Supplementary Appendix for
"Fast Estimation of Ideal Points with Massive Data"

Kosuke Imai James Lo Jonathan Olmsted

In this appendix, we derive the variational Expectation-Maximization (EM) algorithm for the Bayesian ideal point models. For completeness, we begin by describing variational inference for the standard ideal point model (Appendix A) and the model with an ordinal outcome (Appendix B). In Appendix A, we also briefly explain variational inference in the context of the standard ideal point model for the readers who are not familiar with it. We then derive the variational EM algorithms for the dynamic and hierarchical ideal point models (Appendices C and D), which represent the main contributions of the paper. Finally, we show derivations for of variational EM algorithms for the generalized Wordfish and network models (Appendices E and F).

A Variational Inference for the Standard Ideal Point Model

We begin by deriving the variational EM algorithm for the standard ideal point model described. The key idea of variational inference is to come up with the best approximation to the posterior distribution under a certain factorization assumption. Under the standard ideal point model, we consider the approximating distribution that satisfies the following independence relationship,

$$q(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^J) \;=\; q(\mathbf{Y}^*)\; q(\mathbf{x}_1, \ldots, \mathbf{x}_N)\; q(\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_J). \tag{81}$$

Under this assumption, we find the optimal variational distribution that best approximates the true posterior distribution. We do this by minimizing the following Kullback-Leibler divergence, which is a measure of similarity of two distributions,

$$KL(q||p) \;=\; \mathbb{E}_q \left\{ \log \frac{q(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^J)}{p(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^J \mid \mathbf{Y})} \right\} \tag{82}$$

where the expectation is taken with respect to the approximating distribution q.

It is well known that minimizing the Kullback-Leibler divergence given in equation (82) is equivalent to maximizing the lower bound of the marginal log-likelihood function, which is called the evidence lower bound in the literature. This can be shown by the following equality,

$$\log p(\mathbf{Y}) \;=\; \underbrace{\mathbb{E}_q \left[\log \left\{ \frac{p(\mathbf{Y}, \mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^J)}{q(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^J)} \right\} \right]}_{\text{evidence lower bound}} + KL(q||p) \tag{83}$$

Therefore, we can use the Expectation-Maximization (EM) algorithm of Dempster, Laird and Rubin (1977) to maximize the evidence lower bound. It is important to note that the proposed algorithm is derived without making an additional assumption other than the factorization assumption given in equation (81). For example, we do not assume that q belongs to a certain family of distributions.

Variational Distribution

We outline the proposed variational EM algorithm. As shown below, each step resembles the Gibbs sampler algorithm used to estimate the standard Bayesian ideal point model. There are a total of three steps in this EM algorithm (latent propensities, ideal points, and item parameters) and we repeat these steps until convergence.

Latent Propensities. We expand the joint density given in equation (4) and apply the logarithm to the product. Collecting the terms that involve $\mathbf{Y}^*$, we obtain,

$$\begin{aligned}\log q(\mathbf{Y}^*) =& \ \mathbb{E}_{\tilde{\boldsymbol{\beta}},\mathbf{x}}\left[\log p(\mathbf{Y} \mid \mathbf{Y}^*) + \log p(\mathbf{Y}^* \mid \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^J)\right] + \text{const.} \\ =& \ \sum_{i=1}^N \sum_{j=1}^J \left[\log\left(\mathbf{1}\{y_{ij}^* > 0\}\mathbf{1}\{y_{ij}=1\} + \mathbf{1}\{y_{ij}^* \le 0\}\mathbf{1}\{y_{ij}=0\}\right) - \frac{1}{2}\left\{{y_{ij}^*}^2 - 2y_{ij}^*\mathbb{E}(\tilde{\mathbf{x}}_i)^\top \mathbb{E}(\tilde{\boldsymbol{\beta}}_j)\right\}\right] + \text{const.}\end{aligned}$$

We recognize this as the product of $N \times J$ truncated Normal distributions, which is given as,

$$q(\mathbf{Y}^*) = \prod_{i=1}^N \prod_{j=1}^J q(y_{ij}^*) \quad \text{where} \quad q(y_{ij}^*) = \begin{cases} \mathcal{TN}(m_{ij}, 1, 0, \infty) & \text{if } y_{ij}=1 \\ \mathcal{TN}(m_{ij}, 1, -\infty, 0) & \text{if } y_{ij}=0 \end{cases}$$

where $m_{ij} = \mathbb{E}(\tilde{\mathbf{x}}_i)^\top \mathbb{E}(\tilde{\boldsymbol{\beta}}_j)$. Abstention is treated as missing at random and so the variational distribution in that case is $q(y_{ij}^*) = \mathcal{N}(m_{ij}, 1)$. Given this result, we update the mean of y_{ij}^* as,

$$\mathbb{E}(y_{ij}^*) = \begin{cases} m_{ij} + \frac{\phi(m_{ij})}{\Phi(m_{ij})} & \text{if } y_{ij}=1 \\ m_{ij} - \frac{\phi(m_{ij})}{1-\Phi(m_{ij})} & \text{if } y_{ij}=0 \end{cases}$$

Ideal Points. For the ideal points, we have,

$$\begin{aligned}\log q(\mathbf{x}_1, \dots, \mathbf{x}_N) =& \ \mathbb{E}_{\tilde{\boldsymbol{\beta}},\mathbf{y}^*}[\log p(\mathbf{Y}^* \mid \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^J) + \log p(\mathbf{x}_1, \dots, \mathbf{x}_N)] + \text{const.} \\ =& \ \sum_{i=1}^N \sum_{j=1}^J \mathbb{E}_{\tilde{\boldsymbol{\beta}}_j, y_{ij}^*}\left[\log \phi_1(y_{ij}^*; \tilde{\mathbf{x}}_i^\top \tilde{\boldsymbol{\beta}}_j, 1)\right] + \sum_{j=1}^J \log \phi_K(\mathbf{x}_i; \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) + \text{const.}\end{aligned}$$

We apply the standard result of the Bayesian linear regression to obtain,

$$q(\mathbf{x}_1, \dots, \mathbf{x}_N) = \prod_{i=1}^N q(\mathbf{x}_i) \quad \text{where} \quad q(\mathbf{x}_i) = \mathcal{N}(\mathbf{A}^{-1}\mathbf{a}_i, \ \mathbf{A}^{-1})$$

where $\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} + \sum_{j=1}^J \mathbb{E}(\boldsymbol{\beta}_j \boldsymbol{\beta}_j^\top)$ and $\mathbf{a}_i = \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\mu}_{\mathbf{x}} + \sum_{j=1}^J \mathbb{E}(\boldsymbol{\beta}_j)\mathbb{E}(y_{ij}^*) - \mathbb{E}(\boldsymbol{\beta}_j \alpha_j)$. Thus, we update the required moments as $\mathbb{E}(\mathbf{x}_i) = \mathbf{A}^{-1}\mathbf{a}_i$, and

$$\mathbb{E}(\tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^\top) = \begin{bmatrix} 1 & \mathbf{a}_i^\top \mathbf{A}^{-1} \\ \mathbf{A}^{-1}\mathbf{a}_i & \mathbf{A}^{-1}\mathbf{a}_i\mathbf{a}_i^\top \mathbf{A}^{-1} + \mathbf{A}^{-1} \end{bmatrix}$$

Item Parameters. The derivation for item parameters proceeds as follows,

$$\begin{aligned}\log q(\tilde{\boldsymbol\beta}_1,\ldots,\tilde{\boldsymbol\beta}_J) &= \mathbb{E}_{\mathbf{x},\mathbf{y}^*}[\log p(\mathbf{Y}^* \mid \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol\beta}_j\}_{j=1}^J) + \log p(\tilde{\boldsymbol\beta}_1,\ldots,\tilde{\boldsymbol\beta}_J)] + \text{const.}\\ &= \sum_{j=1}^J\sum_{i=1}^N \mathbb{E}_{\mathbf{x}_i, y^*_{ij}}\left[\log \phi_1(y^*_{ij} \mid \tilde{\mathbf{x}}_i^\top \tilde{\boldsymbol\beta}_j, 1)\right] + \sum_{j=1}^J \log p(\tilde{\boldsymbol\beta}_j) + \text{const.}\end{aligned}$$

Again, using the standard Bayesian linear regression result, we obtain,

$$q(\tilde{\boldsymbol\beta}_1,\ldots,\tilde{\boldsymbol\beta}_J) = \prod_{j=1}^J q(\tilde{\beta}_j) \quad \text{where} \quad q(\tilde{\boldsymbol\beta}_j) = \mathcal{N}(\mathbf{B}^{-1}\mathbf{b}_j,\ \mathbf{B}^{-1})$$

where $\mathbf{B} = \Sigma_{\tilde{\boldsymbol\beta}}^{-1} + \sum_{i=1}^N \mathbb{E}(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top)$ and $\mathbf{b}_j = \boldsymbol\Sigma_{\tilde{\boldsymbol\beta}}^{-1}\boldsymbol\mu_{\tilde{\boldsymbol\beta}} + \sum_{i=1}^N \mathbb{E}(\tilde{\mathbf{x}}_i)\mathbb{E}(y^*_{ij})$. This variational distribution implies the following required moments, i.e., $\mathbb{E}(\tilde{\boldsymbol\beta}_j) = \mathbf{B}^{-1}\mathbf{b}_j$, $\mathbb{E}(\tilde{\boldsymbol\beta}_j\tilde{\boldsymbol\beta}_j^\top) = \mathbf{B}^{-1}\mathbf{b}_j\mathbf{b}_j^\top\mathbf{B}^{-1} + \mathbf{B}^{-1}$ whose lower-right $K \times K$ block matrix and lower-left K dimensional column vector are equal to $\mathbb{E}(\boldsymbol\beta_j\boldsymbol\beta_j^\top)$ and $\mathbb{E}(\boldsymbol\beta_j\alpha_j)$, respectively.

