SDS 383D: Final Project Notes

May 15, 2017

Professor Scott

Spencer Woody

Let y_{nrt} be the read count of gene $n \in \{1, 2, \dots N\}$ in replicate $r \in \{1, 2, \dots, R_n\}$ at continuous time $t \in \mathbf{t}_{nr}$, where $\mathbf{t}_{nr}$ is a vector of length T_{nr}. We use a negative-binomial regression model,

$$(y_{nrt}|\psi_{nr}(t)) \sim \text{NB}\left(\alpha_n, \frac{\exp\left[\psi_{nr}(t)\right]}{1 + \exp\left[\psi_{nr}(t)\right]}\right),$$

with a hierarchical Gaussian process prior on $\psi_{nr}(t)$,

$$\begin{aligned} \psi_{nr}(t) &\sim \text{GP}(g_n(t), k_\psi(t, t')) \\ g_n(t) &\sim \text{GP}(0, k_g(t, t')), \end{aligned}$$

for some covariance functions $k_\psi(t, t')$ and $k_g(t, t')$, which respectively depend on hyperparameters θ_ψ and θ_g (which might both be vectors). Notice that the expectation of y_{nrt} is

$$\mathbb{E}(y_{nrt}|\psi_{nr}(t)) = \alpha_n \cdot \exp\left[\psi_{nr}(t)\right]$$

Then introduce a Polya-Gamma latent variable,

$$\omega_{nrt} \sim \text{PG}(y_{nrt} + \alpha_n, 0),$$

whose expectation is

$$\mathbb{E}_{\omega_{nrt}}\left[\exp\left(-\omega_{nrt}\left[\psi_{nr}(t)\right]^2/2\right)\right] = \cosh^{-(y_{nrt}+\alpha_n)}(\psi_{nr}(t)/2).$$

The joint likelihood may be written as

$$\begin{aligned} p(y_{nrt}|\psi_{nr}(t), \omega_{nrt}) &\propto \frac{\left(\exp\left[\psi_{nr}(t)\right]\right)^{y_{nrt}}}{\left(1 + \exp\left[\psi_{nr}(t)\right]\right)^{\alpha_n + y_{nrt}}} \\ &= \frac{2^{-(y_{nrt}+\alpha_n)} \cdot \exp\left(\frac{y_{nrt}-\alpha_n}{2}\psi_{nr}(t)\right)}{\cosh^{y_{nrt}+\alpha_n}(\psi_{nr}(t)/2)} \\ &\propto \exp\left(\frac{y_{nrt} - \alpha_n}{2}\psi_{nr}(t)\right)\mathbb{E}_{\omega_{nrt}}\left[\exp\left(-\omega_{nrt}\left[\psi_{nr}(t)\right]^2/2\right)\right]. \end{aligned}$$

Suppose we have observations from times $\mathbf{t}_{nr}$, so the data vector is $\mathbf{y}_{nr} = \{y_{nrt}\}_{t \in \mathbf{t}_{nr}}$ which is associated with draws from the GP $\boldsymbol{\psi}_{nr} = \{\psi_{nr}(t)\}_{t \in \mathbf{t}_{nr}}$. Then there is the latent variable vector $\boldsymbol{\omega}_{nr} = \{\omega_{nrt}\}_{t \in \mathbf{t}_{nr}}$, and also define the diagonal matrix $\boldsymbol{\Omega}_{nr} = \text{diag}\left(\boldsymbol{\omega}_{nr}\right)$. Finally define the vector $\mathbf{g}_n$ be a vector of draws from the GP $g_n(t, t')$ at times $\mathbf{t}_{nr}$ and the matrix $\mathbf{K}_\psi(\mathbf{t}_{nr}, \mathbf{t}_{nr'})$ such that it's (i, j) element is $k_\psi(\mathbf{t}_{nr}[i], \mathbf{t}_{nr'}[j])$ and $\mathbf{K}_g(\mathbf{t}_{nr}, \mathbf{t}_{nr'})$ is defined similarly. Now we can find the marginal prior of distribution of $\boldsymbol{\psi}_{nr}$ with

