Incorporating Random effect into GP Sparse Approximation

Daeyoung Lim*
Department of Statistics
Korea University

March 9, 2016

1 GP linear mixed model

1.1 Model specifications

The model specifications remain the same except that the random effect term has been added.

$$
\begin{aligned}
y|\theta &\sim \mathcal{N}\left(Z\alpha + A\beta, \gamma^2 I_n\right), (n \times n) \\
\alpha|\sigma^2 &\sim \mathcal{N}\left(0, \frac{\sigma^2}{m} I_{2m}\right), (2m \times 1) \\
\beta &\sim \mathcal{N}\left(0, \Sigma_\beta\right), (s \times 1) \\
\lambda &\sim \mathcal{N}\left(\mu_\lambda, \Sigma_\lambda\right), (d \times 1) \\
\sigma &\sim \text{half-Cauchy}\left(A_\sigma\right) \\
\gamma &\sim \text{half-Cauchy}\left(A_\gamma\right)
\end{aligned}
$$

where A is the design matrix for the random effects and β is the parameter vector of the random effects.

1.2 Lower bound

$$
p\left(y, \theta\right) = \mathcal{N}\left(y|Z\alpha + A\beta, \gamma^2 I_n\right) \mathcal{N}\left(\alpha \middle| 0, \frac{\sigma^2}{m} I_{2m}\right) \mathcal{N}\left(\beta|0, \Sigma_\beta\right) \mathcal{N}\left(\lambda|\mu_\lambda, \Sigma_\lambda\right) \mathrm{HC}\left(A_\sigma\right) \mathrm{HC}\left(A_\gamma\right)
$$

*Prof. Taeryon Choi

$$
\begin{aligned}
(1)\ \mathbb{E}\left[\log p\left(y|\theta\right)\right] &= -\frac{n}{2}\log\left(2\pi\right) - \frac{n}{2}\mathbb{E}\left[\log\gamma^2\right] \\
&\quad - \frac{1}{2}\mathbb{E}\left[\frac{1}{\gamma^2}\right]\Big\{y^Ty - 2\left(\mathbb{E}\left[Z\right]m_\alpha + Am_\beta\right)^T y + \mathrm{Tr}\left(\mathbb{E}\left[Z^TZ\right]S_\alpha\right) + m_\alpha^T\mathbb{E}\left[Z^TZ\right]m_\alpha \\
&\quad + 2m_\beta^TA^T\mathbb{E}\left[Z\right]m_\alpha + \mathrm{Tr}\left(A^TAS_\beta\right) + m_\beta^TA^TAm_\beta\Big\} \\
(2)\ \mathbb{E}\left[\log p\left(\alpha|\sigma\right)\right] &= -m\log\left(2\pi\right) - m\mathbb{E}\left[\log\sigma^2\right] + m\log m - \frac{m}{2}\mathbb{E}\left[\frac{1}{\sigma^2}\right]\mathrm{Tr}\left(S_\alpha + m_\alpha m_\alpha^T\right) \\
(3)\ \mathbb{E}\left[\log p\left(\beta\right)\right] &= -\frac{s}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left|\Sigma_\beta\right| - \frac{1}{2}\left\{\mathrm{Tr}\left(\Sigma_\beta^{-1}S_\beta\right) + m_\beta^T\Sigma_\beta^{-1}m_\beta\right\} \\
(4)\ \mathbb{E}\left[\log p\left(\lambda\right)\right] &= -\frac{d}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left|\Sigma_\lambda\right| - \frac{1}{2}\left(m_\lambda - \mu_\lambda\right)^T\Sigma_\lambda^{-1}\left(m_\lambda - \mu_\lambda\right) - \frac{1}{2}\mathrm{Tr}\left(\Sigma_\lambda^{-1}S_\lambda\right) \\
(5)\ \mathbb{E}\left[\log p\left(\sigma\right)\right] &= \log\left(2A_\sigma\right) - \log\pi - \mathbb{E}\left[\log\left(A_\sigma^2 + \sigma^2\right)\right] \\
(6)\ \mathbb{E}\left[\log p\left(\gamma\right)\right] &= \log\left(2A_\gamma\right) - \log\pi - \mathbb{E}\left[\log\left(A_\gamma^2 + \gamma^2\right)\right] \\
(1)\ \mathbb{E}\left[\log q\left(\alpha\right)\right] &= -m\log\left(2\pi\right) - \frac{1}{2}\log\left|S_\alpha\right| - m \\
(2)\ \mathbb{E}\left[\log q\left(\lambda\right)\right] &= -\frac{d}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left|S_\lambda\right| - \frac{d}{2} \\
(3)\ \mathbb{E}\left[\log q\left(\beta\right)\right] &= -\frac{s}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left|S_\beta\right| - \frac{s}{2} \\
(4)\ \mathbb{E}\left[\log q\left(\sigma\right)\right] &= -C_\sigma\frac{\mathcal{H}\left(2m, C_\sigma, A_\sigma^2\right)}{\mathcal{H}\left(2m-2, C_\sigma, A_\sigma^2\right)} - \log\mathcal{H}\left(2m-2, C_\sigma, A_\sigma^2\right) - 2m\mathbb{E}\left[\log\sigma\right] - \mathbb{E}\left[\log\left(A_\sigma^2 + \sigma^2\right)\right] \\
(5)\ \mathbb{E}\left[\log q\left(\gamma\right)\right] &= -C_\gamma\frac{\mathcal{H}\left(n, C_\gamma, A_\gamma^2\right)}{\mathcal{H}\left(n-2, C_\gamma, A_\gamma^2\right)} - \log\mathcal{H}\left(n-2, C_\gamma, A_\gamma^2\right) - n\mathbb{E}\left[\log\gamma\right] - \mathbb{E}\left[\log\left(A_\gamma^2 + \gamma^2\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L} &= (1) + (2) + (3) + (4) + (5) + (6) - (1) - (2) - (3) - (4) - (5) \\
&= -\frac{n}{2}\log\left(2\pi\right) + m\log m - \frac{m}{2}\mathbb{E}\left[\frac{1}{\sigma^2}\right]\left(\mathrm{Tr}\left(S_\alpha\right) + m_\alpha' m_\alpha\right) + \frac{1}{2}\log\left|\Sigma_\beta^{-1}S_\beta\right| + \frac{1}{2}\log\left|\Sigma_\lambda^{-1}S_\lambda\right| + \frac{1}{2}\log\left|S_\alpha\right| \\
&\quad - \frac{1}{2}\left\{\mathrm{Tr}\left(\Sigma_\beta^{-1}S_\beta\right) + m_\beta'\Sigma_\beta^{-1}m_\beta\right\} - \frac{1}{2}\left(m_\lambda - \mu_\lambda\right)'\Sigma_\lambda^{-1}\left(m_\lambda - \mu_\lambda\right) + \log\left(\frac{4A_\sigma A_\gamma}{\pi^2}\right) + m + \frac{d}{2} + \frac{s}{2} \\
&\quad + C_\sigma\frac{\mathcal{H}\left(2m, C_\sigma, A_\sigma^2\right)}{\mathcal{H}\left(2m-2, C_\sigma, A_\sigma^2\right)} + C_\gamma\frac{\mathcal{H}\left(n, C_\gamma, A_\gamma^2\right)}{\mathcal{H}\left(n-2, C_\gamma, A_\gamma^2\right)} + \log\mathcal{H}\left(2m-2, C_\sigma, A_\sigma^2\right) + \log\mathcal{H}\left(n-2, C_\gamma, A_\gamma^2\right) \\
&\quad - \frac{1}{2}\mathbb{E}\left[\frac{1}{\gamma^2}\right]\left\{y'y - 2\left(\mathbb{E}\left[Z\right]m_\alpha + Am_\beta\right)'y + \mathrm{Tr}\left(\mathbb{E}\left[Z'Z\right]\left(S_\alpha + m_\alpha m_\alpha'\right)\right) + 2m_\beta'A'\mathbb{E}\left[Z\right]m_\alpha\right. \\
&\quad \left. + \mathrm{Tr}\left(A'AS_\beta\right) + m_\beta'A'Am_\beta\right\}
\end{aligned}
$$

