Web-based Supplementary Materials for Gaussian Process-Based Bayesian Nonparametric Inference of Population Size Trajectories from Gene Genealogies

by Julia A. Palacios and Vladimir N. Minin

1. Prior Sensitivity

In all our examples, we placed a Gamma prior on the precision parameter θ with parameters $\alpha = 0.001$ and $\beta = 0.001$. This precision parameter, unknown to us *a priori*, controls the smoothness of the GP prior. We investigate the sensitivity of our results to the Gamma prior specification using the Egyptian HCV data. In the first plot of Figure 1, we show the prior and posterior distributions of θ under our default prior. The difference in densities suggests that prior choices do not have an impact on the posterior distribution. Since the mean of a Gamma distributed random variable is α/β, we investigate the sensitivity by fixing $\beta = .001$ and setting the value of α to 0.001, 0.002, 0.005, 0.01 and 0.1, corresponding to prior means 1, 2, 5, 10 and 100 and variances 1000, 2000, 5000, 10000 and 100000, and by trying two extremes: $\alpha = 1$, $\beta = .0001$ and $\alpha = .001$, $\beta = 1$, to examine the posterior distribution of θ under these priors. The posterior sample boxplots displayed in Figure 6 demonstrate that our results are fairly robust to different choices of α.

2. Sensitivity to the Order of the Gaussian Process

We evaluate our GP-based method for three different Gaussian Process priors for the Egyptian HCV genealogy. In Figure 2, we show the recovered trajectories for Brownian Motion (BM), Ornstein-Uhlenbeck (OU) and approximated Integrated Brownian motion (IBM) (Lindgren and Rue, 2008). The common characteristic of these three priors is the sparsity of their precision matrices (inverse covariance matrix), allowing for computational tractability. Figure 2 shows that the order of the process does make a difference, but only in regions with large posterior uncertainty, where prior influence is more pronounced.

Appendix A: Coalescent Simulation Algorithms

Proposition 1: Algorithm 1 generates $t_n < t_{n-1} < \cdots < t_1$, such that

$$P(t_{k-1} > t | t_k) = \exp\left[-\int_{t_k}^{t} \frac{C_k dx}{N_e(x)}\right], \tag{A.1}$$

where $N_e(t)$ is known deterministically.

Proof. Let $T_i = t_k + E_1 + \ldots + E_i$, where $\{E_i\}_{i=1}^{\infty}$ are iid exponential $Exp(C_k\lambda)$ random numbers. Given t_k, Algorithm 1 generates and accumulates iid exponential random numbers until T_i is accepted with probability $1/\lambda N_e(T_i)$, in which case, T_i is labeled t_{k-1}. Let $N(t_k, t] = \#\{i \geq 1 : t_k < T_i \leq t\}$ denote the number of iid exponential random numbers generated

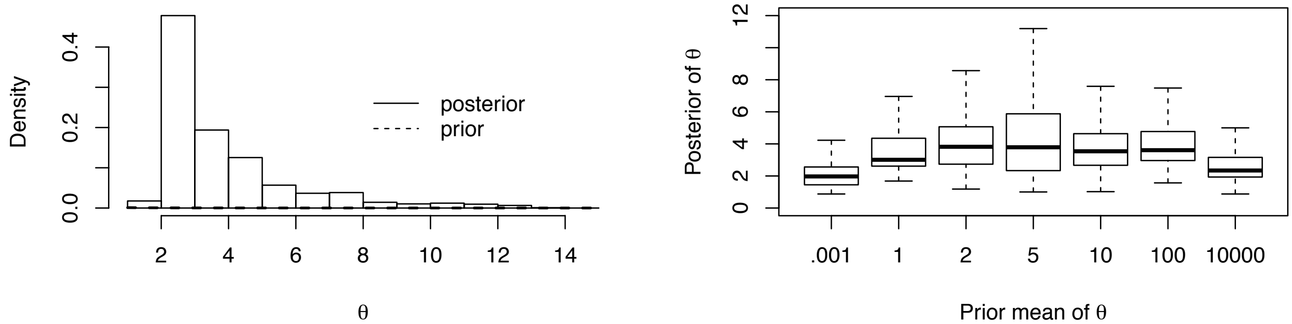

Figure 1. Prior sensitivity on the GP precision parameter. Left plot shows the prior and posterior distributions represented by dashed line and vertical bars respectively. Right plot shows the boxplots of the posterior distributions of the precision parameter when the prior distributions differ in mean and variance of the precision parameter θ. These plots are based on the Egyptian HCV data.

