Cosine basis

Daeyoung Lim*
Department of Statistics
Korea University

July 2, 2016

1 Criteo Dataset

1.1 File descriptions

- **train.csv** - The training set consists of a portion of Criteo's traffic over a period of 7 days. Each row corresponds to a display ad served by Criteo. Positive (clicked) and negatives (non-clicked) examples have both been subsampled at different rates in order to reduce the dataset size. The examples are chronologically ordered.
- **test.csv** - The test set is computed in the same way as the training set but for events on the day following the training period.
- **random_submission.csv** - A sample submission file in the correct format.

1.2 Data fields

- **Label** - Target variable that indicates if an ad was clicked (1) or not (0).
- **I1 - I13** - A total of 13 columns of integer features (mostly count features).
- **C1 - C26** - A total of 26 columns of categorical features. The values of these features have been hashed onto 32 bits for anonymization purposes.

The semantics of the features is undisclosed. When a value is missing, the field is empty.

2 Normal

2.1 Model specifications

$$
\begin{aligned}
y_i &= w_i'\beta + f\left(x_i\right) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right) \\
\theta_j | \sigma, \tau, \gamma &\sim \mathcal{N}\left(0, \sigma^2\tau^2 \exp\left[-j\gamma\right]\right) \\
\tau^2 &\sim \mathrm{IG}\left(\frac{r_{0,\tau}}{2}, \frac{s_{0,\tau}}{2}\right) \\
\sigma^2 &\sim \mathrm{IG}\left(\frac{r_{0,\sigma}}{2}, \frac{s_{0,\sigma}}{2}\right) \\
\beta &\sim \mathcal{N}\left(\mu_\beta^0, \Sigma_\beta^0\right) \\
\gamma &\sim \mathrm{Exp}\left(\omega_0\right) \\
|\psi| &= \gamma, \quad \psi \sim \mathrm{DE}\left(0, \omega_0\right) \\
\varphi_j\left(x\right) &= \sqrt{2}\cos\left(\pi j x\right)
\end{aligned}
$$

*Prof. Taeryon Choi

Joint density:

$$p\left(y,\Theta\right)=\mathcal{N}\left(y\middle|W\beta+f_J,\sigma^2 I_n\right)\left\{\prod_{j=1}^{J}\mathcal{N}\left(\theta_j\middle|0,\sigma^2\tau^2\exp\left[-j\left|\psi\right|\right]\right)\right\}\mathrm{IG}\left(\tau^2\middle|\frac{r_{0,\tau}}{2},\frac{s_{0,\tau}}{2}\right)\mathrm{IG}\left(\sigma^2\middle|\frac{r_{0,\sigma}}{2},\frac{s_{0,\sigma}}{2}\right)\mathcal{N}\left(\beta\middle|\mu_\beta^0,\Sigma_\beta^0\right)$$
$$\mathrm{DE}\left(\psi|0,\omega_0\right)$$

We will use the joint density to derive the LB and updating algorithm. The variational distributions are

$$\begin{aligned}
q_1\left(\beta\right)&=\mathcal{N}\left(\mu_\beta^q,\Sigma_\beta^q\right)\\
q_2\left(\theta_J\right)&=\mathcal{N}\left(\mu_\theta^q,\Sigma_\theta^q\right)\\
q_3\left(\sigma^2\right)&=\mathrm{IG}\left(\frac{r_{q,\sigma}}{2},\frac{s_{q,\sigma}}{2}\right)\\
q_4\left(\tau^2\right)&=\mathrm{IG}\left(\frac{r_{q,\tau}}{2},\frac{s_{q,\tau}}{2}\right)\\
q_5\left(\psi\right)&=\mathcal{N}\left(\mu_\psi^q,\sigma_\psi^{2\,q}\right)\quad\text{(NCVMP)}.
\end{aligned}$$

2.2 Lower bound

2.2.1 LB: $\mathrm{E}\left[\ln p\left(y|\Theta\right)\right]$

$$\begin{aligned}
\mathrm{E}\left[\ln p\left(y|\Theta\right)\right]&=-\frac{n}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2}\mathrm{E}\left[\left(y-W\beta-\varphi_J\theta\right)'\left(y-W\beta-\varphi_J\theta\right)\right]\\
&=-\frac{n}{2}\ln\left(2\pi\sigma^2\right)-\frac{1}{2}\left(y-W\mu_\beta^q-\varphi_J\mu_\theta^q\right)'\left(y-W\mu_\beta^q-\varphi_J\mu_\theta^q\right)-\frac{1}{2}\left(\mathrm{Tr}\left(W'W\Sigma_\beta^q\right)+\mathrm{Tr}\left(\varphi_J'\varphi_J\Sigma_\theta^q\right)\right)
\end{aligned}$$

2.3 LB: $\mathrm{E}\left[\ln p\left(\theta_j|\sigma,\tau,\psi\right)\right]$

$$\sum_{j=1}^{J}\mathrm{E}\left[\ln p\left(\theta_j|\sigma,\tau,\psi\right)\right]=\sum_{j=1}^{J}\mathrm{E}\left[-\frac{1}{2}\ln\left(2\pi\right)+\ln\frac{1}{\sigma}+\ln\frac{1}{\tau}+\frac{j}{2}\left|\psi\right|-\frac{\theta_j^2e^{j|\psi|}}{2\sigma^2\tau^2}\right]$$

Let's note the following fact: if $X\sim\mathcal{N}\left(\mu,\sigma^2\right)$, then $|X|\sim$ folded-Normal$\left(\mu,\sigma^2\right)$. Then,

$$\begin{aligned}
\mathrm{E}\left|X\right|&=\sigma\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}+\mu\left(1-2\Phi\left(\frac{-\mu}{\sigma}\right)\right)\\
&=\sigma\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}-\mu\,\mathrm{erf}\left(\frac{-\mu}{\sqrt{2\sigma^2}}\right)\\
\mathrm{E}\,e^{t|X|}&=\exp\left\{\frac{\sigma^2t^2}{2}+\mu t\right\}\left[1-\Phi\left(-\frac{\mu}{\sigma}-\sigma t\right)\right]+\exp\left\{\frac{\sigma^2t^2}{2}-\mu t\right\}\left[1-\Phi\left(\frac{\mu}{\sigma}-\sigma t\right)\right].
\end{aligned}$$

3 Probit: No Restriction

3.1 Model specifications

$$
\begin{aligned}
\Pr\left(y_i = 1 | f, \beta\right) &= \Phi\left(w_i'\beta + f\left(x_i\right)\right) \\
y_i^* &= w_i'\beta + f\left(x_i\right) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}\left(0, 1\right) \\
y_i &= \begin{cases} 1, & \text{if } y_i^* \geq 0 \\ 0, & \text{if } y_i^* < 0 \end{cases} \\
\theta_j | \sigma, \tau, \gamma &\sim \mathcal{N}\left(0, \sigma^2\tau^2 \exp\left[-j\gamma\right]\right) \\
\tau^2 &\sim \text{IG}\left(\frac{r_{0,\tau}}{2}, \frac{s_{0,\tau}}{2}\right) \\
\sigma^2 &\sim \text{IG}\left(\frac{r_{0,\sigma}}{2}, \frac{s_{0,\sigma}}{2}\right) \\
\beta &\sim \mathcal{N}\left(\mu_\beta^0, \sigma^2\Sigma_\beta^0\right) \qquad (p \times 1) \\
\gamma &\sim \text{Exp}\left(\omega_0\right) \\
|\psi| &= \gamma, \quad \psi \sim \text{DE}\left(0, \omega_0\right) \\
\varphi_j\left(x\right) &= \sqrt{2}\cos\left(\pi j x\right)
\end{aligned}
$$

Joint density:

$$
\begin{aligned}
p\left(y, y^*, \Theta\right) = C &\left\{\prod_{j=1}^{J} \mathcal{N}\left(\theta_j \middle| 0, \sigma^2\tau^2 \exp\left[-j\left|\psi\right|\right]\right)\right\} \text{IG}\left(\tau^2 \middle| \frac{r_{0,\tau}}{2}, \frac{s_{0,\tau}}{2}\right) \text{IG}\left(\sigma^2 \middle| \frac{r_{0,\sigma}}{2}, \frac{s_{0,\sigma}}{2}\right) \mathcal{N}\left(\beta \middle| \mu_\beta^0, \sigma^2\Sigma_\beta^0\right) \\
&\text{DE}\left(\psi | 0, \omega_0\right) \left\{\prod_{i=1}^{n} \left(1\left[y_i^* \geq 0\right] 1\left[y_i = 1\right] + 1\left[y_i^* < 0\right] 1\left[y_i = 0\right]\right) \phi\left(y_i^* - w_i'\beta - \varphi_i'\theta_J\right)\right\}
\end{aligned}
$$

where C is the normalizing constant. The variational distributions are

$$
\begin{aligned}
q_1\left(\beta\right) &= \mathcal{N}\left(\mu_\beta^q, \Sigma_\beta^q\right) \\
q_2\left(\theta_J\right) &= \mathcal{N}\left(\mu_\theta^q, \Sigma_\theta^q\right) \\
q_3\left(\sigma^2\right) &= \text{IG}\left(\frac{r_{q,\sigma}}{2}, \frac{s_{q,\sigma}}{2}\right) \\
q_4\left(\tau^2\right) &= \text{IG}\left(\frac{r_{q,\tau}}{2}, \frac{s_{q,\tau}}{2}\right) \\
q_5\left(\psi\right) &= \mathcal{N}\left(\mu_\psi^q, \sigma_\psi^{2\,q}\right) \qquad \text{(NCVMP)} \\
q_6\left(y^*\right) &= \mathcal{TN}\left(\mu_{y^*}^q, I_n, 0\right)
\end{aligned}
$$

