1 Notations

We would like to fix the notation before we proceed to the models. Any bold symbol stands for a vector. $[N]$ is defined as $\{1, 2, \ldots, N\}$ for all $N \in \mathbb{Z}^+$.

- There are $S \in \mathbb{Z}^+$ different schools in our model. Denote each school as $s \in [S]$.

- There are $N \in \mathbb{Z}^+$ different respondents in our model. Denote each respondent as $i \in [N]$ and the school that student i belongs to $s_i \in [S]$. We also denote the index set of all students in school s as I_s with $|I_s| = N_s \in \mathbb{Z}^+$. We have $\sum_{s\in[S]} N_s = N$.

- There are $J \in \mathbb{Z}^+$ questions in our model. Denote each question as $j \in [J]$.

- The response matrix is denoted as $Y \in \{0,1\}^{N\times J}$. Denote each element as $y_{ij} \in \{0, 1\}$, which indicates whether the i^{th} respondent answers the question j correctly. We consider Y as a random matrix and y_{ij} as the observed data of Y_{ij}.

- There are $K \in \mathbb{Z}^+$ possible attributes for each question. Denote each attribute as $k \in [K]$.

- Q-matrix $Q \in \{0, 1\}^{J\times K}$ establishes the relationship between questions and skills, such that an element $q_{jk} \in \{0, 1\}$ indicates whether skill k is required on the question j. For purpose of this paper, Q is known and held constant.

- We use $\mathbf{z}_i \in \{0, 1\}^K$ to denote the attribute profile of respondent i. We use $\mathcal{Z}$ to denote the collection of all attribute vectors $\mathbf{z}_i$. Notice that there is a natural bijection between attribute vector $\mathbf{z}_i$ and integer classes $c \in [2^K]$. Define vector $\mathbf{v} = [2^{K-1}, \ldots, 2^0]$, then we can see that $\mathbf{v}'\mathbf{z}_i + 1 \in [2^K]$ for any $\mathbf{z}_i$. With a slight abuse of notation, we use $\mathbf{v}^{-1}(c)$ to denote the corresponding attribute vector associated with integer class c. We would also like to use $\mathbf{v}_k^{-1}(c)$ to denote the k^{th} element of $\mathbf{v}^{-1}(c)$. For example, suppose $K = 3$, then $\mathbf{v}^{-1}(1) = [0, 0, 0]$ and $\mathbf{v}^{-1}(8) = [1, 1, 1]$. Also, we have that $\mathbf{v}_1^{-1}(1) = 0$ and $\mathbf{v}_1^{-1}(8) = 1$.

- For each question j, we define a matrix named $\Delta^{(j)} \in \{0, 1\}^{2^K\times 2^K}$. The c^{th} row vector of $\Delta^{(j)}$ represents the design vector for attribute profile $\mathbf{v}^{-1}(c)$. Denote the c^{th} row vector of $\Delta^{(j)}$ as $\boldsymbol{\delta}_c^{(j)} = [\delta_{cl}^{(j)}]_{l\in[2^K]}$, where $\delta_{cl}^{(j)} = 0$ if question j is not testing any of all attributes in $\mathbf{v}^{-1}(l)$ or $\mathbf{v}^{-1}(l)$ contains any attribute that $\mathbf{v}^{-1}(c)$ does not have. Otherwise, $\delta_{cl}^{(j)} = 1$. We should notice that $\Delta^{(j)}$ may contain columns that are 0 for all entries. In the following discussions, $\Delta^{(j)}$ does not contain those 0 columns. That is to say, $\Delta^{(j)} \in \{0, 1\}^{2^K\times 2^{\sum_{k=1}^K q_{jk}}}$. Let Δ denote the collection of all $\Delta^{(j)}$.

- $U \in \mathbb{R}^{S\times D}$ is the design matrix for school-level item-response effects. For example, the possible covariates can be average class size or that if the school is public or private. We use $\mathbf{u}_s \in \mathbb{R}^D$ to denote the design vector of school s.

- $X \in \mathbb{R}^{N\times M}$ is the design matrix for individual-level item-response effects. For example, the possible covariates can be gender or annual household income. We use $\mathbf{x}_i \in \mathbb{R}^m$ to denote the design vector of respondent i. The first column of X is a column of 1's, which is the intercept term.

- For question j, we have the corresponding item parameters $\boldsymbol{\beta}_j = [\beta_{jp}]_{p\in[P_j]} \in \mathbb{R}^{P_j}$, where $P_j = 2^{\sum_{k=1}^K q_{jk}}$.

- We denote the standard logistic function as $g(x) = \frac{1}{1+e^{-x}}$.

2 Model Formulation

There are two parts of our full model. The first part is the item-response model and the second part is the attribute profile. In a Bayesian way, we can interpret the multilevel item-response model as the following:

$$
\begin{aligned}
y_{ijt}|\mathbf{z}_{it} = \mathbf{v}^{-1}(l), \boldsymbol{\beta}_j &\sim \text{Bernoulli}(g(\boldsymbol{\beta}_j' \boldsymbol{\delta}_l^{(j)})) && \text{conditionally independent for all } i, j, t \\
\boldsymbol{\beta}_j &\sim \mathcal{N}(\boldsymbol{\mu}_j^{(\beta)}, \Sigma_j^{(\beta)}) && \text{independent for all } j \\
z_{itk}|z_{i(t-1)k} = z, \boldsymbol{\gamma}_{(s_i)tk}^{(z)} &\sim \text{Bernoulli}(g(\boldsymbol{\gamma}_{(s_i)tk}^{(z)}{}' \mathbf{x}_{itk})) && \text{conditionally independent for all } z \in \{0,1\}, t \geq 2, i, k \\
z_{i1k} &\sim \text{Bernoulli}(g(\boldsymbol{\gamma}_{(s_i)1k}{}' \mathbf{x}_{i1k})) && \text{conditionally independent for all } t = 1, i, k \\
\boldsymbol{\gamma}_{(s)tk}^{(z)}|\Omega_{tk}^{(z)}, \boldsymbol{\tau}_{tk}^{(z)} &\sim \mathcal{N}(\Omega_{tk}^{(z)}{}' \mathbf{u}_{tk}^{(z)}, \text{diag}(\boldsymbol{\tau}_{tk}^{(z)})) && \text{conditionally independent for all } z \in \{0,1\}, t, k \\
\boldsymbol{\omega}_{tkm}^{(z)} &\sim \mathcal{N}(\boldsymbol{\mu}_{tkm}^{(\omega^z)}, \Sigma_{tkm}^{(\omega^z)}) && \text{independent for all } z \in \{0,1\}, m \in [M_{tk}], t, k \\
\tau_{tkm}^{(z)} &\sim \text{IG}(a_{tkm}^{(\tau^z)}, b_{tkm}^{(\tau^z)}) && \text{independent for all } z \in \{0,1\}, m \in [M_{tk}], t, k
\end{aligned}
$$

Denote $\boldsymbol{\tau}_{tk}^{(1)} = [\tau_{1tk}^{(1)}, \ldots, \tau_{tkm}^{(1)}]$ and $\boldsymbol{\tau}_{tk}^{(0)} = [\tau_{1tk}^{0}, \ldots, \tau_{tkm}^{(0)}]$. Denote $Z = \{\boldsymbol{z}_i : i \in [N]\}$ as the collection of all attribute profiles of N students. Denote $\boldsymbol{\beta} = \{\boldsymbol{\beta}_j : j \in [J]\}$ as the collection of $\boldsymbol{\beta}_j$. $\gamma \in \mathbb{R}^{S \times M}$ is the collection of $\boldsymbol{\gamma}_{(s)}$, with each row vector as $\boldsymbol{\gamma}_{(s)} \in \mathbb{R}^M$. $\Omega \in \mathbb{R}^{D \times M}$ is the collection of $\boldsymbol{\omega}_m$, with each column vector as $\boldsymbol{\omega}_m \in \mathbb{R}^D$.

2.1 Mean Field Approximation

2.2 Likelihood Function

We can then get the joint likelihood function

$$
\begin{aligned}
& p(Y, Z, \boldsymbol{\beta}, \Gamma, \Omega, \boldsymbol{\tau}) \\
&= p(Y|Z, \boldsymbol{\beta}) p(\boldsymbol{\beta}) p(Z|\Gamma) p(\Gamma|\Omega, \boldsymbol{\tau}) p(\Omega) p(\boldsymbol{\tau}) \qquad (1) \\
&= \left(\prod_{i=1}^N \prod_{t=1}^T \prod_{j=1}^J \prod_{l=1}^{2^K} p(y_{ijt}|\mathbf{z}_{it} = \mathbf{v}^{-1}(l), \boldsymbol{\beta}_j) \right) \left(\prod_{j=1}^J p(\boldsymbol{\beta}_j) \right) \left(\prod_{i=1}^N \prod_{k=1}^K \left(\left(\prod_{t=2}^T p(z_{itk}|z_{i(t-1)k} = z, \boldsymbol{\gamma}_{(s_i)tk}^{(z)}) \right) p(z_{i1k}|\boldsymbol{\gamma}_{(s_i)1k}) \right) \right) \times \\
& \left(\prod_{k=1}^K \prod_{t=1}^T \prod_{z=0}^1 p(\Gamma_{tk}^{(z)}|\Omega_{tk}^{(z)}, \boldsymbol{\tau}_{tk}^{(z)}) \right) \left(\prod_{k=1}^K \prod_{t=1}^T \prod_{z=0}^1 p(\Omega_{tk}^{(z)}) \right) \left(\prod_{k=1}^K \prod_{t=1}^T \prod_{z=0}^1 \prod_{m=1}^{M_{tk}} p(\tau_{tkm}^{(z)}) \right) \qquad (2) \\
&\propto \left(\prod_{i=1}^N \prod_{t=1}^T \prod_{j=1}^J \prod_{l=1}^{2^K} g\left((2y_{ijt} - 1) \boldsymbol{\beta}_j' \boldsymbol{\delta}_l^{(j)} \right)^{\mathbb{1}(\boldsymbol{z}_{it} = \mathbf{v}^{-1}(l))} \right) \left(\prod_{j=1}^J \exp\left(-\frac{1}{2} (\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta)})' \Sigma_j^{(\beta)^{-1}} (\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta)}) \right) \right) \\
& \left(\prod_{i=1}^N \prod_{k=1}^K \left(\left(\prod_{t=2}^T \prod_{z=0}^1 g\left((2z_{itk} - 1) \boldsymbol{\gamma}_{(s_i)tk}^{(z)}{}' \mathbf{x}_{itk} \right)^{\mathbb{1}(\boldsymbol{z}_{i(t-1)k} = z)} \right) g\left((2z_{i1k} - 1) \boldsymbol{\gamma}_{(s_i)1k}{}' \mathbf{x}_{i1k} \right) \right) \right) \times \\
& \left(\prod_{k=1}^K \prod_{t=1}^T \prod_{s=1}^S \prod_{z=0}^1 \prod_{m=1}^{M_{tk}} \tau_{tkm}^{(z)}{}^{-\frac{1}{2}} \exp\left(-\frac{1}{2\tau_{tkm}^{(z)}} (\gamma_{(s)tkm}^{(z)} - \mathbf{u}_{(s)tk}^{(z)}{}' \boldsymbol{\omega}_{tkm}^{(z)})^2 \right) \right) \times \\
& \left(\prod_{k=1}^K \prod_{t=1}^T \prod_{m=1}^{M_{tk}} \prod_{z=0}^1 \exp\left(-\frac{1}{2} \left(\boldsymbol{\omega}_{tkm}^{(z)} - \boldsymbol{\mu}_{tkm}^{(\omega^z)} \right)' \Sigma_{tkm}^{(\omega^z)^{-1}} \left(\boldsymbol{\omega}_{tkm}^{(z)} - \boldsymbol{\mu}_{tkm}^{(\omega^z)} \right) \right) \right) \times
\end{aligned}
$$

