1 Probability Model

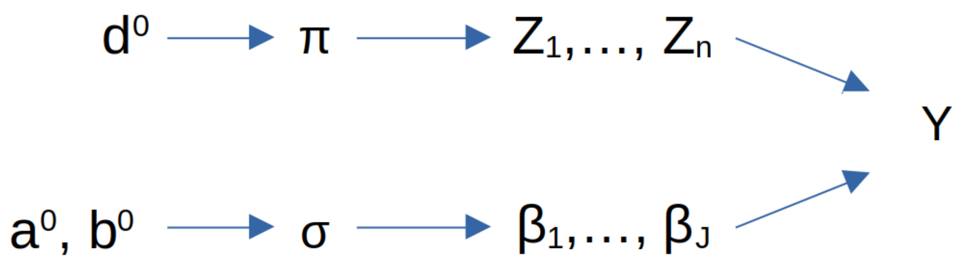

Figure 1: The probability model with hidden variables.

- $\sigma^2 \sim IG(a^0, b^0)$. We have the following :

$$P(\sigma^2|a^0, b^0) = \frac{1}{\Gamma(a^0)} {b^0}^{a^0} (\sigma^2)^{-a^0-1} e^{-\frac{b^0}{\sigma^2}}. \tag{1}$$

- $\boldsymbol{\beta}_j \in \mathbb{R}^L$, $\boldsymbol{\beta}_j|\sigma^2 \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2 \boldsymbol{I}_{L\times L})$. $P(\boldsymbol{\beta}_j|\sigma^2) = (\frac{1}{2\pi\sigma^2})^{\frac{L}{2}} e^{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j}$. We have the following:

$$P(\boldsymbol{\beta}|\sigma^2) = \prod_{j=1}^{J} P(\boldsymbol{\beta}_j|\sigma^2) = (\frac{1}{2\pi\sigma^2})^{\frac{JL}{2}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{J}\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j} \tag{2}$$

- $\boldsymbol{\pi} \in [0,1]^L$, $\boldsymbol{\pi}|\mathbf{d}^0 \sim \text{Dirichlet}(\mathbf{d}^0 = (d_1^0, ..., d_L^0))$. We have the following:

$$P(\boldsymbol{\pi}|\mathbf{d}^0) = \frac{1}{Beta(\mathbf{d}^0)} \prod_{l=1}^{L} \pi_l^{d_l^0 - 1} \tag{3}$$

- $\mathbf{Z}_i = [z_{i1}, ..., z_{iL}]^T \in \{0,1\}^L$, $\mathbf{Z}_i|\boldsymbol{\pi} \sim \text{Categorical}(\boldsymbol{\pi})$. $P(\mathbf{Z}_i|\boldsymbol{\pi}) = \prod_{l=1}^{L} \pi_l^{z_{il}}$. We have the following:

$$P(\mathbf{Z}|\boldsymbol{\pi}) = \prod_{i=1}^{N} P(\mathbf{Z}_i|\boldsymbol{\pi}) = \prod_{i=1}^{N}\prod_{l=1}^{L} \pi_l^{z_{il}} \tag{4}$$

- $P(Y_{ij} = y_{ij}|\mathbf{z}_i, \boldsymbol{\beta}_j) = \prod_{l=1}^{L} [\sigma((2y_{ij}-1)\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl})]^{z_{il}}$, where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function. $\boldsymbol{\Delta}_j = [\boldsymbol{\delta}_{j1}, ..., \boldsymbol{\delta}_{jL}]^T$. We have the following:

$$P(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\beta}, \Delta) = \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{l=1}^{L} [\sigma((2y_{ij}-1)\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl})]^{z_{il}} \tag{5}$$

2 Variational Bayesian Inference

We can then get the likelihood function:

$$
\begin{aligned}
P(\mathbf{Y},\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2|\boldsymbol{\Delta},\mathbf{d^0},a^0,b^0) &= P(\mathbf{Y}|\mathbf{Z},\boldsymbol{\beta},\Delta)P(\mathbf{Z}|\boldsymbol{\pi})P(\boldsymbol{\pi}|\mathbf{d}^0)P(\boldsymbol{\beta}|\sigma^2)P(\sigma^2|a^0,b^0) \\
&= \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{l=1}^{L}[\sigma((2y_{ij}-1)\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl})]^{z_{il}} \times \prod_{i=1}^{N}\prod_{l=1}^{L}\pi_l^{z_{il}} \times \frac{1}{Beta(\mathbf{d}^0)}\prod_{l=1}^{L}\pi_l^{d_l^0-1} \\
&\times (\frac{1}{2\pi\sigma^2})^{\frac{JL}{2}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{J}\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j} \times \frac{1}{\Gamma(a^0)} b^{0^{a^0}} (\sigma^2)^{-a^0-1} e^{-\frac{b^0}{\sigma^2}}
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
\log P(\mathbf{Y}|\boldsymbol{\Delta},\mathbf{d^0},a^0,b^0) &= \log\int\int\int\sum_{\mathbf{Z}} P(\mathbf{Y},\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2|\boldsymbol{\Delta},\mathbf{d^0},a^0,b^0)d\boldsymbol{\beta}d\boldsymbol{\pi}d\sigma^2 \\
&= \log\int\int\int\sum_{\mathbf{Z}} P(\mathbf{Y}|\mathbf{Z},\boldsymbol{\beta},\Delta)P(\mathbf{Z}|\boldsymbol{\pi})P(\boldsymbol{\pi}|\mathbf{d}^0)P(\boldsymbol{\beta}|\sigma^2)P(\sigma^2|a^0,b^0)d\boldsymbol{\beta}d\boldsymbol{\pi}d\sigma^2 \\
&\geq \int\int\int\sum_{\mathbf{Z}} Q(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2)\log\frac{P(\mathbf{Y}|\mathbf{Z},\boldsymbol{\beta},\Delta)P(\mathbf{Z}|\boldsymbol{\pi})P(\boldsymbol{\pi}|\mathbf{d}^0)P(\boldsymbol{\beta}|\sigma^2)P(\sigma^2|a^0,b^0)}{Q(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2)}d\boldsymbol{\beta}d\boldsymbol{\pi}d\sigma^2 \\
&= \mathcal{L}(Q)
\end{aligned} \tag{7}
$$

We restrict the space of the approximate densities Q by employing the mean-field variational family of approximate densities. Therefore, we assume the following equation holds:

$$
\begin{aligned}
Q(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2) &= q_1(\mathbf{Z})q_2(\boldsymbol{\beta})q_3(\boldsymbol{\pi})q_4(\sigma^2) \\
&= (\prod_{i=1}^{N} q_{1i}(\mathbf{Z}_i))(\prod_{j=1}^{J} q_{2j}(\boldsymbol{\beta}_j))(q_3(\boldsymbol{\pi}))(q_4(\sigma^2))
\end{aligned} \tag{8}
$$

Due to the existence of the sigmoid link function, the posterior is not conjugate to Gaussian distribution at this point, we need to manipulate the sigmoid function before conducting further steps. Here, we want to employ the tangent transformation approach. A common method is proposed by Jaakkola and Jordan [2000]. The main result is the following equation:

$$
\sigma(x) \geq \sigma(\xi)e^{(x-\xi)/2-\lambda(\xi)(x^2-\xi^2)}, \tag{9}
$$

where $\lambda(\xi) = \frac{1}{2\xi}(\sigma(\xi)-1/2) = \frac{1}{4\xi}tanh(\xi/2)$. This lower bound is attainable by optimizing ξ.

We can then plug (9) into (5) to get a lower bound:

$$
\log P(\mathbf{Y}|\mathbf{Z},\boldsymbol{\beta}) \geq \log h(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\xi}) = \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L} z_{il}[\log\sigma(\xi_{ijl}) + \frac{1}{2}((2y_{ij}-1)\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl} - \xi_{ijl}) - \lambda(\xi_{ijl})(\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl}\boldsymbol{\delta}_{jl}^T\boldsymbol{\beta}_j - \xi_{ijl}^2)] \tag{10}
$$

Denote $\tilde{P}(\mathbf{Y},\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2|\boldsymbol{\Delta},\mathbf{d^0},a^0,b^0) = h(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\xi})P(\mathbf{Z}|\boldsymbol{\pi})P(\boldsymbol{\pi}|\mathbf{d}^0)P(\beta|\sigma^2)P(\sigma^2|a^0,b^0)$. Now we have a new variational lower bound:

$$
\begin{aligned}
\mathcal{L}(Q) &= \int\int\int \sum_{\mathbf{Z}} Q(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2)\log\frac{P(\mathbf{Y}|\mathbf{Z},\boldsymbol{\beta},\Delta)P(\mathbf{Z}|\boldsymbol{\pi})P(\boldsymbol{\pi}|\mathbf{d}^0)P(\boldsymbol{\beta}|\sigma^2)P(\sigma^2|a^0,b^0)}{Q(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2)}d\boldsymbol{\beta}d\boldsymbol{\pi}d\sigma^2 \\
&\geq \int\int\int \sum_{\mathbf{Z}} Q(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2)\log\frac{h(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\xi})P(\mathbf{Z}|\boldsymbol{\pi})P(\boldsymbol{\pi}|\mathbf{d}^0)P(\boldsymbol{\beta}|\sigma^2)P(\sigma^2|a^0,b^0)}{Q(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2)}d\boldsymbol{\beta}d\boldsymbol{\pi}d\sigma^2 \\
&= \tilde{\mathcal{L}}(Q,\boldsymbol{\xi})
\end{aligned} \tag{11}
$$

3 Coordinate Ascent Variational Inference(CAVI)

A common approach is to use the coordinate ascent method. We optimize the variational approximation of each latent variable while holding others fixed. (Bishop 2006)