3.2 Lower bound

3.2.1 LB: $\mathrm{E}\left[\ln p\left(y^*|\mathbf{rest}\right)\right] + \mathrm{H}\left[y^*\right]$

$$
\begin{aligned}
\mathrm{E}\left[\ln p\left(y^*|\mathrm{rest}\right)\right] + \mathrm{H}\left[y^*\right] &= \sum_{i=1}^{n} \mathrm{E}\left[\ln\phi\left(y_i^* - w_i'\beta - \varphi_i'\theta_J\right) - \ln\phi\left(y_i^* - w_i'\mu_\beta^q - \varphi_i'\mu_\theta^q\right)\right] \\
&\quad + \sum_{i=1}^{n} \ln\left(\left\{\Phi\left(w_i'\mu_\beta^q + \varphi_i'\mu_\theta^q\right)\right\}^{y_i}\left\{1 - \Phi\left(w_i'\mu_\beta^q + \varphi_i'\mu_\theta^q\right)\right\}^{1-y_i}\right) \\
&= -\frac{1}{2}\left(\mathrm{Tr}\left(W'W\Sigma_\beta^q\right) + \mathrm{Tr}\left(\varphi_J'\varphi_J\Sigma_\theta^q\right)\right) \\
&\quad + \sum_{i=1}^{n} \ln\left(\left\{\Phi\left(w_i'\mu_\beta^q + \varphi_i'\mu_\theta^q\right)\right\}^{y_i}\left\{1 - \Phi\left(w_i'\mu_\beta^q + \varphi_i'\mu_\theta^q\right)\right\}^{1-y_i}\right)
\end{aligned}
$$

3.2.2 LB: $\mathrm{E}\left[\ln p\left(\theta_j|\sigma,\tau,\psi\right)\right]$

Let's note the following fact: if $X \sim \mathcal{N}\left(\mu,\sigma^2\right)$, then $|X| \sim \text{folded-Normal}\left(\mu,\sigma^2\right)$. Then,

$$
\begin{aligned}
\mathrm{E}\,|X| &= \sigma\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} + \mu\left(1 - 2\Phi\left(\frac{-\mu}{\sigma}\right)\right) \\
&= \sigma\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} - \mu\,\mathrm{erf}\left(\frac{-\mu}{\sqrt{2\sigma^2}}\right) \\
\mathrm{E}\,e^{t|X|} &= \exp\left\{\frac{\sigma^2t^2}{2} + \mu t\right\}\left[1 - \Phi\left(-\frac{\mu}{\sigma} - \sigma t\right)\right] + \exp\left\{\frac{\sigma^2t^2}{2} - \mu t\right\}\left[1 - \Phi\left(\frac{\mu}{\sigma} - \sigma t\right)\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sum_{j=1}^{J}\mathrm{E}\left[\ln p\left(\theta_j|\sigma,\tau,\psi\right)\right] + \mathrm{H}\left[\theta_J\right] &= \sum_{j=1}^{J}\mathrm{E}\left[-\frac{1}{2}\ln\left(2\pi\right) + \frac{1}{2}\ln\frac{1}{\sigma^2} + \frac{1}{2}\ln\frac{1}{\tau^2} + \frac{j}{2}|\psi| - \frac{\theta_j^2 e^{j|\psi|}}{2\sigma^2\tau^2}\right] + \mathrm{H}\left[\theta_J\right] \\
&= -\frac{J}{2}\left\{\ln\left(2\pi\right) - \left(\mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right) - \ln\left(\frac{s_{q,\sigma}}{2}\right)\right) - \left(\mathrm{di}\left(\frac{r_{q,\tau}}{2}\right) - \ln\left(\frac{s_{q,\tau}}{2}\right)\right)\right\} \\
&\quad + \frac{J\left(J+1\right)}{4}\left\{\sigma_\psi^q\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\mu_\psi^{q\,2}}{2\sigma_\psi^{q\,2}}\right) + \mu_\psi^q\left(1 - 2\Phi\left(\frac{-\mu_\psi^q}{\sigma_\psi^q}\right)\right)\right\} \\
&\quad - \frac{1}{2}\frac{r_{q,\sigma}}{s_{q,\sigma}}\frac{r_{q,\tau}}{s_{q,\tau}}\left(\mathrm{Tr}\left(\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\mu_\theta^q\right)\sum_{j=1}^{J}Q_j\left(\mu_\psi^q,\sigma_\psi^{q\,2}\right) + \frac{J}{2}\left(1 + \ln\left(2\pi\right)\right) + \frac{1}{2}\ln\left|\Sigma_\theta^q\right|
\end{aligned}
$$

where

$$
\begin{aligned}
Q_j\left(\mu_\psi^q,\sigma_\psi^{q\,2}\right) &= \mathrm{E}\,e^{j|\psi|} \\
&= \exp\left\{\frac{\sigma_\psi^{q\,2}j^2}{2} + \mu_\psi^q j\right\}\left[1 - \Phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j\right)\right] + \exp\left\{\frac{\sigma_\psi^{q\,2}j^2}{2} - \mu_\psi^q j\right\}\left[1 - \Phi\left(\frac{\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j\right)\right].
\end{aligned}
$$

3.2.3 LB: $\mathrm{E}\left[\ln p\left(\tau^2\right)\right] + \mathrm{H}\left[\tau^2\right]$

$$
\begin{aligned}
\mathrm{E}\left[\ln p\left(\tau^2\right)\right] + \mathrm{H}\left[\tau^2\right] = {}& \frac{r_{0,\tau}}{2}\ln\left(\frac{s_{0,\tau}}{2}\right) - \ln\Gamma\left(\frac{r_{0,\tau}}{2}\right) + \left(\frac{r_{0,\tau}}{2} - \frac{r_{q,\tau}}{2}\right)\mathrm{di}\left(\frac{r_{q,\tau}}{2}\right) - \left(\frac{r_{0,\tau}}{2} + 1\right)\ln\left(\frac{s_{q,\tau}}{2}\right) \\
& - \frac{s_{0,\tau}}{2}\frac{r_{q,\tau}}{s_{q,\tau}} + \frac{r_{q,\tau}}{2} + \ln\left(\frac{s_{q,\tau}}{2}\right) + \ln\Gamma\left(\frac{r_{q,\tau}}{2}\right)
\end{aligned}
$$

3.2.4 LB: $\mathrm{E}\left[\ln p\left(\sigma^2\right)\right] + \mathrm{H}\left[\sigma^2\right]$

$$
\begin{aligned}
\mathrm{E}\left[\ln p\left(\sigma^2\right)\right] + \mathrm{H}\left[\sigma^2\right] = \frac{r_{0,\sigma}}{2} \ln\left(\frac{s_{0,\sigma}}{2}\right) - \ln\Gamma\left(\frac{r_{0,\sigma}}{2}\right) + \left(\frac{r_{0,\sigma}}{2} - \frac{r_{q,\sigma}}{2}\right) \mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right) - \left(\frac{r_{0,\sigma}}{2} + 1\right) \ln\left(\frac{s_{q,\sigma}}{2}\right) \\
- \frac{s_{0,\sigma}}{2} \frac{r_{q,\sigma}}{s_{q,\sigma}} + \frac{r_{q,\sigma}}{2} + \ln\left(\frac{s_{q,\sigma}}{2}\right) + \ln\Gamma\left(\frac{r_{q,\sigma}}{2}\right)
\end{aligned}
$$

3.2.5 LB: $\mathrm{E}\left[\ln p\left(\beta\right)\right] + \mathrm{H}\left[\beta\right]$

$$
\begin{aligned}
\mathrm{E}\left[\ln p\left(\beta\right)\right] + \mathrm{H}\left[\beta\right] = \frac{p+1}{2} + \frac{p+1}{2}\left(\mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right) - \ln\left(\frac{s_{q,\sigma}}{2}\right)\right) + \frac{1}{2}\ln\left|\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right| \\
- \frac{1}{2}\frac{r_{q,\sigma}}{s_{q,\sigma}}\left\{\mathrm{Tr}\left(\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right) + \left(\mu_\beta^q - \mu_\beta^0\right)' \Sigma_\beta^{0\,-1}\left(\mu_\beta^q - \mu_\beta^0\right)\right\}
\end{aligned}
$$

3.2.6 LB: $\mathrm{E}\left[\ln p\left(\psi\right)\right] + \mathrm{H}\left[\psi\right]$

$$
\begin{aligned}
\mathrm{E}\left[\ln p\left(\psi\right)\right] + \mathrm{H}\left[\psi\right] = \ln\frac{\omega_0}{2} - \omega_0\left\{\sigma_\psi^q\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\mu_\psi^{q\,2}}{2\sigma_\psi^{q\,2}}\right) + \mu_\psi^q\left(1 - 2\Phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q}\right)\right)\right\} \\
+ \frac{1}{2}\ln\left(2\pi\sigma_\psi^{q\,2}\right) - \frac{1}{2}
\end{aligned}
$$

3.2.7 LB

$$
\begin{aligned}
\mathcal{L} = & -\frac{1}{2}\left(\mathrm{Tr}\left(W'W\Sigma_\beta^q\right) + \mathrm{Tr}\left(\varphi_J'\varphi_J\Sigma_\theta^q\right)\right) + \sum_{i=1}^n \ln\left(\left\{\Phi\left(w_i'\mu_\beta^q + \varphi_i'\mu_\theta^q\right)\right\}^{y_i}\left\{1 - \Phi\left(w_i'\mu_\beta^q + \varphi_i'\mu_\theta^q\right)\right\}^{1-y_i}\right) \\
& - \frac{J}{2}\left\{\ln\left(2\pi\right) - \left(\mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right) - \ln\left(\frac{s_{q,\sigma}}{2}\right)\right) - \left(\mathrm{di}\left(\frac{s_{q,\tau}}{2}\right) - \ln\left(\frac{s_{q,\tau}}{2}\right)\right)\right\} \\
& S_2\left(\mu_\psi^q, \sigma_\psi^{q\,2}\right) + \frac{J}{2}\left(1 + \ln\left(2\pi\right)\right) + \frac{1}{2}\ln\left|\Sigma_\theta^q\right| \\
& + \frac{r_{0,\tau}}{2}\ln\left(\frac{s_{0,\tau}}{2}\right) - \ln\Gamma\left(\frac{r_{0,\tau}}{2}\right) + \left(\frac{r_{0,\tau}}{2} - \frac{r_{q,\tau}}{2}\right)\mathrm{di}\left(\frac{r_{q,\tau}}{2}\right) - \left(\frac{r_{0,\tau}}{2} + 1\right)\ln\left(\frac{s_{q,\tau}}{2}\right) \\
& - \frac{s_{0,\tau}}{2}\frac{r_{q,\tau}}{s_{q,\tau}} + \frac{r_{q,\tau}}{2} + \ln\left(\frac{s_{q,\tau}}{2}\right) + \ln\Gamma\left(\frac{r_{q,\tau}}{2}\right) \\
& + \frac{r_{0,\sigma}}{2}\ln\left(\frac{s_{0,\sigma}}{2}\right) - \ln\Gamma\left(\frac{r_{0,\sigma}}{2}\right) + \left(\frac{r_{0,\sigma}}{2} - \frac{r_{q,\sigma}}{2}\right)\mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right) - \left(\frac{r_{0,\sigma}}{2} + 1\right)\ln\left(\frac{s_{q,\sigma}}{2}\right) \\
& - \frac{s_{0,\sigma}}{2}\frac{r_{q,\sigma}}{s_{q,\sigma}} + \frac{r_{q,\sigma}}{2} + \ln\left(\frac{s_{q,\sigma}}{2}\right) + \ln\Gamma\left(\frac{r_{q,\sigma}}{2}\right) \\
& + \frac{p+1}{2} + \frac{p+1}{2}\left(\mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right) - \ln\left(\frac{s_{q,\sigma}}{2}\right)\right) + \frac{1}{2}\ln\left|\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right| \\
& - \frac{1}{2}\frac{r_{q,\sigma}}{s_{q,\sigma}}\left\{\mathrm{Tr}\left(\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right) + \left(\mu_\beta^q - \mu_\beta^0\right)'\Sigma_\beta^{0\,-1}\left(\mu_\beta^q - \mu_\beta^0\right)\right\} \\
& + \ln\left(\frac{\omega_0}{2}\right) + S_1\left(\mu_\psi^q, \sigma_\psi^{q\,2}\right) + \frac{1}{2}\ln\left(2\pi\sigma_\psi^{q\,2}\right) - \frac{1}{2}
\end{aligned}
$$