$$\left(\prod_{k=1}^{K}\prod_{t=1}^{T}\prod_{z=0}^{1}\left(\prod_{m=1}^{M_{tk}}(\tau_{tkm}^{(z)})^{-a_{tkm}^{(\gamma^z)}-1}\exp\left(-\frac{b_{tkm}^{(\tau^z)}}{\tau_{tkm}^{(z)}}\right)\right)\right) \tag{3}$$

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 inequality:

$$\log g(x) \geq \log \sigma(\xi) + (x-\xi)/2 - h(\xi)(x^2-\xi^2), \tag{4}$$

where $h(\xi) = \frac{\sigma(\xi)-\frac{1}{2}}{2\xi}$.

Take the logarithm of the joint likelihood and apply Jaakkola and Jordan's lower bound

$$\begin{aligned}
&\log p(Y, Z, \boldsymbol{\beta}, \Gamma, \Omega, \Sigma^{(\Gamma)}, \boldsymbol{\tau}) \\
=&\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{l=1}^{2^K}\mathbb{1}(\boldsymbol{z}_{it}=\mathbf{v}^{-1}(l))\log g\left((2y_{ijt}-1)\boldsymbol{\beta}_j'\boldsymbol{\delta}_l^{(j)}\right)+\sum_{j=1}^{J}-\frac{1}{2}(\boldsymbol{\beta}_j-\boldsymbol{\mu}_j^{(\beta)})'{\Sigma_j^{(\beta)}}^{-1}(\boldsymbol{\beta}_j-\boldsymbol{\mu}_j^{(\beta)}) \\
&+\sum_{i=1}^{N}\sum_{k=1}^{K}\left(\sum_{t=2}^{T}\sum_{z=0}^{1}\mathbb{1}(\boldsymbol{z}_{i(t-1)k}=z)\log g\left((2z_{itk}-1){\boldsymbol{\gamma}_{(s_i)tk}^{(z)}}'\mathbf{x}_{itk}\right)+\log g\left((2z_{i1k}-1){\boldsymbol{\gamma}_{(s_i)1k}}'\mathbf{x}_{i1k}\right)\right) \\
&+\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{z=0}^{1}\sum_{m=1}^{M_{tk}}\left(-\frac{1}{2}\log\tau_{tkm}^{(z)}-\frac{1}{2\tau_{tkm}^{(z)}}(\gamma_{(s)tkm}^{(z)}-{\mathbf{u}_{(s)tk}^{(z)}}'\boldsymbol{\omega}_{tkm}^{(z)})^2\right) \\
&+\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{m=1}^{M_{tk}}\sum_{z=0}^{1}\left(-\frac{1}{2}\left(\boldsymbol{\omega}_{tkm}^{(z)}-\boldsymbol{\mu}_{tkm}^{(\omega^z)}\right)'{\Sigma_{tkm}^{(\omega^z)}}^{-1}\left(\boldsymbol{\omega}_{tkm}^{(z)}-\boldsymbol{\mu}_{tkm}^{(\omega^z)}\right)\right)+\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{z=0}^{1}\sum_{m=1}^{M_{tk}}\left(\left(-a_{tkm}^{(\gamma^z)}-1\right)\log\tau_{tkm}^{(z)}-\frac{b_{tkm}^{(\tau^z)}}{\tau_{tkm}^{(z)}}\right) \\
\geq&\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{l=1}^{2^K}\mathbb{1}(\boldsymbol{z}_{it}=\mathbf{v}^{-1}(l))\left(\log\sigma(\xi_{jl})+(y_{ijt}-\frac{1}{2})\boldsymbol{\beta}_j'\boldsymbol{\delta}_l^{(j)}-\frac{\xi_{jl}}{2}-h(\xi_{jl})\left((\boldsymbol{\beta}_j'\boldsymbol{\delta}_l^{(j)})^2-\xi_{jl}^2\right)\right) \\
&+\sum_{i=1}^{N}\sum_{k=1}^{K}\left(\sum_{t=2}^{T}\sum_{z=0}^{1}\mathbb{1}(\boldsymbol{z}_{i(t-1)k}=z)\left(\log\sigma(\eta_{itk}^{(z)})+(z_{itk}-\frac{1}{2}){\boldsymbol{\gamma}_{(s_i)tk}^{(z)}}'\mathbf{x}_{itk}-\frac{\eta_{itk}^{(z)}}{2}-h(\eta_{itk}^{(z)})\left(({\boldsymbol{\gamma}_{(s_i)tk}^{(z)}}'\mathbf{x}_{itk})^2-{\eta_{itk}^{(z)}}^2\right)\right)\right) \\
&+\sum_{i=1}^{N}\sum_{k=1}^{K}\left(\log\sigma(\eta_{i1k})+(z_{i1k}-\frac{1}{2}){\boldsymbol{\gamma}_{(s_i)1k}}'\mathbf{x}_{i1k}-\frac{\eta_{i1k}}{2}-h(\eta_{i1k})\left(({\boldsymbol{\gamma}_{(s_i)1k}}'\mathbf{x}_{i1k})^2-\eta_{i1k}^2\right)\right) \\
&+\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{z=0}^{1}\sum_{m=1}^{M_{tk}}\left(-\frac{1}{2}\log\tau_{tkm}^{(z)}-\frac{1}{2\tau_{tkm}^{(z)}}(\gamma_{(s)tkm}^{(z)}-{\mathbf{u}_{(s)tk}^{(z)}}'\boldsymbol{\omega}_{tkm}^{(z)})^2\right) \\
&+\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{m=1}^{M_{tk}}\sum_{z=0}^{1}\left(-\frac{1}{2}\left(\boldsymbol{\omega}_{tkm}^{(z)}-\boldsymbol{\mu}_{tkm}^{(\omega^z)}\right)'{\Sigma_{tkm}^{(\omega^z)}}^{-1}\left(\boldsymbol{\omega}_{tkm}^{(z)}-\boldsymbol{\mu}_{tkm}^{(\omega^z)}\right)\right)+\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{z=0}^{1}\sum_{m=1}^{M_{tk}}\left(\left(-a_{tkm}^{(\gamma^z)}-1\right)\log\tau_{tkm}^{(z)}-\frac{b_{tkm}^{(\tau^z)}}{\tau_{tkm}^{(z)}}\right)
\end{aligned} \tag{5}$$