Evidence Lower Bound

We can also derive the expression for the evidence lower bound, $\mathcal{L}(q)$ given in equation (83). This lower bound can be decomposed as

$$\begin{aligned}\mathcal{L}(q) &= \mathbb{E}\left[\log p(\mathbf{Y}, \mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol\beta}_j\}_{j=1}^J)\right] - \mathbb{E}\left[\log q(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol\beta}_j\}_{j=1}^J)\right]\\ &= \sum_{i=1}^N\sum_{j=1}^J \mathbb{E}\left[\log p(y_{ij} \mid y^*_{ij}) + \log p(y^*_{ij} \mid \mathbf{x}_i, \tilde{\boldsymbol\beta}_j)\right] + \sum_{i=1}^N \mathbb{E}\left[\log p(\mathbf{x}_i)\right] + \sum_{j=1}^J \mathbb{E}\left[\log p(\tilde{\boldsymbol\beta}_j)\right]\\ &\quad - \sum_{i=1}^N\sum_{j=1}^J \mathbb{E}\left[\log q(y^*_{ij})\right] - \sum_{i=1}^N \mathbb{E}\left[\log q(\mathbf{x}_i)\right] - \sum_{j=1}^J \mathbb{E}\left[\log q(\tilde{\boldsymbol\beta}_j)\right]\end{aligned}$$

First, we need to compute the entropy for each variational distribution, which is given below,

$$\begin{aligned}\mathbb{E}[\log q(y^*_{ij})] &= \frac{1}{2}\log(2\pi e) + y_{ij}\log\Phi(m_{ij}) + (1-y_{ij})\log\Phi\left(-m_{ij}\right) + \left(\frac{-m_{ij}\phi(-m_{ij})}{2\Phi(m_{ij})}\right)^{y_{ij}}\left(\frac{m_{ij}\phi(-m_{ij})}{2\Phi(-m_{ij})}\right)^{1-y_{ij}}\\ \mathbb{E}\left[\log q(\mathbf{x}_i)\right] &= \frac{K}{2}\log(2\pi e) - \frac{1}{2}\log|\mathbf{A}|\\ \mathbb{E}[\log q(\tilde{\boldsymbol\beta}_j)] &= \frac{K+1}{2}\log(2\pi e) - \frac{1}{2}\log|\mathbf{B}|\end{aligned}$$

Second, we need to evaluate the expectation of log-likelihood and log prior density using the variational distribution. We begin by noting $\mathbb{E}[\log p(y_{ij} \mid y^*_{ij})] = 0$. For the log-likelihood, therefore, we have,

$$\mathbb{E}\left[\log p(y^*_{ij} \mid \mathbf{x}_i, \tilde{\boldsymbol\beta}_j)\right] = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\left\{\mathbb{E}\left[(y^*_{ij})^2\right] - 2\mathbb{E}(y^*_{ij})\mathbb{E}(\tilde{\mathbf{x}}_i)^\top\mathbb{E}(\tilde{\boldsymbol\beta}_j) + \mathbb{E}\left[(\tilde{\mathbf{x}}_i^\top\tilde{\boldsymbol\beta}_j)^2\right]\right\}$$

where $\mathbb{E}[(\tilde{\mathbf{x}}_i^\top\tilde{\boldsymbol\beta}_j)^2] = \mathbb{E}[\mathrm{tr}(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top\tilde{\boldsymbol\beta}_j\tilde{\boldsymbol\beta}_j^\top)] = \mathrm{tr}[\mathbb{E}(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top)\mathbb{E}(\tilde{\boldsymbol\beta}_j\tilde{\boldsymbol\beta}_j^\top)]$. The expectation of the log prior density is given by,

$$\mathbb{E}[\log p(\mathbf{x}_i)] = -\frac{K}{2}\log(2\pi) - \frac{1}{2}|\boldsymbol\Sigma_{\mathbf{x}}| - \frac{1}{2}\left\{\mathbb{E}(\mathbf{x}_i^\top\boldsymbol\Sigma_{\mathbf{x}}^{-1}\mathbf{x}_i) - 2\boldsymbol\mu_{\mathbf{x}}^\top\boldsymbol\Sigma_{\mathbf{x}}^{-1}\mathbb{E}(\mathbf{x}_i) + \boldsymbol\mu_{\mathbf{x}}^\top\boldsymbol\Sigma_{\mathbf{x}}^{-1}\boldsymbol\mu_{\mathbf{x}}\right\}$$

where $\mathbb{E}\left[\mathbf{x}_i^\top \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_i\right] = \text{tr}\{\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbb{E}(\mathbf{x}_i\mathbf{x}_i)^\top\}$. Moreover,

$$\mathbb{E}[\log(\tilde{\boldsymbol{\beta}}_j)] \;=\; -\frac{K+1}{2}\log(2\pi) - \frac{1}{2}\left\{\mathbb{E}\left(\tilde{\boldsymbol{\beta}}_j^\top \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}_j}^{-1}\tilde{\boldsymbol{\beta}}_j\right) - 2\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}_j}^\top \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}_j}^{-1}\mathbb{E}(\tilde{\boldsymbol{\beta}}_j) + \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}_j}^\top \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}_j}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}_j}\right\}$$

where $\mathbb{E}(\tilde{\boldsymbol{\beta}}_j^\top \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\tilde{\boldsymbol{\beta}}_j) = \text{tr}\{\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\mathbb{E}(\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^\top)\}$. Finally, we note that $\mathbb{E}(\mathbf{x}_i\mathbf{x}_i^\top) = \mathbf{A}^{-1}\mathbf{a}_i\mathbf{a}_i^\top\mathbf{A}^{-1} + \mathbf{A}^{-1}$ and $\mathbb{E}(\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^\top) = \mathbf{B}^{-1}\mathbf{b}_j\mathbf{b}_j^\top\mathbf{B}^{-1} - \mathbf{B}^{-1}$.

B Variational Inference for the Model with an Ordinal Outcome

We derive the variational distribution for the ideal point with a three-category ordinal outcome. We use the reparameterized model. The joint posterior of the reparameterized model is given by,

$$\begin{aligned}
& p(\mathbf{Z}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tau_j^2, \tilde{\boldsymbol{\beta}}_j\}_{j=1}^J \mid \mathbf{Y}) \\
\propto & \prod_{i=1}^N\prod_{j=1}^J \left[\mathbf{1}\{z_{ij}^* < 0\}\mathbf{1}\{y_{ij}=0\} + \mathbf{1}\{0 \le z_{ij}^* < 1\}\mathbf{1}\{y_{ij}=1\} + \mathbf{1}\{z_{ij}^* \ge 1\}\mathbf{1}\{y_{ij}=2\}\right]\phi_1(z_{ij}^*; \tilde{\mathbf{x}}_i^\top\tilde{\boldsymbol{\beta}}_j, \tau_j^{-2}) \\
& \times \prod_{j=1}^J \phi_{K+1}(\tilde{\boldsymbol{\beta}}_j; \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}})\,\mathcal{G}\left(\tau_j^2; \frac{\nu_\tau}{2}, \frac{s_\tau^2}{2}\right)\prod_{i=1}^N \phi_K(\mathbf{x}_i; \mu_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}})
\end{aligned}$$

We derive the variational EM algorithm in a manner similar to the one used for the standard item response model with a binary outcome variable. The key difference is that the current model has an additional parameter τ_j.