3.3 Update

3.3.1 θ_j

$$
\begin{aligned}
\Sigma_\theta^q &\leftarrow \left(\varphi_J'\varphi_J + \frac{r_{q,\sigma}}{s_{q,\sigma}}\frac{r_{q,\tau}}{s_{q,\tau}}\mathrm{Dg}\left(Q_{1:J}\right)\right)^{-1} \\
\mu_\theta^q &\leftarrow \Sigma_\theta^q\varphi_J'\left(\mu_{y^*}^q - W\mu_\beta^q\right)
\end{aligned}
$$

3.3.2 τ^2

$$
\begin{aligned}
r_{q,\tau} &\leftarrow r_{0,\tau} + J \\
s_{q,\tau} &\leftarrow s_{0,\tau} + \frac{r_{q,\sigma}}{s_{q,\sigma}} \sum_{j=1}^{J} \left(\Sigma_{\theta,jj}^{q} + {\mu_{\theta,j}^{q}}^{2} \right) Q_j
\end{aligned}
$$

3.3.3 σ^2

$$
\begin{aligned}
r_{q,\sigma} &\leftarrow r_{0,\sigma} + J + p + 1 \\
s_{q,\sigma} &\leftarrow s_{0,\sigma} + \frac{r_{q,\tau}}{s_{q,\tau}} \sum_{j=1}^{J} \left(\Sigma_{\theta,jj}^{q} + {\mu_{\theta,j}^{q}}^{2} \right) Q_j + \operatorname{Tr}\left({\Sigma_{\beta}^{0}}^{-1} \Sigma_{\beta}^{q} \right) + \left(\mu_{\beta}^{q} - \mu_{\beta}^{0} \right)' {\Sigma_{\beta}^{0}}^{-1} \left(\mu_{\beta}^{q} - \mu_{\beta}^{0} \right)
\end{aligned}
$$

3.3.4 β

$$
\begin{aligned}
\Sigma_{\beta}^{q} &\leftarrow \left(\frac{r_{q,\sigma}}{s_{q,\sigma}} {\Sigma_{\beta}^{0}}^{-1} + W'W \right)^{-1} \\
\mu_{\beta}^{q} &\leftarrow \Sigma_{\beta}^{q} \left(\frac{r_{q,\sigma}}{s_{q,\sigma}} {\Sigma_{\beta}^{0}}^{-1} \mu_{\beta}^{0} + W' \left(\mu_{y^*}^{q} - \varphi_J \mu_{\theta}^{q} \right) \right)
\end{aligned}
$$

3.3.5 ψ

$$\frac{\partial S_1}{\partial \mu_\psi^q} = -\omega_0 \left\{ -\frac{1}{\sigma_\psi^q}\sqrt{\frac{2}{\pi}} \exp\left(-\frac{\mu_\psi^{q\,2}}{2\sigma_\psi^{q\,2}} \right) + 1 - 2\Phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q} \right) + 2\frac{\mu_\psi^q}{\sigma_\psi^q}\phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q} \right) \right\}$$

$$\frac{\partial S_1}{\partial \sigma_\psi^{q\,2}} = -\omega_0 \left\{ \left(\frac{1}{\sqrt{2\pi}\sigma_\psi^q} + \frac{\mu_\psi^{q\,2}}{\sqrt{\pi}\left(\sigma_\psi^{q\,2}\right)^{3/2}} \right) \exp\left(-\frac{\mu_\psi^{q\,2}}{\sigma_\psi^{q\,2}} \right) - \phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q} \right) \frac{\mu_\psi^{q\,2}}{\left(\sigma_\psi^{q\,2}\right)^{3/2}} \right\}$$

$$\begin{aligned}
\frac{\partial Q_j}{\partial \mu_\psi^q} = {} & j\exp\left(\frac{\sigma_\psi^{q\,2} j^2}{2} + \mu_\psi^q j \right) \left\{ 1 - \Phi\left(\frac{-\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j \right) \right\} + \frac{1}{\sigma_\psi^q} \exp\left(\frac{\sigma_\psi^{q\,2} j^2}{2} + \mu_\psi^q j \right) \phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j \right) \\
& - j\exp\left(\frac{\sigma_\psi^{q\,2} j^2}{2} - \mu_\psi^q j \right) \left\{ 1 - \Phi\left(\frac{\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j \right) \right\} - \frac{1}{\sigma_\psi^q} \exp\left(\frac{\sigma_\psi^{q\,2} j^2}{2} - \mu_\psi^q j \right) \phi\left(\frac{\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Q_j}{\partial \sigma_\psi^{q\,2}} = {} & j\exp\left(\frac{\sigma_\psi^{q\,2} j^2}{2} + \mu_\psi^q j \right) \left\{ 1 - \Phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j \right) \right\} + \left(\frac{1}{2\sigma_\psi^q} j - \frac{\mu_\psi^q}{2\left(\sigma_\psi^{q\,2}\right)^{3/2}} \right) \phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j \right) \exp\left(\frac{\sigma_\psi^{q\,2} j^2}{2} + \mu_\psi^q j \right) \\
& + j\exp\left(\frac{\sigma_\psi^{q\,2} j^2}{2} - \mu_\psi^q j \right) \left\{ 1 - \Phi\left(\frac{\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j \right) \right\} + \left(\frac{1}{2\sigma_\psi^q} j + \frac{\mu_\psi^q}{2\left(\sigma_\psi^{q\,2}\right)^{3/2}} \right) \phi\left(\frac{\mu_\psi^q}{\sigma_\psi^q} - \sigma_\psi^q j \right) \exp\left(\frac{\sigma_\psi^{q\,2} j^2}{2} - \mu_\psi^q j \right)
\end{aligned}$$

$$\frac{\partial S_2}{\partial \mu_\psi^q} = -\frac{1}{2}\frac{r_{q,\sigma}}{s_{q,\sigma}}\frac{r_{q,\tau}}{s_{q,\tau}} \left(\mathrm{Tr}\left(\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\mu_\theta^q \right) \sum_{j=1}^{J} \frac{\partial Q_j}{\partial \mu_\psi^q} - \frac{J(J+1)}{4\omega_0}\frac{\partial S_1}{\partial \mu_\psi^q}$$

$$\frac{\partial S_2}{\partial \sigma_\psi^{q\,2}} = -\frac{1}{2}\frac{r_{q,\sigma}}{s_{q,\sigma}}\frac{r_{q,\tau}}{s_{q,\tau}} \left(\mathrm{Tr}\left(\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\mu_\theta^q \right) \sum_{j=1}^{J} \frac{\partial Q_j}{\partial \sigma_\psi^{q\,2}} - \frac{J(J+1)}{4\omega_0}\frac{\partial S_1}{\partial \sigma_\psi^{q\,2}}$$

$$\sigma_\psi^{q\,2} \leftarrow -\frac{1}{2}\left\{ \frac{\partial S_1}{\partial \sigma_\psi^{q\,2}} + \frac{\partial S_2}{\partial \sigma_\psi^{q\,2}} \right\}^{-1}$$

$$\mu_\psi^q \leftarrow \mu_\psi^q + \sigma_\psi^{q\,2}\left\{ \frac{\partial S_1}{\partial \mu_\psi^q} + \frac{\partial S_2}{\partial \mu_\psi^q} \right\}$$

4 Probit: No restriction, with normal random effects

4.1 Model

$$
\begin{aligned}
\mathrm{P}\left(y_i=1|f,\beta\right) &= \Phi\left(w_i^\top\beta + Zu + f\left(x_i\right)\right)\\
y_i^* &= w_i^\top\beta + Zu + f\left(x_i\right) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}\left(0,1\right)\\
y_i &= \begin{cases} 1, & \text{if } y_i^* \geq 0\\ 0, & \text{if } y_i^* < 0\end{cases}\\
\theta_j|\sigma,\tau,\gamma &\sim \mathcal{N}\left(0, \sigma^2\tau^2\exp\left[-j\gamma\right]\right)\\
\tau^2 &\sim \mathrm{IG}\left(\frac{r_{0,\tau}}{2}, \frac{s_{0,\tau}}{2}\right)\\
\sigma^2 &\sim \mathrm{IG}\left(\frac{r_{0,\sigma}}{2}, \frac{s_{0,\sigma}}{2}\right)\\
\beta &\sim \mathcal{N}\left(\mu_\beta^0, \sigma^2\Sigma_\beta^0\right) \quad (p\times 1)\\
u &\sim \mathcal{N}\left(\mu_u^0, \Sigma_u^0\right), \quad (s\times 1)\\
\gamma &\sim \mathrm{Exp}\left(\omega_0\right)\\
|\psi| &= \gamma, \quad \psi \sim \mathrm{DE}\left(0,\omega_0\right)\\
\varphi_j\left(x\right) &= \sqrt{2}\cos\left(\pi j x\right)
\end{aligned}
$$

Joint density:

$$
\begin{aligned}
p\left(y,y^*,\Theta\right) = C &\left\{\prod_{j=1}^J \mathcal{N}\left(\theta_j\middle|0,\sigma^2\tau^2\exp\left[-j\left|\psi\right|\right]\right)\right\}\mathrm{IG}\left(\tau^2\middle|\frac{r_{0,\tau}}{2},\frac{s_{0,\tau}}{2}\right)\mathrm{IG}\left(\sigma^2\middle|\frac{r_{0,\sigma}}{2},\frac{s_{0,\sigma}}{2}\right)\mathcal{N}\left(\beta\middle|\mu_\beta^0,\sigma^2\Sigma_\beta^0\right)\\
&\mathcal{N}\left(u|\mu_u^0,\Sigma_u^0\right)\mathrm{DE}\left(\psi|0,\omega_0\right)\left\{\prod_{i=1}^n\left(1\left[y_i^*\geq 0\right]1\left[y_i=1\right] + 1\left[y_i^*<0\right]1\left[y_i=0\right]\right)\phi\left(y_i^* - w_i'\beta - z_i'u - \varphi_i'\theta_J\right)\right\}
\end{aligned}
$$

where C is the normalizing constant.