$$\begin{aligned} p(\boldsymbol{\psi}_{nr}|\mathbf{g}_n, \theta_\psi) &\sim \mathcal{N}\left(\mathbf{f}_n, \mathbf{K}_\psi(\mathbf{t}_{nr}, \mathbf{t}_{nr})\right) \\ p(\mathbf{f}_n|\theta_g) &\sim \mathcal{N}\left(\mathbf{0}, \mathbf{K}_g(\mathbf{t}_{nr}, \mathbf{t}_{nr})\right) \\ \Rightarrow p(\boldsymbol{\psi}_{nr}|\theta_\psi, \theta_g) &\sim \mathcal{N}\left(\mathbf{0}, \mathbf{K}_\psi(\mathbf{t}_{nr}, \mathbf{t}_{nr}) + \mathbf{K}_g(\mathbf{t}_{nr}, \mathbf{t}_{nr})\right). \end{aligned}$$

Define the vector $\boldsymbol{\theta} = (\theta_\psi, \theta_g)^T$ to contain the hyperparameters of both covariance functions, $k_g(\cdot, \cdot)$ and $k_\psi(\cdot, \cdot)$. We can now write the prior of the concatenated vector $\boldsymbol{\psi}_n = \{\boldsymbol{\psi}_{nr}\}_{r=1}^{R_n}$ as

$$p(\boldsymbol{\psi}_n|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{0}, \mathbf{K}_n)$$

where the matrix $\mathbf{K}_n$ is a $R_n \times R_n$ arrangement of matrices, each of which has dimension $T_{nr} \times T_{nr'}$ and is

$$\mathbf{K}_n[r, r'] = \text{cov}(\boldsymbol{\psi}_{nr}, \boldsymbol{\psi}_{nr'}) = \begin{cases} \mathbf{K}_g(\mathbf{t}_{nr}, \mathbf{t}_{nr}) + \mathbf{K}_\psi(\mathbf{t}_{nr}, \mathbf{t}_{nr}) & \text{if } r = r' \\ \mathbf{K}_g(\mathbf{t}_{nr}, \mathbf{t}_{nr'}) & \text{otherwise} \end{cases}$$

The conditional posterior of $\boldsymbol{\psi}_n$, given the values of $\boldsymbol{\omega}_n$ and the data vector $\mathbf{y}_n = \{\mathbf{y}_{nr}\}_{r=1}^{R_n}$ is

$$
\begin{aligned}
p(\boldsymbol{\psi}_n|\mathbf{y}_n, \boldsymbol{\omega}_n, \boldsymbol{\theta}), &\propto p(\boldsymbol{\psi}_n|\boldsymbol{\theta}) \prod_{r=1}^{R_n} \prod_{t \in \mathbf{t}_{nr}} p(y_{nrt}|\psi_{nr}(t), \omega_{nrt}) \\
&\propto p(\boldsymbol{\psi}_n|\boldsymbol{\theta}) \prod_{r=1}^{R_n} \prod_{t \in \mathbf{t}_{nr}} \exp\left[-\frac{\omega_{nrt}}{2}\left(\psi_{nr}(t) - \frac{y_{nrt}-\alpha_n}{2\omega_{nrt}}\right)^2\right], \text{ define } z_{nrt} = \frac{y_{nrt}-\alpha_n}{2}, \\
&\propto p(\boldsymbol{\psi}_n|\boldsymbol{\theta}) \cdot \exp\left[-\frac{1}{2}\left(\boldsymbol{\psi}_n - \boldsymbol{\Omega}_n^{-1}\mathbf{z}_n\right)^T \boldsymbol{\Omega}_n \left(\boldsymbol{\psi}_n - \boldsymbol{\Omega}_n^{-1}\mathbf{z}_n\right)\right] \\
&\propto \mathcal{N}\left(\boldsymbol{\psi}_n|\boldsymbol{\Sigma}_n\mathbf{z}_n, \boldsymbol{\Sigma}_n\right), \text{ with } \boldsymbol{\Sigma}_n = \left(\mathbf{K}_n^{-1} + \boldsymbol{\Omega}_n\right)^{-1}.
\end{aligned}
$$