1.3 Variational update

1.3.1 $q\left(\alpha\right)$

$$
\begin{aligned}
\log q\left(\alpha\right) &\propto \left\langle \log\mathcal{N}\left(y|Z\alpha + A\beta, \gamma^2 I_n\right)\mathcal{N}\left(\alpha\middle|0, \frac{\sigma^2}{m}I_{2m}\right)\right\rangle \\
&\propto \mathbb{E}\left[\frac{1}{\gamma^2}\right]\left(\mathbb{E}\left[Z'\right]y - \mathbb{E}\left[Z'\right]Am_\beta\right)\alpha - \frac{1}{2}\alpha'\left(m\mathbb{E}\left[\frac{1}{\sigma^2}\right]I_{2m} + \mathbb{E}\left[\frac{1}{\gamma^2}\right]\mathbb{E}\left[Z'Z\right]\right)\alpha \\
S_\alpha &= \left(\frac{\mathcal{H}\left(n, C_\gamma, A_\gamma^2\right)}{\mathcal{H}\left(n-2, C_\gamma, A_\gamma^2\right)}\mathbb{E}\left[Z'Z\right] + m\frac{\mathcal{H}\left(2m, C_\sigma, A_\sigma^2\right)}{\mathcal{H}\left(2m-2, C_\sigma, A_\sigma^2\right)}I_{2m}\right)^{-1} \\
m_\alpha &= \frac{\mathcal{H}\left(n, C_\gamma, A_\gamma^2\right)}{\mathcal{H}\left(n-2, C_\gamma, A_\gamma^2\right)}S_\alpha\mathbb{E}\left[Z'\right]\left(y - Am_\beta\right)
\end{aligned}
$$

1.3.2 $q(\beta)$

$$
\begin{aligned}
\log q(\beta) &\propto \log \mathcal{N}\left(y|Z\alpha + A\beta, \gamma^2 I_n\right) + \log \mathcal{N}\left(\beta|0, \Sigma_\beta\right) \\
&\propto -\frac{1}{2}\mathbb{E}\left[\frac{1}{\gamma^2}\right]\left(-2y'A\beta + 2m_\alpha'\mathbb{E}\left[Z'\right]A\beta + \beta'A'A\beta\right) - \frac{1}{2}\beta'\Sigma_\beta^{-1}\beta \\
&\propto -\frac{1}{2}\left\{\beta'\left(\mathbb{E}\left[\frac{1}{\gamma^2}\right]A'A + \Sigma_\beta^{-1}\right)\beta - 2\left(A'y - A'\mathbb{E}\left[Z\right]m_\alpha\right)\beta\right\} \\
S_\beta &= \left(\frac{\mathcal{H}\left(n, C_\gamma, A_\gamma^2\right)}{\mathcal{H}\left(n-2, C_\gamma, A_\gamma^2\right)}A'A + \Sigma_\beta^{-1}\right)^{-1} \\
m_\beta &= S_\beta A'\left(y - \mathbb{E}\left[Z\right]m_\alpha\right)
\end{aligned}
$$

1.3.3 $q(\sigma)$

$$
\begin{aligned}
\log q(\sigma) &\propto \log \mathcal{N}\left(\alpha \middle| 0, \frac{\sigma^2}{m}I_{2m}\right)\mathrm{HC}\left(A_\sigma\right) \\
&\propto -m\log\sigma^2 - \frac{m}{2}\left(\mathrm{Tr}\left(S\right) + m_\alpha'm_\alpha\right)/\sigma^2 - \log\left(A_\sigma^2 + \sigma^2\right) \\
q(\alpha) &\propto \frac{\exp\left\{-\frac{m}{2}\left(\mathrm{Tr}\left(S_\alpha\right) + m_\alpha'm_\alpha\right)/\sigma^2\right\}}{\sigma^{2m}\left(A_\sigma^2 + \sigma^2\right)}
\end{aligned}
$$

1.3.4 $q(\gamma)$

$$
\begin{aligned}
\log q(\gamma) &\propto \log \mathcal{N}\left(y|Z\alpha + A\beta, \gamma^2 I_n\right) + \log \mathrm{HC}\left(A_\gamma\right) \\
&\propto -n\log\gamma - \log\left(A_\beta^2 + \gamma^2\right) \\
&\quad - \frac{1}{2\gamma^2}\left(y'y - 2y'\left(\mathbb{E}\left[Z\right]m_\alpha + Am_\beta\right) + \mathrm{Tr}\left(\mathbb{E}\left[Z'Z\right]\left(S_\alpha + m_\alpha m_\alpha'\right)\right) + 2m_\beta'A'\mathbb{E}\left[Z\right]m_\alpha + m_\beta'A'Am_\beta + \mathrm{Tr}\left(S_\beta\right)\right) \\
q(\gamma) &\propto \frac{\exp\left\{-C_\gamma/\gamma^2\right\}}{\gamma^n\left(A_\gamma^2 + \gamma^2\right)} \\
C_\gamma &= \frac{1}{2}\left(y'y - 2y'\left(\mathbb{E}\left[Z\right]m_\alpha + Am_\beta\right) + \mathrm{Tr}\left(\mathbb{E}\left[Z'Z\right]\left(S_\alpha + m_\alpha m_\alpha'\right)\right) + 2m_\beta'A'\mathbb{E}\left[Z\right]m_\alpha + m_\beta'A'Am_\beta + \mathrm{Tr}\left(S_\beta\right)\right)
\end{aligned}
$$

1.3.5 $q(\lambda)$

To perform nonconjugate variational message passing, we first need to compute $\mathbb{E}\left[\log p\left(y, \theta\right)\right]$ leaving only the parts that contain $\left(m_\lambda, S_\lambda\right)$ since every other element will be eliminated once we differentiate it with regard to the mean and covariance of lambda.