4.2 Lower Bound

4.2.1 LB: $\mathrm{E}\left(\ln p\left(y^*|\mathbf{rest}\right)\right) + \mathrm{H}\left(y^*\right)$

$$
\begin{aligned}
\mathrm{E}\left(\ln p\left(y^*|\mathrm{rest}\right)\right) + \mathrm{H}\left(y^*\right) &= \sum_{i=1}^n \mathrm{E}\left(\ln\phi\left(y_i^* - w_i'\beta - z_i'u - \varphi_i'\theta_J\right) - \ln\phi\left(y_i^* - w_i'\mu_\beta^q - z_i'\mu_u^q - \varphi_i'\mu_\theta^q\right)\right)\\
&\quad + \sum_{i=1}^n \ln\left(\left\{\Phi\left(w_i'\mu_\beta^q + z_i'\mu_u^q + \varphi_i'\mu_\theta^q\right)\right\}^{y_i}\left\{1-\Phi\left(w_i'\mu_\beta^q + z_i'\mu_u^q + \varphi_i'\mu_\theta^q\right)\right\}^{1-y_i}\right)\\
&= -\frac{1}{2}\left(\mathrm{Tr}\left(W'W\Sigma_\beta^q\right) + \mathrm{Tr}\left(Z'Z\Sigma_u^q\right) + \mathrm{Tr}\left(\varphi_J'\varphi_J\Sigma_\theta^q\right)\right)\\
&\quad + \sum_{i=1}^n \ln\left(\left\{\Phi\left(w_i'\mu_\beta^q + z_i'\mu_u^q + \varphi_i'\mu_\theta^q\right)\right\}^{y_i}\left\{1-\Phi\left(w_i'\mu_\beta^q + z_i'\mu_u^q + \varphi_i'\mu_\theta^q\right)\right\}^{1-y_i}\right)
\end{aligned}
$$

4.2.2 LB: $\mathrm{E}\left(\ln p\left(u\right)\right) + \mathrm{H}\left(u\right)$

$$
\mathrm{E}\left(\ln p\left(u\right)\right) + \mathrm{H}\left(u\right) = \frac{1}{2}\ln\left|\Sigma_u^{0^{-1}}\Sigma_u^q\right|
$$

Everything else is the same as the one without random effects.

4.2.3 LB: $\mathrm{E}\left[\ln p\left(\theta_j|\sigma,\tau,\psi\right)\right]$

$$
\begin{aligned}
\sum_{j=1}^{J}\mathrm{E}\left[\ln p\left(\theta_j|\sigma,\tau,\psi\right)\right]+\mathrm{H}\left[\theta_J\right] &= \sum_{j=1}^{J}\mathrm{E}\left[-\frac{1}{2}\ln\left(2\pi\right)+\frac{1}{2}\ln\frac{1}{\sigma^2}+\frac{1}{2}\ln\frac{1}{\tau^2}+\frac{j}{2}\left|\psi\right|-\frac{\theta_j^2 e^{j|\psi|}}{2\sigma^2\tau^2}\right]+\mathrm{H}\left[\theta_J\right]\\
&= -\frac{J}{2}\left\{\ln\left(2\pi\right)-\left(\mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right)-\ln\left(\frac{s_{q,\sigma}}{2}\right)\right)-\left(\mathrm{di}\left(\frac{r_{q,\tau}}{2}\right)-\ln\left(\frac{s_{q,\tau}}{2}\right)\right)\right\}\\
&\quad+\frac{J\left(J+1\right)}{4}\left\{\sigma_\psi^q\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\mu_\psi^{q\,2}}{2\sigma_\psi^{q\,2}}\right)+\mu_\psi^q\left(1-2\Phi\left(\frac{-\mu_\psi^q}{\sigma_\psi^q}\right)\right)\right\}\\
&\quad-\frac{1}{2}\frac{r_{q,\sigma}}{s_{q,\sigma}}\frac{r_{q,\tau}}{s_{q,\tau}}\left(\mathrm{Tr}\left(\Sigma_\theta^q\right)+\mu_\theta^{q\prime}\mu_\theta^q\right)\sum_{j=1}^{J}Q_j\left(\mu_\psi^q,\sigma_\psi^{q\,2}\right)+\frac{J}{2}\left(1+\ln\left(2\pi\right)\right)+\frac{1}{2}\ln\left|\Sigma_\theta^q\right|
\end{aligned}
$$

where

$$
\begin{aligned}
Q_j\left(\mu_\psi^q,\sigma_\psi^{q\,2}\right) &= \mathrm{E}\,e^{j|\psi|}\\
&= \exp\left\{\frac{\sigma_\psi^{q\,2}j^2}{2}+\mu_\psi^q j\right\}\left[1-\Phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q}-\sigma_\psi^q j\right)\right]+\exp\left\{\frac{\sigma_\psi^{q\,2}j^2}{2}-\mu_\psi^q j\right\}\left[1-\Phi\left(\frac{\mu_\psi^q}{\sigma_\psi^q}-\sigma_\psi^q j\right)\right].
\end{aligned}
$$

4.2.4 LB: $\mathrm{E}\left[\ln p\left(\tau^2\right)\right]+\mathrm{H}\left[\tau^2\right]$

$$
\begin{aligned}
\mathrm{E}\left[\ln p\left(\tau^2\right)\right]+\mathrm{H}\left[\tau^2\right] &= \frac{r_{0,\tau}}{2}\ln\frac{s_{0,\tau}}{2}-\ln\Gamma\left(\frac{r_{0,\tau}}{2}\right)+\left(\frac{r_{0,\tau}}{2}+1\right)\left\{\mathrm{di}\left(\frac{r_{q,\tau}}{2}\right)-\ln\left(\frac{s_{q,\tau}}{2}\right)\right\}-\frac{s_{0,\tau}}{2}\frac{r_{q,\tau}}{s_{q,\tau}}\\
&\quad+\frac{r_{q,\tau}}{2}+\ln\frac{s_{q,\tau}}{2}+\ln\Gamma\left(\frac{r_{q,\tau}}{2}\right)-\left(1+\frac{r_{q,\tau}}{2}\right)\mathrm{di}\left(\frac{r_{q,\tau}}{2}\right)
\end{aligned}
$$

4.2.5 LB: $\mathrm{E}\left[\ln p\left(\sigma^2\right)\right]+\mathrm{H}\left[\sigma^2\right]$

$$
\begin{aligned}
\mathrm{E}\left[\ln p\left(\sigma^2\right)\right]+\mathrm{H}\left[\sigma^2\right] &= \frac{r_{0,\sigma}}{2}\ln\frac{s_{0,\sigma}}{2}-\ln\Gamma\left(\frac{r_{0,\sigma}}{2}\right)+\left(\frac{r_{0,\sigma}}{2}+1\right)\left\{\mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right)-\ln\left(\frac{s_{q,\sigma}}{2}\right)\right\}-\frac{s_{0,\sigma}}{2}\frac{r_{q,\sigma}}{s_{q,\sigma}}\\
&\quad+\frac{r_{q,\sigma}}{2}+\ln\frac{s_{q,\sigma}}{2}+\ln\Gamma\left(\frac{r_{q,\sigma}}{2}\right)-\left(1+\frac{r_{q,\sigma}}{2}\right)\mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right)
\end{aligned}
$$

4.2.6 LB: $\mathrm{E}\left[\ln p\left(\beta\right)\right]+\mathrm{H}\left[\beta\right]$

$$
\begin{aligned}
\mathrm{E}\left[\ln p\left(\beta\right)\right]+\mathrm{H}\left[\beta\right] &= \frac{p+1}{2}+\frac{1}{2}\left(\mathrm{di}\left(\frac{r_{q,\sigma}}{2}\right)-\ln\left(\frac{s_{q,\sigma}}{2}\right)\right)+\frac{1}{2}\ln\left|\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right|\\
&\quad-\frac{1}{2}\frac{r_{q,\sigma}}{s_{q,\sigma}}\left\{\mathrm{Tr}\left(\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right)+\left(\mu_\beta^q-\mu_\beta^0\right)'\Sigma_\beta^{0\,-1}\left(\mu_\beta^q-\mu_\beta^0\right)\right\}
\end{aligned}
$$

4.2.7 LB: $\mathrm{E}\left[\ln p\left(\psi\right)\right]+\mathrm{H}\left[\psi\right]$

$$
\begin{aligned}
\mathrm{E}\left[\ln p\left(\psi\right)\right]+\mathrm{H}\left[\psi\right] &= \ln\frac{\omega_0}{2}-\omega_0\left\{\sigma_\psi^q\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\mu_\psi^{q\,2}}{2\sigma_\psi^{q\,2}}\right)+\mu_\psi^q\left(1-2\Phi\left(-\frac{\mu_\psi^q}{\sigma_\psi^q}\right)\right)\right\}\\
&\quad+\frac{1}{2}\ln\left(2\pi\sigma_\psi^{q\,2}\right)-\frac{1}{2}
\end{aligned}
$$