3 Optimization and Estimation

3.1 CAVI

3.2 Our Model

We have that $q(\boldsymbol{\beta}_j) \sim \mathcal{N}(\boldsymbol{\mu}_j^{(\beta*)}, \Sigma_j^{(\beta*)})$,

$$\text{with} \quad \Sigma_j^{(\beta*)} = \left({\Sigma_j^{(\beta)}}^{-1} + 2 \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{l=1}^{2^K} \mathbb{E}_q[\mathbb{1}(\boldsymbol{z}_{it} = \mathbf{v}^{-1}(l))] h(\xi_{jl}) \boldsymbol{\delta}_l^{(j)} {\boldsymbol{\delta}_l^{(j)}}' \right)^{-1}, \tag{6}$$

$$\text{and} \quad \boldsymbol{\mu}_j^{(\beta*)} = \Sigma_j^{(\beta*)} \left({\Sigma_j^{(\beta)}}^{-1} \boldsymbol{\mu}_j^{(\beta)} + \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{l=1}^{2^K} \mathbb{E}_q[\mathbb{1}(\boldsymbol{z}_{it} = \mathbf{v}^{-1}(l))] \left(y_{ijt} - \frac{1}{2} \right) \boldsymbol{\delta}_l^{(j)} \right). \tag{7}$$

We have that $q(\boldsymbol{\gamma}_{(s)1k}) \sim \mathcal{N}(\boldsymbol{\mu}_{(s)1k}^{(\gamma*)}, \Sigma_{(s)1k}^{(\gamma*)})$,

$$\text{with} \quad \Sigma_{(s)1k}^{(\gamma*)} = \left(\operatorname{diag}\left(\mathbb{E}[{\boldsymbol{\tau}_{1k}}^{-1}]\right) + 2 \sum_{i \in I_s} h(\eta_{i1k}) \right)^{-1}, \tag{8}$$

$$\text{and} \quad \boldsymbol{\mu}_{(s)1k}^{(\gamma*)} = \Sigma_{(s)1k}^{(\gamma*)} \left(\operatorname{diag}\left(\mathbb{E}[{\boldsymbol{\tau}_{1k}}^{-1}]\right) \mathbb{E}[\Omega_{1k}]' \mathbf{u}_{(s)1k} + \sum_{i \in I_s} \mathbb{E}_q \left[z_{i1k} - \frac{1}{2} \right] \right). \tag{9}$$

We have that $q(\boldsymbol{\gamma}_{(s)tk}^{(z)}) \sim \mathcal{N}(\boldsymbol{\mu}_{(s)tk}^{(\gamma^{(z)}*)}, \Sigma_{(s)tk}^{(\gamma^{(z)}*)})$,

$$\text{with} \quad \Sigma_{(s)tk}^{(\gamma^{(z)}*)} = \left(\operatorname{diag}\left(\mathbb{E}[{\boldsymbol{\tau}_{tk}^{(z)}}^{-1}] \right) + 2 \sum_{i \in I_s} \mathbb{E}[\mathbb{1}(\boldsymbol{z}_{i(t-1)k} = z)] h(\eta_{itk}^{(z)}) \mathbf{x}_{itk} \mathbf{x}_{itk}' \right)^{-1}, \tag{10}$$

$$\text{and} \quad \boldsymbol{\mu}_{(s)tk}^{(\gamma^{(z)}*)} = \Sigma_{(s)tk}^{(\gamma^{(z)}*)} \left(\operatorname{diag}\left(\mathbb{E}[{\boldsymbol{\tau}_{tk}^{(z)}}^{-1}] \right) \mathbb{E}[\Omega_{tk}^{(z)}]' \mathbf{u}_{(s)tk} + \sum_{i \in I_s} \mathbb{E}_q \left[\mathbb{1}(\boldsymbol{z}_{i(t-1)k} = z) \left(z_{itk} - \frac{1}{2} \right) \right] \mathbf{x}_{itk} \right). \tag{11}$$