Variational Distribution

Latent Propensities. We begin by deriving the variational distribution for the latent propensities as follows,

$$\begin{aligned}
\log q(z_{ij}^*) &= \mathbb{E}_{\tilde{\boldsymbol{\beta}}_j, \mathbf{x}_i, \tau_j}[\log p(y_{ij} \mid z_{ij}^*) + \log p(z_{ij}^* \mid \mathbf{x}_i, \tilde{\boldsymbol{\beta}}, \tau_j^2)] + \text{const.} \\
&= \log\left[\mathbf{1}\{z_{ij}^* < 0\}\mathbf{1}\{y_{ij}=1\} + \mathbf{1}\{0 \le z_{ij}^* < 1\}\mathbf{1}\{y_{ij}=2\} + \mathbf{1}\{z_{ij}^* \ge 1\}\mathbf{1}\{y_{ij}=3\}\right] \\
&\quad - \frac{\mathbb{E}(\tau_j^2)}{2}\left\{(z_{ij}^*)^2 - 2z_{ij}^*\mathbb{E}(\tilde{\mathbf{x}}_i)^\top\mathbb{E}(\tilde{\boldsymbol{\beta}}_j)\right\} + \text{const.}
\end{aligned}$$

We recognize this as the following truncated normal distribution,

$$q(\mathbf{Z}^*) \;=\; \prod_{i=1}^N\prod_{j=1}^J q(z_{ij}^*) \quad \text{where} \quad q(z_{ij}^*) \;=\; \begin{cases} \mathcal{TN}(m_{ij},\; w_j^{-2},\; -\infty,\; 0) & \text{if } y_{ij}=0 \\ \mathcal{TN}(m_{ij},\; w_j^{-2},\; 0,\; 1) & \text{if } y_{ij}=1 \\ \mathcal{TN}(m_{ij},\; w_j^{-2},\; 1,\; \infty) & \text{if } y_{ij}=2 \end{cases}$$

where $m_{ij} = \mathbb{E}(\tilde{\mathbf{x}}_i)^\top\mathbb{E}(\tilde{\boldsymbol{\beta}}_j)$ and $w_j^2 = \mathbb{E}(\tau_j^2)$. We then update the mean of z_{ij}^* as,

$$\mathbb{E}(z_{ij}^*) \;=\; \begin{cases} m_{ij} - \frac{\phi(m_{ij}w_j)}{1-\Phi(m_{ij}w_j)}w_j^{-1} & \text{if } y_{ij}=0 \\ m_{ij} + \frac{\phi(m_{ij}w_j)-\phi((1-m_{ij})w_j)}{\Phi((1-m_{ij})w_j)+\Phi(m_{ij}w_j)-1}w_j^{-1} & \text{if } y_{ij}=1 \\ m_{ij} + \frac{\phi((1-m_{ij})w_j)}{1-\Phi((1-m_{ij})w_j)}w_j^{-1} & \text{if } y_{ij}=2 \end{cases}$$

For abstention, we have $q(z_{ij}^*) = \mathcal{N}(m_{ij},\; w_j^{-2})$ and hence $\mathbb{E}(z_{ij}^*) = m_{ij}$.

Ideal Points. We expand the joint density p and apply the logarithm to the product.

$$\begin{aligned}\log q(\mathbf{x}_i) &= \mathbb{E}_{z^*,\tilde{\boldsymbol{\beta}}^*,\tau}\left[\sum_{j=1}^{J}\log p(z^*_{ij} \mid \mathbf{x}_i, \tilde{\boldsymbol{\beta}}^*_j, \tau_j^2)\right] + \log p(\mathbf{x}_i) + \text{const.}\\ &= -\sum_{j=1}^{J}\frac{\mathbb{E}(\tau_j^2)}{2}\left\{\mathbf{x}_i^\top \mathbb{E}\left(\boldsymbol{\beta}^*_j\boldsymbol{\beta}^{*\top}_j\right)\mathbf{x}_i - 2\boldsymbol{\beta}^{*\top}_j(z^*_{ij}-\alpha^*_j)\mathbf{x}_i\right\} - \frac{1}{2}\left(\mathbf{x}_i^\top\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_i - 2\boldsymbol{\mu}_{\mathbf{x}}^\top\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_i\right) + \text{const.}\end{aligned}$$

Since this is a normal distribution density, we have,

$$p(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \prod_{i=1}^{N} q(\mathbf{x}_i) \quad \text{where} \quad q(\mathbf{x}_i) = \mathcal{N}(\mathbf{A}^{-1}\mathbf{a}_i,\ \mathbf{A}^{-1})$$

with $\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} + \sum_{j=1}^{J}\mathbb{E}(\tau_j^2)\mathbb{E}\left(\boldsymbol{\beta}^*_j\boldsymbol{\beta}^{*\top}_j\right)$ and $\mathbf{a}_i = \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\mu}_{\mathbf{x}} + \sum_{j=1}^{J}\mathbb{E}(\tau_j^2)\left\{\mathbb{E}(\boldsymbol{\beta}^*_j)^\top\mathbb{E}(z^*_{ij}) - \mathbb{E}\left(\boldsymbol{\beta}^{*\top}_j\alpha^*_j\right)\right\}$. Given this result, we update the required moments, i.e., $\mathbb{E}(\mathbf{x}_i) = \mathbf{A}^{-1}\mathbf{a}_i$ and $\mathbb{E}(\mathbf{x}_i\mathbf{x}_i^\top) = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{a}_i\mathbf{a}_i^\top\mathbf{A}^{-1}$.

Item Parameters. For the item parameters, the variational distribution can be derived as follows,

$$\begin{aligned}\log q(\tilde{\boldsymbol{\beta}}_j) &= \mathbb{E}_{\mathbf{z}^*_i,\mathbf{x},\tau_j}\left[\sum_{i=1}^{N}\log p(z^* \mid \mathbf{x}_i, \tilde{\boldsymbol{\beta}}_j, \tau_j)\right] + \log p(\tilde{\boldsymbol{\beta}}_j))] + \text{const.}\\ &= -\frac{\mathbb{E}(\tau_j^2)}{2}\sum_{i=1}^{N}\left\{\tilde{\boldsymbol{\beta}}_j^\top\mathbb{E}(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top)\tilde{\boldsymbol{\beta}}_j - 2z^*_{ij}\tilde{\mathbf{x}}_i^\top\tilde{\boldsymbol{\beta}}_j\right\} - \frac{1}{2}\left(\tilde{\boldsymbol{\beta}}_j^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\tilde{\boldsymbol{\beta}}_j - \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\tilde{\boldsymbol{\beta}}_j\right) + \text{const.}\end{aligned}$$

This is another Normal distribution, which is given by,

$$q(\tilde{\boldsymbol{\beta}}_1,\ldots,\tilde{\boldsymbol{\beta}}_J) = \prod_{j=1}^{J} q(\tilde{\boldsymbol{\beta}}_j) \quad \text{where} \quad q(\tilde{\boldsymbol{\beta}}_j) = \mathcal{N}(\mathbf{B}_j^{-1}\mathbf{b}_j,\ \mathbf{B}_j^{-1})$$

where $\mathbf{B}_j = \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1} + \mathbb{E}(\tau_j^2)\sum_{i=1}^{N}\mathbb{E}(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top)$ and $\mathbf{b}_j = \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}} + \mathbb{E}(\tau_j^2)\sum_{i=1}^{N}\mathbb{E}(\tilde{\mathbf{x}}_i)^\top\mathbb{E}(z^*_{ij})$. We update the required moment as $\mathbb{E}(\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^\top) = \mathbf{B}_j^{-1} + \mathbf{B}_j^{-1}\mathbf{b}_j\mathbf{b}_j^\top\mathbf{B}_j^{-1}$.

Variance Parameters. Finally, we derive the variational distribution for the variance parameters,

$$\begin{aligned}\log q(\tau_j^2) &= \mathbb{E}_{\mathbf{z}^*_j,\mathbf{x},\tilde{\boldsymbol{\beta}}^*_j}\left[\sum_{i=1}^{N}\log p(z^*_{ij} \mid \mathbf{x}_i, \tilde{\boldsymbol{\beta}}_j, \tau_j^2)\right] + \log p(\tau_j^2) + \text{const.}\\ &= \frac{N}{2}\log\tau_j^2 - \frac{\tau_j^2}{2}\sum_{i=1}^{N}\left[\mathbb{E}({z^*_{ij}}^2) - 2\mathbb{E}(z^*_{ij})\mathbb{E}(\tilde{\mathbf{x}}_i)^\top\mathbb{E}(\tilde{\boldsymbol{\beta}}_j) + \text{tr}\{\mathbb{E}(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top)\mathbb{E}(\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^\top)\}\right]\\ &\quad + \left(\frac{\nu_\tau}{2} - 1\right)\log\tau_j^2 - \frac{s_\tau\tau_j^2}{2} + \text{const.}\end{aligned}$$

We recognize this as a Gamma distribution and thus the variational distribution is given by,

$$q(\tau_1^2,\ldots,\tau_J^2) = \prod_{j=1}^{J} q(\tau_j) \quad \text{where} \quad q(\tau_j^2) = \mathcal{G}\left(\frac{c_j}{2},\ \frac{d_j}{2}\right)$$

with $c_j = N + \nu_\tau$ and $d_j = s_\tau + \sum_{i=1}^{N}\mathbb{E}({z^*_{ij}}^2) - 2\sum_{i=1}^{N}\mathbb{E}(z^*_{ij})\mathbb{E}(\tilde{\mathbf{x}}_i)^\top\mathbb{E}(\tilde{\boldsymbol{\beta}}_j) + \sum_{i=1}^{N}\text{tr}\{\mathbb{E}(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top)\mathbb{E}(\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^\top)\}$. The required moment is given by $\mathbb{E}(\tau_j^2) = c_j/d_j$.