$$
\begin{aligned}
\mathbb{E}\left[\log p\left(y,\theta\right)\right] \propto\ & \mathbb{E}\left[\frac{1}{\gamma^2}\right] y'\mathbb{E}\left[Z\right] m_\alpha - \frac{1}{2}\mathbb{E}\left[\frac{1}{\gamma^2}\right]\left(\mathrm{Tr}\left(\mathbb{E}\left[Z'Z\right]\left(S_\alpha + m_\alpha m_\alpha'\right)\right)\right. \\
& - \mathbb{E}\left[\frac{1}{\gamma^2}\right] m_\beta' A'\mathbb{E}\left[Z\right] m_\alpha - \frac{1}{2}\left(m_\lambda - \mu_\lambda\right)'\Sigma_\lambda^{-1}\left(m_\lambda - \mu_\lambda\right) - \frac{1}{2}\mathrm{Tr}\left(\Sigma_\lambda^{-1} S_\lambda\right) \\
\propto\ & \mathbb{E}\left[\frac{1}{\gamma^2}\right]\left\{y - Am_\beta\right\}'\mathbb{E}\left[Z\right] m_\alpha - \frac{1}{2}\mathbb{E}\left[\frac{1}{\gamma^2}\right]\left(\mathrm{Tr}\left(\mathbb{E}\left[Z'Z\right]\left(S_\alpha + m_\alpha m_\alpha'\right)\right)\right. \\
& - \frac{1}{2}\left(m_\lambda - \mu_\lambda\right)'\Sigma_\lambda^{-1}\left(m_\lambda - \mu_\lambda\right) - \frac{1}{2}\mathrm{Tr}\left(\Sigma_\lambda^{-1} S_\lambda\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{\gamma^2}\right]\left\{y - Am_\beta\right\}'\mathbb{E}\left[Z\right] m_\alpha &= \mathbb{E}\left[\frac{1}{\gamma^2}\right]\sum_{i=1}^{n}\sum_{j=1}^{2m}\left\langle Z\right\rangle_{ij} e_i m_j^\alpha, \qquad e = y - Am_\beta \\
&= \mathbb{E}\left[\frac{1}{\gamma^2}\right]\sum_{i=1}^{n}\sum_{j=1}^{m} e_i \exp\left\{-\frac{1}{2}t_{ij}' S_\lambda t_{ij}\right\}\left(m_j^\alpha \cos\left(t_{ij}' m_\lambda\right) + m_{j+m}^\alpha \sin\left(t_{ij}' m_\lambda\right)\right) \\
F_3 &= \frac{\partial}{\partial m_\lambda}\mathbb{E}\left[\frac{1}{\gamma^2}\right]\left\{y - Am_\beta\right\}'\mathbb{E}\left[Z\right] m_\alpha \\
&= \mathbb{E}\left[\frac{1}{\gamma^2}\right]\sum_{i=1}^{n}\sum_{j=1}^{m} e_i \exp\left\{-\frac{1}{2}t_{ij}' S_\lambda t_{ij}\right\}\left(m_{j+m}^\alpha \cos\left(t_{ij}' m_\lambda\right) - m_j^\alpha \sin\left(t_{ij}' m_\lambda\right)\right) t_{ij} \\
F_1 &= \frac{\partial}{\partial S_\lambda}\mathbb{E}\left[\frac{1}{\gamma^2}\right]\left\{y - Am_\beta\right\}'\mathbb{E}\left[Z\right] m_\alpha \\
&= -\frac{1}{2}\mathbb{E}\left[\frac{1}{\gamma^2}\right]\sum_{i=1}^{n}\sum_{j=1}^{m} e_i \exp\left\{-\frac{1}{2}t_{ij}' S_\lambda t_{ij}\right\}\left(m_j^\alpha \cos\left(t_{ij}' m_\lambda\right) + m_{j+m}^\alpha \sin\left(t_{ij}' m_\lambda\right)\right) t_{ij} t_{ij}'
\end{aligned}
$$

2 GP Logistic Model

2.1 Model specifications

For logistic models, we first postulate a link function, $g\left(\cdot\right)$ for the predictors.

$$
\begin{aligned}
y &= g^{-1}\left(\eta\right) + \epsilon, \quad \epsilon \sim \mathcal{N}\left(0, \gamma^2 I_n\right) \\
g\left(\mathbb{E}\left[y\right]\right) &= \eta \\
\eta &= f\left(x\right) + A\beta \\
g^{-1}\left(x\right) &= e^x / \left(1 + e^x\right) \\
f &\sim \mathcal{GP}\left(m\left(\cdot\right), \kappa\left(\cdot,\cdot\right)\right)
\end{aligned}
$$

Since y is a Bernoulli random variable, $\mathbb{E}\left[y\right] = \mathbb{P}\left(y = 1\right)$. Without loss of generality, we will assume the mean function to be zero and that the Gaussian process has a sparse approximation representation as in Tan & Nott. Therefore,

$$
f\left(x\right) \approx \sum_{r=1}^{m}\left[a_r \cos\left\{\left(s_r \odot x\right)^T \lambda\right\} + b_r \sin\left\{\left(s_r \odot x\right)^T \lambda\right\}\right]
$$

and this could further be represented in matrix form which reduces this to a linear model.

$$
f\left(x\right) = Z\alpha
$$

$$
\begin{aligned}
\eta &= Z\alpha + A\beta \\
y &= \exp\left\{\left(Z\alpha + A\beta\right) - \log\left(\mathbf{1} + \exp\left\{Z\alpha + A\beta\right\}\right)\right\} + \epsilon
\end{aligned}
$$

Every scalar function applied to a vector or a matrix is done so elementwise. We think of the following priors:

$$
\begin{aligned}
\alpha|\sigma &\sim \mathcal{N}\left(0, \frac{\sigma^2}{m} I_{2m}\right) \\
\beta &\sim \mathcal{N}\left(\mu_\beta, \Sigma_\beta\right) \\
\lambda &\sim \mathcal{N}\left(\mu_\lambda, \Sigma_\lambda\right) \\
\sigma &\sim \text{half-Cauchy}\left(A_\sigma\right) \\
\gamma &\sim \text{half-Cauchy}\left(A_\gamma\right) \\
\theta &= (\alpha, \beta, \lambda, \sigma, \gamma)
\end{aligned}
$$

$$
\begin{aligned}
\log p\left(y, \theta\right) = y^T\left(Z\alpha + A\beta\right) - \mathbf{1}_n^T \log\left(\mathbf{1}_n + \exp\left\{Z\alpha + A\beta\right\}\right) - \left(m + \frac{s+d}{2}\right)\log\left(2\pi\right) - m\log\sigma^2 + m\log m \\
- \frac{m}{2\sigma^2}\alpha^T\alpha - \frac{1}{2}\log\left|\Sigma_\beta\right| - \frac{1}{2}\left(\beta - \mu_\beta\right)^T\Sigma_\beta^{-1}\left(\beta - \mu_\beta\right) - \frac{1}{2}\log\left|\Sigma_\lambda\right| - \frac{1}{2}\left(\lambda - \mu_\lambda\right)^T\Sigma_\lambda^{-1}\left(\lambda - \mu_\lambda\right) \\
+ \log\left(2A_\sigma\right) + \log\left(2A_\gamma\right) 2\log\pi - \log\left(A_\sigma^2 + \sigma^2\right) - \log\left(A_\gamma^2 + \gamma^2\right)
\end{aligned}
$$

Because $-\mathbf{1}_n^T \log\left(\mathbf{1}_n + \exp\left\{Z\alpha + A\beta\right\}\right)$ is analytically intractable for expectation which is essentially integration, we come up with the following approximation:

$$
\begin{aligned}
-\log\left(1 + e^x\right) &= \max_{\xi\in\mathbb{R}}\left\{B\left(\xi\right)x^2 - \frac{1}{2}x + C\left(\xi\right)\right\}, \quad \forall x \in \mathbb{R} \\
B\left(\xi\right) &= -\tanh\left(\xi/2\right)/\left(4\xi\right) \\
C\left(\xi\right) &= \xi/2 - \log\left(1 + e^\xi\right) + \xi\tanh\left(\xi/2\right)/4
\end{aligned}
$$

then

$$
\begin{aligned}
-\mathbf{1}_n^T \log\left\{\mathbf{1}_n^T + \exp\left(Z\alpha + A\beta\right)\right\} &\geq \mathbf{1}_n^T\left\{B\left(\xi\right)\odot\left(Z\alpha + A\beta\right)^2 - \frac{1}{2}\left(Z\alpha + A\beta\right) + C\left(\xi\right)\right\} \\
&= \left(Z\alpha + A\beta\right)^T \mathrm{Dg}\left\{B\left(\xi\right)\right\}\left(Z\alpha + A\beta\right) - \frac{1}{2}\mathbf{1}_n^T\left(Z\alpha + A\beta\right) + \mathbf{1}_n^T C\left(\xi\right),
\end{aligned}
$$

where $\xi = (\xi_1, \ldots, \xi_n)$.