We have that $q(\boldsymbol{\omega}_{tkm}^{(z)}) \sim N(\boldsymbol{\mu}_{\text{tkm}}^{(\omega^{(z)}*)}, \Sigma_{tkm}^{(\omega^{(z)}*)})$,

$$\text{with} \quad \Sigma_{tkm}^{(\omega^{(z)}*)} = \left({\Sigma_{tkm}^{(\omega^{(z)})}}^{-1} + \mathbb{E}[{\tau_{tkm}^{(z)}}^{-1}] U_{tk}' U_{tk} \right)^{-1}, \tag{12}$$

$$\text{and} \quad \boldsymbol{\mu}_{\text{tkm}}^{(\omega^{(z)}*)} = \Sigma_{tkm}^{(\omega^{(z)}*)} \left[{\Sigma_{tkm}^{(\omega^{(z)})}}^{-1} \boldsymbol{\mu}_{tkm}^{(\omega^{(z)})} + \mathbb{E}[{\tau_{tkm}^{(z)}}^{-1}] U_{tk}' \mathbb{E}[\boldsymbol{\gamma}_{tkm}^{(z)}] \right]. \tag{13}$$

We have that $\tau_{tkm}^{(z)} \sim \text{Inverse-Gamma}(a_{tkm}^{(\tau^{(z)}*)}, b_{tkm}^{(\tau^{(z)}*)})$,

$$\text{with} \quad a_{tkm}^{(\tau^{(z)}*)} = a_{tkm}^{(\tau^{(z)})} + \frac{S}{2}, \tag{14}$$

$$\text{and} \quad b_{tkm}^{(\tau^{(z)}*)} = b_{tkm}^{(\tau^{(z)})} + \frac{1}{2} \sum_{s=1}^{S} \mathbb{E}[(\gamma_{(s)tkm}^{(z)} - {\mathbf{u}_{(s)tk}^{(z)}}' \boldsymbol{\omega}_{tkm}^{(z)})^2]. \tag{15}$$

We have that $q(\mathbf{z}_i) \sim \text{Categorical}(\boldsymbol{\pi}_i^*)$,

$$\text{with} \quad \pi_{il}^* = \frac{\exp\left(\mathbb{E}_q[\log \pi_l] + \sum_{j=1}^{J} \phi_{ijl}\right)}{\sum_{l=1}^{2^K} \exp\left(\mathbb{E}_q[\log \pi_l] + \sum_{j=1}^{J} \phi_{ijl}\right)}, \tag{16}$$

$$\text{and} \quad \phi_{ijl} = \log g(\xi_{ijl}) + \frac{(2y_{ij}-1)\mathbb{E}_q\left[\left(\boldsymbol{\gamma}'_{(s_i)}\mathbf{x}_i + \boldsymbol{\beta}'_j\boldsymbol{\delta}_l^{(j)}\right)\right] - \xi_{ijl}}{2} - h(\xi_{ijl})(\mathbb{E}_q\left[\left(\boldsymbol{\gamma}'_{(s_i)}\mathbf{x}_i + \boldsymbol{\beta}'_j\boldsymbol{\delta}_l^{(j)}\right)^2\right] - \xi_{ijl}^2). \tag{17}$$

3.3 ELBO

The ELBO is $\mathbb{E}_q[\log p(x,z) - \log q(z)]$. In our case, ELBO is the following