Evidence Lower Bound

The lower bound is given by,

$$
\begin{aligned}
\mathcal{L}(q) &= \mathbb{E}\left[\log p(\mathbf{Y}, \mathbf{Z}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tau_j^2, \tilde{\boldsymbol{\beta}}_j\}_{j=1}^J)\right] - \mathbb{E}\left[\log q(\mathbf{Z}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tau_j^2, \tilde{\boldsymbol{\beta}}_j\}_{j=1}^J)\right] \\
&= \sum_{i=1}^N\sum_{j=1}^J \mathbb{E}\left[\log p(y_{ij} \mid z_{ij}^*) + \log p(z_{ij}^* \mid \mathbf{x}_i, \tilde{\boldsymbol{\beta}}_j, \tau_j^2)\right] + \sum_{i=1}^N \mathbb{E}[\log p(\mathbf{x}_i)] + \sum_{j=1}^J \mathbb{E}[\log p(\tilde{\boldsymbol{\beta}}_j)] + \sum_{j=1}^J \mathbb{E}[\log p(\tau_j^2)] \\
&\quad - \sum_{i=1}^N\sum_{j=1}^J \mathbb{E}\left[\log q(z_{ij}^*)\right] - \sum_{i=1}^N \mathbb{E}\left[\log q(\mathbf{x}_i)\right] - \sum_{j=1}^J \mathbb{E}\left[\log q(\tilde{\boldsymbol{\beta}}_j)\right] - \sum_{j=1}^J \mathbb{E}\left[\log q(\tau_j^2)\right]
\end{aligned}
$$

We begin by noting $\mathbb{E}[\log p(y_{ij} \mid z_{ij}^*)] = 0$. Next, we compute the entropy of variational distribution for the latent propensity as follows,

$$
\begin{aligned}
&\mathbb{E}[\log q(z_{ij}^*)] \\
&= \frac{1}{2}\{\log(2\pi e) - \log \mathbb{E}(\tau_j^2)\} + \mathbf{1}\{y_{ij}=0\}\log\Phi\left(m_{ij}^*\right) + \mathbf{1}\{y_{ij}=1\}\log\{\Phi(\tilde{m}_{ij}) - \Phi(m_{ij}^*)\} + \mathbf{1}\{y_{ij}=2\}\log\Phi(-\tilde{m}_{ij}) \\
&\quad + \left(-\frac{m_{ij}^*\phi(m_{ij}^*)}{2\Phi(m_{ij}^*)}\right)^{\mathbf{1}\{y_{ij}=0\}} \left(\frac{m_{ij}^*\phi(m_{ij}^*) - \tilde{m}_{ij}\phi(\tilde{m}_{ij})}{2\{\Phi(\tilde{m}_{ij}) - \Phi(m_{ij}^*)\}}\right)^{\mathbf{1}\{y_{ij}=1\}} \left(\frac{\tilde{m}_{ij}\phi(\tilde{m}_{ij})}{2\Phi(-\tilde{m}_{ij})}\right)^{\mathbf{1}\{y_{ij}=2\}}
\end{aligned}
$$

where $m_{ij}^* = -\mathbb{E}(\tilde{\mathbf{x}}_i)^\top\mathbb{E}(\tilde{\boldsymbol{\beta}}_j)/\sqrt{\mathbb{E}(\tau_j^2)}$ and $\tilde{m}_{ij} = \{1 - \mathbb{E}(\tilde{\mathbf{x}}_i)^\top\mathbb{E}(\tilde{\boldsymbol{\beta}}_j)\}/\sqrt{\mathbb{E}(\tau_j^2)}$. In addition, we have the following entropies,

$$
\begin{aligned}
\mathbb{E}[\log q(\mathbf{x}_i)] &= \frac{K}{2}\log(2\pi e) - \frac{1}{2}\log|\mathbf{A}| \\
\mathbb{E}[\log q(\tilde{\boldsymbol{\beta}}_j)] &= \frac{K+1}{2}\log(2\pi e) - \frac{1}{2}\log|\mathbf{B}_j| \\
\mathbb{E}[\log q(\tau_j^2)] &= \frac{c_j}{2} - \log\frac{d_j}{2} + \log\left[\Gamma\left(\frac{c_j}{2}\right)\right] + \left(1 - \frac{c_j}{2}\right)\psi\left(\frac{c_j}{2}\right)
\end{aligned}
$$

where $\Gamma(\cdot)$ and $\psi(\cdot)$ are the gamma and digamma functions, respectively.

Next, the expected log-likelihood can be calculated as,

$$
\mathbb{E}[\log p(z_{ij}^* \mid \mathbf{x}_i, \tilde{\boldsymbol{\beta}}_j, \tau_j^2)] = -\frac{1}{2}\left\{\log(2\pi) - \mathbb{E}(\log\tau_j^2)\right\} - \frac{\mathbb{E}(\tau_j^2)}{2}\left\{\mathbb{E}[(z_{ij}^*)^2] - 2\mathbb{E}(z_{ij}^*)\mathbb{E}(\tilde{\mathbf{x}}_i)^\top\mathbb{E}(\tilde{\boldsymbol{\beta}}_j) + \mathbb{E}(\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top\tilde{\boldsymbol{\beta}}_j)\right\}
$$

where $\mathbb{E}(\log\tau_j^2) = \psi(c_j/2) - \log(d_j/2)$ and $\mathbb{E}(\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top\tilde{\boldsymbol{\beta}}_j) = \text{tr}\{\mathbb{E}(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top)\mathbb{E}(\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^\top)\}$.

Finally, we compute the expected log-prior as follows,

$$
\begin{aligned}
\mathbb{E}[\log p(\mathbf{x}_i)] &= -\frac{K}{2}\{\log(2\pi) + \log|\boldsymbol{\Sigma}_{\mathbf{x}}|\} - \frac{1}{2}\{\mathbb{E}(\mathbf{x}_i^\top\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_i) - 2\mathbb{E}(\mathbf{x}_i)^\top\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{x}}^\top\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\mu}_{\mathbf{x}}\} \\
\mathbb{E}[\log p(\tilde{\boldsymbol{\beta}}_i)] &= -\frac{K+1}{2}\{\log(2\pi) + \log|\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}|\} - \frac{1}{2}\{\mathbb{E}(\tilde{\boldsymbol{\beta}}_j^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\tilde{\boldsymbol{\beta}}_j) - 2\mathbb{E}(\tilde{\boldsymbol{\beta}}_j)^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}} + \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}\} \\
\mathbb{E}[\log p(\tau_j^2)] &= \frac{c_j}{2}\log\frac{d_j}{2} - \log\Gamma\left(\frac{c_j}{2}\right) + \left(\frac{c_j}{2} - 1\right)\mathbb{E}(\log\tau_j^2) - \frac{d_j}{2}\mathbb{E}(\tau_j^2)
\end{aligned}
$$

where $\mathbb{E}(\mathbf{x}_i^\top\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbf{x}_i) = \text{tr}\{\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\mathbb{E}(\mathbf{x}_i\mathbf{x}_i^\top)\}$ and $\mathbb{E}(\tilde{\boldsymbol{\beta}}_j^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\tilde{\boldsymbol{\beta}}_j) = \text{tr}\{\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\mathbb{E}(\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^\top)\}$.

C Variational Inference for the Dynamic Ideal Point Model

For the dynamic model, the joint posterior distribution is given by equation (31). We derive the variational EM algorithm under the factorization assumption given in equation (32).

Latent Propensities. For the latent propensities, we derive the variational distribution as follows,

$$\begin{aligned}\log q(y_{ijt}) &= \mathbb{E}_{x_i,\tilde{\boldsymbol{\beta}}_j}[\log p(y_{ijt} \mid y^*_{ijt}) + \log p(y_{ijt} \mid x_{it}, \tilde{\boldsymbol{\beta}}_{jt})] + \text{const.} \\ &= \log[\mathbf{1}\{y_{ijt}=1\}\mathbf{1}\{y^*_{ijt}>0\} + \mathbf{1}\{y_{ijt}=0\}\mathbf{1}\{y_{ijt}\leq 0\}] - \frac{1}{2}\left({y^*_{ijt}}^2 - 2y^*_{ijt}\mathbb{E}(\tilde{\mathbf{x}}_{it})^\top\mathbb{E}(\tilde{\boldsymbol{\beta}}_{jt})\right) + \text{const.}\end{aligned}$$

This is a truncated normal distribution, and hence the approximating distribution is given by equation (60).

Item Parameters. The variational distribution for item parameters is given by,

$$\begin{aligned}\log(\tilde{\boldsymbol{\beta}}_{jt}) &= \sum_{i\in\mathcal{I}_t}\mathbb{E}_{y^*_{ijt},\mathbf{x}}[\log p(y^*_{ijt} \mid x_{it}, \tilde{\boldsymbol{\beta}}_{jt})] + \log p(\tilde{\boldsymbol{\beta}}) + \text{const.} \\ &= -\frac{1}{2}\sum_{i\in\mathcal{I}_t}\left\{\tilde{\boldsymbol{\beta}}_{jt}^\top\mathbb{E}(\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{it}^\top)\tilde{\boldsymbol{\beta}}_{jt} - 2\mathbb{E}(y^*_{ijt})\mathbb{E}(\tilde{\mathbf{x}}_{it})^\top\tilde{\boldsymbol{\beta}}_{jt}\right\} - \frac{1}{2}(\tilde{\boldsymbol{\beta}}_{jt}^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\tilde{\boldsymbol{\beta}}_{jt} - 2\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\tilde{\boldsymbol{\beta}}_{jt}) + \text{const.}\end{aligned}$$

where $\mathcal{I}_t = \{i : \underline{T}_i \leq t \leq \overline{T}_i\}$. Recognizing this as a Normal distribution, we have the resulting variational distribution given in equation (62).