$$
\begin{aligned}
\log \underline{p}\left(y, \theta; \xi\right) = y^T\left(Z\alpha + A\beta\right) + \left(Z\alpha + A\beta\right)^T \mathrm{Dg}\left\{B\left(\xi\right)\right\}\left(Z\alpha + A\beta\right) - \frac{1}{2}\mathbf{1}_n^T\left(Z\alpha + A\beta\right) + \mathbf{1}_n^T C\left(\xi\right) \\
- \left(m + \frac{s+d}{2}\right)\log\left(2\pi\right) - m\log\sigma^2 + m\log m - \frac{m}{2\sigma^2}\alpha^T\alpha - \frac{1}{2}\log\left|\Sigma_\beta\right| - \frac{1}{2}\left(\beta - \mu_\beta\right)^T\Sigma_\beta^{-1}\left(\beta - \mu_\beta\right) \\
- \frac{1}{2}\log\left|\Sigma_\lambda\right| - \frac{1}{2}\left(\lambda - \mu_\lambda\right)^T\Sigma_\lambda^{-1}\left(\lambda - \mu_\lambda\right) + \log\left(2A_\sigma\right) + \log\left(2A_\gamma\right) 2\log\pi - \log\left(A_\sigma^2 + \sigma^2\right) \\
- \log\left(A_\gamma^2 + \gamma^2\right)
\end{aligned}
$$

3 Sparse GP Probit Regression

This section will extensively use the results of [1] Tan, Linda S. L., Ong, Victor M. H., Nott, David J., and Jasra, Ajay. "Variational Inference for Sparse Spectrum Gaussian Process Regression," May 9, 2013.

3.1 Bayesian probit regression without random effects

The ordinary formulation goes like this:

$$
\mathbb{P}\left(y_i = 1|x_i, \beta\right) = \Phi\left(x_i^T\beta\right),
$$

where Φ is the cumulative distribution of standard Gaussian distribution. Since the following likelihood

$$\mathbb{P}(Y = y|X,\beta) = \prod_{i=1}^{n} \Phi\left(x_i^T\beta\right)^{y_i}\left[1-\Phi\left(x_i^T\beta\right)\right]^{1-y_i}$$

hinders the tractable calculation of the posterior, we devise the following latent variable:

$$Z_i = x_i^T\beta + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0,1)$$

for $i = 1, \dots, n$ and let

$$y_i = \begin{cases} 1, & \text{if } Z_i \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

It automatically follows that

$$\begin{aligned} Z_i|y_i = 0, x_i, \beta &\sim \mathcal{N}\left(x_i^T\beta, 1\right)\mathbf{1}\left[Z_i < 0\right] \\ Z_i|y_i = 1, x_i, \beta &\sim \mathcal{N}\left(x_i^T\beta, 1\right)\mathbf{1}\left[Z_i \geq 0\right] \end{aligned}$$

suggesting a truncated Gaussian for each Z_i depending on what value y_i takes on. According to probit regression without random effects, the introduction of latent variables enables the tractable computation of the posterior with regards to the parameters.

$$\pi(Z,y,\beta) = C\pi(\beta)\prod_{i=1}^{n}\left\{1\left[Z_i \geq 0\right]1\left[y_i = 1\right] + 1\left[Z_i < 0\right]1\left[y_i = 0\right]\right\}\phi\left(Z_i - x_i^T\beta\right)$$

where C is the proportionality constant and ϕ is the density of standard Gaussian. Following the mean-field approximation and computing the optimal density for Z_i, it becomes a truncated normal about zero with mean $x_i^T\beta$ and variance 1.

3.2 Incorporating random effects and nonparametric statistics

Taking one step forward, the random effects can be taken into consideration in many situations. Furthermore, we often do not know exactly what functional form the parameters should take on. Such situations greatly call for the use of nonparametric statistics.

$$\mathbb{P}(y_i = 1|f) = \Phi\left(f\left(x_i\right)\right).$$

In the previous section where linear probit regression was considered, the functional form was assumed to be linear, i.e. $x_i^T\beta$. However, we now insist that f could be any function. Adopting Gaussion process prior for f and using sparse approximation,

$$\mathbb{P}(y_i = 1|f) = \Phi\left(Z_i^T\alpha\right).$$

Note that Z_i is different from the one we used in the previous section. Moreover, we will assume there exist additive random effects:

$$\mathbb{P}(y_i = 1|f,\alpha,\beta) = \Phi\left(z_i^T\alpha + a_i^T\beta\right).$$

Assume latent variables

$$y_i^* = z_i^T\alpha + a_i^T\beta + \epsilon, \quad \epsilon \sim \mathcal{N}(0,1).$$

Let

$$y_i = \begin{cases} 1, & \text{if } y_i^* \geq 0 \\ 0, & \text{if } y_i^* < 0. \end{cases}$$

Let's assume priors

$$
\begin{aligned}
\alpha|\sigma &\sim \mathcal{N}\left(0, \frac{\sigma^2}{m} I_{2m}\right) \\
\beta &\sim \mathcal{N}\left(\mu_\beta, \Sigma_\beta\right) \\
\lambda &\sim \mathcal{N}\left(\mu_\lambda, \Sigma_\lambda\right) \\
\sigma &\sim \text{half-Cauchy}\left(A_\sigma\right) \\
\theta &= \left(\alpha, \beta, \lambda, \sigma\right)
\end{aligned}
$$

$$
\pi\left(y, y^*, \theta\right) = C\pi\left(\alpha|\sigma\right)\pi\left(\beta\right)\pi\left(\lambda\right)\pi\left(\sigma\right)\prod_{i=1}^{n}\left\{1\left[y_i^* \geq 0\right]1\left[y_i = 1\right] + 1\left[y_i^* < 0\right]1\left[y_i = 0\right]\right\}\phi\left(y_i^* - z_i^T\alpha_i - a_i^T\beta\right)
$$

3.2.1 $q\left(\alpha\right)$

Let $y^* = \begin{bmatrix} y_1^* & \dots & y_n^* \end{bmatrix}^T$.