$$
\begin{aligned}
ELBO &= \mathbb{E}_q[\log p(Y, Z, \boldsymbol{\beta}, \gamma, \Omega, \Sigma^{(\gamma)}, \boldsymbol{\pi}) - \log q(Z, \boldsymbol{\beta}, \gamma, \Omega, \Sigma^{(\gamma)}, \boldsymbol{\pi})] && (18)\\
&\geq \mathbb{E}_q[\log p(Y, Z, \boldsymbol{\beta}, \gamma, \Omega, \Sigma^{(\gamma)}, \boldsymbol{\pi}, \Xi) - \log q(Z, \boldsymbol{\beta}, \gamma, \Omega, \Sigma^{(\gamma)}, \boldsymbol{\pi})] && (19)\\
&= \mathbb{E}_q[\log p(Y|Z, \boldsymbol{\beta}, \gamma, \Xi) + \log p(\boldsymbol{\beta}) + \log p(\gamma|\Omega, \Sigma^{(\gamma)}) + \log p(\Sigma^{(\gamma)}) + \log p(\Omega) + \log p(Z|\boldsymbol{\pi}) + \log p(\boldsymbol{\pi})]\\
&- \mathbb{E}_q[\log q(Z) + \log q(\boldsymbol{\beta}) + \log q(\gamma) + \log q(\Omega) + \log q(\Sigma^{(\gamma)}) + \log q(\boldsymbol{\pi})] && (20)\\
&= \mathbb{E}_q\Bigg[\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{2^K} \mathbb{1}(\boldsymbol{z}_i = \mathbf{v}^{-1}(l))\left(\log g(\xi_{ijl}) + \frac{(2y_{ij}-1)\left(\boldsymbol{\gamma}'_{(s_i)}\mathbf{x}_i + \boldsymbol{\beta}'_j\boldsymbol{\delta}_l^{(j)}\right) - \xi_{ijl}}{2} - h(\xi_{ijl})(\left(\boldsymbol{\gamma}'_{(s_i)}\mathbf{x}_i + \boldsymbol{\beta}'_j\boldsymbol{\delta}_l^{(j)}\right)^2 - \xi_{ijl}^2)\right)\\
&+ \sum_{j=1}^{J} -\frac{1}{2}(\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta)})'{\Sigma_j^{(\beta)}}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta)}) + \left(\sum_{s=1}^{S} -\frac{1}{2}\log\det(\Sigma^{(\gamma)}) - \frac{1}{2}(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)'{\Sigma^{(\gamma)}}^{-1}(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)\right)\\
&- \frac{\nu^{(\gamma)} + M + 1}{2}\log\det\Sigma^{(\gamma)} - \frac{1}{2}\operatorname{tr}\left(\Psi^{(\gamma)}{\Sigma^{(\gamma)}}^{-1}\right) + \sum_{m=1}^{M} -\frac{1}{2}(\boldsymbol{\omega}_m - \boldsymbol{\mu}_m^{(\omega)})'{\Sigma_m^{(\omega)}}^{-1}(\boldsymbol{\omega}_m - \boldsymbol{\mu}_m^{(\omega)})\\
&+ \sum_{i=1}^{N}\sum_{l=1}^{2^K} \mathbb{1}(\boldsymbol{z}_i = \mathbf{v}^{-1}(l))\log\pi_l + \sum_{l=1}^{2^K}(d_l - 1)\log\pi_l - \sum_{i=1}^{N}\sum_{l=1}^{2^K} \mathbb{1}(\boldsymbol{z}_i = \mathbf{v}^{-1}(l))\log\pi_{il}^* - \sum_{l=1}^{2^K}(d_l^* - 1)\log\pi_l\\
&- \sum_{j=1}^{J}\left(-\frac{1}{2}\log\det\Sigma_j^{(\beta*)} - \frac{1}{2}(\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta*)})'{\Sigma_j^{(\beta*)}}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta*)})\right) + \frac{\nu^{(\gamma*)} + M + 1}{2}\log\det\Sigma^{(\gamma)} + \frac{1}{2}\operatorname{tr}\left(\Psi^{(\gamma*)}{\Sigma^{(\gamma)}}^{-1}\right)\\
&- \frac{\nu^{(\gamma*)}}{2}\log\det\Psi^{(\gamma*)} - \sum_{s=1}^{S}\left(-\frac{1}{2}\log\det(\Sigma^{(\gamma*)}) - \frac{1}{2}(\boldsymbol{\gamma}_{(s)} - \boldsymbol{\mu}_s^{(\gamma*)})'{\Sigma^{(\gamma*)}}^{-1}(\boldsymbol{\gamma}_{(s)} - \boldsymbol{\mu}_s^{(\gamma*)})\right) + \log\mathcal{B}(d^*)\\
&- \sum_{m=1}^{M}\left(-\frac{1}{2}\det(\Sigma_m^{\omega*}) - \frac{1}{2}(\boldsymbol{\omega}_m - \boldsymbol{\mu}_m^{(\omega*)})'{\Sigma_m^{(\omega*)}}^{-1}(\boldsymbol{\omega}_m - \boldsymbol{\mu}_m^{(\omega*)})\right)\Bigg] + \text{constant} && (21)\\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{l=1}^{2^K}\pi_{il}^*\left(\log g(\xi_{ijl}) + \frac{(2y_{ij}-1)\left({\boldsymbol{\mu}_{(s_i)}^{(\gamma*)}}'\mathbf{x}_i + {\boldsymbol{\mu}_j^{(\beta*)}}'\boldsymbol{\delta}_l^{(j)}\right) - \xi_{ijl}}{2} - h(\xi_{ijl})(\mathbb{E}_q[\left(\boldsymbol{\gamma}'_{(s_i)}\mathbf{x}_i + \boldsymbol{\beta}'_j\boldsymbol{\delta}_l^{(j)}\right)^2] - \xi_{ijl}^2)\right)\\
&- \frac{\nu^{(\gamma*)}}{2}\log\det\Psi^{(\gamma*)} + \frac{\nu^{(\gamma*)} - \nu^{(\gamma)} - S}{2}\mathbb{E}_q[\log\det\Sigma^{(\gamma)}] + \frac{1}{2}\mathbb{E}_q[\operatorname{tr}\left((\Psi^{(\gamma*)} - \Psi^{(\gamma)}){\Sigma^{(\gamma)}}^{-1}\right)] + \sum_{l=1}^{2^K}(d_l - d_l^*)\mathbb{E}_q[\log\pi_l]\\
&+ \sum_{m=1}^{M}\frac{1}{2}\left(\log\det\Sigma_m^{(\omega*)} - \operatorname{tr}\left({\Sigma_m^{(\omega)}}^{-1}\mathbb{E}_q[(\boldsymbol{\omega}_m - \boldsymbol{\mu}_m^{(\omega)})(\boldsymbol{\omega}_m - \boldsymbol{\mu}_m^{(\omega)})']\right)\right) + \sum_{i=1}^{N}\sum_{l=1}^{2^K}\pi_{il}^*(\mathbb{E}_q[\log\pi_l] - \log\pi_{il}^*)\\
&+ \sum_{j=1}^{J}\frac{1}{2}\left(\log\det\Sigma_j^{(\beta*)} - \operatorname{tr}\left({\Sigma_j^{(\beta)}}^{-1}\mathbb{E}_q[(\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta)})(\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta)})']\right)\right) + \log\mathcal{B}(d^*)\\
&+ \sum_{s=1}^{S}\frac{1}{2}\left(\log\det\Sigma_s^{(\gamma*)} - \mathbb{E}_q[(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)'{\Sigma^{(\gamma)}}^{-1}(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)]\right) + \text{constant} && (22)
\end{aligned}
$$