Ideal Points. For notational simplicity and without loss of generality, let $\underline{T}_i = 1$ and $\overline{T}_i = T$. Then, we have,

$$\begin{aligned}&\log q(x_{i1},\ldots,x_{iT}) \\ &= \sum_{t=1}^{T}\sum_{j=1}^{J_t}\mathbb{E}_{y^*_{ijt},\tilde{\boldsymbol{\beta}}}[\log p(y^*_{ijt} \mid x_{it}, \tilde{\boldsymbol{\beta}}_{jt})] + \sum_{t=2}^{T}\log p(x_{it} \mid x_{i,t-1}) + \log p(x_{it}) + \text{const.} \\ &= -\frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\left[\mathbb{E}(\beta_{jt}^2)x_{it}^2 - 2\{\mathbb{E}(y^*_{ijt})\mathbb{E}(\beta_{jt}) - \mathbb{E}(\beta_{jt}\alpha_{jt})\}x_{it}\right] - \frac{1}{2\omega_x^2}\sum_{t=2}^{T}(x_{it}-x_{i,t-1})^2 - \frac{1}{2\Sigma_x}(x_i-\mu_x)^2 + \text{const.}\end{aligned}$$

We recognize that this expression resembles the posterior distribution of the following dynamic linear model, i.e., $p(x_{i1},\ldots,x_{iT} \mid \ddot{y}_{i1},\ldots,\ddot{y}_{iT},\ddot{\beta}_1,\ldots,\ddot{\beta}_T)$,

$$\begin{aligned}\ddot{y}_{it} &= \ddot{\beta}_t x_{it} + \ddot{\epsilon}_{it} \\ x_{it} &= x_{i,t-1} + \eta_{it}\end{aligned}$$

where $\ddot{\beta}_t = \sqrt{\sum_{j=1}^{J_t}\mathbb{E}(\beta_{jt}^2)}$, $\ddot{y}_{it} = \{\sum_{j=1}^{J_t}\mathbb{E}(y^*_{ijt})\mathbb{E}(\beta_{jt}) - \mathbb{E}(\beta_{jt}\alpha_{jt})\}/\ddot{\beta}_t$, $\ddot{\epsilon}_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $\eta_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\omega_x^2)$, and $x_{i1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_x,\Sigma_x)$.

Thus, we can apply the standard Kalman filtering results to obtain the necessary moments. We begin by applying the forward recursion relationship,

$$p(x_{it} \mid \ddot{y}_{i1},\ldots,\ddot{y}_{it}) = \phi(x_{it}; c_{it}, C_{it})$$

where $c_{it} = c_{i,t-1} + K_t(\ddot{y}_{it} - \ddot{\beta}_t c_{i,t-1})$ and $C_{it} = (1 - K_t \ddot{\beta}_t)\Omega_t$ with $\Omega_{it} = \omega_{x,i}^2 + C_{i,t-1}$, $K_t = \ddot{\beta}_t \Omega_{it}/S_{it}$ and $S_t = \ddot{\beta}_t^2 \Omega_{it} + 1$. We recursively compute these quantities by setting $c_{i0} = \mu_{x,i}$ and $C_{i0} = \Sigma_{x,i}$. We then use the backward recursion to obtain,

$$p(x_{it} \mid \ddot{y}_{i1}, \ldots, \ddot{y}_{iT}) \quad = \quad \phi_1(x_{it}; d_{it}, D_{it})$$

where $d_{it} = c_{it} + J_{it}(d_{i,t+1} - c_{i,t+1})$ and $D_{it} = C_{it} + J_{it}^2(D_{i,t+1} - \Omega_{i,t+1})$ with $J_{it} = C_{it}/\Omega_{i,t+1}$. Again, the recursive computation is done by setting $d_{iT} = c_{iT}$ and $D_{iT} = C_{iT}$.

D Variational Inference for the Hierarchical Ideal Point Model

For the hierarchical model, the joint posterior distribution is given by equation (39). We derive the variational EM algorithm under the factorization assumption given in equation (40).

Latent Propensities. The latent propensities can be derived as,

$$\begin{aligned}
\log q(y_\ell^*) &= \mathbb{E}_{\tilde{\boldsymbol{\beta}},\boldsymbol{\gamma},\eta}[\log p(y_\ell \mid y_\ell^*) + \log p(y_\ell^* \mid \tilde{\boldsymbol{\beta}}_{j[\ell]}, \boldsymbol{\gamma}_{g[i[\ell]]}, \eta_{i[\ell]}, \mathbf{z}_{i[\ell]})] + \text{const.} \\
&= \log[\mathbf{1}\{y_\ell^* > 0, y_\ell = 1\} + \mathbf{1}\{y_\ell^* \leq 0, y_\ell = 0\}] \\
&\quad - \frac{1}{2}\mathbb{E}_{\tilde{\boldsymbol{\beta}},\boldsymbol{\gamma},\eta}[(y_\ell^* - \alpha_{j[\ell]} - \beta_{j[\ell]}\boldsymbol{\gamma}_{g[i[\ell]]}^\top \mathbf{z}_{i[\ell]} - \beta_{j[\ell]}\eta_{i[\ell]})^2] + \text{const.} \\
&= \log[\mathbf{1}\{y_\ell^* > 0, y_\ell = 1\} + \mathbf{1}\{y_\ell^* \leq 0, y_\ell = 0\}] \\
&\quad - \frac{1}{2}\left[y_\ell^{*2} - 2y_\ell^*\left\{\mathbb{E}(\alpha_{j[\ell]}) + \mathbb{E}(\beta_{j[l]})\mathbb{E}(\boldsymbol{\gamma}_{g[i[\ell]]})^\top \mathbf{z}_{i[\ell]} + \mathbb{E}(\eta_{i[\ell]})\mathbb{E}(\beta_{j[\ell]})\right\}\right] + \text{const.}
\end{aligned}$$

We recognize that this is a truncated normal distribution, and therefore, we obtain the variational distribution given in equation (65).

Ideal Point Error Terms. We derive the variational distribution for the ideal point error terms as follows,

$$\begin{aligned}
&\log q(\eta_n) \\
&= \mathbb{E}_{y^*,\tilde{\boldsymbol{\beta}},\boldsymbol{\gamma},\sigma^2}\left[\log p(\eta_n \mid \sigma_{g[n]}^2) + \sum_{\ell=1}^{L}\mathbf{1}\{i[\ell] = n\}\log p(y_\ell^* \mid \tilde{\boldsymbol{\beta}}_{j[\ell]}, \boldsymbol{\gamma}_{g[n]}, \mathbf{z}_n, \eta_n)\right] + \text{const.} \\
&= -\frac{1}{2}\mathbb{E}_{y^*,\tilde{\boldsymbol{\beta}},\boldsymbol{\gamma},\sigma^2}\left[\frac{\eta_n^2}{\sigma_{g[n]}^2} + \sum_{\ell=1}^{L}\mathbf{1}\{i[\ell] = n\}(y_\ell^* - \alpha_{j[\ell]} - \beta_{j[\ell]}\boldsymbol{\gamma}_{g[n]}^\top \mathbf{z}_n - \beta_{j[\ell]}\eta_n)^2\right] + \text{const.} \\
&= -\frac{1}{2}\mathbb{E}_{y^*,\tilde{\boldsymbol{\beta}},\boldsymbol{\gamma},\sigma^2}\left[\left(\sigma_{g[n]}^{-2} + \sum_{\ell=1}^{L}\mathbf{1}\{i[\ell] = n\}\beta_{j[\ell]}^2\right)\eta_n^2 - 2\sum_{\ell=1}^{L}\mathbf{1}\{i[\ell] = n\}(y_\ell^*\beta_{j[\ell]} - \alpha_{j[\ell]}\beta_{j[\ell]} - \beta_{j[\ell]}^2\boldsymbol{\gamma}_{g[n]}^\top \mathbf{z}_n)\eta_n\right] \\
&\quad + \text{const.}
\end{aligned}$$

Thus, the approximating distribution is again a normal distribution and is given by equation (67).

Item Parameters. The variational distribution for item parameters can be derived as follows,

$$\log q(\tilde{\boldsymbol{\beta}}_k)$$

$$
\begin{aligned}
&= \mathbb{E}_{y^*,\eta,\boldsymbol{\gamma}}\left[\log p(\tilde{\boldsymbol{\beta}}_k) + \sum_{\ell=1}^{L} \mathbf{1}\{j[\ell]=k\}\log p(y_\ell^* \mid \tilde{\boldsymbol{\beta}}_k, \boldsymbol{\gamma}_{g[i[\ell]]}, \eta_{i[\ell]})\right] + \text{const.}\\
&= -\frac{1}{2}\mathbb{E}_{y^*,\eta,\boldsymbol{\gamma}}\left[(\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}})^\top \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}(\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}) + \sum_{\ell=1}^{L} \mathbf{1}\{j[\ell]=k\}\{y_\ell^* - \alpha_k - \beta_k(\boldsymbol{\gamma}_{g[i[\ell]]}^\top \mathbf{z}_{i[\ell]} + \eta_{i[\ell]})\}^2\right] + \text{const.}\\
&= -\frac{1}{2}\mathbb{E}_{y^*,\eta,\boldsymbol{\gamma}}\left[(\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}})^\top \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}(\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}) + \sum_{\ell=1}^{L} \mathbf{1}\{j[\ell]=k\}(y_\ell^* - \tilde{\boldsymbol{\beta}}_k^\top \tilde{\mathbf{x}}_{i[\ell]})^2\right] + \text{const.}\\
&= -\frac{1}{2}\mathbb{E}_{y^*,\eta,\boldsymbol{\gamma}}\left[\tilde{\boldsymbol{\beta}}_k^\top\left(\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1} + \sum_{\ell=1}^{L}\mathbf{1}\{j[\ell]=k\}\tilde{\mathbf{x}}_{i[\ell]}\tilde{\mathbf{x}}_{i[\ell]}^\top\right)\tilde{\boldsymbol{\beta}}_k - 2\tilde{\boldsymbol{\beta}}_k^\top\left(\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}} + \sum_{\ell=1}^{L}\mathbf{1}\{j[\ell]=k\}y_\ell^*\tilde{\mathbf{x}}_{i[\ell]}\right)\right] + \text{const.}
\end{aligned}
$$

where $\tilde{\mathbf{x}}_{i[\ell]} = (1, \boldsymbol{\gamma}_{g[i[\ell]]}^\top \mathbf{z}_{i[\ell]} + \eta_{i[\ell]})^\top$ is a 2 dimensional vector. Thus, the variational distribution is given by equation (68).