$$
\begin{aligned}
\log q\left(\alpha\right) &\propto \left\langle \log\pi\left(\alpha|\sigma\right)\prod_{i=1}^{n}\phi\left(y_i^* - z_i^T\alpha - a_i^T\beta\right)\right\rangle \\
&\propto -\frac{m}{2}\left\langle\frac{1}{\sigma^2}\right\rangle\alpha^T\alpha + \left\langle\sum_{i=1}^{n}\left[-\frac{1}{2}\left\{y_i^{*2} - 2\left(z_i^T\alpha + a_i^T\beta\right)y_i^* + \left(z_i^T\alpha + a_i^T\beta\right)^2\right\}\right]\right\rangle \\
&\propto -\frac{m}{2}\left\langle\frac{1}{\sigma^2}\right\rangle\alpha^T\alpha + \left\langle-\frac{1}{2}\sum_{i=1}^{n}\left[-2y_i^*z_i^T\alpha + \alpha^Tz_iz_i^T\alpha + 2\beta^Ta_iz_i^T\alpha\right]\right\rangle \\
&\propto -\frac{m}{2}\left\langle\frac{1}{\sigma^2}\right\rangle\alpha^T\alpha - \frac{1}{2}\left\langle-2y^{*T}Z\alpha + \alpha^T\left(\sum_{i=1}^{n}z_iz_i^T\right)\alpha + 2\beta^T\left(\sum_{i=1}^{n}a_iz_i^T\right)\alpha\right\rangle \\
&\propto -\frac{m}{2}\left\langle\frac{1}{\sigma^2}\right\rangle\alpha^T\alpha - \frac{1}{2}\left\{-2\left\langle y^{*T}\right\rangle\left\langle Z\right\rangle\alpha + \alpha^T\left\langle Z^TZ\right\rangle\alpha + 2\left\langle\beta\right\rangle^TA^T\left\langle Z\right\rangle\alpha\right\} \\
&\propto -\frac{1}{2}\left\{\alpha^T\left(m\left\langle\frac{1}{\sigma^2}\right\rangle I_{2m} + \left\langle Z^TZ\right\rangle\right)\alpha - 2\left(\left\langle Z\right\rangle^T\left\langle y^*\right\rangle - \left\langle Z\right\rangle^TA\left\langle\beta\right\rangle\right)^T\alpha\right\} \\
q\left(\alpha\right) &= \mathcal{N}\left(\mu_{q(\alpha)}, \Sigma_{q(\alpha)}\right) \\
\mu_{q(\alpha)} &= \Sigma_{q(\alpha)}\left(\left\langle Z\right\rangle^T\left\langle y^*\right\rangle - \left\langle Z\right\rangle^TA\left\langle\beta\right\rangle\right) \\
\Sigma_{q(\alpha)} &= \left(m\left\langle\frac{1}{\sigma^2}\right\rangle I_{2m} + \left\langle Z^TZ\right\rangle\right)^{-1}
\end{aligned}
$$

3.2.2 $q(\beta)$

$$
\begin{aligned}
\log q(\beta) &\propto \log \pi(\beta) \prod_{i=1}^{n} \phi\left(y_i^* - z_i^T \alpha - a_i^T \beta\right) \\
&\propto -\frac{s}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_\beta| - \frac{1}{2}(\beta-\mu_\beta)^T \Sigma_\beta^T (\beta-\mu_\beta) \\
&\quad + \left\langle \sum_{i=1}^{T} \left[-\frac{1}{2}\left(y_i^* - 2\left(z_i^T\alpha + a_i^T\beta\right) y_i^* + \left(z_i^T\alpha + a_i^T\beta\right)^2 \right) \right] \right\rangle \\
&\propto -\frac{1}{2}\left(\beta^T\Sigma_\beta^{-1}\beta - 2\mu_\beta^T\Sigma_\beta^{-1}\beta\right) + \left\langle \sum_{i=1}^{n} \left[-\frac{1}{2}\left(-2y_i^* a_i^T\beta + 2\alpha^T z_i a_i^T\beta + \beta^T a_i a_i^T \beta\right) \right] \right\rangle \\
&\propto -\frac{1}{2}\left(\beta^T\Sigma_\beta^{-1}\beta - 2\mu_\beta^T\Sigma_\beta^{-1}\beta\right) - \frac{1}{2}\left(-2\left\langle y^{*T}\right\rangle A\beta + 2\left\langle \alpha^T\right\rangle\left\langle Z^T\right\rangle A\beta + \beta^T A^T A\beta\right) \\
&\propto -\frac{1}{2}\left\{\beta^T\Sigma_\beta^{-1}\beta + \beta^T A^T A\beta - 2\mu_\beta^T\Sigma_\beta^{-1}\beta + 2\left\langle \alpha^T\right\rangle\left\langle Z^T\right\rangle A\beta - 2\left\langle y^{*T}\right\rangle A\beta\right\} \\
&\propto -\frac{1}{2}\left\{\beta^T\left(\Sigma_\beta^{-1} + A^T A\right)\beta - 2\left(\Sigma_\beta^{-1}\mu_\beta + A^T\langle y^*\rangle - A^T\langle Z\rangle\langle\alpha\rangle\right)^T\beta\right\} \\
q(\beta) &= \mathcal{N}\left(\mu_{q(\beta)}, \Sigma_{q(\beta)}\right) \\
\mu_{q(\beta)} &= \Sigma_{q(\beta)}\left(\Sigma_\beta^{-1}\mu_\beta + A^T\langle y^*\rangle - A^T\langle Z\rangle\langle\alpha\rangle\right) \\
\Sigma_{q(\beta)} &= \left(\Sigma_\beta^{-1} + A^T A\right)^{-1}
\end{aligned}
$$

3.2.3 $q(y^*)$

$$
\begin{aligned}
\log q(y^*) &\propto \left\langle \log \prod_{i=1}^{n} \left\{1[y_i^* \geq 0]\,1[y_i = 1] + 1[y_i^* < 0]\,1[y_i = 0]\right\} \phi\left(y_i^* - z_i^T\alpha - a_i^T\beta\right) \right\rangle \\
&\propto \sum_{i=1}^{n} \log\left\{1[y_i^* \geq 0]\,1[y_i = 1] + 1[y_i^* < 0]\,1[y_i = 0]\right\} \\
&\quad + \left\langle \sum_{i=1}^{n} \left[-\frac{1}{2}\left(y_i^{*2} - 2\left(z_i^T\alpha + a_i^T\beta\right) y_i^* + \left(z_i^T\alpha + a_i^T\beta\right)^2 \right) \right] \right\rangle \\
&\propto \sum_{i=1}^{n} \log\left\{1[y_i^* \geq 0]\,1[y_i = 1] + 1[y_i^* < 0]\,1[y_i = 0]\right\} \\
&\quad - \frac{1}{2}\left(y^{*T}y^* - 2\left(\langle Z\rangle\langle\alpha\rangle + A\langle\beta\rangle\right)^T y^* \right) \\
q(y^*) &= \mathcal{TN}\left(\langle Z\rangle\langle\alpha\rangle + A\langle\beta\rangle, I_n\right)
\end{aligned}
$$

where $\mathcal{TN}$ indicates truncated normal distribution, in this case multivariate. Each element of y^* is truncated at zero. The direction of truncation remains the same.

$$
\mu_{q(y^*)} = Z\mu_{q(\alpha)} + A\mu_{q(\beta)} + \frac{\phi\left(Z\mu_{q(\alpha)} + A\mu_{q(\beta)}\right)}{\left\{\Phi\left(Z\mu_{q(\alpha)} + A\mu_{q(\beta)}\right)\right\}^{y}\left\{\Phi\left(Z\mu_{q(\alpha)} + A\mu_{q(\beta)}\right) - 1_n\right\}^{1_n - y}}
$$

3.2.4 $q(\sigma)$

$$
\begin{aligned}
\log q(\sigma) &\propto \langle \log \pi(\alpha|\sigma)\, \pi(\sigma) \rangle \\
&\propto -m \log \sigma^2 - \frac{m}{2} \left\langle \alpha^T \alpha \right\rangle / \sigma^2 - \log\left(A_\sigma^2 + \sigma^2\right) \\
q(\sigma) &\propto \frac{\exp\left(-C_\sigma / \sigma^2\right)}{\sigma^{2m}\left(A_\sigma^2 + \sigma^2\right)} \\
C_\sigma &= \frac{m}{2}\left(\mathrm{Tr}\left(\Sigma_{q(\alpha)}\right) + \mu_{q(\alpha)}^T \mu_{q(\alpha)}\right)
\end{aligned}
$$