To compute the ELBO, we need to compute the following expectations:

$$\mathbb{E}_q[\log \pi_l] = \psi(d_l^*) - \psi(\sum_{l=1}^{2^K} d_l^*) \tag{23}$$

$$\begin{aligned}
\mathbb{E}_q[(\boldsymbol{\omega}_m - \boldsymbol{\mu}_m^{(\omega)})(\boldsymbol{\omega}_m - \boldsymbol{\mu}_m^{(\omega)})'] &= \mathbb{E}_q[\boldsymbol{\omega}_m \boldsymbol{\omega}_m'] - \mathbb{E}_q[\boldsymbol{\omega}_m \boldsymbol{\mu}_m^{(\omega)'}] - \mathbb{E}_q[\boldsymbol{\mu}_m^{(\omega)} \boldsymbol{\omega}_m'] + \boldsymbol{\mu}_m^{(\omega)} \boldsymbol{\mu}_m^{(\omega)'} && (24)\\
&= \Sigma_m^{(\omega*)} + \boldsymbol{\mu}_m^{(\omega*)} \boldsymbol{\mu}_m^{(\omega*)'} - \boldsymbol{\mu}_m^{(\omega*)} \boldsymbol{\mu}_m^{(\omega)'} - \boldsymbol{\mu}_m^{(\omega)} \boldsymbol{\mu}_m^{(\omega*)'} + \boldsymbol{\mu}_m^{(\omega)} \boldsymbol{\mu}_m^{(\omega)'} && (25)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_q[(\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta)})(\boldsymbol{\beta}_j - \boldsymbol{\mu}_j^{(\beta)})'] &= \mathbb{E}_q[\boldsymbol{\beta}_j \boldsymbol{\beta}_j'] - \mathbb{E}_q[\boldsymbol{\beta}_j \boldsymbol{\mu}_j^{(\beta)'}] - \mathbb{E}_q[\boldsymbol{\mu}_j^{(\beta)} \boldsymbol{\beta}_j'] + \boldsymbol{\mu}_j^{(\beta)} \boldsymbol{\mu}_j^{(\beta)'} && (26)\\
&= \Sigma_j^{(\beta*)} + \boldsymbol{\mu}_j^{(\beta*)} \boldsymbol{\mu}_j^{(\beta*)'} - \boldsymbol{\mu}_j^{(\beta*)} \boldsymbol{\mu}_j^{(\beta)'} - \boldsymbol{\mu}_j^{(\beta)} \boldsymbol{\mu}_j^{(\beta*)'} + \boldsymbol{\mu}_j^{(\beta)} \boldsymbol{\mu}_j^{(\beta)'} && (27)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_q[\log\det \Sigma^{(\gamma)}] &= -\mathbb{E}_q[\log\det {\Sigma^{(\gamma)}}^{-1}] && (28)\\
&= -\psi_M(\frac{\nu^*}{2}) - M\log 2 - \log\det {\Psi^{(\gamma*)}}^{-1} && (29)\\
&= -\psi_M(\frac{\nu^*}{2}) - M\log 2 + \log\det \Psi^{(\gamma*)} && (30)\\
\mathbb{E}_q[{\Sigma^{(\gamma)}}^{-1}] &= M\Psi^{(\gamma*)-1} && (31)
\end{aligned}$$

We can see that $q(\Omega)$ follows a matrix normal distribution $\mathcal{MN}_{M,D}(M^{(\Omega*)}, U^{(\Omega*)}, V^{(\Omega*)})$ with parameters