Group-level Coefficients. The derivation of the variational distribution for group-level coefficients is given as follows,

$$
\begin{aligned}
&\log q(\boldsymbol{\gamma}_m)\\
&= \mathbb{E}_{y^*,\eta,\tilde{\boldsymbol{\beta}}}\left[\log p(\boldsymbol{\gamma}_m) + \sum_{\ell=1}^{L}\mathbf{1}\{g[i[\ell]]=m\}\log p(y_\ell^* \mid \tilde{\boldsymbol{\beta}}_{j[\ell]}, \boldsymbol{\gamma}_m, \eta_{i[\ell]})\right] + \text{const.}\\
&= -\frac{1}{2}\mathbb{E}_{y^*,\eta,\tilde{\boldsymbol{\beta}}}\left[(\boldsymbol{\gamma}_m - \boldsymbol{\mu}_{\boldsymbol{\gamma}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}(\boldsymbol{\gamma}_m - \boldsymbol{\mu}_{\boldsymbol{\gamma}}) + \sum_{\ell=1}^{L} 1\{g[i[\ell]]=m\}\left\{y_\ell^* - \alpha_{j[\ell]} - \beta_{j[\ell]}(\boldsymbol{\gamma}_m^\top \mathbf{z}_{i[\ell]} + \eta_{i[\ell]})\right\}^2\right] + \text{const.}\\
&= -\frac{1}{2}\mathbb{E}_{y^*,\eta,\tilde{\boldsymbol{\beta}}}\left[\boldsymbol{\gamma}_m^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} + \sum_{\ell=1}^{L} 1\{g[i[\ell]]=m\}\beta_{j[\ell]}^2 \mathbf{z}_{i[\ell]}\mathbf{z}_{i[\ell]}^\top\right)\boldsymbol{\gamma}_m\right.\\
&\qquad\left. -2\boldsymbol{\gamma}_m^\top\left\{\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\gamma}} + \sum_{\ell=1}^{L} 1\{g[i[\ell]]=m\}\mathbf{z}_{i[\ell]}\beta_{j[\ell]}(y_\ell^* - \alpha_{j[\ell]} - \beta_{j[\ell]}\eta_{i[\ell]})\right\}\right] + \text{const.}
\end{aligned}
$$

Thus, the variational distribution is given by equation (70).

Group-level Variances. Finally, we derive the variational distribution for the group-level variance parameter σ_m^2.

$$
\begin{aligned}
\log q(\sigma_m^2) &= \mathbb{E}_\eta\left[\log p(\sigma_m^2) + \sum_{n=1}^{N}\mathbf{1}\{g[n]=m\}\log p(\eta_n \mid \sigma_m^2)\right] + \text{const.}\\
&= \mathbb{E}_\eta\left[-\left(\frac{\nu_\sigma + \sum_{n=1}^{N}\mathbf{1}\{g[n]=m\}}{2} + 1\right)\log\sigma_m^2 - \frac{1}{2\sigma_m^2}\left(s_\sigma^2 + \sum_{n=1}^{N}\mathbf{1}\{g[n]=m\}\eta_n^2\right)\right] + \text{const.}
\end{aligned}
$$

Thus, the approximating distribution is the inverse-gamma distribution given in equation (71).

E Local Variational Inference for the Generalized Wordfish Model

For the generalized Wordfish model, the joint posterior distribution is given by equation (46). We derive the variational EM algorithm under the factorization assumption given in equation (47). One feature of this variational EM algorithm, when compared with those for the other models considered in this paper, is the use of local variational inference. As a result, we have an additional local variational parameter ξ_{jk} to be optimized in our algorithm. For computational efficiency, we update this local variational parameter only once at each iteration. While this approximation can be optimized when updating each parameter, we find that this makes little improvement in practice.

Document verboseness. We derive the variational distribution for the document verboseness parameter ψ_k.

$$\log q(\psi_k \mid \mathbf{Y}) \;=\; \mathbb{E}_{\tilde{\boldsymbol{\beta}}_j, x_{i[k]}}\left[\sum_{j=1}^{J}\left\{y_{jk}\psi_k - \exp\left(\psi_k + \tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{x}}_{i[k]}\right)\right\} - \frac{1}{2\sigma_\psi^2}(\psi_k^2 - 2\psi_k\mu_\psi)\right] + \text{const.}$$

where $\tilde{\mathbf{x}}_{i[k]} = (1, x_{i[k]})$. We observe that the existence of the exponential function makes it impossible to obtain a closed form expression for the integral. We first employ the local variational inference by introducing an additional unidentified parameter ξ_{jk},

$$\log q(\psi_k) \;=\; \mathbb{E}_{\tilde{\boldsymbol{\beta}}_j, x_{i[k]}}\left[\sum_{j=1}^{J}\left\{y_{jk}\psi_k - \exp(\xi_{jk})\exp\left(-\xi_{jk} + \psi_k + \tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{x}}_{i[k]}\right)\right\} - \frac{1}{2\sigma_\psi^2}(\psi_k^2 - 2\psi_k\mu_\psi)\right] + \text{const.}$$

We then apply the second-order approximation to the exponential function, i.e., $\exp(x) \approx 1+x+x^2/2$. Note that this approximation is most accurate around $x = 0$. Hence, we set this local variational parameter ξ_{jk} to the current value of $\psi_k + \tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{x}}_{i[k]}$. This yields,

$$\begin{aligned}
&\log q(\psi_k)\\
\approx\;& \mathbb{E}_{\tilde{\boldsymbol{\beta}}_j, x_{i[k]}}\Bigg[\sum_{j=1}^{J}\Bigg\{y_{jk}\psi_k - \exp(\xi_{jk})\Bigg(1 - \xi_{jk} + \psi_k + \tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{x}}_{i[k]} + \frac{\xi_{jk}^2}{2} - \xi_{jk}\psi_k - \xi_{jk}\tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{x}}_{i[k]} + \frac{\psi_k^2}{2} + \psi_k\tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{x}}_{i[k]}\\
&\qquad + \frac{\tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{x}}_{i[k]}\tilde{\mathbf{x}}_{i[k]}^\top \tilde{\boldsymbol{\beta}}_j}{2}\Bigg)\Bigg\} - \frac{1}{2\sigma_\psi^2}(\psi_k^2 - 2\psi_k\mu_\psi)\Bigg] + \text{const.}\\
=\;& \mathbb{E}_{\tilde{\boldsymbol{\beta}}_j, x_{i[k]}}\left[-\frac{1}{2}\sum_{j=1}^{J}\left\{\exp(\xi_{jk})\psi_k^2 - 2\psi_k\left[y_{jk} - \exp(\xi_{jk})\left(1 - \xi_{jk} + \tilde{\boldsymbol{\beta}}_j^\top \tilde{\mathbf{x}}_{i[k]}\right)\right]\right\} - \frac{1}{2\sigma_\psi^2}(\psi_k^2 - 2\psi_k\mu_\psi)\right] + \text{const.}
\end{aligned}$$

Thus, the variational distribution is given by,

$$q(\{\psi_k\}_{k=1}^K) \;=\; \prod_{k=1}^{K} q(\psi_k) \;\approx\; \prod_{k=1}^{K}\mathcal{N}(A_k^{-1}a_k, A_k^{-1})$$

where $a_k = \sum_{j=1}^{J}[y_{jk} - \exp(\xi_{jk})\{1 - \xi_{jk} + \mathbb{E}(\alpha_j) + \mathbb{E}(\beta_j)\mathbb{E}(x_{i[k]})\}] + \sigma_\psi^{-2}\mu_\psi$ and $A_k = \sum_{j=1}^{J}\exp(\xi_{jk}) + \sigma_\psi^{-2}$.