Algorithm 1: Coordinate ascent variational inference (CAVI)

Input: A model $p(\mathbf{x},\mathbf{z})$, a dataset $\mathbf{x}$
Output: A variational density $q(\mathbf{z}) = \prod_{j=1}^{m} q_j(z_j)$
while *ELBO has not converged* **do**
  for $j \in \{1, ..., m\}$ **do**
    Set $q_j(z_j) \propto \exp\{\mathbb{E}_{-j}[\log p(z_j|\mathbf{z_{-j}},\mathbf{x})]\} \propto \exp\{\mathbb{E}_{-j}[\log p(\mathbf{z},\mathbf{x})]\}$
  end
  Compute $ELBO(q) = \mathbb{E}_q[\log p(\mathbf{x},\mathbf{z})] - \mathbb{E}_q[\log q(\mathbf{z})]$
end

$$
\begin{aligned}
\log\tilde{P}(\mathbf{Y},\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\pi},\sigma^2|\boldsymbol{\Delta},\mathbf{d^0},a^0,b^0) &= \log[h(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\xi})P(\mathbf{Z}|\boldsymbol{\pi})P(\boldsymbol{\pi}|\mathbf{d}^0)P(\boldsymbol{\beta}|\sigma^2)P(\sigma^2|a^0,b^0)] \\
&= \log h(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\xi}) + \log P(\mathbf{Z}|\boldsymbol{\pi}) + \log P(\boldsymbol{\pi}|\mathbf{d}^0) + \log P(\boldsymbol{\beta}|\sigma^2) + \log P(\sigma^2|a^0,b^0) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L} z_{il}[\log\sigma(\xi_{ijl}) + \frac{1}{2}((2y_{ij}-1)\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl} - \xi_{ijl}) - \lambda(\xi_{ijl})(\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl}\boldsymbol{\delta}_{jl}^T\boldsymbol{\beta}_j - \xi_{ijl}^2)] \\
&+ \sum_{i=1}^{N}\sum_{l=1}^{L} z_{il}\log\pi_l - \log Beta(\mathbf{d}^0) + \sum_{l=1}^{L}(d_l^0-1)\log\pi_l - \frac{JL}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{J}\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j \\
&- \log(\Gamma(a^0)) + a^0\log b^0 - b^0/\sigma^2 - (a^0+1)\log(\sigma^2) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L} z_{il}[\log\sigma(\xi_{ijl}) + \frac{1}{2}((2y_{ij}-1)\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl} - \xi_{ijl}) - \lambda(\xi_{ijl})(\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl}\boldsymbol{\delta}_{jl}^T\boldsymbol{\beta}_j - \xi_{ijl}^2)] \\
&+ \sum_{i=1}^{N}\sum_{l=1}^{L} z_{il}\log\pi_l + \sum_{l=1}^{L}(d_l^0-1)\log\pi_l - \frac{JL}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{J}\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j - \frac{b^0}{\sigma^2} \\
&- (a^0+1)\log(\sigma^2) + constant
\end{aligned} \tag{12}
$$

- Variational approximate distribution of $\mathbf{Z}$:

$$
\begin{aligned}
\log q_{1i}(\mathbf{Z}_i) &\propto \mathbb{E}_{-q_{1i}(\mathbf{Z}_i)}[\log \tilde{P}(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \sigma^2 | \boldsymbol{\Delta}, \mathbf{d^0}, a^0, b^0)] \\
&\propto \mathbb{E}_{-q_{1i}(\mathbf{Z}_i)}[\log h(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\xi}) + \log P(\mathbf{Z}|\boldsymbol{\pi})] \\
&= \mathbb{E}_{q_2(\boldsymbol{\beta})q_3(\boldsymbol{\pi})}[\sum_{l=1}^{L} z_{il}(\log \pi_l + \sum_{j=1}^{J}(\log \sigma(\xi_{ijl}) + \frac{1}{2}((2y_{ij}-1)\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl} - \xi_{ijl}) - \lambda(\xi_{ijl})(\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl}\boldsymbol{\delta}_{jl}^T\boldsymbol{\beta}_j - \xi_{ijl}^2)))] \\
&= \sum_{l=1}^{L} z_{il}[\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log \pi_l] + \sum_{j=1}^{J}(\log \sigma(\xi_{ijl}) + \frac{1}{2}((2y_{ij}-1)\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j^T]\boldsymbol{\delta}_{jl} - \xi_{ijl}) \\
&- \lambda(\xi_{ijl})(\boldsymbol{\delta}_{jl}^T\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T]\boldsymbol{\delta}_{jl} - \xi_{ijl}^2))]
\end{aligned} \tag{13}
$$