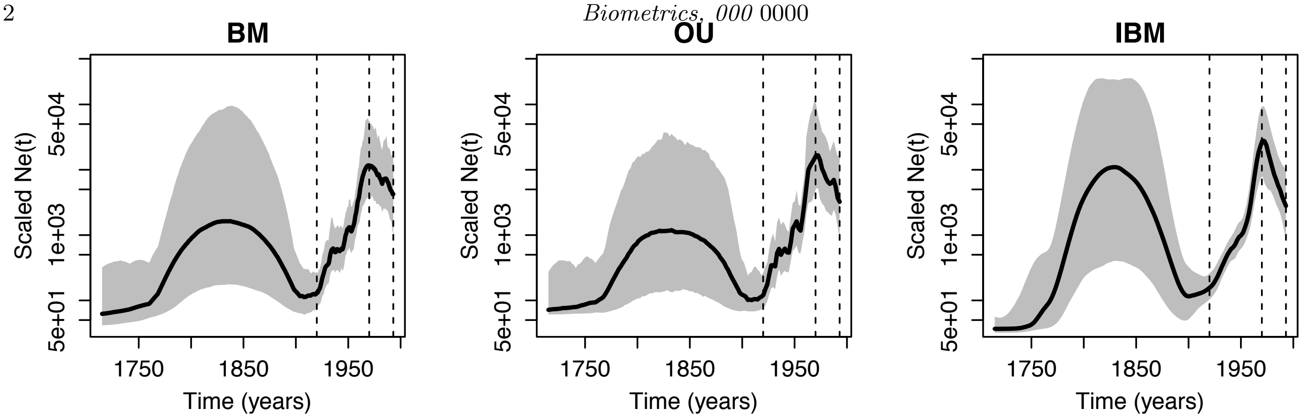

Figure 2. Egyptian HCV recoverd by placing three different Gaussian process priors. The first plot (left to right) corresponds to a Brownian motion (BM), the second – to Ornstein-Uhlenbeck (OU) and the last one – to the approximated integrated Brownian motion (IBM).

in $(t_k, t]$. Then, $\{N(t_k, t], t > t_k\}$ constitutes a Poisson process with intensity $C_k\lambda$. Then, given $N(t_k, t] = 1$, the conditional density of a point x in $(t_k, t]$ is $1/(t-t_k)$ and the probability of accepting such a point as a coalescent time point with variable population size is $1/\lambda N_e(x)$. Hence

$$P(t_{k-1} \leq t | t_k, N(t_k, t] = 1) = \frac{1}{\lambda(t-t_k)} \int_{t_k}^{t} \frac{dx}{N_e(x)}, \tag{A.2}$$

and

$$P(t_{k-1} > t | t_k, N(t_k, t] = m) = \left(1 - \frac{1}{\lambda(t-t_k)} \int_{t_k}^{t} \frac{dx}{N_e(x)}\right)^m. \tag{A.3}$$

Then,

$$P(t_{k-1} > t | t_k) = \sum_{m=1}^{\infty} P(t_{k-1} > t | t_k, N(t_k, t] = m) P(N(t_k, t] = m)$$

$$= \sum_{m=1}^{\infty} \left(1 - \frac{1}{\lambda(t-t_k)} \int_{t_k}^{t} \frac{dx}{N_e(x)}\right)^m \frac{(C_k\lambda(t-t_k))^m \exp\left[-C_k\lambda(t-t_k)\right]}{m!}$$

$$= \exp\left[-C_k\lambda(t-t_k)\right] \sum_{m=1}^{\infty} \frac{\left(C_k\lambda(t-t_k) - C_k \int_{t_k}^{t} \frac{dt_{k-1}}{N_e(t_{k-1})}\right)^m}{m!}$$

$$= \exp\left[-\int_{t_k}^{t} \frac{C_k dx}{N_e(x)}\right].$$

Algorithm 2 Simulation of isochronous coalescent times by thinning with $f(t) \sim \mathcal{GP}(\mathbf{0}, \mathbf{C}(\theta))$