4.3 Update

4.3.1 u

$q(u) = \mathcal{N}(\mu_u^q, \Sigma_u^q)$ where

$$\Sigma_u^q = \left(Z'Z + \Sigma_u^{0\,-1}\right)^{-1}$$
$$\mu_u^q = \Sigma_u^q\left(\Sigma_u^{0\,-1}\mu_u^0 + Z'\left(\mu_{y^*}^q - W\mu_\beta^q - \varphi_J\mu_\theta^q\right)\right)$$

4.3.2 y_i^*

$$q(y_i^*) = \mathcal{TN}\left(\mu_{y_i^*}, 1\right)$$

truncated at zero where

$$\mu_{y_i^*} = w_i'\mu_\beta^q + z_i'\mu_u^q + \varphi_i'\mu_\theta^q.$$

To distinguish $\mu_{y_i^*}$ which is the parameter of the variational distribution from the expected value, we will denote the expectation by $\mu_{y_i^*}^q$ where

$$\mathrm{E}(y_i^*) = \mu_{y_i^*}^q = \mu_{y_i^*} + \frac{\phi\left(\mu_{y_i^*}\right)}{\Phi\left(\mu_{y_i^*}\right)^{y_i}\left(\Phi\left(\mu_{y_i^*}\right) - 1\right)^{1-y_i}}$$

4.3.3 θ_j

$q(\theta_J) = \mathcal{N}\left(\mu_\theta^q, \Sigma_\theta^q\right)$ where

$$\Sigma_\theta^q = \left(\frac{r_{q,\tau}}{s_{q,\tau}}\frac{r_{q,\sigma}}{s_{q,\sigma}}\,\mathrm{Dg}\left(Q_{1:J}\right) + \varphi_J'\varphi_J\right)^{-1}$$
$$\mu_\theta^q = \Sigma_\theta^q\varphi_J'\left(\mu_{y_i}^q - W\mu_\beta^q - Z\mu_u^q\right).$$

4.3.4 β

$q(\beta) = \mathcal{N}\left(\mu_\beta^q, \Sigma_\beta^q\right)$ where

$$\Sigma_\beta^q = \left(\frac{r_{q,\sigma}}{s_{q,\sigma}}\Sigma_\beta^{0\,-1} + W'W\right)^{-1}$$
$$\mu_\beta^q = \Sigma_\beta^q\left(\Sigma_\beta^{0\,-1}\mu_\beta^0 + W'\left(\mu_{y^*}^q - Z\mu_u^q - \varphi_J\mu_\theta^q\right)\right).$$

4.3.5 τ^2

$$r_{q,\tau} \leftarrow r_{0,\tau} + J$$
$$s_{q,\tau} \leftarrow s_{0,\tau} + \frac{r_{q,\sigma}}{s_{q,\sigma}}\sum_{j=1}^{J}\left(\Sigma_{\theta,jj}^q + {\mu_{\theta,j}^q}^2\right)Q_j$$

4.3.6 σ^2

$$r_{q,\sigma} \leftarrow r_{0,\sigma} + J + p + 1$$
$$s_{q,\sigma} \leftarrow s_{0,\sigma} + \frac{r_{q,\tau}}{s_{q,\tau}}\sum_{j=1}^{J}\left(\Sigma_{\theta,jj}^q + {\mu_{\theta,j}^q}^2\right)Q_j + \mathrm{Tr}\left(\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right) + \left(\mu_\beta^q - \mu_\beta^0\right)'\Sigma_\beta^{0\,-1}\left(\mu_\beta^q - \mu_\beta^0\right)$$

5 Probit: Monotone Shape Restriction

5.1 Model

$$
\begin{aligned}
\mathrm{P}\left(y_i=1|\beta,\theta_J\right) &= \Phi\left(w_i'\beta+\delta\theta_J'\varphi_J^a\left(x_i\right)\theta_J\right)\\
y_i^* &= w_i'\beta+\delta\theta_J'\varphi_J^a\left(x_i\right)\theta_J+\epsilon_i, \quad \epsilon_i\sim\mathcal{N}\left(0,1\right)\\
y_i &= \begin{cases}1, & \text{if } y_i^*\geq 0\\ 0, & \text{if } y_i^*<0\end{cases}\\
\theta_j|\sigma,\tau,\gamma &\sim \mathcal{N}\left(0,\sigma\tau^2\exp\left[-j\gamma\right]\right), \quad \text{for } j\geq 2\\
\theta_0|\sigma &\sim \mathcal{N}\left(0,\sigma\sigma_0^2\right)\\
\tau^2 &\sim \mathrm{IG}\left(\frac{r_{0,\tau}}{2},\frac{s_{0,\tau}}{2}\right)\\
\sigma^2 &\sim \mathrm{IG}\left(\frac{r_{0,\sigma}}{2},\frac{s_{0,\sigma}}{2}\right)\\
\beta &\sim \mathcal{N}\left(\mu_\beta^0,\sigma^2\Sigma_\beta^0\right) \qquad (p\times 1)\\
\gamma &\sim \mathrm{Exp}\left(\omega_0\right)\\
|\psi| &= \gamma, \quad \psi\sim\mathrm{DE}\left(0,\omega_0\right)\\
\varphi_{0,0}^a\left(x\right) &= x-0.5\\
\varphi_{0,j}^a\left(x\right) &= \varphi_{j,0}^a\left(x\right)=\frac{\sqrt{2}}{\pi j}\sin\left(\pi jx\right)-\frac{\sqrt{2}}{\left(\pi j\right)^2}\left[1-\cos\left(\pi j\right)\right] \text{ for } j\geq 1,\\
\varphi_{j,j}^a\left(x\right) &= \frac{\sin\left(2\pi jx\right)}{2\pi j}+x-0.5 \text{ for } j\geq 1,\\
\varphi_{j,k}^a\left(x\right) &= \frac{\sin\left[\pi\left(j+k\right)x\right]}{\pi\left(j+k\right)}+\frac{\sin\left[\pi\left(j-k\right)x\right]}{\pi\left(j-k\right)}\\
&\quad -\frac{1-\cos\left[\pi\left(j+k\right)\right]}{\left[\pi\left(j+k\right)\right]^2}-\frac{1-\cos\left[\pi\left(j-k\right)\right]}{\left[\pi\left(j-k\right)\right]^2}\\
&\quad \text{for } j\neq k \text{ and } j,k\geq 1.
\end{aligned}
$$

Joint density:

$$
\begin{aligned}
p\left(y,y^*,\Theta\right) \propto{}& \mathcal{N}\left(\theta_0|0,\sigma\sigma_0^2\right)\left\{\prod_{j=1}^{J}\mathcal{N}\left(\theta_j|0,\sigma\tau^2\exp\left[-j\left|\psi\right|\right]\right)\right\}\mathrm{IG}\left(\tau^2\middle|\frac{r_{0,\tau}}{2},\frac{s_{0,\tau}}{2}\right)\mathrm{IG}\left(\sigma^2\middle|\frac{r_{0,\sigma}}{2},\frac{s_{0,\sigma}}{2}\right)\mathcal{N}\left(\beta|\mu_\beta^0,\sigma^2\Sigma_\beta^0\right)\\
&\times\mathrm{DE}\left(\psi|0,\omega_0\right)\left\{\prod_{i=1}^{n}\left(1\left[y_i^*\geq 0\right]1\left[y_i=1\right]+1\left[y_i^*<0\right]1\left[y_i=0\right]\right)\phi\left(y_i^*-w_i'\beta-\delta\theta_J'\varphi_J^a\left(x_i\right)\theta_J\right)\right\}
\end{aligned}
$$

5.2 Update

5.2.1 y_i^*

$$\begin{aligned}
\ln q\left(y_i^*\right) &\propto -\frac{1}{2}\,\mathrm{E}_{-y_i^*}\left(y_i^{*2}-2y_i^*\left(w_i'\beta+\delta\theta_J'\varphi_J^a\left(x_i\right)\theta_J\right)+\left(w_i'\beta+\delta\theta_J'\varphi_J^a\left(x_i\right)\theta_J\right)^2\right)\\
&\propto -\frac{1}{2}\left(y_i^{*2}-2y_i^*\left(w_i'\mu_\beta^q+\delta\left(\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)+\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\right)\right.\\
&\quad+\mathrm{Tr}\left(w_iw_i'\Sigma_\beta^q\right)+\mu_\beta^{q\prime}w_iw_i'\mu_\beta^q+2\delta w_i'\mu_\beta^q\left(\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)+\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\\
&\quad\left.+2\delta^2\,\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)^2+4\delta^2\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right)\mu_\theta^q+\delta^2\left(\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)+\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)^2\right)\\
&\propto -\frac{1}{2}\left(y_i^*-w_i'\mu_\beta^q-\delta\left(\mathrm{Tr}\left(\varphi_J\left(x_i\right)\Sigma_\theta^q\right)+\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta\right)\right)^2
\end{aligned}$$

Therefore,

$$q\left(y_i^*\right)\sim\mathcal{TN}\left(\mu_{y_i^*},1\right)$$

truncated at zero where

$$\mu_{y_i^*}=w_i'\mu_\beta^q+\delta\,\mathrm{Tr}\left(\varphi_J\left(x_i\right)\Sigma_\theta^q\right)+\delta\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q.$$

To distinguish the mean of the normal distribution and the mean of the truncated normal distribution, the latter will be denoted $\mu_{y_i^*}^q$ where

$$\mathrm{E}\left(y_i^*\right)=\mu_{y_i^*}^q=\mu_{y_i^*}+\frac{\phi\left(\mu_{y_i^*}\right)}{\Phi\left(\mu_{y_i^*}\right)^{y_i}\left(\Phi\left(\mu_{y_i^*}\right)-1\right)^{1-y_i}}.$$

