Supplemental Material of "A Multitask Point Process Predictive Model"

Wenzhao Lian[1] WL89@DUKE.EDU
Ricardo Henao[1] R.HENAO@DUKE.EDU
Vinayak Rao[2] VARAO@PURDUE.EDU
Joseph Lucas[3] JOE@STAT.DUKE.EDU
Lawrence Carin[1] LCARIN@DUKE.EDU
[1]Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA
[2]Department of Statistics, Purdue University, West Lafayette, IN 47907, USA
[3]Center for Predictive Medicine, Duke Clinical Research Institute, Durham, NC 27708, USA

1. Key quantities in variational inference

Variational E-step:

$$\frac{d\mathcal{F}_1}{d\boldsymbol{\mu}^u} = \sum_{u=1}^{U}\sum_{i=1}^{N^u}\Big\{-\Delta_i^u\Big[2b_i^u \boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM,i\cdot}^{u}{}^{\top}\Big] + \mathbb{I}(t_i^u\in\mathcal{T}^u)\left[\tilde{G}'\left(-\frac{(b_i^u)^2}{2B_{ii}^u}\right)\frac{b_i^u}{B_{ii}^u}\boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM,i\cdot}^{u}{}^{\top}\right]\Big\}\,,$$

$$\frac{d\mathcal{F}_1}{d\boldsymbol{\Sigma}^u} = \sum_{u=1}^{U}\sum_{i=1}^{N^u}\Big\{-\Delta_i^u[\boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM,i\cdot}^{u}{}^{\top}\boldsymbol{K}_{NM,i\cdot}^{u}\boldsymbol{K}_{MM}^{-1}] + \mathbb{I}(t_i^u\in\mathcal{T}^u)\left[\left(-\tilde{G}'\left(-\frac{(b_i^u)^2}{2B_{ii}^u}\right)\frac{(b_i^u)^2}{2(B_{ii}^u)^2}+\frac{1}{B_{ii}^u}\right)\boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM,i\cdot}^{u}{}^{\top}\boldsymbol{K}_{NM,i\cdot}^{u}\boldsymbol{K}_{MM}^{-1}\right]\Big\}\,,$$

$$\begin{aligned}
\frac{d\mathcal{F}_2}{d\boldsymbol{\mu}^u} &= -\boldsymbol{K}_{MM}^{-1}(\boldsymbol{\mu}^u-\boldsymbol{\mu}_M^0)\,,\\
\frac{d\mathcal{F}_2}{d\boldsymbol{\Sigma}^u} &= -\frac{1}{2}\boldsymbol{K}_{MM}^{-1}+\frac{1}{2}\boldsymbol{\Sigma}^{u\,-1}\,,\\
\frac{d\mathcal{F}_3}{d\boldsymbol{\mu}^u} &= 0\,,\\
\frac{d\mathcal{F}_3}{d\boldsymbol{\Sigma}^u} &= 0\,.
\end{aligned}$$

Variational M-step:

Updates for $\boldsymbol{\mu}_M^0$ can be derived in closed form as follows,

$$\hat{\boldsymbol{\mu}}_M^0 = \frac{1}{\xi+U}\left(\xi\boldsymbol{g}+\sum_{u=1}^{U}\boldsymbol{\mu}^u\right)\,.$$

Gradient methods are needed for updating other parameters (denoted as θ_k), including pseudo input positions and GP hyper-parameters, with key quantities summarized below:

$$\frac{d\mathcal{F}_1}{d\theta_k} = \sum_{u=1}^{U}\sum_{i=1}^{N^u}\Big\{\mathbb{I}(t_i^u\in\mathcal{T}^u)\Big[-\tilde{G}'\left(-\frac{(b_i^u)^2}{2B_{ii}^u}\right)\left(-\frac{1}{2(B_{ii}^u)^2}\right)\times\left\{2B_{ii}^u b_i^u(\boldsymbol{\mu}^u-\boldsymbol{g})^T\frac{d\boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM,i\cdot}^{u}{}^{\top}}{d\theta_k}-(b_i^u)^2\frac{dB_{ii}^u}{d\theta_k}\right\} + \frac{1}{B_{ii}^u}\frac{dB_{ii}^u}{d\theta_k}\Big]-\Delta_i^u\left[2b_i^u(\boldsymbol{\mu}^u-\boldsymbol{g})^T\frac{d\boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM,i\cdot}^{u}{}^{\top}}{d\theta_k}+\frac{dB_{ii}^u}{d\theta_k}\right]\Big\}\,,$$

$$\frac{dB_{ii}^u}{d\theta_k} = \frac{d\boldsymbol{K}_{NN,ii}}{d\theta_k}-\frac{d\boldsymbol{K}_{NM,i\cdot}^{u}\boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM,i\cdot}^{u}{}^{\top}}{d\theta_k} + 2(\boldsymbol{\Sigma}^u\boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM,i\cdot}^{u}{}^{\top})^{\top}\frac{d\boldsymbol{K}_{MM}^{-1}\boldsymbol{K}_{NM,i\cdot}^{u}{}^{\top}}{d\theta_k}\,,$$