Input: $k = n$, $t_n = 0$, $t = 0$, $i_j = 0$, $m_j = 0$, $j = 2, \ldots, n$, λ
Output: $\mathcal{T} = \{t_k\}_{k=1}^{n}$, $\mathcal{N} = \{\{t_{k,i}\}_{i=1}^{m_k}\}_{k=2}^{n}$, $\mathbf{f}_{\mathcal{T},\mathcal{N}}$
while $k > 1$ **do**
  Sample $E \sim \mathit{Exponential}(C_k\lambda)$ and $U \sim U(0,1)$
  t=t+E
  Sample $f(t) \sim P(f(t) | \{f(t_l)\}_{l=k}^{n}, \{\{f(t_{l,i})\}_{i=1}^{m_l}\}_{l=k}^{n}; \theta)$
  if $U \leq \frac{1}{1+\exp(-f(t))}$ **then**
    $k \leftarrow k-1$, $t_k \leftarrow t$
  else
    $i_k \leftarrow i_k + 1$, $m_k \leftarrow m_k + 1$, $t_{k,i_k} \leftarrow t$
  end if
end while

Algorithm 3 and 4 are analogous heterochronous versions of Algorithm 1 and 2.

An R implementation of these algorithms is available at
http://www.stat.washington.edu/people/jpalacio.

Algorithm 3 Simulation of heterochronous coalescent by thinning - $N_e(t)$ is a deterministic function

Input: $n_1, n_2, \dots, n_m$, $s_1, \dots, s_m$, $1/N_e(t) \le \lambda$, $N_e(t)$, m
Output: for $n = \sum_{j=1}^{m} n_j$, $\mathcal{T} = \{t_k\}_{k=1}^{n}$

```
i = 1, j = n − 1, n = n_1, t = t_n = s_1
while i < m + 1 do
    Sample E ~ Exp(binom(n,2) λ) and U ~ U(0,1)
    if U ≤ 1/(N_e(t+E) λ) then
        if t + E < s_{i+1} then
            t_j ← t ← t + E
            j ← j − 1, n ← n − 1
            if n > 1 then
                go to 2
            else
                go to 14
            end if
        else
            i ← i + 1, t ← s_i, n ← n + n_i
        end if
    else
        t ← t + E
    end if
end while
```

Algorithm 4 Simulation of heterochronous coalescent by thinning with $f(t) \sim \mathcal{GP}(\mathbf{0}, \mathbf{C}(\theta))$

Input: $n_1, n_2, \dots, n_m$, $s_1 = 0, \dots, s_m$, $i_j = 0$, $m_j = 0$, $j = 2, \dots, n$, λ, m
Output: for $n = \sum_{j=1}^{m} n_j$, $\mathcal{T} = \{t_k\}_{k=1}^{n}$, $\mathcal{N} = \{\{t_{k,i}\}_{i=1}^{m_k}\}_{k=2}^{n}$, $\mathbf{f}_{\mathcal{T},\mathcal{N}}$

```
i = 1, j = n − 1, n = n_1, t = t_n = s_1
while i < m + 1 do
    Sample E ~ Exp(binom(n,2) λ) and U ~ U(0,1)
    Sample f(t+E) ~ P(f(t+E) | {f(t_l)}_{l=k}^{n}, {{f(t_{l,i})}_{i=1}^{m_l}}_{l=k}^{n}; θ)
    if U ≤ 1/(1 + exp(−f(t+E))) then
        if t + E < s_{i+1} then
            t_j ← t ← t + E
            j ← j − 1, n ← n − 1
            if n > 1 then
                go to 2
            else
                go to 14
            end if
        else
            i ← i + 1, t ← s_i, n ← n + n_i
        end if
    else
        if t + E < s_{i+1} then
            t_{j+1, i_{j+1}} ← t + E, i_{j+1} ← i_{j+1} + 1
        end if
        t ← t + E
    end if
end while
```

Appendix B: MCMC Sampling

Since the coalescent under isochronous sampling is a particular case of the coalescent model under heterochronous sampling, we employ here the notation of the heterochronous coalescent, understanding that $C_{0,k} = C_k$, $I_{0,k} = (t_k, t_{k-1}]$ and $i = 0$ for isochronous data.