The variational approximate distribution of $\mathbf{Z_i}$ follows a categorical distribution with parameter $\boldsymbol{\pi}_i^* = (\pi_{i1}^*, ..., \pi_{iL}^*)$, where $\pi_{il}^* = \frac{e^{\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log \pi_l] + \sum_{j=1}^{J}(\log \sigma(\xi_{ijl}) + \frac{1}{2}((2y_{ij}-1)\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j^T]\boldsymbol{\delta}_{jl} - \xi_{ijl}) - \lambda(\xi_{ijl})(\boldsymbol{\delta}_{jl}^T\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T]\boldsymbol{\delta}_{jl} - \xi_{ijl}^2))}}{\sum_{l=1}^{L} e^{\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log \pi_l] + \sum_{j=1}^{J}(\log \sigma(\xi_{ijl}) + \frac{1}{2}((2y_{ij}-1)\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j^T]\boldsymbol{\delta}_{jl} - \xi_{ijl}) - \lambda(\xi_{ijl})(\boldsymbol{\delta}_{jl}^T\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T]\boldsymbol{\delta}_{jl} - \xi_{ijl}^2))}}$.

- Variational approximate distribution of $\boldsymbol{\beta}$:

$$
\begin{aligned}
\log q_{2j}(\boldsymbol{\beta}_j) &\propto \mathbb{E}_{-q_{2j}(\boldsymbol{\beta}_j)}[\log \tilde{P}(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \sigma^2 | \boldsymbol{\Delta}, \mathbf{d^0}, a^0, b^0)] \\
&\propto \mathbb{E}_{-q_{2j}(\boldsymbol{\beta}_j)}[\log h(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\xi}) + \log P(\boldsymbol{\beta}|\sigma^2)] \\
&\propto \mathbb{E}_{q_1(\mathbf{Z})q_4(\sigma^2)}[\sum_{i=1}^{N}\sum_{l=1}^{L} z_{il}[(y_{ij} - \frac{1}{2})\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl} - \lambda(\xi_{ijl})\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl}\boldsymbol{\delta}_{jl}^T\boldsymbol{\beta}_j] - \frac{1}{2\sigma^2}\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j] \\
&= \mathbb{E}_{q_1(\mathbf{Z})q_4(\sigma^2)}[\boldsymbol{\beta}_j^T(\sum_{i=1}^{N}\sum_{l=1}^{L} z_{il}(y_{ij} - \frac{1}{2})\boldsymbol{\delta}_{jl}) - \frac{1}{2}\boldsymbol{\beta}_j^T(\frac{1}{\sigma^2}\boldsymbol{I}_{L\times L} + 2(\sum_{i=1}^{N}\sum_{l=1}^{L} z_{il}\lambda(\xi_{ijl})\boldsymbol{\delta}_{jl}\boldsymbol{\delta}_{jl}^T))\boldsymbol{\beta}_j] \\
&= \boldsymbol{\beta}_j^T(\sum_{i=1}^{N}\sum_{l=1}^{L}\mathbb{E}_{q_1(\mathbf{Z})}[z_{il}](y_{ij} - \frac{1}{2})\boldsymbol{\delta}_{jl}) - \frac{1}{2}\boldsymbol{\beta}_j^T(\mathbb{E}_{q_4(\sigma^2)}[\frac{1}{\sigma^2}]\boldsymbol{I}_{L\times L} + 2(\sum_{i=1}^{N}\sum_{l=1}^{L}\mathbb{E}_{q_1(\mathbf{Z})}[z_{il}]\lambda(\xi_{ijl})\boldsymbol{\delta}_{jl}\boldsymbol{\delta}_{jl}^T))\boldsymbol{\beta}_j
\end{aligned} \tag{14}
$$

The variational approximate distribution of $\boldsymbol{\beta}_j$ follows a Gaussian distribution $N(\boldsymbol{\mu}_j^*, \mathbf{V}_j^*)$, where $\boldsymbol{\mu}_j^* = \mathbf{V}_j^*(\sum_{i=1}^{N}\sum_{l=1}^{L}\mathbb{E}_{q_1(\mathbf{Z})}[z_{il}](y_{ij} - \frac{1}{2})\boldsymbol{\delta}_{jl}) = \mathbf{V}_j^*\boldsymbol{\delta}_j^T\mathbb{E}_{q_1(\mathbf{Z})}[\mathbf{Z}]^T(\mathbf{y_j} - \frac{1}{2}\times\mathbf{1}_N)$, and $\mathbf{V}_j^* = [\mathbb{E}_{q_4(\sigma^2)}[\frac{1}{\sigma^2}]\boldsymbol{I}_{L\times L} + 2(\sum_{i=1}^{N}\sum_{l=1}^{L}\mathbb{E}_{q_1(\mathbf{Z})}[z_{il}]\lambda(\xi_{ijl})\boldsymbol{\delta}_{jl}\boldsymbol{\delta}_{jl}^T)]^{-1}$.
Note that $\boldsymbol{\delta}_j \in \{0,1\}^{L\times L}$, $\mathbb{E}_{q_1(\mathbf{Z})}[\mathbf{Z}] \in \mathbb{R}^{N\times L}$, and $\mathbf{y_j} \in \mathbb{R}^N$.