The conditional posterior of each ω_{nrt} is

$$
\begin{aligned}
p(\omega_{nrt}|y_{nrt}, \psi_{nr}(t), \mathbf{t}_{nr}) &\propto \left[\exp\left(-\omega_{nrt}\left[\psi_{nr}(t)\right]^2/2\right)\right] \cdot \mathrm{PG}(\omega_{nrt}|y_{nrt}+\alpha_n, 0) \\
&\propto \mathrm{PG}(\omega_{nrt}|y_{nrt}+\alpha_n, \psi_{nr}(t)).
\end{aligned}
$$

Prediction

Now suppose that we want to infer the underlying time series of both the gene-level function and each replicate-level function, i.e. $\mathbf{g}_n^\star$ which is $g_n(t)$ at times $\mathbf{t}_n^\star$ and $\boldsymbol{\psi}_{nr}^\star$ which is $\psi_{nr}(t)$ at times $\mathbf{t}_{nr}^\star$. The respective joint distributions between these vectors and $\boldsymbol{\psi}_n$ are

$$
\begin{aligned}
\begin{bmatrix} \boldsymbol{\psi}_n \\ \mathbf{g}_n^\star \end{bmatrix} &\sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_n & \mathbf{K}_{n\star}^T \\ \mathbf{K}_{n\star} & \mathbf{K}_{n\star\star} \end{bmatrix}\right) \\
\begin{bmatrix} \boldsymbol{\psi}_n \\ \boldsymbol{\psi}_{nr}^\star \end{bmatrix} &\sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_n & \mathbf{K}_{nr\star}^T \\ \mathbf{K}_{nr\star} & \mathbf{K}_{nr\star\star} \end{bmatrix}\right)
\end{aligned}
$$

with $\mathbf{K}_{n\star}$ and $\mathbf{K}_{nr\star}$ are defined element-wise such that

$$
\begin{aligned}
\mathbf{K}_{n\star}[i,j] &= \mathrm{cov}\left(\mathbf{g}_n^\star[i], \boldsymbol{\psi}_n[j]\right) = k_g(\mathbf{t}_n^\star[i], \mathbf{t}_n[j]) \\
\mathbf{K}_{nr\star}[i,j] &= \mathrm{cov}\left(\boldsymbol{\psi}_{nr}^\star[i], \boldsymbol{\psi}_n[j] \in \boldsymbol{\psi}_{nr'}\right) = \begin{cases} k_g(\mathbf{t}_{nr}^\star[i], \mathbf{t}_n[j]) + k_\psi(\mathbf{t}_{nr}^\star[i], \mathbf{t}_n[j]) & \text{if } r = r' \\ k_g(\mathbf{t}_{ni}^\star[i], \mathbf{t}_n[j]) & \text{otherwise} \end{cases},
\end{aligned}
$$

and the matrices $\mathbf{K}_{n\star\star}$ and $\mathbf{K}_{nr\star\star}$ are

$$
\begin{aligned}
\mathbf{K}_{n\star\star} &= \mathbf{K}_g(\mathbf{t}_n^\star, \mathbf{t}_n^\star) \\
\mathbf{K}_{n\star\star} &= \mathbf{K}_g(\mathbf{t}_n^\star, \mathbf{t}_n^\star) + \mathbf{K}_\psi(\mathbf{t}_n^\star, \mathbf{t}_n^\star).
\end{aligned}
$$

The conditional distribution of $\mathbf{g}_n^\star$ given $\boldsymbol{\psi}_n$ is

$$
(\mathbf{g}_n^\star|\boldsymbol{\psi}_n, \boldsymbol{\theta}) \sim \mathcal{N}\left(\mathbf{K}_{n\star}\mathbf{K}_n^{-1}\boldsymbol{\psi}_n,\ \mathbf{K}_{n\star\star} - \mathbf{K}_{n\star}\mathbf{K}_n^{-1}\mathbf{K}_{n\star}^T\right).
$$

