

# Cosine basis

Daeyoung Lim\*  
Department of Statistics  
Korea University

June 30, 2016

## 1 Normal

### 1.1 Model specifications

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

Joint density:

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

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

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

---

\*Prof. Taeryon Choi

## 1.2 Lower bound

### 1.2.1 LB: $E[\ln p(y|\Theta)]$

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

### 1.3 LB: $E[\ln p(\theta_j|\sigma, \tau, \psi)]$

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

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

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

## 2 Probit: No Restriction

### 2.1 Model specifications

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

Joint density:

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

where  $C$  is the normalizing constant. The variational distributions are

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

## 2.2 Lower bound

### 2.2.1 LB: $\mathbb{E}[\ln p(y^*|\text{rest})] + \mathbb{H}[y^*]$

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

### 2.2.2 LB: $\mathbb{E}[\ln p(\theta_j|\sigma, \tau, \psi)]$

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

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

Therefore,

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

where

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

### 2.2.3 LB: $\mathbb{E} [\ln p(\tau^2)] + \mathbb{H} [\tau^2]$

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

### 2.2.4 LB: $\mathbb{E} [\ln p(\sigma^2)] + \mathbb{H} [\sigma^2]$

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

### 2.2.5 LB: $\mathbb{E} [\ln p(\beta)] + \mathbb{H} [\beta]$

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

### 2.2.6 LB: $\mathbb{E} [\ln p(\psi)] + \mathbb{H} [\psi]$

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

## 2.3 Update

### 2.3.1 $\theta_j$

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

### 2.3.2 $\tau^2$

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

### 2.3.3 $\sigma^2$

$$r_{q,\sigma} \leftarrow r_{0,\sigma} + J + p + 1$$

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

### 2.3.4 $\beta$

$$\Sigma_{\beta}^q \leftarrow \left( \frac{r_{q,\sigma}}{s_{q,\sigma}} \Sigma_{\beta}^{0^{-1}} + W'W \right)^{-1}$$

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

### 2.3.5 $\psi$

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

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

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

$$- j \exp \left( \frac{\sigma_{\psi}^{q^2} j^2}{2} - \mu_{\psi}^q j \right) \left\{ 1 - \Phi \left( \frac{\mu_{\psi}^q}{\sigma_{\psi}^q} - \sigma_{\psi}^q j \right) \right\} - \frac{1}{\sigma_{\psi}^q} \exp \left( \frac{\sigma_{\psi}^{q^2} j^2}{2} - \mu_{\psi}^q j \right) \phi \left( \frac{\mu_{\psi}^q}{\sigma_{\psi}^q} - \sigma_{\psi}^q j \right)$$

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

$$+ j \exp \left( \frac{\sigma_{\psi}^{q^2} j^2}{2} - \mu_{\psi}^q j \right) \left\{ 1 - \Phi \left( \frac{\mu_{\psi}^q}{\sigma_{\psi}^q} - \sigma_{\psi}^q j \right) \right\} + \left( \frac{1}{2\sigma_{\psi}^q} j + \frac{\mu_{\psi}^q}{2(\sigma_{\psi}^{q^2})^{3/2}} \right) \phi \left( \frac{\mu_{\psi}^q}{\sigma_{\psi}^q} - \sigma_{\psi}^q j \right) \exp \left( \frac{\sigma_{\psi}^{q^2} j^2}{2} - \mu_{\psi}^q j \right)$$

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

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

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

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

### 3 Probit: No restriction, with normal random effects

#### 3.1 Model

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

Joint density:

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

where  $C$  is the normalizing constant.

#### 3.2 Lower Bound

##### 3.2.1 LB: $E(\ln p(y^* | \text{rest})) + H(y^*)$

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

##### 3.2.2 LB: $E(\ln p(u)) + H(u)$

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

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

### 3.2.3 LB: $\mathbb{E} [\ln p(\theta_j | \sigma, \tau, \psi)]$

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

where

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

### 3.2.4 LB: $\mathbb{E} [\ln p(\tau^2)] + \mathbb{H} [\tau^2]$

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

### 3.2.5 LB: $\mathbb{E} [\ln p(\sigma^2)] + \mathbb{H} [\sigma^2]$

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

### 3.2.6 LB: $\mathbb{E} [\ln p(\beta)] + \mathbb{H} [\beta]$

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

### 3.2.7 LB: $\mathbb{E} [\ln p(\psi)] + \mathbb{H} [\psi]$

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

### 3.3 Update

#### 3.3.1 $u$

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

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

#### 3.3.2 $y_i^*$

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

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

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

#### 3.3.3 $\theta_j$

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

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

#### 3.3.4 $\beta$

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

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

#### 3.3.5 $\tau^2$

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

#### 3.3.6 $\sigma^2$

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



## 4 Probit: Monotone Shape Restriction

### 4.1 Model

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

Joint density:

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

## 4.2 Update

### 4.2.1 $y_i^*$

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

Therefore,

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

truncated at zero where

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

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

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

### 4.2.2 $\theta$

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

Then by NCVMP introduced in Wand (2014),

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

By matrix differential calculus,

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

#### 4.2.3 $\beta$

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

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

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

#### 4.2.4 $\tau^2$

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

#### 4.2.5 $\sigma^2$

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

where

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

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

#### 4.2.6 $\psi$

Same as before except that  $r_{q,\sigma}/s_{q,\sigma}$  in  $S_2(\mu_\psi^q, \sigma_\psi^{q2})$  is replaced by  $\mathbf{E}(1/\sigma)$ .

### 4.3 Lower Bound

#### 4.3.1 $E(\ln p(y^*|\text{rest})) + H(y^*)$

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

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

#### 4.3.2 $E(\ln p(\theta)) + H(\theta)$

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

#### 4.3.3 $E(\ln p(\beta)) + H(\beta)$

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

#### 4.3.4 $E(\ln p(\tau^2)) + H(\tau^2)$

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

#### 4.3.5 $E(\ln p(\sigma^2)) + H(\sigma^2)$

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

### 4.3.6 $E(\ln p(\psi)) + H(\psi)$

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

### 4.3.7 LB

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

## 4.4 Some Calculation

### 4.4.1 $E(1/\sigma)$

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

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

for  $\text{Re}(r) > 0 \wedge \text{Re}(p) > -1$ . 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  $x \rightarrow 0$ ,  $\sigma \rightarrow \infty$ ; as  $x \rightarrow \infty$ ,  $\sigma \rightarrow 0$ . Therefore,

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

Now we can compute the expectations.

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

#### 4.4.2 $\phi(x)/\Phi(x)$ when $x$ is very small

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

$$\begin{aligned} E(y_i^*) &= \mu_{y_i^*} + \frac{\phi(\mu_{y_i^*})}{\Phi(\mu_{y_i^*})^{y_i} (\Phi(\mu_{y_i^*}) - 1)^{1-y_i}} \\ \mu_{y_i^*} &= w'_i \mu_\beta^q + \delta \text{Tr}(\varphi_J^a(x_i) \Sigma_\theta^q) + \delta \mu_\theta^{q'} \varphi_J^a(x_i) \mu_\theta^q \end{aligned}$$

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

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

This becomes clearer when we look at the Laurent series of  $\phi(t)/\Phi(t)$ :

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

For such a reason, even for as fairly reasonably small a value as  $-40$ ,  $\text{dnorm}(-40)/\text{pnorm}(-40) = \text{NaN}$ .

## 4.5 Corrections

### 4.5.1 Original Paper: Linear model

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

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

where

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

Arranging the terms,

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

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.

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

The original paper gives the following derivative:

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

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

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

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

Therefore, the derivative should be as follows:

$$\begin{aligned}
\frac{\partial S_2}{\partial \Sigma_\theta^q} &= -\frac{1}{2} \text{E} \left( \frac{1}{\sigma^2} \right) \sum_{i=1}^n \left\{ \underbrace{6}_{(1)} \varphi_J^a(x_i) \Sigma_\theta^q \varphi_J^a(x_i) + 4\varphi_J^a(x_i) \mu_\theta^q \mu_\theta^{q'} \varphi_J^a(x_i) \right. \\
&\quad \left. - 2\delta \left( y_i - w'_i \mu_\beta^q - \underbrace{\delta \mu_\theta^{q'} \varphi_J^a(x_i) \mu_\theta^q}_{(2)} \right) \varphi_J^a(x_i) \right\}
\end{aligned}$$

## 5 Probit: Monotone Convex/Concave Shape Restriction

### 5.1 Model