- Variational approximate distribution of $\boldsymbol{\pi}$:

$$
\begin{aligned}
\log q_3(\boldsymbol{\pi}) &\propto \mathbb{E}_{-q_3(\boldsymbol{\pi})}[\log \tilde{P}(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \sigma^2 | \boldsymbol{\Delta}, \mathbf{d^0}, a^0, b^0)] \\
&\propto \mathbb{E}_{q_1(\mathbf{Z})q_2(\boldsymbol{\beta})q_4(\sigma^2)}[\log P(\boldsymbol{\pi}|\mathbf{d}^0) + \log P(\mathbf{Z}|\boldsymbol{\pi})] \\
&= \mathbb{E}_{q_1(\mathbf{Z})}[\sum_{i=1}^{N}\sum_{l=1}^{L} z_{il}\log \pi_l + \sum_{l=1}^{L}(d_l^0 - 1)\log \pi_l] \\
&= \mathbb{E}_{q_1(\mathbf{Z})}[\sum_{l=1}^{L}(\sum_{i=1}^{N} z_{il} + d_l^0 - 1)\log \pi_l] \\
&= \sum_{l=1}^{L}(\sum_{i=1}^{N}\mathbb{E}_{q_1(\mathbf{Z})}[z_{il}] + d_l^0 - 1)\log \pi_l
\end{aligned} \tag{15}
$$

The variational approximate distribution of $\boldsymbol{\pi}$ follows a Dirichlet distribution with parameter $\mathbf{d}^* = (d_1^*, ..., d_L^*)$, where $d_l^* = d_l^0 + \sum_{i=1}^N \mathbb{E}_{q_1(\mathbf{Z})}[z_{il}]$.

- Variational approximate distribution of σ^2:

$$
\begin{aligned}
\log q_4(\sigma^2) &\propto \mathbb{E}_{-q_4(\boldsymbol{\sigma^2})}[\log \tilde{P}(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \sigma^2 | \boldsymbol{\Delta}, \mathbf{d^0}, a^0, b^0)] \\
&\propto \mathbb{E}_{q_1(\mathbf{Z})q_2(\boldsymbol{\beta})q_3(\boldsymbol{\pi})}[\log P(\boldsymbol{\beta}|\sigma^2) + \log P(\sigma^2|a^0, b^0)] \\
&= \mathbb{E}_{q_2(\boldsymbol{\beta})}[-\frac{JL}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^J \boldsymbol{\beta}_j^T\boldsymbol{\beta}_j - \frac{b^0}{\sigma^2} - (a^0+1)\log(\sigma^2)] \\
&= \mathbb{E}_{q_2(\boldsymbol{\beta})}[-(\frac{JL}{2} + a^0 + 1)\log(\sigma^2) - \frac{1}{\sigma^2}(\frac{1}{2}\sum_{j=1}^J \boldsymbol{\beta}_j^T\boldsymbol{\beta}_j + b^0)] \\
&= -(\frac{1}{2}JL + a^0 + 1)\log(\sigma^2) - \frac{1}{\sigma^2}(\frac{1}{2}\sum_{j=1}^J \mathbb{E}_{q_2(\boldsymbol{\beta})}[\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j] + b^0)
\end{aligned}
\tag{16}
$$

The variational approximate distribution of $\boldsymbol{\sigma}^2$ follows an inverse-gamma distribution with parameters $a^* = a^0 + \frac{1}{2}JL$, and $b^* = b^0 + \frac{1}{2}\sum_{j=1}^J \mathbb{E}_{q_2(\boldsymbol{\beta})}[\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j]$.

To update parameters of variational approximate distributions, we need to calculate some expectations:

- $\mathbb{E}_{q_1(\mathbf{Z})}[z_{il}] = \pi_{il}^*$
- $\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j] = \boldsymbol{\mu}_j^*$
- $\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j] = \mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[tr(\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T)] = tr(\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T]) = tr(\mathbf{V}_j^*) + \boldsymbol{\mu}_j^{*T}\boldsymbol{\mu}_j^*$
- $\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T] = \mathbf{V}_j^* + \boldsymbol{\mu}_j^*\boldsymbol{\mu}_j^{*T}$
- $\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log \pi_l] = \psi(d_l^*) - \psi(\sum_{l=1}^L d_l^*)$, where $\psi(\cdot)$ represents the digamma function.
- $\mathbb{E}_{q_4(\sigma^2)}[\frac{1}{\sigma^2}] = a^*/b^*$

We can calculate the ELBO now.