Given the fact that the marginal posterior of $\boldsymbol{\psi}_n$ is

$$
(\boldsymbol{\psi}_n|\mathbf{y}_n, \boldsymbol{\omega}_n, \boldsymbol{\theta}) \sim \mathcal{N}\left(\boldsymbol{\Sigma}_n\mathbf{z}_n, \boldsymbol{\Sigma}_n\right),
$$

and using Lemma 0.1 we can write the marginal posterior of $\mathbf{g}_n^\star$ as

$$
(\mathbf{g}_n^\star|\mathbf{y}_n, \boldsymbol{\omega}_n, \boldsymbol{\theta}) \sim \mathcal{N}\left(\mathbf{K}_{n\star}\mathbf{K}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{z}_n,\ \mathbf{K}_{n\star}\mathbf{K}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{K}_n^{-1}\mathbf{K}_{n\star}^T + \mathbf{K}_{n\star\star} - \mathbf{K}_{n\star}\mathbf{K}_n^{-1}\mathbf{K}_{n\star}^T\right).
$$

Similarly, the marginal posterior of $\boldsymbol{\psi}_{nr}^\star$ is

$$
(\boldsymbol{\psi}_{nr}^\star|\mathbf{y}_n, \boldsymbol{\omega}_n, \boldsymbol{\theta}) \sim \mathcal{N}\left(\mathbf{K}_{nr\star}\mathbf{K}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{z}_n,\ \mathbf{K}_{nr\star}\mathbf{K}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{K}_n^{-1}\mathbf{K}_{nr\star}^T + \mathbf{K}_{nr\star\star} - \mathbf{K}_{nr\star}\mathbf{K}_n^{-1}\mathbf{K}_{nr\star}^T\right).
$$

Covariance matrix function

We choose the Matérn(5/2) covariance function,

$$k(t,t') = \tau_1^2 \exp\left\{1 + \sqrt{5}\cdot\frac{d}{b} + \frac{5}{3}\cdot\frac{d^2}{b^2}\right\}\exp\left\{-\sqrt{5}\cdot\frac{d}{b}\right\},\ d = \|t-t'\|, \tag{1}$$

so the parameters are $\theta = (b, \tau_1^2, \tau_2^2)^T$, and we refer to b as the *relative length* parameter, τ_1^2 is the *amplitude* parameter, and τ_1^2 is the *nugget* parameter.

Lemma 0.1. *Define the random vectors x and γ such that the conditional distribution of x given γ and the marginal distribution of γ are, respectively,*

$$\begin{aligned}(x|\gamma) &\sim \mathcal{N}_n(A\gamma, \Sigma)\\ \gamma &\sim \mathcal{N}_p(m, V)\end{aligned}$$

where A is a $n \times p$ matrix. Then the joint distribution of (x, γ) is

$$\begin{bmatrix} x \\ \gamma \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} Am \\ m \end{bmatrix}, \begin{bmatrix} AVA^T + \Sigma & AV \\ VA^T & \Sigma \end{bmatrix}\right). \tag{2}$$

Proof. Equivalently, x may be written as

$$x = A\gamma + \epsilon,\ \ \epsilon \sim \mathcal{N}_n(0, \Sigma)$$

and then $(x, \gamma)^T$ is multivariate normal because it can be written as an affine transformation of univariate normal variables,

$$\begin{bmatrix} x \\ \gamma \end{bmatrix} = \begin{bmatrix} A \\ \mathcal{I}_p \end{bmatrix}\gamma + \begin{bmatrix} \mathcal{I}_n \\ \mathcal{O}_{p\times n} \end{bmatrix}\epsilon.$$

From this, the mean and covariance matrix in (2) may be derived from properties of the multivatiate normal distribution. □