$$\begin{aligned}
M^{(\Omega*)} &= [\boldsymbol{\mu}_1^{(\omega*)'}, \ldots, \boldsymbol{\mu}_M^{(\omega*)'}] && (32)\\
U^{(\Omega*)} &= \text{diag}[\text{tr}(\Sigma_1^{(\omega*)}), \ldots, \text{tr}(\Sigma_M^{(\omega*)})]/(\sum_{m=1}^{M} \text{tr}(\Sigma_m^{(\omega*)})) && (33)\\
V^{(\Omega*)} &= \sum_{m=1}^{M} \Sigma_m^{(\omega*)} && (34)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_q[(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)'] && (35)\\
&= \mathbb{E}_q[\boldsymbol{\gamma}_{(s)}\boldsymbol{\gamma}_{(s)}'] - \mathbb{E}_q[\boldsymbol{\gamma}_{(s)}\mathbf{u}_s'\Omega'] - \mathbb{E}_q[\Omega\mathbf{u}_s]\boldsymbol{\gamma}_{(s)}' + \mathbb{E}_q[\Omega\mathbf{u}_s\mathbf{u}_s'\Omega'] && (36)\\
&= \Sigma_{(s)}^{(\gamma*)} + \boldsymbol{\mu}_{(s)}^{(\gamma*)}\boldsymbol{\mu}_{(s)}^{(\gamma*)'} - \boldsymbol{\mu}_{(s)}^{(\gamma*)}\mathbf{u}_s'\mathbb{E}_q[\Omega'] - \mathbb{E}_q[\Omega]\mathbf{u}_s\boldsymbol{\mu}_{(s)}^{(\gamma*)'} + \mathbb{E}_q[\Omega\mathbf{u}_s\mathbf{u}_s'\Omega'] && (37)\\
&= \Sigma_{(s)}^{(\gamma*)} + \boldsymbol{\mu}_{(s)}^{(\gamma*)}\boldsymbol{\mu}_{(s)}^{(\gamma*)'} - \boldsymbol{\mu}_{(s)}^{(\gamma*)}\mathbf{u}_s' M^{(\Omega*)'} - M^{(\Omega*)}\mathbf{u}_s\boldsymbol{\mu}_{(s)}^{(\gamma*)'} + M^{(\Omega*)}\mathbf{u}_s\mathbf{u}_s' M^{(\Omega*)} + U^{(\Omega*)}\,\text{tr}(\mathbf{u}_s\mathbf{u}_s' V^{(\Omega*)}) && (38)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_q[(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)'{\Sigma^{(\gamma)}}^{-1}(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)] && (39)\\
&= \mathbb{E}_q[\mathrm{tr}\left((\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)'{\Sigma^{(\gamma)}}^{-1}(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)\right)] && (40)\\
&= \mathrm{tr}\left(\mathbb{E}_q[(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)'{\Sigma^{(\gamma)}}^{-1}(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)]\right) && (41)\\
&= \mathrm{tr}\left(\mathbb{E}_q[{\Sigma^{(\gamma)}}^{-1}(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)']\right) && (42)\\
&= \mathrm{tr}\left(\mathbb{E}_q[{\Sigma^{(\gamma)}}^{-1}]\mathbb{E}_q[(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)(\boldsymbol{\gamma}_{(s)} - \Omega\mathbf{u}_s)']\right) && (43)\\
&= \mathrm{tr}\left(M\Psi^{(\gamma*)-1}\left(\Sigma^{(\gamma*)}_{(s)} + \boldsymbol{\mu}^{(\gamma*)}_{(s)}{\boldsymbol{\mu}^{(\gamma*)}_{(s)}}' - \boldsymbol{\mu}^{(\gamma*)}_{(s)}\mathbf{u}_s'{M^{(\Omega*)}}' - M^{(\Omega*)}\mathbf{u}_s{\boldsymbol{\mu}^{(\gamma*)}_{(s)}}' + M^{(\Omega*)}\mathbf{u}_s\mathbf{u}_s'M^{(\Omega*)} + U^{(\Omega*)}\,\mathrm{tr}(\mathbf{u}_s\mathbf{u}_s'V^{(\Omega*)})\right)\right) && (44)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_q[\left(\boldsymbol{\gamma}_{(s_i)}'\mathbf{x}_i + \boldsymbol{\beta}_j'\boldsymbol{\delta}_l^{(j)}\right)^2] && (45)\\
&= \mathbb{E}_q[\mathbf{x}_i'\boldsymbol{\gamma}_{(s_i)}\boldsymbol{\gamma}_{(s_i)}'\mathbf{x}_i + 2\mathbf{x}_i'\boldsymbol{\gamma}_{(s_i)}\boldsymbol{\beta}_j'\boldsymbol{\delta}_l^{(j)} + {\boldsymbol{\delta}_l^{(j)}}'\boldsymbol{\beta}_j\boldsymbol{\beta}_j'\boldsymbol{\delta}_l^{(j)}] && (46)\\
&= \mathbf{x}_i'\mathbb{E}_q[\boldsymbol{\gamma}_{(s_i)}\boldsymbol{\gamma}_{(s_i)}']\mathbf{x}_i + 2\mathbf{x}_i'\mathbb{E}_q[\boldsymbol{\gamma}_{(s_i)}]\mathbb{E}_q[\boldsymbol{\beta}_j]'\boldsymbol{\delta}_l^{(j)} + {\boldsymbol{\delta}_l^{(j)}}'\mathbb{E}_q[\boldsymbol{\beta}_j\boldsymbol{\beta}_j']\boldsymbol{\delta}_l^{(j)} && (47)\\
&= \mathbf{x}_i'\left(\Sigma^{(\gamma*)}_{(s_i)} + \boldsymbol{\mu}^{(\gamma*)}_{(s_i)}{\boldsymbol{\mu}^{(\gamma*)}_{(s_i)}}'\right)\mathbf{x}_i + 2\mathbf{x}_i'\boldsymbol{\mu}^{(\gamma*)}_{(s_i)}{\boldsymbol{\mu}^{(\beta*)}_j}'\boldsymbol{\delta}_l^{(j)} + {\boldsymbol{\delta}_l^{(j)}}'\left(\Sigma^{(\beta*)}_j + \boldsymbol{\mu}^{(\beta*)}_j{\boldsymbol{\mu}^{(\beta*)}_j}'\right)\boldsymbol{\delta}_l^{(j)} && (48)
\end{aligned}$$