Sampling the number of latent points. A reversible jump algorithm is constructed for the number of "rejected" points. We propose to add or remove points with equal probability in each interval defined by Equations (3) and (4). When adding a point in a particular interval, we propose a location uniformly from the interval and its predicted function value $f(t^*) \sim P(f(t^*)|\mathbf{f}_{\mathcal{T},\mathcal{N}}, \theta)$. When removing a point, we propose to remove a point selected uniformly from the pool of rejected points in that particular interval. We add points with proposal distributions $q_{up}^{i,k}$ and remove points with proposal distributions $q_{down}^{i,k}$. Then,

$$q_{up}^{i,k} = \frac{P(f(t^*)|\mathcal{T}, \mathcal{N}, \theta)}{2l(I_{i,k})}, \tag{B.1}$$

$$q_{down}^{i,k} = \frac{1}{2m_{i,k}}, \tag{B.2}$$

and the acceptance probabilities are:

$$a_{up}^{i,k} = \frac{l(I_{i,k})\lambda C_{i,k}}{(m_{i,k}+1)(1+e^{f(t^*)})}, \tag{B.3}$$

$$a_{down}^{i,k} = \frac{m_{i,k}(1+e^{f(t^*)})}{l(I_{i,k})\lambda C_{i,k}}. \tag{B.4}$$

Sampling locations of latent points. We use a Metropolis-Hastings algorithm to update the locations of latent points. We first choose an interval defined by Equations (3) and (4) with probability proportional to its length and we then propose point locations uniformly at random in that interval together with their predictive function values $\mathbf{f}_{\mathbf{t}^*} \sim P(\mathbf{f}_{\mathbf{t}^*}|\mathbf{f}_{\mathcal{T},\mathcal{N}}, \theta)$. Since the proposal distributions are symmetric, the acceptance probabilities are:

$$a^{i,k} = \frac{1+e^{f(t)}}{1+e^{f(t^*)}}. \tag{B.5}$$

Sampling transformed effective population size values. We use an elliptical slice sampling proposal described in (Murray et al., 2010). In both cases, isochronous or heterochronous, the full conditional distribution of the function values $\mathbf{f}_{\mathcal{T},\mathcal{N}}$ is proportional to the product of a Gaussian density and the thinning acceptance and rejection probabilities:

$$P(\mathbf{f}_{\mathcal{T},\mathcal{N}}|\mathcal{T}, \mathcal{N}, \lambda, \theta) \propto P(\mathbf{f}_{\mathcal{T},\mathcal{N}}|\theta)L(\mathbf{f}_{\mathcal{T},\mathcal{N}}), \tag{B.6}$$

where

$$L(\mathbf{f}_{\mathcal{T},\mathcal{N}}) = \prod_{k=2}^{n}\left(\frac{1}{1+e^{-f(t_{k-1})}}\right)\prod_{i=1}^{m_k}\frac{1}{1+e^{f(t_{k,i})}}. \tag{B.7}$$

Sampling hyperparameters. The full conditional of the precision parameter θ is a Gamma distribution. Therefore, we update θ by drawing from its full conditional:

$$\theta|\mathbf{f}_{\mathcal{T},\mathcal{N}}, \mathcal{T}, \mathcal{N} \sim Gamma\left(\alpha^* = \alpha + \frac{\#\{\mathcal{N}\cup\mathcal{T}\}}{2}, \beta^* = \beta + \frac{\mathbf{f}_{\mathcal{T},\mathcal{N}}^t Q \mathbf{f}_{\mathcal{T},\mathcal{N}}}{2}\right), \tag{B.8}$$

where $Q = \frac{1}{\theta}C^{-1}$.

For the upper bound λ on $N_e(t)^{-1}$, we use the Metropolis-Hastings update by proposing new values using a uniform proposal reflected at 0. That is, we propose λ^* from $U(\lambda - a, \lambda + a)$. If the proposed value λ^* is negative, we flip its sign. Since the proposal distribution is symmetric, the acceptance probability is:

$$a = \frac{P(\lambda^*)}{P(\lambda)}\left(\frac{\lambda^*}{\lambda}\right)^{\#\{\mathcal{N}\cup\mathcal{T}\}} \exp\left[-(\lambda^* - \lambda)\sum_{k=2}^{n}\sum_{i=1}^{m_k} C_{i,k}l(I_{i,k})\right], \tag{B.9}$$

where $P(\lambda)$ is defined in Equation (7).

References

Lindgren, F. and Rue, H. (2008). On the second-order random walk model for irregular locations. *Scandinavian Journal of Statistics* **35,** 691–700.

Murray, I., Adams, R. P., and MacKay, D. J. (2010). Elliptical slice sampling. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, volume 9, pages 541–548.