5.2.2 θ

$$\begin{aligned}
S_1\left(\mu_\theta^q,\Sigma_\theta^q\right)&=\mathrm{E}\left(\ln p\left(\theta_J|\sigma^2,\sigma^2,\psi\right)\right)\\
&=-\frac{1}{2}\,\mathrm{E}\left(\frac{1}{\sigma}\right)\mathrm{Tr}\left\{\left(\Sigma_\theta^q+\mu_\theta^q\mu_\theta^{q\prime}\right)\mathrm{Dg}\left(\mathrm{E}\left(\Upsilon^{-1}\right)\right)\right\}\\
S_2\left(\mu_\theta^q,\Sigma_\theta^q\right)&=\mathrm{E}\left(\ln p\left(y^*|\beta,\theta_J,\sigma^2\right)\right)\\
&=-\frac{1}{2}\sum_{i=1}^n\left\{\left(\mu_{y_i^*}^q-w_i'\mu_\beta^q-\delta\,\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)-\delta\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)^2\right.\\
&\quad\left.+2\,\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)+4\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right)\mu_\theta^q\right\}
\end{aligned}$$

Then by NCVMP introduced in Wand (2014),

$$\begin{aligned}
\Sigma_\theta^q&\leftarrow-\frac{1}{2}\left\{\sum_{a=1}^2\frac{\partial S_a\left(\mu_\theta^q,\Sigma_\theta^q\right)}{\partial\Sigma_\theta^q}\right\}^{-1}\\
\mu_\theta^q&\leftarrow\mu_\theta^q+\Sigma_\theta^q\left\{\sum_{a=1}^2\frac{\partial S_a\left(\mu_\theta^q,\Sigma_\theta^q\right)}{\partial\mu_\theta^q}\right\}.
\end{aligned}$$

By matrix differential calculus,

$$
\begin{aligned}
\frac{\partial S_1}{\partial \Sigma_\theta^q} &= -\frac{1}{2}\,\mathrm{E}\left(\frac{1}{\sigma}\right)\mathrm{Dg}\left(\mathrm{E}\left(\Upsilon^{-1}\right)\right) \\
\frac{\partial S_2}{\partial \Sigma_\theta^q} &= -\frac{1}{2}\sum_{i=1}^{n}\Big\{4\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right) + 4\varphi_J^a\left(x_i\right)\mu_\theta^q\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right) \\
&\quad -2\delta\left(\mu_{y_i^*}^q - w_i'\mu_\beta^q - \delta\,\mathrm{Tr}\left(\Sigma_\theta^q\varphi_J^a\left(x_i\right)\right) - \delta\mu_\theta^{q\prime}\Sigma_\theta^q\mu_\theta^q\right)\varphi_J^a\left(x_i\right)\Big\} \\
\frac{\partial S_1}{\partial \mu_\theta^q} &= -\,\mathrm{E}\left(\frac{1}{\sigma}\right)\mathrm{Dg}\left(\mathrm{E}\left(\Upsilon^{-1}\right)\right)\mu_\theta^q \\
\frac{\partial S_2}{\partial \mu_\theta^q} &= -\frac{1}{2}\sum_{i=1}^{n}\Big\{8\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right)\mu_\theta^q - 4\delta\left(\mu_{y_i^*}^q - w_i'\mu_\beta^q - \delta\,\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)\right. \\
&\quad \left. -\delta\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\varphi_J^a\left(x_i\right)\mu_\theta^q\Big\}
\end{aligned}
$$

5.2.3 β

$$
\begin{aligned}
\ln q\left(\beta\right) &\propto \mathrm{E}_{-\beta}\left(-\frac{1}{2\sigma^2}\left(\beta-\mu_\beta^0\right)'\Sigma_\beta^{0\,-1}\left(\beta-\mu_\beta^0\right) - \frac{1}{2}\sum_{i=1}^{n}\left(y_i^* - w_i'\beta - \delta\theta_J'\varphi_J^a\left(x_i\right)\theta_J\right)^2\right) \\
&\propto -\frac{1}{2}\left(\beta'\left(\mathrm{E}\left(\frac{1}{\sigma^2}\right)\Sigma_\beta^{0\,-1} + W'W\right)\beta\right. \\
&\quad \left. -2\left(\mathrm{E}\left(\frac{1}{\sigma^2}\right)\beta'\Sigma_\beta^{0\,-1}\mu_\beta^0 + \beta'W'\mu_{y^*}^q - \beta'\delta\sum_{i=1}^{n}w_i\left(\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\right)\right)
\end{aligned}
$$

Therefore, $q\left(\beta\right) = \mathcal{N}\left(\mu_\beta^q, \Sigma_\beta^q\right)$ where

$$
\begin{aligned}
\Sigma_\beta^q &\leftarrow \left(\mathrm{E}\left(\frac{1}{\sigma^2}\right)\Sigma_\beta^{0\,-1} + W'W\right)^{-1} \\
\mu_\beta^q &\leftarrow \Sigma_\beta^q\left(\mathrm{E}\left(\frac{1}{\sigma^2}\right)\Sigma_\beta^{0\,-1}\mu_\beta^0 + W'\mu_{y^*}^q - \delta\sum_{i=1}^{n}w_i\left(\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\right)
\end{aligned}
$$

5.2.4 τ^2

$$
\begin{aligned}
r_{q,\tau} &\leftarrow r_{0,\tau} + J \\
s_{q,\tau} &\leftarrow s_{0,\tau} + \mathrm{E}\left(\frac{1}{\sigma}\right)\mathrm{Tr}\left(\left(\Sigma_\theta^{q*} + \mu_\theta^{q*}\mu_\theta^{q*\prime}\right)\mathrm{Dg}\left(Q_{1:J}\right)\right)
\end{aligned}
$$

5.2.5 σ^2

$$
\begin{aligned}
\ln q\left(\sigma^2\right) &\propto -\frac{1}{2}\ln\sigma - \frac{J}{2}\ln\sigma - \frac{1}{2}\frac{\theta_0^2}{\sigma_0^2}\frac{1}{\sigma} - \frac{1}{2}\sum_{j=1}^{J}\frac{r_{q,\tau}}{s_{q,\tau}}e^{j|\psi|}\theta_j^2\frac{1}{\sigma} \\
&\quad + \left(\frac{r_{0,\sigma}}{2}+1\right)\ln\frac{1}{\sigma^2} - \frac{s_{0,\sigma}}{2}\frac{1}{\sigma^2} - \frac{p+1}{2}\ln\sigma^2 \\
&\quad - \frac{1}{2\sigma^2}\left\{\left(\mu_\beta^q-\mu_\beta^0\right)'\Sigma_\beta^{0\,-1}\left(\mu_\beta^q-\mu_\beta^0\right) + \mathrm{Tr}\left(\Sigma_{\beta^0}^{-1}\Sigma_\beta^q\right)\right\} \\
&\propto \left(\frac{J+1}{4}+\frac{r_{0,\sigma}+p+1}{2}+1\right)\ln\frac{1}{\sigma^2} - \frac{1}{2}\frac{1}{\sigma}\,\mathrm{Tr}\left\{\left(\Sigma_\theta^q+\mu_\theta^q\mu_\theta^{q\prime}\right)\mathrm{Dg}\left(\mathrm{E}\left(\Upsilon^{-1}\right)\right)\right\} \\
&\quad - \frac{1}{2}\frac{1}{\sigma^2}\left\{s_{0,\sigma} + \left(\mu_\beta^q-\mu_\beta^0\right)'\Sigma_\beta^{0\,-1}\left(\mu_\beta^q-\mu_\beta^0\right) + \mathrm{Tr}\left(\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right)\right\} \\
q\left(\sigma^2\right) &\propto \left(\frac{1}{\sigma}\right)^{2a}\exp\left(\frac{b}{\sigma}-\frac{c}{\sigma^2}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
a &\leftarrow \frac{J+1}{4}+\frac{r_{0,\sigma}+p+1}{2}+1 \\
b &\leftarrow -\frac{1}{2}\,\mathrm{Tr}\left\{\left(\Sigma_\theta^q+\mu_\theta^q\mu_\theta^{q\prime}\right)\mathrm{Dg}\left(\mathrm{E}\left(\Upsilon^{-1}\right)\right)\right\} \\
c &\leftarrow \frac{1}{2}\left\{s_{0,\sigma} + \left(\mu_\beta^q-\mu_\beta^0\right)'\Sigma_\beta^{0\,-1}\left(\mu_\beta^q-\mu_\beta^0\right) + \mathrm{Tr}\left(\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right)\right\}
\end{aligned}
$$

and $\Upsilon = \left(\sigma_0^2, \tau^2e^{-\gamma}, \ldots, \tau^2e^{-J\gamma}\right)'$ which gives $\mathrm{E}\left(\Upsilon^{-1}\right) = \left(1/\sigma_0^2, r_{q,\tau}/s_{q,\tau}\,\mathrm{E}\left(\Gamma^{-1}\right)\right)$. Lastly, $\mathrm{E}\left(\Gamma^{-1}\right) = Q_{1:J}$.

5.2.6 ψ

Same as before except that $r_{q,\sigma}/s_{q,\sigma}$ in $S_2\left(\mu_\psi^q, \sigma_\psi^{q\,2}\right)$ is replaced by $\mathrm{E}\left(1/\sigma\right)$.

