$$\begin{aligned}
p(\boldsymbol{W}_i \mid \text{rest}) &\propto \; p(\boldsymbol{W}_i) \times \prod_{n=1}^{N_i} p(\lambda_{in} \mid \boldsymbol{W}_i) \\
&\propto \; p(\boldsymbol{W}_i) \times \prod_{n=1}^{N_i} \prod_{k=1}^{K} W_{ik}^{\mathbb{1}\{\lambda_{in}=k\}} \\
&\propto \; \prod_{k=1}^{K} W_{ik}^{d/K-1} \times \prod_{n=1}^{N_i} \prod_{k=1}^{K} W_{ik}^{\mathbb{1}\{\lambda_{in}=k\}} \\
&\propto \; \prod_{k=1}^{K} W_{ik}^{\left(d/K+\sum_{n=1}^{N_i} \mathbb{1}\{\lambda_{in}=k\}\right)-1}.
\end{aligned}$$

Therefore,

$$\boldsymbol{W}_i \mid \boldsymbol{y}, \text{rest} \;\sim\; \text{Dirichlet}\left(d/K + \sum_{n=1}^{N_i} \mathbb{1}\left\{\lambda_{i,n} = 1\right\}, ..., d/K + \sum_{n=1}^{N_i} \mathbb{1}\left\{\lambda_{i,n} = K\right\} \right).$$

Consequently, the full conditional for $\boldsymbol{W}_i$ can be sampled from directly from a Dirichlet distribution of the form above.

5.13 Posterior Estimate for Z, W, and λ

Obtaining posterior estimates and quantifying uncertainty for Z, W, and λ can be a challenge since they depend directly on K and are susceptible to label-switching. Therefore, we describe a way to select point-estimates for Z, W, and λ from their posterior samples.

Suppose we obtain B samples from the posterior distribution of θ. Let $\theta^{(b)}$ denote sample b in that sample for $b \in \{1, ..., B\}$. Then, for each posterior sample of Z and W_i (i.e. $Z^{(b)}$ and

$W_i^{(b)}$) let $A_i^{(b)}$ be a $J \times J$ adjacency-matrix;

$$A_{i,j,j'}^{(b)} = \sum_{k=1}^{K} W_{i,k}^{(b)} \mathbb{1}\left\{Z_{j,k}^{(b)} = 1\right\} \mathbb{1}\left\{Z_{j',k}^{(b)} = 1\right\}.$$

Then, compute the mean adjacency-matrix as $\bar{A}_i = \sum_{b=1}^{B} A_i^{(b)}/B$. The posterior estimate for $\boldsymbol{Z}_i$ is then

$$\hat{\boldsymbol{Z}}_i = \text{argmin}_{\boldsymbol{Z}} \sum_{j,j'} (A_{i,j,j'}^{(b)} - \bar{A}_{i,j,j'})^2).$$

We search for such $\hat{\boldsymbol{Z}}_i$ using the Monte Carlo method. Suppose $\hat{\boldsymbol{Z}}_i = \boldsymbol{Z}^{(b)}$. We then let $\widehat{\boldsymbol{W}}_i = \boldsymbol{W}_i^{(b)}$, and $\hat{\boldsymbol{\lambda}}_i = \boldsymbol{\lambda}_i^{(b)}$.