$$
\begin{aligned}
\tilde{\mathcal{L}}(Q, \boldsymbol{\xi}) = &\int\int\int\sum_{\mathbf{Z}} Q(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \sigma^2) \log \frac{h(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\xi})P(\mathbf{Z}|\boldsymbol{\pi})P(\boldsymbol{\pi}|\mathbf{d}^0)P(\boldsymbol{\beta}|\sigma^2)P(\sigma^2|a^0, b^0)}{Q(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \sigma^2)} d\boldsymbol{\beta}d\boldsymbol{\pi}d\sigma^2 \\
&= \mathbb{E}_Q[\log h(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\xi})P(\mathbf{Z}|\boldsymbol{\pi})P(\boldsymbol{\pi}|\mathbf{d}^0)P(\boldsymbol{\beta}|\sigma^2)P(\sigma^2|a^0, b^0) - \log Q(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \sigma^2)] \\
&= \mathbb{E}_{q_1(\mathbf{Z})q_2(\boldsymbol{\beta})}[\log h(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\xi})] + \mathbb{E}_{q_1(\mathbf{Z})q_3(\boldsymbol{\pi})}[\log P(\mathbf{Z}|\boldsymbol{\pi})] + \mathbb{E}_{q_3(\boldsymbol{\pi})}[\log P(\boldsymbol{\pi}|\mathbf{d}^0)] + \mathbb{E}_{q_2(\boldsymbol{\beta})q_4(\sigma^2)}[\log P(\boldsymbol{\beta}|\sigma^2)] \\
&+ \mathbb{E}_{q_4(\sigma^2)}[\log P(\sigma^2|a^0, b^0)] - \mathbb{E}_{q_1(\mathbf{Z})}[\log q_1(\mathbf{Z})] - \mathbb{E}_{q_2(\boldsymbol{\beta})}[\log q_2(\boldsymbol{\beta})] - \mathbb{E}_{q_3(\boldsymbol{\pi})}[\log q_3(\boldsymbol{\pi})] - \mathbb{E}_{q_4(\sigma^2)}[\log q_4(\sigma^2)]
\end{aligned}
\tag{17}
$$

- Calculate $\mathbb{E}_{q_1(\mathbf{Z})q_2(\boldsymbol{\beta})}[\log h(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\xi})]$.

$$\begin{aligned}
\mathbb{E}_{q_1(\mathbf{Z})q_2(\boldsymbol{\beta})}[\log h(\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\xi})] &= \mathbb{E}_{q_1(\mathbf{Z})q_2(\boldsymbol{\beta})}[\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L} z_{il}[\log\sigma(\xi_{ijl}) + (y_{ij}-\frac{1}{2})\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl} - \frac{1}{2}\xi_{ijl} - \lambda(\xi_{ijl})(\boldsymbol{\beta}_j^T\boldsymbol{\delta}_{jl}\boldsymbol{\delta}_{jl}^T\boldsymbol{\beta}_j - \xi_{ijl}^2)]] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L}\mathbb{E}_{q_1(\mathbf{Z})}[z_{il}][\log\sigma(\xi_{ijl}) + (y_{ij}-\frac{1}{2})\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j]^T\boldsymbol{\delta}_{jl} - \frac{1}{2}\xi_{ijl} \\
&- \lambda(\xi_{ijl})(\boldsymbol{\delta}_{jl}^T\mathbb{E}_{q_{2j}(\boldsymbol{\beta}_j)}[\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T]\boldsymbol{\delta}_{jl} - \xi_{ijl}^2)] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L}\pi_{il}^*[\log\sigma(\xi_{ijl}) + (y_{ij}-\frac{1}{2})\boldsymbol{\mu}_j^{*T}\boldsymbol{\delta}_{jl} - \frac{1}{2}\xi_{ijl} - \lambda(\xi_{ijl})(\boldsymbol{\delta}_{jl}^T(\mathbf{V}_j^* + \boldsymbol{\mu}_j^*\boldsymbol{\mu}_j^{*T})\boldsymbol{\delta}_{jl} - \xi_{ijl}^2)]
\end{aligned} \tag{18}$$

- Calculate $\mathbb{E}_{q_1(\mathbf{Z})q_3(\boldsymbol{\pi})}[\log P(\mathbf{Z}|\boldsymbol{\pi})]$.

$$\begin{aligned}
\mathbb{E}_{q_1(\mathbf{Z})q_3(\boldsymbol{\pi})}[\log P(\mathbf{Z}|\boldsymbol{\pi})] &= \mathbb{E}_{q_1(\mathbf{Z})q_3(\boldsymbol{\pi})}[\sum_{i=1}^{N}\sum_{l=1}^{L} z_{il}\log\pi_l] \\
&= \sum_{i=1}^{N}\sum_{l=1}^{L}\mathbb{E}_{q_1(\mathbf{Z})}[z_{il}]\mathbb{E}_{q_3(\pi)}[\log\pi_l] \\
&= \sum_{i=1}^{N}\sum_{l=1}^{L}\pi_{il}^*[\psi(d_l^*) - \psi(\sum_{l=1}^{L} d_l^*)]
\end{aligned} \tag{19}$$

- Calculate $\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log P(\boldsymbol{\pi}|\mathbf{d}^0)]$.