5.3 Lower Bound

5.3.1 $\mathrm{E}(\ln p(y^*|\mathbf{rest})) + \mathrm{H}(y^*)$

$$
\begin{aligned}
\mathrm{E}(\ln p(y^*|\mathrm{rest})) + \mathrm{H}(y^*) &= \sum_{i=1}^{n} \mathrm{E}\left(\ln \phi\left(y_i^* - w_i'\beta - \delta\theta_J'\varphi_J^a(x_i)\theta_J\right) - \ln \phi\left(y_i^* - \mu_{y_i^*}\right)\right) \\
&\quad + \sum_{i=1}^{n} \ln\left(\Phi\left(\mu_{y_i^*}\right)^{y_i}\left(1-\Phi\left(\mu_{y_i^*}\right)\right)^{1-y_i}\right) \\
&= -\frac{1}{2}\sum_{i=1}^{n}\Big[-2\mu_{y_i^*}^q\left(w_i'\mu_\beta^q + \delta\,\mathrm{Tr}\left(\varphi_J^a(x_i)\Sigma_\theta^q\right) + \delta\mu_\theta^{q\prime}\varphi_J^a(x_i)\mu_\theta^q\right) \\
&\quad + \mathrm{Tr}\left(w_i w_i'\Sigma_\beta^q\right) + \mu_\beta^{q\prime}w_i w_i'\mu_\beta^q + 2w_i'\mu_\beta^q\left(\delta\,\mathrm{Tr}\left(\varphi_J^a(x_i)\Sigma_\theta^q\right) + \delta\mu_\theta^{q\prime}\varphi_J^a(x_i)\mu_\theta^q\right) \\
&\quad + \delta^2\left\{2\,\mathrm{Tr}\left(\varphi_J^a(x_i)\Sigma_\theta^q\right)^2 + 4\mu_\theta^{q\prime}\varphi_J^a(x_i)\Sigma_\theta^q\varphi_J^a(x_i)\mu_\theta^q + \left(\mathrm{Tr}\left(\varphi_J^a(x_i)\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\varphi_J^a(x_i)\mu_\theta^q\right)^2\right\} \\
&\quad + 2\mu_{y_i^*}^q\mu_{y_i^*} - \mu_{y_i^*}^2\Big] + \sum_{i=1}^{n}\ln\left(\Phi\left(\mu_{y_i^*}\right)^{y_i}\left(1-\Phi\left(\mu_{y_i^*}\right)\right)^{1-y_i}\right) \\
&= -\frac{1}{2}\sum_{i=1}^{n}\left[\mathrm{Tr}\left(w_i w_i'\Sigma_\beta^q\right) + \delta^2\left\{2\,\mathrm{Tr}\left(\varphi_J^a(x_i)\Sigma_\theta^q\right)^2 + 4\mu_\theta^{q\prime}\varphi_J^a(x_i)\Sigma_\theta^q\varphi_J^a(x_i)\mu_\theta^q\right\}\right] \\
&\quad + \sum_{i=1}^{n}\ln\left(\Phi\left(\mu_{y_i^*}\right)^{y_i}\left(1-\Phi\left(\mu_{y_i^*}\right)\right)^{1-y_i}\right)
\end{aligned}
$$

Note that $\sum_{i=1}^{n}\mathrm{Tr}\left(w_i w_i'\Sigma_\beta^q\right) = \mathrm{Tr}\left(W'W\Sigma_\beta^q\right)$.

5.3.2 $\mathrm{E}(\ln p(\theta)) + \mathrm{H}(\theta)$

$$
\mathrm{E}(\ln p(\theta|\mathrm{rest})) + \mathrm{H}(\theta) = -\frac{J+1}{2}\mathrm{E}(\ln\sigma) - \frac{1}{2}\mathrm{E}\left(\frac{1}{\sigma}\right)\mathrm{Tr}\left\{\left(\Sigma_\theta^q + \mu_\theta^q\mu_\theta^{q\prime}\right)\mathrm{Dg}\left(\mathrm{E}\left(\Upsilon^{-1}\right)\right)\right\} + \frac{1}{2}\ln\left|\Sigma_\theta^q\right|
$$

5.3.3 $\mathrm{E}(\ln p(\beta)) + \mathrm{H}(\beta)$

$$
\mathrm{E}(\ln p(\beta|\mathrm{rest})) + \mathrm{H}(\beta) = -\frac{p+1}{2}\mathrm{E}\left(\ln\sigma^2\right) - \frac{1}{2}\mathrm{E}\left(\frac{1}{\sigma^2}\right)\left\{\left(\mu_\beta^q - \mu_\beta^0\right)'\Sigma_\beta^{0\,-1}\left(\mu_\beta^q - \mu_\beta^0\right) + \mathrm{Tr}\left(\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right)\right\} + \frac{1}{2}\ln\left|\Sigma_\beta^q\right|
$$

5.3.4 $\mathrm{E}\left(\ln p\left(\tau^2\right)\right) + \mathrm{H}\left(\tau^2\right)$

$$
\begin{aligned}
\mathrm{E}\left(\ln p\left(\tau^2|\mathrm{rest}\right)\right) + \mathrm{H}\left(\tau^2\right) &= -\frac{r_{0,\tau}}{2}\ln\left(\frac{s_{q,\tau}}{2}\right) + \left(\frac{r_{0,\tau}}{2} - \frac{r_{q,\tau}}{2}\right)\mathrm{di}\left(\frac{r_{q,\tau}}{2}\right) + \left(1 - \frac{s_{0,\tau}}{s_{q,\tau}}\right)\frac{r_{q,\tau}}{2} + \ln\Gamma\left(\frac{r_{q,\tau}}{2}\right) \\
&\quad + \frac{r_{0,\tau}}{2}\ln\frac{s_{0,\tau}}{2} - \ln\Gamma\left(\frac{r_{0,\tau}}{2}\right)
\end{aligned}
$$

5.3.5 $\mathrm{E}\left(\ln p\left(\sigma^2\right)\right) + \mathrm{H}\left(\sigma^2\right)$

$$
\begin{aligned}
\mathrm{E}\left(\ln p\left(\sigma^2|\mathrm{rest}\right)\right) + \mathrm{H}\left(\sigma^2\right) &= \frac{r_{0,\sigma}}{2}\ln\left(\frac{s_{0,\sigma}}{2}\right) - \ln\Gamma\left(\frac{r_{0,\sigma}}{2}\right) - \left(\frac{r_{0,\sigma}}{2} + 1\right)\mathrm{E}\left(\ln\sigma^2\right) - \frac{s_{0,\sigma}}{2}\mathrm{E}\left(\frac{1}{\sigma^2}\right) \\
&\quad + 2a\,\mathrm{E}(\ln\sigma) - b\,\mathrm{E}\left(\frac{1}{\sigma}\right) + c\,\mathrm{E}\left(\frac{1}{\sigma^2}\right)
\end{aligned}
$$

5.3.6 $\mathrm{E}(\ln p(\psi)) + \mathrm{H}(\psi)$

$$\mathrm{E}(\ln p(\psi)) + \mathrm{H}(\psi) = \ln\left(\frac{\omega_0}{2}\right) + S_1\left(\mu_\psi^q, \sigma_\psi^{q\,2}\right) + \frac{1}{2}\ln\left(2\pi\sigma_\psi^{q\,2}\right) - \frac{1}{2}$$

5.3.7 LB

$$\begin{aligned}
\mathcal{L} = &-\frac{1}{2}\left\{\mathrm{Tr}\left(W'W\Sigma_\beta^q\right) + \delta^2\sum_{i=1}^n\left(2\,\mathrm{Tr}\left(\varphi_J^a(x_i)\,\Sigma_\theta^q\right)^2 + 4\mu_\theta^{q\prime}\varphi_J^a(x_i)\,\Sigma_\theta^q\varphi_J^a(x_i)\,\mu_\theta^q\right)\right\} \\
&+\sum_{i=1}^n\left\{\ln\left(\Phi\left(\mu_{y_i^*}\right)^{y_i}\left(1-\Phi\left(\mu_{y_i^*}\right)\right)^{1-y_i}\right)\right\} \\
&-\frac{1}{2}\mathrm{E}\left(\frac{1}{\sigma}\right)\mathrm{Tr}\left\{\left(\Sigma_\theta^q + \mu_\theta^q\mu_\theta^{q\prime}\right)\mathrm{Dg}\left(\mathrm{E}\left(\Upsilon^{-1}\right)\right)\right\} + \frac{1}{2}\ln\left|\Sigma_\theta^q\right| \\
&-\frac{1}{2}\mathrm{E}\left(\frac{1}{\sigma^2}\right)\left\{\left(\mu_\beta^q - \mu_\beta^0\right)'\Sigma_\beta^{0\,-1}\left(\mu_\beta^q - \mu_\beta^0\right) + \mathrm{Tr}\left(\Sigma_\beta^{0\,-1}\Sigma_\beta^q\right)\right\} + \frac{1}{2}\ln\left|\Sigma_\beta^q\right| \\
&+\frac{r_{0,\sigma}}{2}\ln\left(\frac{s_{0,\sigma}}{2}\right) - \ln\Gamma\left(\frac{r_{0,\sigma}}{2}\right) - \frac{s_{0,\sigma}}{2}\mathrm{E}\left(\frac{1}{\sigma^2}\right) - b\,\mathrm{E}\left(\frac{1}{\sigma}\right) + c\,\mathrm{E}\left(\frac{1}{\sigma^2}\right) \\
&-\frac{r_{0,\tau}}{2}\ln\left(\frac{s_{q,\tau}}{2}\right) + \left(\frac{r_{0,\tau}}{2} - \frac{r_{q,\tau}}{2}\right)\mathrm{di}\left(\frac{r_{q,\tau}}{2}\right) + \left(1 - \frac{s_{0,\tau}}{s_{q,\tau}}\right)\frac{r_{q,\tau}}{2} + \ln\Gamma\left(\frac{r_{q,\tau}}{2}\right) \\
&+\ln\left(\frac{\omega_0}{2}\right) + S_1\left(\mu_\psi^q, \sigma_\psi^{q\,2}\right) + \frac{1}{2}\ln\left(2\pi\sigma_\psi^{q\,2}\right) - \frac{1}{2}
\end{aligned}$$

5.4 Some Calculation

5.4.1 $\mathrm{E}(1/\sigma)$

As we were going through the variational approximation for the monotone shape restriction model, we have seen the term $\mathrm{E}(1/\sigma)$ and $\mathrm{E}\left(1/\sigma^2\right)$ several times and did not mention how to compute the integral. We give the mathematical consideration. Wolfram Alpha gives

$$\int_0^\infty x^p \exp\left(qx - rx^2\right)\,dx = \frac{1}{2}r^{-p/2-1}\left(q\Gamma\left(\frac{p}{2}+1\right)\,{}_1F_1\left(\frac{p}{2}+1;\frac{3}{2};\frac{q^2}{4r}\right) + \sqrt{r}\Gamma\left(\frac{p+1}{2}\right)\,{}_1F_1\left(\frac{p+1}{2};\frac{1}{2};\frac{q^2}{4r}\right)\right)$$

for $\mathrm{Re}(r) > 0 \wedge \mathrm{Re}(p) > -1$. Note that ${}_1F_1(a;b;x)$ is the *Kummer confluent hypergeometric function.* R package fAsianOptions provides the function `kummerM(x,a,b)` for ${}_1F_1(a;b;x)$.