Ideal Points. The update for ideal points proceeds in the same exact manner as that for the document verboseness parameter. That is, we introduce the local variational parameter ξ_{jk}. Then, we have the following approximation for the variational distribution for ideal points,

$$\log q(x_n)$$

$$\approx \mathbb{E}_{\psi_k,\tilde{\boldsymbol{\beta}}_j}\Bigg[\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbf{1}\{i[k]=n\}\Bigg\{y_{jk}\beta_j x_n - \exp(\xi_{jk})\Bigg(1-\xi_{jk}+\psi_k+\alpha_j+\beta_j x_n+\frac{\xi_{jk}^2}{2}-\xi_{jk}\psi_k$$

$$-\xi_{jk}\alpha_j-\xi_{jk}\beta_j x_n+\frac{\psi_k^2}{2}+\psi_k\alpha_j+\psi_k\beta_j x_n+\frac{\alpha_j^2}{2}+\alpha_j\beta_j x_n+\frac{\beta_j^2 x_n^2}{2}\Bigg)\Bigg\}-\frac{1}{2\sigma_x^2}(x_n^2-2x_n\mu_x)\Bigg]+\text{const.}$$

$$= \mathbb{E}_{\psi_k,\tilde{\boldsymbol{\beta}}_j}\Bigg[-\frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbf{1}\{i[k]=n\}\left\{\exp(\xi_{jk})\beta_j^2 x_n^2-2\beta_j x_n(y_{jk}-\exp(\xi_{jk})(1+\alpha_j-\xi_{jk}+\psi_k))\right\}$$

$$-\frac{1}{2\sigma_x^2}(x_n^2-2x_n\mu_x)\Bigg]+\text{const.}$$

Thus, the variational distribution is given by,

$$q(\{x_n\}_{n=1}^N) \;=\; \prod_{n=1}^{N} q(x_n) \;\approx\; \prod_{n=1}^{N}\mathcal{N}(B_n^{-1}b_n, B_n^{-1})$$

where $b_n = \sum_{j=1}^{J}\sum_{k=1}^{K}\mathbf{1}\{i[k]=n\}\mathbb{E}(\beta_j)\{y_{jk}-\exp(\xi_{jk})(1+\mathbb{E}(\alpha_j)-\xi_{jk}+\mathbb{E}(\psi_k))\}+\sigma_x^{-2}\mu_x$ and $B_n = \sum_{j=1}^{J}\sum_{k=1}^{K}\mathbf{1}\{i[k]=n\}\exp(-\xi_{jk})/\mathbb{E}(\beta_j^2)+\sigma_x^{-2}$.

Term Parameters. Finally, the update for the term parameters $\tilde{\boldsymbol{\beta}}_j$ proceeds similarly. The variational distribution is given by,

$$\log q(\tilde{\boldsymbol{\beta}}_j) \;=\; \mathbb{E}_{\psi_k,x_{i[k]}}\left[\sum_{k=1}^{K}\left\{y_{jk}\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_{i[k]}-\exp\left(\psi_k+\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_{i[k]}\right)\right\}-\frac{1}{2}(\tilde{\boldsymbol{\beta}}_j-\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}})^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}(\tilde{\boldsymbol{\beta}}_j^\top-\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}})\right]+\text{const.}$$

Again, we consider the second-order approximation through local variational inference.

$$\log q(\tilde{\boldsymbol{\beta}}_j)$$

$$\approx \mathbb{E}_{\psi_k,x_{i[k]}}\Bigg[\sum_{k=1}^{K}\Bigg\{y_{jk}\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_{i[k]}-\exp(\xi_{jk})\Bigg(1-\xi_{jk}+\psi_k+\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_{i[k]}+\frac{\xi_{jk}^2}{2}-\xi_{jk}\psi_k-\xi_{jk}\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_{i[k]}+\frac{\psi_k^2}{2}$$

$$+\psi_k\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_{i[k]}+\frac{\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_{i[k]}\tilde{\mathbf{x}}_{i[k]}^\top\tilde{\boldsymbol{\beta}}_j}{2}\Bigg)\Bigg\}-\frac{1}{2}(\tilde{\boldsymbol{\beta}}_j-\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}})^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}(\tilde{\boldsymbol{\beta}}_j^\top-\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}})\Bigg]+\text{const.}$$

$$= \mathbb{E}_{\psi_k,x_{i[k]}}\Bigg[-\frac{1}{2}\sum_{k=1}^{K}\left\{\exp(\xi_{jk})\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_{i[k]}\tilde{\mathbf{x}}_{i[k]}^\top\tilde{\boldsymbol{\beta}}_j-2\left[y_{jk}-\exp(\xi_{jk})\left(1-\xi_{jk}+\psi_k\right)\right]\tilde{\boldsymbol{\beta}}_j^\top\tilde{\mathbf{x}}_{i[k]}\right\}$$

$$-\frac{1}{2}(\tilde{\boldsymbol{\beta}}_j-\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}})^\top\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}(\tilde{\boldsymbol{\beta}}_j-\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}})\Bigg]+\text{const.}$$

Thus, the variational distribution is given by,

$$q(\{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^J) \;=\; \prod_{j=1}^{J} q(\tilde{\boldsymbol{\beta}}_j) \;\approx\; \prod_{j=1}^{J}\mathcal{N}(\mathbf{C}_j^{-1}\mathbf{c}_j, \mathbf{C}_j^{-1})$$

where $\mathbf{c}_j = \{y_{jk}-\exp(\xi_{jk})(1-\xi_{jk}+\mathbb{E}(\psi_k))\}\mathbb{E}(\tilde{\mathbf{x}}_{i[k]})+\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}$ and $\mathbf{C}_j = \sum_{k=1}^{K}\exp(\xi_{jk})\mathbb{E}(\tilde{\mathbf{x}}_{i[k]}\tilde{\mathbf{x}}_{i[k]}^\top)+\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}$.

F Approximate Variational Inference for the Network Ideal Point Model

Latent Propensity. This update step is similar to the latent propensity update step for the standard ideal point model (see Appendix A).

$$
\begin{aligned}
& \log q(\mathbf{Y}^*) \\
= \; & \mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{x},\mathbf{z}} \left[\sum_{i=1}^{N}\sum_{j=1}^{J} \log\left(\mathbf{1}\{y_{ij}^* > 0\}\mathbf{1}\{y_{ij}=1\} + \mathbf{1}\{y_{ij}^* \leq 0\}\mathbf{1}\{y_{ij}=0\}\right) - \frac{1}{2}\left\{y_{ij}^* - \alpha_j - \beta_i + (x_i - z_j)^2\right\}^2 \right] \\
& +\text{const.} \\
= \; & \sum_{i=1}^{N}\sum_{j=1}^{J} \log\left(\mathbf{1}\{y_{ij}^* > 0\}\mathbf{1}\{y_{ij}=1\} + \mathbf{1}\{y_{ij}^* \leq 0\}\mathbf{1}\{y_{ij}=0\}\right) \\
& -\frac{1}{2}\left[(y_{ij}^*)^2 - 2y_{ij}^*\{\mathbb{E}(\alpha_j) + \mathbb{E}(\beta_i) - \mathbb{E}(x_i^2) - \mathbb{E}(z_j^2) + 2\mathbb{E}(x_i)\mathbb{E}(z_j)\}\right] + \text{const.}
\end{aligned}
$$

Thus, the variational distribution is given by the following truncated normal distribution,

$$
q(\mathbf{Y}^*) \;=\; \prod_{i=1}^{N}\prod_{j=1}^{J} q(y_{ij}^*) \quad \text{where} \quad q(y_{ij}^*) \;=\; \begin{cases} \mathcal{TN}(m_{ij}, 1, 0, \infty) & \text{if } y_{ij} = 1 \\ \mathcal{TN}(m_{ij}, 1, -\infty, 0) & \text{if } y_{ij} = 0 \end{cases}
$$

where $m_{ij} = \mathbb{E}(\alpha_j) + \mathbb{E}(\beta_i) - \mathbb{E}(x_i^2) - \mathbb{E}(z_j^2) + 2\mathbb{E}(x_i)\mathbb{E}(z_j)$.

User-specific Intercept.

$$
\begin{aligned}
& \log q(\beta_i) \\
= \; & \mathbb{E}_{\mathbf{Y}^*,\boldsymbol{\alpha},\mathbf{x},\mathbf{z}} \left[\sum_{j=1}^{J} -\frac{1}{2}\left\{\beta_i - y_{ij}^* + \alpha_j - (x_i - z_j)^2\right\}^2 - \frac{1}{2\sigma_\beta^2}(\beta_i - \mu_\beta)^2 \right] + \text{const.} \\
= \; & -\frac{1}{2}\left[\sum_{j=1}^{J} \left\{\beta_i^2 - 2\beta_i\left(\mathbb{E}(y_{ij}^*) - \mathbb{E}(\alpha_j) + \mathbb{E}(x_i^2) - 2\mathbb{E}(x_i)\mathbb{E}(z_j) + \mathbb{E}(z_j^2)\right)\right\} + \frac{1}{\sigma_\beta^2}\left(\beta_i^2 - 2\beta_i\mu_\beta\right) \right] + \text{const.} \\
= \; & -\frac{1}{2}\left[\left(J + \frac{1}{\sigma_\beta^2}\right)\beta_i^2 - 2\beta_i \left\{ \frac{\mu_\beta}{\sigma_\beta^2} + \sum_{j=1}^{J}\left(\mathbb{E}(y_{ij}^*) - \mathbb{E}(\alpha_j) + \mathbb{E}(x_i^2) - 2\mathbb{E}(x_i)\mathbb{E}(z_j) + \mathbb{E}(z_j^2)\right) \right\} \right] + \text{const.}
\end{aligned}
$$

Thus, the variational distribution is given by,

$$
q(\boldsymbol{\beta}) \;=\; \prod_{i=1}^{N} q(\beta_i) \quad \text{where} \quad \mathcal{N}(B^{-1}b_i,\ B^{-1})
$$

where $B = J + 1/\sigma_\beta^2$ and $b_i = \mu_\beta/\sigma_\beta^2 + \sum_{j=1}^{J}\left(\mathbb{E}(y_{ij}^*) - \mathbb{E}(\alpha_j) + \mathbb{E}(x_i^2) - 2\mathbb{E}(x_i)\mathbb{E}(z_j) + \mathbb{E}(z_j^2)\right)$.