$$\begin{aligned}
\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log P(\boldsymbol{\pi}|\mathbf{d}^0)] &= \mathbb{E}_{q_3(\boldsymbol{\pi})}[\sum_{l=1}^{L}(d_l^0-1)\log\pi_l - \log Beta(\mathbf{d}^0)] \\
&= \sum_{l=1}^{L}(d_l^0-1)\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log\pi_l] - \log Beta(\mathbf{d}^0) \\
&= \sum_{l=1}^{L}(d_l^0-1)[\psi(d_l^*) - \psi(\sum_{l=1}^{L} d_l^*)] - \log Beta(\mathbf{d}^0)
\end{aligned} \tag{20}$$

- Calculate $\mathbb{E}_{q_2(\boldsymbol{\beta})q_4(\sigma^2)}[\log P(\boldsymbol{\beta}|\sigma^2)]$.

$$\begin{aligned}
\mathbb{E}_{q_2(\boldsymbol{\beta})q_4(\sigma^2)}[\log P(\boldsymbol{\beta}|\sigma^2)] &= \mathbb{E}_{q_2(\boldsymbol{\beta})q_4(\sigma^2)}[-\frac{JL}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{J}\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j] \\
&= -\frac{JL}{2}\mathbb{E}_{q_4(\sigma^2)}[\log(2\pi\sigma^2)] - \frac{1}{2}\mathbb{E}_{q_4(\sigma^2)}[\frac{1}{\sigma^2}]\sum_{j=1}^{J}\mathbb{E}_{q_2(\boldsymbol{\beta})}[\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j] \\
&= \frac{JL}{2}[\psi(a^*) - \log b^* - \log 2\pi] - \frac{a^*}{2b^*}\sum_{j=1}^{J}[tr(\mathbf{V}_j^*) + \boldsymbol{\mu}_j^{*T}\boldsymbol{\mu}_j^*]
\end{aligned} \tag{21}$$

- Calculate $\mathbb{E}_{q_4(\sigma^2)}[\log P(\sigma^2|a^0,b^0)]$.

$$\begin{aligned}\mathbb{E}_{q_4(\sigma^2)}[\log P(\sigma^2|a^0,b^0)] &= \mathbb{E}_{q_4(\sigma^2)}[-\log(\Gamma(a^0)) + a^0\log b^0 - b^0/\sigma^2 - (a^0+1)\log(\sigma^2)] \\ &= -\log(\Gamma(a^0)) + a^0\log b^0 - b^0 \times \frac{a^*}{b^*} - (a^0+1)(\log(b^*) - \psi(a^*))\end{aligned} \tag{22}$$

- Calculate $\mathbb{E}_{q_1(\mathbf{Z})}[\log q_1(\mathbf{Z})]$.

$$\begin{aligned}\mathbb{E}_{q_1(\mathbf{Z})}[\log q_1(\mathbf{Z})] &= \mathbb{E}_{q_1(\mathbf{Z})}[\sum_{i=1}^{N}\log q_{1i(\mathbf{z}_i)}(\mathbf{z})] \\ &= \sum_{i=1}^{N}\mathbb{E}_{q_1(\mathbf{Z})}[\sum_{l=1}^{L} z_{il}\log(\pi^*_{il})] \\ &= \sum_{i=1}^{N}\sum_{l=1}^{L}\pi^*_{il}\log(\pi^*_{il})\end{aligned} \tag{23}$$

- Calculate $\mathbb{E}_{q_2(\boldsymbol{\beta})}[\log q_2(\boldsymbol{\beta})]$.

$$\begin{aligned}\mathbb{E}_{q_2(\boldsymbol{\beta})}[\log q_2(\boldsymbol{\beta})] &= \mathbb{E}_{q_2(\boldsymbol{\beta})}[\sum_{j=1}^{J}\log q_{2j(\boldsymbol{\beta}_j)}(\boldsymbol{\beta}_j)] \\ &= \sum_{j=1}^{J}\mathbb{E}_{q_2(\boldsymbol{\beta})}[\log q_{2j(\boldsymbol{\beta}_j)}(\boldsymbol{\beta}_j)] \\ &= -\frac{JL}{2}(1+\log 2\pi) - \sum_{j=1}^{J}\frac{1}{2}\log|\mathbf{V}^*_{\mathbf{j}}|\end{aligned} \tag{24}$$

- Calculate $\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log q_3(\boldsymbol{\pi})]$.

$$\begin{aligned}\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log q_3(\boldsymbol{\pi})] &= \mathbb{E}_{q_3(\boldsymbol{\pi})}[\log q_3(\boldsymbol{\pi})] \\ &= \mathbb{E}_{q_3(\boldsymbol{\pi})}[\frac{1}{Beta(\mathbf{d}^*)}\prod_{l=1}^{L}{\pi^*_l}^{d^*_l-1}] \\ &= -\log Beta(\mathbf{d}^*) + \sum_{l=1}^{L}(d^*_l-1)\mathbb{E}_{q_3(\boldsymbol{\pi})}[\log\pi^*_l] \\ &= -\log Beta(\mathbf{d}^*) + \sum_{l=1}^{L}(d^*_l-1)[\psi(d^*_l) - \psi(\sum_{l=1}^{L}d^*_l)]\end{aligned} \tag{25}$$