3.2.5 LB: $\mathbb{E}\left[\log p(y, y^* | \theta)\right]$

Recall that

$$
\begin{aligned}
\pi(y, y^*, \theta) &= C \pi(\alpha|\sigma)\, \pi(\beta)\, \pi(\lambda)\, \pi(\sigma) \prod_{i=1}^{n} \left\{1\left[y_i^* \geq 0\right] 1\left[y_i = 1\right] + 1\left[y_i^* < 0\right] 1\left[y_i = 0\right]\right\} \phi\left(y_i^* - z_i^T \alpha - a_i^T \beta\right) \\
\mathcal{L} &= \mathbb{E}\left[\log \pi(y, y^*, \theta)\right] - \mathbb{E}\left[\log q(y^*)\, q(\theta)\right] \\
&= \mathbb{E}\left[\log p(y | y^*, \theta)\right] + \mathbb{E}\left[\log p(y^* | \theta)\right] + \mathbb{E}\left[\log p(\theta)\right] - \mathbb{E}\left[\log q(\theta)\right]
\end{aligned}
$$

The above equation includes all the nodes in the graphical model. Therefore, we compute the expecation of the logarithm of each node.

$$
\begin{aligned}
\log p(y, y^* | \theta) &= \log \prod_{i=1}^{n} \left\{1\left[y_i^* \geq 0\right] 1\left[y_i = 1\right] + 1\left[y_i^* < 0\right] 1\left[y_i = 0\right]\right\} \phi\left(y_i^* - z_i^T \alpha - a_i^T \beta\right) \\
&\Longrightarrow \mathbb{E}\left[\sum_{i=1}^{n} \log\left\{1\left[y_i^* \geq 0\right] 1\left[y_i = 1\right] + 1\left[y_i^* < 0\right] 1\left[y_i = 0\right]\right\}\right] + \sum_{i=1}^{n} \mathbb{E}\left[\log \phi\left(y_i^* - z_i^T \alpha - a_i^T \beta\right)\right]
\end{aligned}
$$

If we set $U = \{1\left[y_i^* \geq 0\right] 1\left[y_i = 1\right] + 1\left[y_i^* < 0\right] 1\left[y_i = 0\right]\}$ and let it be a new random variable, it is always 1 with probability 1. Therefore, the logarithm of U, i.e. $\log U$ is always 0 which terminates the need for calculating the expected value since it's 0 regardless. We proceed to the remaining terms.

$$
\sum_{i=1}^{n} \mathbb{E}\left[\log \phi\left(y_i^* - z_i^T \alpha - a_i^T \beta\right)\right] \qquad \cdots (1)
$$

The entropy of the variational distribution of $q(y_i^*)$ should be coupled with eqn (1) to simplify the calculation.

$$
\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\left[\log \phi\left(y_i^* - z_i^T \alpha - a_i^T \beta\right) - \log \phi\left(y_i^* - \left\langle z_i^T \right\rangle \mu_{q(\alpha)} - a_i^T \mu_{q(\beta)}\right)\right] \\
&+ \sum_{i=1}^{n} \log\left(\left\{\Phi\left(\langle z_i \rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right)\right\}^{y_i} \left\{1 - \Phi\left(\langle z_i \rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right)\right\}^{1-y_i}\right) \\
&= \sum_{i=1}^{n} \mathbb{E}\left[-\frac{1}{2}\left(-2\left(z_i^T \alpha + a_i^T \beta\right) y_i^* + \left(z_i^T \alpha + a_i^T \beta\right)^2 + 2\left(\langle z_i \rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right) y_i^*\right)\right] \\
&\quad + \frac{1}{2} \sum_{i=1}^{n} \left(\left(\langle z_i \rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right)^2\right) \\
&\quad + \sum_{i=1}^{n} \log\left(\left\{\Phi\left(\langle z_i \rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right)\right\}^{y_i} \left\{1 - \Phi\left(\langle z_i \rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right)\right\}^{1-y_i}\right)
\end{aligned}
$$

$$\boxed{\begin{aligned}&= -\frac{1}{2}\left(\mathrm{Tr}\left(\left\langle Z^T Z\right\rangle \Sigma_{q(\alpha)}\right) + \mu_{q(\alpha)}^T \left(\left\langle Z^T Z\right\rangle - \langle Z\rangle^T \langle Z\rangle\right) \mu_{q(\alpha)} + \mathrm{Tr}\left(A^T A \Sigma_{q(\beta)}\right)\right)\\ &\quad + \sum_{i=1}^{n} \log\left(\left\{\Phi\left(\langle z_i\rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right)\right\}^{y_i} \left\{1 - \Phi\left(\langle z_i\rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right)\right\}^{1-y_i}\right)\end{aligned}}$$

3.2.6 LB: $\mathbb{E}\left[\log \pi\left(\alpha|\sigma\right)\right] - \mathbb{E}\left[\log q\left(\alpha\right)\right]$

$$\begin{aligned}\mathbb{E}\left[\log \pi\left(\alpha|\sigma\right)\right] - \mathbb{E}\left[\log q\left(\alpha\right)\right] &= -m\left(\log\left(2\pi\right) + \left\langle \log \sigma^2\right\rangle - \log m\right) - \left\langle \frac{m}{2\sigma^2}\alpha^T\alpha\right\rangle\\ &\quad - \left\langle\left(-m\log\left(2\pi\right) - \frac{1}{2}\log\left|\Sigma_{q(\alpha)}\right| - \frac{1}{2}\left(\alpha - \mu_{q(\alpha)}\right)^T \Sigma_{q(\alpha)}^{-1}\left(\alpha - \mu_{q(\alpha)}\right)\right)\right\rangle\\ &= -m\left\langle \log\sigma^2\right\rangle + m\log m - \frac{m}{2}\left\langle\frac{1}{\sigma^2}\right\rangle\left(\mathrm{Tr}\left(\Sigma_{q(\alpha)}\right) + \mu_{q(\alpha)}^T\mu_{q(\alpha)}\right) + \frac{1}{2}\log\left|\Sigma_{q(\alpha)}\right| + m\end{aligned}$$

3.2.7 LB: $\mathbb{E}\left[\log \pi\left(\beta\right)\right] - \mathbb{E}\left[\log q\left(\beta\right)\right]$

$$\begin{aligned}\mathbb{E}\left[\log \pi\left(\beta\right)\right] - \mathbb{E}\left[\log q\left(\beta\right)\right] &= -\frac{1}{2}\log\left|2\pi\Sigma_\beta\right| - \left\langle\frac{1}{2}\left(\beta - \mu_\beta\right)^T \Sigma_\beta^{-1}\left(\beta - \mu_\beta\right)\right\rangle\\ &\quad - \left(-\frac{1}{2}\log\left|2\pi\Sigma_{q(\beta)}\right| - \left\langle\frac{1}{2}\left(\beta - \mu_{q(\beta)}\right)^T \Sigma_{q(\beta)}^{-1}\left(\beta - \mu_{q(\beta)}\right)\right\rangle\right)\\ &= -\frac{1}{2}\log\left|2\pi\Sigma_\beta\right| - \frac{1}{2}\left\{\mathrm{Tr}\left(\Sigma_\beta^{-1}\Sigma_{q(\beta)}\right) + \left(\mu_{q(\beta)} - \mu_\beta\right)^T \Sigma_\beta^{-1}\left(\mu_{q(\beta)} - \mu_\beta\right)\right\}\\ &\quad - \left(-\frac{1}{2}\log\left|2\pi\Sigma_{q(\beta)}\right| - \frac{s}{2}\right)\end{aligned}$$