$$\frac{d\mathcal{F}_2}{d\theta^k} = -\frac{U}{2}tr\left(\boldsymbol{K}_{MM}^{-1}\frac{d\boldsymbol{K}_{MM}}{d\theta_k}\right)+\frac{1}{2}tr\Big(\boldsymbol{K}_{MM}^{-1}\times\sum_{u=1}^{U}\left(\boldsymbol{\mu}^u\boldsymbol{\mu}^{u\top}+\boldsymbol{\Sigma}^u+\boldsymbol{\mu}_M^0\boldsymbol{\mu}_M^{0\,\top}-2\boldsymbol{\mu}^u\boldsymbol{\mu}_M^{0\,\top}\right)\boldsymbol{K}_{MM}^{-1}\frac{d\boldsymbol{K}_{MM}}{d\theta_k}\Big)\,,$$

$$\frac{d\mathcal{F}_3}{d\theta^k} = -\frac{1}{2}tr\left(\boldsymbol{K}_{MM}^{-1}\frac{d\boldsymbol{K}_{MM}}{d\theta_k}\right) + \frac{1}{2}tr\left(\xi\boldsymbol{K}_{MM}^{-1}(\boldsymbol{\mu}_M^0-\boldsymbol{g})(\boldsymbol{\mu}_M^0-\boldsymbol{g})^{\top}\boldsymbol{K}_{MM}^{-1}\frac{d\boldsymbol{K}_{MM}}{d\theta_k}\right)\,.$$

2. Hierarchical Gaussian process construction

Using the construction in (2) and (3), we can integrate over $\boldsymbol{\mu}_N^0$ to obtain the marginal prior distribution for $\boldsymbol{f}_N^u$ as

$$\boldsymbol{f}_N^u \sim \mathcal{N}\left(\boldsymbol{g},\left(1+\frac{1}{\xi}\right)\boldsymbol{K}_{NN}\right)\,. \quad \text{(S1)}$$

Using the construction in (9)-(11), we can integrate out $\boldsymbol{f}^u_M$ using (10) and (11), obtaining,

$$\boldsymbol{f}^u_N | \boldsymbol{\mu}^u_M \sim \mathcal{N}\Bigg(\boldsymbol{g} + \boldsymbol{K}^u_{NM}\boldsymbol{K}^{-1}_{MM}(\boldsymbol{\mu}^0_M - \boldsymbol{g}), \boldsymbol{K}^u_{NN} + \frac{1}{\xi}\left(\boldsymbol{K}^u_{NN} - \boldsymbol{K}^u_{NM}\boldsymbol{K}^{-1}_{MM}{\boldsymbol{K}^u_{NM}}^\top\right)\Bigg) .$$

Then, integrating out $\boldsymbol{\mu}^0_M$ using (9), we can get the same marginal prior as in (S1),

$$\boldsymbol{f}^u_N \sim \mathcal{N}\left(\boldsymbol{g}, \left(1 + \frac{1}{\xi}\right)\boldsymbol{K}^u_{NN}\right) . \tag{S2}$$

The only difference between (S1) and (S2) is that we need a realization of the function at possible features from all tasks in (S1), while only the features from task u need to be specified in (S2).

3. Confluent hypergeometric functions

When $|x|$ is small, for example $|x| \leq 30$, we use the power series to compute the confluent hypergeometric function:

$${}_1F_1(a, b, x) = \sum_{k=0}^{\infty} \frac{(a)_k x^k}{(b)_k k!} ,$$

where $(a)_0 = 1$, $(a)_k = \prod_{j=0}^{k-1}(a + j)$. In practice, the summation can be terminated at a sufficiently large number to achieve a given error tolerance level.

When $|x|$ is large, for example $|x| > 30$, we use the following computation (Thompson, 1997):

$${}_1F_1(a, b, x) = \frac{\Gamma(b)(-x)^{-a}}{\Gamma(b-1)} \sum_{k=0}^{\infty} \frac{(a)_k(a + a - b)_k}{k!(-x)^k} .$$

Having ${}_1F_1(a, b, x)$, the gradient $\tilde{G}(x) = G(0, \frac{1}{2}, z)$ can be numerically computed, where $G(a, b, x) = \frac{\partial_1 F_1(a,b,x)}{\partial a}$ (Ancarani & Gasaneo, 2008).

Finally, the expectation needed to compute $\mathcal{F}_1$ during variational E-step can be estimated as (Lloyd et al., 2014):

$$\mathbb{E}[\log(f^u_{N,i})^2] = -\tilde{G}(-\frac{{b^u_i}^2}{2B^u_{ii}}) + \log(\frac{B^u_{ii}}{2}) - const .$$

References

Ancarani, L. U. and Gasaneo, G. Derivatives of any order of the confluent hypergeometric function. *Journal of Mathematical Physics*, 49, 2008.

Lloyd, Chris, Gunter, Tom, Osborne, Michael A., and Roberts, Stephen J. Variational inference for Gaussian process modulated Poisson processes. *arXiv:1411.0254*, 2014.

Thompson, William J. *Atlas for Computing Mathematical Functions: An Illustrated Guide for Practitioners with Programs in FORTRAN and Mathematica with Cdrom.* John Wiley & Sons, Inc., 1997.