Ideal Points for Users. This update step does not correspond to a standard distribution. Therefore, we employ the approximation via Taylor expansion.

$$
\log q(x_i) \;=\; -\frac{1}{2}\mathbb{E}_{\mathbf{Y}^*,\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{z}} \left[\sum_{j=1}^{J} \left\{y_{ij}^* - \alpha_j - \beta_i + (x_i - z_j)^2\right\}^2 + \frac{1}{\sigma_x^2}(x_i^2 - 2x_i\mu_x) \right] + \text{const.}
$$

Apply the second-order Taylor expansion around the current value of x_i denoted by $\hat{x}_i$, yielding,

$$\begin{aligned}\log q(x_i) &\approx -\frac{1}{2}\mathbb{E}\left[f(\hat{x}_i)+f'(\hat{x}_i)(x_i-\hat{x}_i)+\frac{1}{2}f''(\hat{x}_i)(x_i-\hat{x}_i)^2\right]+\text{const.}\\ &= -\frac{1}{2}\mathbb{E}\left[\frac{1}{2}f''(\hat{x}_i)x_i^2-\left(f''(\hat{x}_i)\hat{x}_i-f'(\hat{x}_i)\right)x_i\right]+\text{const.}\end{aligned}$$

where

$$\begin{aligned}f(x) &= \frac{1}{\sigma_x^2}(x^2-2x\mu_x)+\sum_{j=1}^{J}\{y_{ij}^*-\alpha_j-\beta_i+(x-z_j)^2\}^2\\ f'(x) &= \frac{2}{\sigma_x^2}(x-\mu_x)+4\sum_{j=1}^{J}(x-z_j)\left\{y_{ij}^*-\alpha_j-\beta_i+(x-z_j)^2\right\}\\ f''(x) &= \frac{2}{\sigma_x^2}+4\sum_{j=1}^{J}\left\{y_{ij}^*-\alpha_j-\beta_i+3(x-z_j)^2\right\}\end{aligned}$$

Therefore, we arrive at the following variational distribution,

$$q(\mathbf{x}) = \prod_{i=1}^{N}q(x_i) \approx \prod_{i=1}^{N}\mathcal{N}(D_i^{-1}d_i, D_i^{-1})$$

where $D_i = \mathbb{E}(f''(\hat{x}_i))/2$ and $d_i = \{\mathbb{E}(f''(\hat{x}_i))\hat{x}_i - \mathbb{E}(f'(\hat{x}_i))\}/2$. The relevant expectations are given by,

$$\begin{aligned}\mathbb{E}(f'(x)) &= \frac{2}{\sigma_x^2}(x-\mu_x)\\ &\quad +4\sum_{j=1}^{J}\left\{(x-\mathbb{E}(z_j))\left(\mathbb{E}(y_{ij}^*)-\mathbb{E}(\alpha_j)-\mathbb{E}(\beta_i)\right)+\left(x^3-3x^2\mathbb{E}(z_j)+3x\mathbb{E}(z_j^2)-\mathbb{E}(z_j^3)\right)\right\}\\ \mathbb{E}(f''(x)) &= \frac{2}{\sigma_x^2}+4\sum_{j=1}^{J}\left\{\left(\mathbb{E}(y_{ij}^*)-\mathbb{E}(\alpha_j)-\mathbb{E}(\beta_i)\right)+3\left(x^2-2x\mathbb{E}(z_j)+\mathbb{E}(z_j^2)\right)\right\}\end{aligned}$$

Politician-specific Intercept. This step is analogous to the user-specific intercept update.

$$\begin{aligned}&\log q(\alpha_j)\\ &= -\frac{1}{2}\mathbb{E}_{\mathbf{Y}^*,\boldsymbol{\beta},\mathbf{x},\mathbf{z}}\left[\sum_{i=1}^{N}\left\{\alpha_j-\left(y_{ij}^*-\beta_i+(x_i-z_j)^2\right)\right\}^2+\frac{1}{\sigma_\alpha^2}(\alpha_j-\mu_\alpha)^2\right]+\text{const.}\\ &= -\frac{1}{2}\sum_{i=1}^{N}\left[\alpha_j^2-2\alpha_j\left\{\mathbb{E}(y_{ij}^*)-\mathbb{E}(\beta_i)+(\mathbb{E}(x_i^2)-2\mathbb{E}(x_i)\mathbb{E}(z_j)+\mathbb{E}(z_j^2))\right\}\right]-\frac{1}{2\sigma_\alpha^2}\left(\alpha_j^2-2\alpha_j\mu_\alpha\right)+\text{const.}\\ &= -\frac{1}{2}\left[\left(N+\frac{1}{\sigma_\alpha^2}\right)\alpha_j^2-2\alpha_j\left(\frac{\mu_\alpha}{\sigma_\alpha^2}+\sum_{i=1}^{N}\left(\mathbb{E}(y_{ij}^*)-\mathbb{E}(\beta_i)+(\mathbb{E}(x_i^2)-2\mathbb{E}(x_i)\mathbb{E}(z_j)+\mathbb{E}(z_j^2))\right)\right)\right]+\text{const.}\end{aligned}$$

Thus, the variational distribution is given by,

$$q(\boldsymbol{\alpha}) = \prod_{j=1}^{J}q(\alpha_j) = \prod_{j=1}^{J}\mathcal{N}(A^{-1}a_j, A^{-1})$$

where $A = N + 1/\sigma_\alpha^2$ and $a_j = \mu_\alpha/\sigma_\alpha^2 + \sum_{i=1}^{N}\left(\mathbb{E}(y_{ij}^*)-\mathbb{E}(\beta_i)+(\mathbb{E}(x_i^2)-2\mathbb{E}(x_i)\mathbb{E}(z_j)+\mathbb{E}(z_j^2))\right)$.

Ideal Points for Politicians. This step is analogous to the update for the ideal points of users and does not correspond to a standard distribution. Therefore, we apply the second-order Taylor series approximation.

$$\log q(z_j) \;=\; -\frac{1}{2}\mathbb{E}_{\mathbf{Y}^*,\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{x}}\left[\sum_{i=1}^{N}\left(y_{ij}^* - \alpha_j - \beta_i + (x_i - z_j)^2\right)^2 + \frac{1}{\sigma_z^2}(z_j^2 - 2z_j\mu_z)\right] + \text{const.}$$

Apply the second-order Taylor expansion around the current value of z_j denoted by $\hat{z}_j$, yielding,

$$\log q(z_j) \;\approx\; -\frac{1}{2}\mathbb{E}\left[\frac{1}{2}g''(\hat{z}_j)z_j^2 - \left(g''(\hat{z}_j)\hat{z}_j - g'(\hat{z}_j)\right)z_j\right] + \text{const.}$$

where

$$\begin{aligned}
g(z) &= \frac{1}{\sigma_z^2}(z^2 - 2z\mu_z) + \sum_{i=1}^{N}\left\{y_{ij}^* - \alpha_j - \beta_i + (x_i - z)^2\right\}^2 \\
g'(z) &= \frac{2}{\sigma_z^2}(z - \mu_z) - 4\sum_{i=1}^{N}(x_i - z)\left\{y_{ij}^* - \alpha_j - \beta_i + (x_i - z)^2\right\} \\
g''(z) &= \frac{2}{\sigma_z^2} + 4\sum_{i=1}^{N}\left\{y_{ij}^* - \alpha_j - \beta_i + 3(x_i - z)^2\right\}
\end{aligned}$$

Therefore, we arrive at the following variational distribution,

$$q(\mathbf{z}) \;=\; \prod_{j=1}^{J} q(z_j) \;\approx\; \prod_{j=1}^{J}\mathcal{N}(E_j^{-1}e_j, E_j^{-1})$$

where $E_j = \mathbb{E}(g''(\hat{z}_j))/2$ and $e_j = \{\mathbb{E}(g''(\hat{z}_j))\hat{z}_j - \mathbb{E}(g'(\hat{z}_j))\}/2$. The relevant expectations are given by,

$$\begin{aligned}
\mathbb{E}(g'(z)) &= \frac{2}{\sigma_z^2}(z - \mu_z) \\
&\quad -4\sum_{i=1}^{N}\left\{(\mathbb{E}(x_i) - z)\left(\mathbb{E}(y_{ij}^*) - \mathbb{E}(\alpha_j) - \mathbb{E}(\beta_i)\right) + \left(\mathbb{E}(x_i^3) - 3\mathbb{E}(x_i^2)z + 3\mathbb{E}(x_i)z^2 - z^3\right)\right\} \\
\mathbb{E}(g''(z)) &= \frac{2}{\sigma_z^2} + 4\sum_{i=1}^{N}\left\{\left(\mathbb{E}(y_{ij}^*) - \mathbb{E}(\alpha_j) - \mathbb{E}(\beta_i)\right) + 3\left(\mathbb{E}(x_i^2) - 2\mathbb{E}(x_i)z + z^2\right)\right\}
\end{aligned}$$