3.2.8 LB: $\mathbb{E}\left[\log \pi\left(\sigma\right)\right] - \mathbb{E}\left[\log q\left(\sigma\right)\right]$

$$\begin{aligned}\mathbb{E}\left[\log \pi\left(\sigma\right)\right] - \mathbb{E}\left[\log q\left(\sigma\right)\right] &= \log\left(2A_\sigma\right) - \log\pi - \left\langle\log\left(A_\sigma^2 + \sigma^2\right)\right\rangle + C_\sigma\frac{\mathcal{H}\left(2m, C_\sigma, A_\sigma^2\right)}{\mathcal{H}\left(2m-2, C_\sigma, A_\sigma^2\right)}\\ &\quad + \log\mathcal{H}\left(2m-2, C_\sigma, A_\sigma^2\right) + 2m\left\langle\log\sigma\right\rangle + \left\langle\log\left(A_\sigma^2 + \sigma^2\right)\right\rangle\end{aligned}$$

3.2.9 LB: $\mathbb{E}\left[\log \pi\left(\lambda\right)\right] - \mathbb{E}\left[\log q\left(\lambda\right)\right]$

$$\begin{aligned}\mathbb{E}\left[\log \pi\left(\lambda\right)\right] - \mathbb{E}\left[\log q\left(\lambda\right)\right] &= -\frac{1}{2}\log\left|\Sigma_\lambda\right| - \frac{1}{2}\left\{\mathrm{Tr}\left(\Sigma_\lambda^{-1}\Sigma_{q(\lambda)}\right) + \left(\mu_{q(\lambda)} - \mu_\lambda\right)^T \Sigma_\lambda^{-1}\left(\mu_{q(\lambda)} - \mu_\lambda\right)\right\}\\ &\quad + \frac{1}{2}\log\left|\Sigma_{q(\lambda)}\right| + \frac{d}{2}\end{aligned}$$

3.2.10 LB

$$\begin{aligned}\mathcal{L} &= -\frac{1}{2}\left\{\mathrm{Tr}\left(\left\langle Z^T Z\right\rangle \Sigma_{q(\alpha)}\right) + \mu_{q(\alpha)}^T \left(\left\langle Z^T Z\right\rangle - \langle Z\rangle^T \langle Z\rangle \mu_{q(\alpha)} + \mathrm{Tr}\left(A^T A \Sigma_{q(\beta)}\right)\right)\right\}\\ &\quad + \sum_{i=1}^{n} \log\left(\left\{\Phi\left(\langle z_i\rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right)\right\}^{y_i} \left\{1 - \Phi\left(\langle z_i\rangle^T \mu_{q(\alpha)} + a_i^T \mu_{q(\beta)}\right)\right\}^{1-y_i}\right)\\ &\quad + m\log m + \frac{1}{2}\log\left|\Sigma_{q(\alpha)}\right| + m - \frac{1}{2}\log\left|\Sigma_\beta\right| - \frac{1}{2}\left\{\mathrm{Tr}\left(\Sigma_\beta^{-1}\Sigma_{q(\beta)}\right) + \left(\mu_{q(\beta)} - \mu_\beta\right)^T \Sigma_\beta^{-1}\left(\mu_{q(\beta)} - \mu_\beta\right)\right\}\\ &\quad + \frac{1}{2}\log\left|\Sigma_{q(\beta)}\right| + \frac{s}{2} + \log\left(2A_\sigma\right) - \log\pi + \log\mathcal{H}\left(2m-2, C_\sigma, A_\sigma^2\right) + \frac{1}{2}\log\left|\Sigma_{q(\lambda)}\right| + \frac{d}{2}\\ &\quad - \frac{1}{2}\log\left|\Sigma_\lambda\right| - \frac{1}{2}\left\{\mathrm{Tr}\left(\Sigma_\lambda^{-1}\Sigma_{q(\lambda)}\right) + \left(\mu_{q(\lambda)} - \mu_\lambda\right)^T \Sigma_\lambda^{-1}\left(\mu_{q(\lambda)} - \mu_\lambda\right)\right\}\end{aligned}$$

3.2.11 $q(\lambda)$

Notice that we have not yet come up with the updating algorithm for λ. This is because of the fact that the full conditional $\lambda|$rest is just the prior density of λ without any dependency on other parameters. Therefore, we instead rely on nonconjugate variational message passing.

$$
\begin{aligned}
q(\lambda) &= \mathcal{N}\left(\mu_{q(\lambda)}, \Sigma_{q(\lambda)}\right) \\
&= \exp\left\{\eta' T(\lambda) - h(\eta)\right\} \\
\mathcal{V}(\eta) &= \frac{\partial^2 h(\eta)}{\partial\eta\partial\eta'} \\
p(y, \lambda) &= \prod_a f_a(y, \lambda) \\
S_a &= \mathbb{E}_q\left[\log f_a(y, \lambda)\right] \\
\nabla_\eta \mathcal{L} &= \mathcal{V}(\eta)^{-1} \sum_{a \in N(\lambda)} \frac{\partial S_a}{\partial \eta} - \eta \\
\eta &\leftarrow \mathcal{V}(\eta)^{-1} \sum_{a \in N(\lambda)} \frac{\partial S_a}{\partial \eta} \\
\Sigma_{q(\lambda)} &\leftarrow -\frac{1}{2}\left[\mathrm{vec}^{-1}\left(\sum_{a \in N(\lambda)} \frac{\partial S_a}{\partial\, \mathrm{vec}\left(\Sigma_{q(\lambda)}\right)}\right)\right]^{-1} \\
\mu_{q(\lambda)} &\leftarrow \mu_{q(\lambda)} + \Sigma_{q(\lambda)} \sum_{a \in N(\lambda)} \frac{\partial S_q}{\partial \mu_{q(\lambda)}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\log p\left(y,\theta\right)\right] = &-m\log\left(\frac{2\pi\sigma^2}{m}\right) - \frac{m}{2}\left\langle\frac{1}{\sigma^2}\right\rangle\left\{\mathrm{Tr}\left(\Sigma_{q(\alpha)}\right) + \mu_{q(\alpha)}'\mu_{q(\alpha)}\right\} - \frac{1}{2}\log\left|2\pi\Sigma_\beta\right| \\
&- \frac{1}{2}\left\{\mathrm{Tr}\left(\Sigma_\beta^{-1}\Sigma_{q(\beta)}\right) + \left(\mu_{q(\beta)} - \mu_\beta\right)'\Sigma_\beta^{-1}\left(\mu_{q(\beta)} - \mu_\beta\right)\right\} - \frac{1}{2}\log\left|2\pi\Sigma_\lambda\right| \\
&- \frac{1}{2}\left\{\mathrm{Tr}\left(\Sigma_\lambda^{-1}\Sigma_{q(\lambda)}\right) + \left(\mu_{q(\lambda)} - \mu_\lambda\right)'\Sigma_\lambda^{-1}\left(\mu_{q(\lambda)} - \mu_\lambda\right)\right\} + \log\left(2A_\sigma\right) - \log\pi - \left\langle\log\left(A_\sigma^2 + \sigma^2\right)\right\rangle \\
&- \frac{n}{2}\log\left(2\pi\right) - \frac{1}{2}\left[\left\langle y^{*\prime}y^*\right\rangle - 2\left(\langle Z\rangle\mu_{q(\alpha)} + A\mu_{q(\beta)}\right)'\mu_{q(y^*)} + \mathrm{Tr}\left\{\left\langle Z'Z\right\rangle\left(\Sigma_{q(\alpha)} + \mu_{q(\alpha)}\mu_{q(\alpha)}'\right)\right\}\right] \\
&- \frac{1}{2}\left[2\mu_{q(\alpha)}'\langle Z\rangle' A\mu_{q(\beta)} + \mathrm{Tr}\left(A'A\Sigma_{q(\beta)}\right) + \mu_{q(\beta)}'A'A\mu_{q(\beta)}\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial\mu_{q(\lambda)}}\mathbb{E}\left[\log p\left(y,\theta\right)\right] = &-\frac{1}{2}\frac{\partial}{\partial\mu_{q(\lambda)}}\left(\mu_{q(\lambda)} - \mu_\lambda\right)'\Sigma_\lambda^{-1}\left(\mu_{q(\lambda)} - \mu_\lambda\right) && \cdots(1) \\
&+ \frac{\partial}{\partial\mu_{q(\lambda)}}\mu_{q(\alpha)}'\langle Z\rangle'\mu_{q(y^*)} && \cdots(2) \\
&- \frac{1}{2}\frac{\partial}{\partial\mu_{q(\lambda)}}\mathrm{Tr}\left\{\left\langle Z'Z\right\rangle\left(\Sigma_{q(\alpha)} + \mu_{q(\alpha)}\mu_{q(\alpha)}'\right)\right\} && \cdots(3) \\
&- \frac{\partial}{\partial\mu_{q(\lambda)}}\mu_{q(\alpha)}'\langle Z\rangle' A\mu_{q(\beta)} && \cdots(4)
\end{aligned}
$$