However, since the variable is not x but rather the reciprocal $1/\sigma$, care must be taken in that the parameters should be changed accordingly. Let $x = 1/\sigma$. Then the differential is $dx = -d\sigma/\sigma^2$ and as $x \to 0$, $\sigma \to \infty$; as $x \to \infty$, $\sigma \to 0$. Therefore,

$$\begin{aligned}
\int_0^\infty x^p \exp\left(qx - rx^2\right)\,dx &= \int_\infty^0 -\left(\frac{1}{\sigma}\right)^p \exp\left(\frac{q}{\sigma} - \frac{r}{\sigma^2}\right)\cdot\frac{1}{\sigma^2}\,d\sigma \\
&= \int_0^\infty \left(\frac{1}{\sigma}\right)^{p+2} \exp\left(\frac{q}{\sigma} - \frac{r}{\sigma^2}\right)\,d\sigma.
\end{aligned}$$

Now we can compute the expectations.

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{\sigma}\right) &= \frac{\int_0^\infty \left(\frac{1}{\sigma}\right)^{p+3}\exp\left(\frac{q}{\sigma} - \frac{r}{\sigma^2}\right)\,d\sigma}{\int_0^\infty \left(\frac{1}{\sigma}\right)^{p+2}\exp\left(\frac{q}{\sigma} - \frac{r}{\sigma^2}\right)\,d\sigma} \\
\mathrm{E}\left(\frac{1}{\sigma^2}\right) &= \frac{\int_0^\infty \left(\frac{1}{\sigma}\right)^{p+4}\exp\left(\frac{q}{\sigma} - \frac{r}{\sigma^2}\right)\,d\sigma}{\int_0^\infty \left(\frac{1}{\sigma}\right)^{p+2}\exp\left(\frac{q}{\sigma} - \frac{r}{\sigma^2}\right)\,d\sigma}
\end{aligned}$$

5.4.2 $\phi(x)/\Phi(x)$ when x is very small

In the variational parameter update section, we encounter the following evaluation:

$$\mathrm{E}\left(y_i^*\right) = \mu_{y_i^*} + \frac{\phi\left(\mu_{y_i^*}\right)}{\Phi\left(\mu_{y_i^*}\right)^{y_i}\left(\Phi\left(\mu_{y_i^*}\right)-1\right)^{1-y_i}}$$
$$\mu_{y_i^*} = w_i'\mu_\beta^q + \delta\,\mathrm{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right) + \delta\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q$$

However, we must note that the expectation behaves extremely pathologically. For example,

$$\lim_{t\to-\infty}\frac{\exp\left(-t^2/2\right)}{\int_{-\infty}^{t}\exp\left(-x^2/2\right)\,dx} = \infty.$$

Even though this does not transpire quite often for in actual computations since we do not deal with infinity in real life. However, we must not simply code `dnorm(x)/pnorm(x)` since `pnorm(x)` returns zero for even as reasonably small a value as -40, notwithstanding the true value of the quotient being $\phi(-40)/\Phi(-40) = 40.02496884$. Therefore, we should use the following identity:

$$\frac{\phi(x)}{\Phi(x)} = \frac{(2/\pi)^{1/2}\exp\left(-x^2/2\right)}{\mathrm{erf}\left(x/\sqrt{2}\right)+1}.$$

The problem is, even the RHS of the identity fails if x is -9. We thus turn to the Laurent series of $\phi(t)/\Phi(t)$:

$$\frac{\exp\left(-t^2/2\right)}{\int_{-\infty}^{t}\exp\left(-x^2/2\right)\,dx} = -t - \frac{1}{t} + \frac{2}{t^3} - \frac{10}{t^5} + \mathcal{O}\left(\left(\frac{1}{t}\right)^6\right) \qquad \text{at } t=-\infty$$

for values smaller than -8.

5.4.3 $\phi(x)/(\Phi(x)-1)$

This is much better than the previous one.

$$\lim_{t\to\infty}\frac{\exp\left(-t^2/2\right)}{\int_{-\infty}^{t}\exp\left(-x^2/2\right)\,dx - 1} = 0.$$

The only problem for this calculation is that when we code the expression, we will be relying on `pnorm` which will give 1 for quite large values, immediately rendering the whole expression indefinite since it makes the denominator zero. Therefore, we again come up with another identity:

$$\frac{\phi(x)}{\Phi(x)-1} = \frac{\exp\left(-x^2/2\right)}{(\pi/2)^{1/2}\left(\mathrm{erf}\left(x/\sqrt{2}\right)+1\right)-1}.$$

Unlike the identity in the previous subsection, we do not have to consider large values since it will tend to zero as x grows larger.

5.5 Corrections

5.5.1 Original Paper: Linear model

There are some mistakes in the original paper so I provide the calculation. $S_2\left(\mu_\theta^q,\Sigma_\theta^q\right) = \mathrm{E}\left(\ln p\left(y|\beta,\theta_J,\sigma^2\right)\right)$. Let's do this.

$$\begin{aligned}
S_2\left(\mu_\theta^q,\Sigma_\theta^q\right) &= -\frac{1}{2}\mathrm{E}\left(\frac{1}{\sigma^2}\right)\sum_{i=1}^{n}\mathrm{E}\left(\left(y_i - w_i'\beta - \delta\theta_J'\varphi_J^a\left(x_i\right)\theta_J\right)^2\right)\\
&= -\frac{1}{2}\mathrm{E}\left(\frac{1}{\sigma^2}\right)\sum_{i=1}^{n}\mathrm{E}\left(y_i^2 - 2\underbrace{y_i\left(w_i'\beta + \delta\theta_J'\varphi_J^a\left(x_i\right)\theta_J\right)}_{T_1} + \underbrace{\left(w_i'\beta + \delta\theta_J'\varphi_J^a\left(x_i\right)\theta_J\right)^2}_{T_2}\right)
\end{aligned}$$

where

$$
\begin{aligned}
\mathrm{E}\left(T_1\right) &= y_i\left(w_i'\mu_\beta^q + \delta\left(\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\right) \\
\mathrm{E}\left(T_2\right) &= \mathrm{E}\left(\left(w_i'\beta\right)^2\right) + 2\delta\left(w_i'\mu_\beta^q\left(\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\right) + \mathrm{E}\left(\left(\theta_J'\varphi_J^a\left(x_i\right)\theta_J\right)\right) \\
&= \mu_\beta^{q\prime}w_iw_i'\mu_\beta^q + \operatorname{Tr}\left(w_iw_i'\Sigma_\beta^q\right) + 2\delta\left(w_i'\mu_\beta^q\left(\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\right) \\
&\quad + \delta^2\left(3\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)^2 + 6\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right)\mu_\theta^q + \left(\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta\right)^2\right)
\end{aligned}
$$

Arranging the terms,

$$
\begin{aligned}
S_2\left(\mu_\theta^q,\Sigma_\theta^q\right) = -\frac{1}{2}\mathrm{E}\left(\frac{1}{\sigma^2}\right)\sum_{i=1}^n &\underbrace{\left(y_i - w_i'\mu_\beta^q - \delta\left(\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\right)^2}_{T_3} \\
&+ \delta^2\left(\underbrace{2\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)^2 + 4\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right)\mu_\theta^q}_{T_4}\right) + \operatorname{Tr}\left(w_iw_i'\Sigma_\beta^q\right).
\end{aligned}
$$

We have represented S_2 with regards to the squared term for the sake of succinctness. However, for matrix differentiation, let's expand it again.

$$
\begin{aligned}
T_3 &= y_i^2 - 2y_i\left(w_i'\mu_\beta^q + \delta\left(\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)\right) + 2\delta w_i'\mu_\beta^q\left(\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right) + \mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right) \\
&\quad + \delta^2\left(\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)^2 + 2\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q + \left(\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\right)^2\right) \\
\frac{\partial T_3}{\partial\Sigma_\theta^q} &= -2\delta y_i\varphi_J^a\left(x_i\right) + 2\delta w_i'\mu_\beta^q\varphi_J^a\left(x_i\right) + 2\delta^2\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right) + 2\delta^2\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q\varphi_J^a\left(x_i\right) \\
\frac{\partial T_4}{\partial\Sigma_\theta^q} &= 4\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right) + 4\varphi_J^a\left(x_i\right)\mu_\theta^q\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)
\end{aligned}
$$

The original paper gives the following derivative:

$$
\begin{aligned}
\frac{\partial S_2}{\partial\Sigma_\theta^q} = -\frac{1}{2}\mathrm{E}\left(\frac{1}{\sigma^2}\right)\sum_{i=1}^n &\Bigg\{4\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right) + 4\varphi_J^a\left(x_i\right)\mu_\theta^q\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right) \\
&-2\delta\left(y_i - w_i'\mu_\beta^q - \delta\underbrace{\operatorname{Tr}\left(\Sigma_\theta^q\varphi_J^a\left(x_i\right)\right)}_{(1)} - \delta\underbrace{\mu_\theta^{q\prime}\Sigma_\theta^q\mu_\theta^q}_{(2)}\right)\varphi_J^a\left(x_i\right)\Bigg\}
\end{aligned}
$$

First, where did (1) come from? Second, (2) should be $\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q$.

(1) is supposed to be $\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right)$ because the derivative of $\operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)^2$ is missing. Let $f = \operatorname{Tr}\left(\varphi_J^a\left(x_i\right)\Sigma_\theta^q\right)^2 = \varphi\Sigma : \varphi\Sigma$.

$$
\begin{aligned}
f &= \varphi\Sigma : \varphi\Sigma \\
\partial f &= 2\left(\varphi\Sigma : \varphi\partial\Sigma\right) \\
&= 2\varphi\Sigma\varphi : \partial\Sigma \\
\frac{\partial f}{\partial\Sigma} &= 2\varphi\Sigma\varphi.
\end{aligned}
$$

Therefore, the derivative should be as follows:

$$\frac{\partial S_2}{\partial \Sigma_\theta^q} = -\frac{1}{2}\,\mathrm{E}\left(\frac{1}{\sigma^2}\right)\sum_{i=1}^{n}\left\{\underbrace{6}_{(1)}\varphi_J^a\left(x_i\right)\Sigma_\theta^q\varphi_J^a\left(x_i\right) + 4\varphi_J^a\left(x_i\right)\mu_\theta^q\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\right.$$

$$\left. -2\delta\left(y_i - w_i'\mu_\beta^q - \delta\underbrace{\mu_\theta^{q\prime}\varphi_J^a\left(x_i\right)\mu_\theta^q}_{(2)}\right)\varphi_J^a\left(x_i\right)\right\}$$

6 Probit: Monotone Convex/Concave Shape Restriction

6.1 Model