- Calculate $\mathbb{E}_{q_4(\sigma^2)}[\log q_4(\sigma^2)]$.

$$\begin{aligned}\mathbb{E}_{q_4(\sigma^2)}[\log q_4(\sigma^2)] &= \mathbb{E}_{q_4(\sigma^2)}[-\log\Gamma(a^*) + a^*\log b^* - (a^*+1)\log(\sigma^2) - b/\sigma^2] \\ &= -\log\Gamma(a^*) + a^*\log b^* - (a^*+1)[\log b^* - \psi(a^*)] - a^* \\ &= -\log\Gamma(a^*) + (a^*+1)\psi(a^*) - \log b^* - a^*\end{aligned} \tag{26}$$

We can then plug (18) - (26) into (17) to calculate ELBO:

$$
\begin{aligned}
ELBO &= \tilde{\mathcal{L}}(Q, \boldsymbol{\xi}) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L}\pi_{il}^{*}[\log\sigma(\xi_{ijl}) + (y_{ij} - \frac{1}{2})\boldsymbol{\mu}_j^{*T}\boldsymbol{\delta}_{jl} - \frac{1}{2}\xi_{ijl} - \lambda(\xi_{ijl})(\boldsymbol{\delta}_{jl}^{T}(\mathbf{V}_j^{*} + \boldsymbol{\mu}_j^{*}\boldsymbol{\mu}_j^{*T})\boldsymbol{\delta}_{jl} - \xi_{ijl}^{2})] \\
&+ \sum_{i=1}^{N}\sum_{l=1}^{L}\pi_{il}^{*}[\psi(d_l^{*}) - \psi(\sum_{l=1}^{L} d_l^{*})] + \sum_{l=1}^{L}(d_l^{0} - 1)[\psi(d_l^{*}) - \psi(\sum_{l=1}^{L} d_l^{*})] - \log Beta(\mathbf{d}^{0}) \\
&+ \frac{JL}{2}[\psi(a^{*}) - \log b^{*} - \log 2\pi] - \frac{a^{*}}{2b^{*}}\sum_{j=1}^{J}[tr(\mathbf{V}_j^{*}) + \boldsymbol{\mu}_j^{*T}\boldsymbol{\mu}_j^{*}] - \log(\Gamma(a^{0})) + a^{0}\log b^{0} - b^{0} \times \frac{a^{*}}{b^{*}} \\
&- (a^{0} + 1)(\log(b^{*}) - \psi(a^{*})) - \sum_{i=1}^{N}\sum_{l=1}^{L}\pi_{il}^{*}\log(\pi_{il}^{*}) + \frac{JL}{2}(1 + \log 2\pi) + \sum_{j=1}^{J}\frac{1}{2}\log|\mathbf{V}_{\mathbf{j}}^{*}| + \log Beta(\mathbf{d}^{*}) \\
&- \sum_{l=1}^{L}(d_l^{*} - 1)[\psi(d_l^{*}) - \psi(\sum_{l=1}^{L} d_l^{*})] + \log\Gamma(a^{*}) - (a^{*} + 1)\psi(a^{*}) + \log b^{*} + a^{*} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{L}\pi_{il}^{*}[\log\sigma(\xi_{ijl}) + (y_{ij} - \frac{1}{2})\boldsymbol{\mu}_j^{*T}\boldsymbol{\delta}_{jl} - \frac{1}{2}\xi_{ijl} - \lambda(\xi_{ijl})(\boldsymbol{\delta}_{jl}^{T}(\mathbf{V}_j^{*} + \boldsymbol{\mu}_j^{*}\boldsymbol{\mu}_j^{*T})\boldsymbol{\delta}_{jl} - \xi_{ijl}^{2})] \\
&+ \sum_{i=1}^{N}\sum_{l=1}^{L}\pi_{il}^{*}[\psi(d_l^{*}) - \psi(\sum_{l=1}^{L} d_l^{*}) - \log(\pi_{il}^{*})] + \sum_{l=1}^{L}(d_l^{0} - d_l^{*})[\psi(d_l^{*}) - \psi(\sum_{l=1}^{L} d_l^{*})] \\
&- \frac{a^{*}}{2b^{*}}\sum_{j=1}^{J}[tr(\mathbf{V}_j^{*}) + \boldsymbol{\mu}_j^{*T}\boldsymbol{\mu}_j^{*}] + \sum_{j=1}^{J}\frac{1}{2}\log|\mathbf{V}_{\mathbf{j}}^{*}| + \log Beta(\mathbf{d}^{*}) + \log\Gamma(a^{*}) + a^{*}(1 - \frac{b^{0}}{b^{*}}) - a^{*}\log(b^{*}) + constant
\end{aligned}
\tag{27}
$$