$$
\begin{aligned}
(1) &= -\frac{1}{2}\frac{\partial}{\partial\mu_{q(\lambda)}}\left(\mu_{q(\lambda)}'\Sigma_\lambda^{-1}\mu_{q(\lambda)} - 2\mu_{q(\lambda)}'\Sigma_\lambda^{-1}\mu_\lambda + \mu_\lambda'\Sigma_\lambda^{-1}\mu_\lambda\right) \\
&= -\frac{1}{2}\left(2\Sigma_\lambda^{-1}\mu_{q(\lambda)} - 2\Sigma_\lambda^{-1}\mu_\lambda\right) \\
&= \Sigma_\lambda^{-1}\left(\mu_\lambda - \mu_{q(\lambda)}\right)
\end{aligned}
$$

$$
(2)+(4) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\mu_i^{q(y^*)} - \mu_i^{A\mu_{q(\beta)}}\right)\exp\left(-\frac{1}{2}t_{ij}'\Sigma_{q(\lambda)}t_{ij}\right)\left\{\mu_{j+m}^{q(\alpha)}\cos\left(t_{ij}'\mu_{q(\lambda)}\right) - \mu_j^{q(\alpha)}\sin\left(t_{ij}'\mu_{q(\lambda)}\right)\right\}t_{ij}
$$

$$
\begin{aligned}
(3) = -\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{r=1}^{m}&\left[\nu_{ijr}^{-}\left\{2B_{jr}\cos\left({t_{ijr}^{-}}'\mu_{q(\lambda)}\right) - \left(A_{jr} + D_{jr}\right)\sin\left({t_{ijr}^{-}}'\mu_{q(\lambda)}\right)\right\}t_{ijr}^{-}\right. \\
&\left.+ \nu_{ijr}^{+}\left\{2B_{jr}\cos\left({t_{ijr}^{+}}'\mu_{q(\lambda)}\right) + \left(D_{jr} - A_{jr}\right)\sin\left({t_{ijr}^{+}}'\mu_{q(\lambda)}\right)\right\}t_{ijr}^{+}\right]
\end{aligned}
$$

$$
\mu_{q(\lambda)} \leftarrow \mu_{q(\lambda)} + \Sigma_{q(\lambda)}\left((1)+(2)+(3)+(4)\right)
$$

$$
\frac{\partial}{\partial\Sigma_{q(\lambda)}}\mathbb{E}\left[\log p\left(y,\theta\right)\right] = -\frac{1}{2}\left(\Sigma_\lambda^{-1} + F_1 + F_2\right)
$$

$$
F_1 = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\mu_i^{q(y^*)} - \mu_i^{A\mu_{q(\beta)}}\right)\exp\left(-\frac{1}{2}t_{ij}'\Sigma_{q(\lambda)}t_{ij}\right)\left\{\mu_j^{q(\alpha)}\cos\left(t_{ij}'\mu_{q(\lambda)}\right) + \mu_{j+m}^{q(\alpha)}\sin\left(t_{ij}'\mu_{q(\lambda)}\right)\right\}t_{ij}t_{ij}'
$$

$$
\begin{aligned}
F_2 = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{r=1}^{m}&\left[\nu_{ijr}^{-}\left\{\left(A_{jr} + D_{jr}\right)\cos\left({t_{ijr}^{-}}'\mu_{q(\lambda)}\right) + 2B_{jr}\sin\left({t_{ijr}^{-}}'\mu_{q(\lambda)}\right)\right\}t_{ijr}^{-}{t_{ijr}^{-}}'\right. \\
&\left.+ \nu_{ijr}^{+}\left\{\left(A_{jr} - D_{jr}\right)\cos\left({t_{ijr}^{+}}'\mu_{q(\lambda)}\right) + 2B_{jr}\sin\left({t_{ijr}^{+}}'\mu_{q(\lambda)}\right)\right\}t_{ijr}^{+}{t_{ijr}^{+}}'\right]
\end{aligned}
$$

$$
\Sigma_{q(\lambda)} \leftarrow \left\{\Sigma_\lambda^{-1} + F_1 + F_2\right\}^{-1}
$$

4 Sparse GP Poisson mixed model

4.1 Model specifications

$$
\begin{aligned}
y &= g^{-1}(\eta) + \epsilon, \quad \epsilon \sim \mathcal{N}\left(0, \gamma^2 I_n\right) \\
g(\mathbb{E}[y]) &= \eta \\
\eta &= f(x) + A\beta \\
g^{-1}(x) &= e^x \\
f &\sim \mathcal{GP}(m(\cdot), \kappa(\cdot,\cdot))
\end{aligned}
$$

The canonical link function for Poisson regression model is $\log x$. Therefore,

$$
\begin{aligned}
y &\sim \text{Poisson}(e^\eta) \\
p(y|\alpha, \beta, X) &= \exp\left\{y'(Z\alpha + A\beta) - 1_n' \exp\{Z\alpha + A\beta\} - 1_n' \log(y!)\right\} \\
\alpha|\sigma &\sim \mathcal{N}\left(0, \frac{\sigma^2}{m} I_{2m}\right) \\
\beta &\sim \mathcal{N}(\mu_\beta, \Sigma_\beta) \\
\lambda &\sim \mathcal{N}(\mu_\lambda, \Sigma_\lambda) \\
\sigma &\sim \text{half-Cauchy}(A_\sigma) \\
\theta &= (\alpha, \beta, \lambda, \sigma)
\end{aligned}
$$

4.2 $\mathbb{E}[\log p(y|\theta)]$

$$
\begin{aligned}
\log p(y|\theta) &= y'(Z\alpha + A\beta) - 1_n' \exp\{Z\alpha + A\beta\} - 1_n' \log(y!) \\
\mathbb{E}\left[1_n' \exp\{Z\alpha + A\beta\}\right] &= \sum_{i=1}^{n} \mathbb{E}\left\{\exp\left[\sum_{r=1}^{m}\left(a_r \cos\left\{(s_r \odot x_i)'\lambda\right\} + b_r \sin\left\{(s_r \odot x_i)'\lambda\right\}\right)\right]\right\} \mathbb{E}\left\{\exp\left(a_i'\beta\right)\right\} \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\exp\left\{z_i'\mu_{q(\alpha)} + \frac{1}{2} z_i'\Sigma_{q(\alpha)} z_i\right\}\right] \exp\left(a_i'\mu_{q(\beta)} + \frac{1}{2} a_i'\Sigma_{q(\beta)} a_i\right)
\end{aligned}
$$