Incorporating Random effect into GP Sparse Approximation

Daeyoung Lim*
Department of Statistics
Korea University

January 29, 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

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

*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^T y - 2\left(\mathbb{E}\left[Z\right]m_\alpha + Am_\beta\right)^T y + \mathrm{Tr}\left(\mathbb{E}\left[Z^T Z\right]S_\alpha\right) + m_\alpha^T\mathbb{E}\left[Z^T Z\right]m_\alpha \\
&\quad + 2m_\beta^T A^T\mathbb{E}\left[Z\right]m_\alpha + \mathrm{Tr}\left(A^T A S_\beta\right) + m_\beta^T A^T A m_\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}
$$

$$\mathcal{L} = (1) + (2) + (3) + (4) + (5) + (6) - (1) - (2) - (3) - (4) - (5)$$

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.

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

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

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

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

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

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 $\xi = \left(\xi_1, \ldots, \xi_n\right)$.

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