Bayesian Nonparametric Methods for Stochastic Epidemic Models

Rowland Seymour, Dr Theodore Kypraios, Prof Philip O'Neill
School of Mathematical Sciences, University of Nottingham

University of Nottingham
UK | CHINA | MALAYSIA

Abstract

Simulating from and making inference for stochastic epidemic models are key strategies for understanding and controlling the spread of infectious diseases. Despite the enormous attention given to methods for parameter estimation, there has been relatively little activity in the area of nonparametric inference. That is, drawing inference for model parameters such as the infection rate, without making specific modelling assumptions about their functional form. Our method enables us to fit heterogeneously mixing models in which the infection rate between two individuals is a function, $f(\cdot)$, of their characteristics, for example location or type. We develop a novel method for inferring $f(\cdot)$ nonparametrically, removing the need to make questionable parametric assumptions about it. We adopt a Bayesian approach by assigning a Gaussian Process (GP) prior distribution to $f(\cdot)$ and then develop efficient data augmentation Markov Chain Monte Carlo methodology to estimate $f(\cdot)$, the GP hyperparameters and the unobserved infection times.

Statistical Epidemiology

Epidemic modelling is used to understand, analyse, predict and prevent outbreaks of infectious diseases. Often it is not clear how a diseases spreads between individuals, and this information can help develop control measures. For example, figure 1 shows an outbreak of Avian Influenza in the Netherlands, where we can see a spatial element to the spread of the virus. But how can we quantify this?

Our research is concerned with inferring model parameters for nonparametric, stochastic epidemic models using a Bayesian framework.

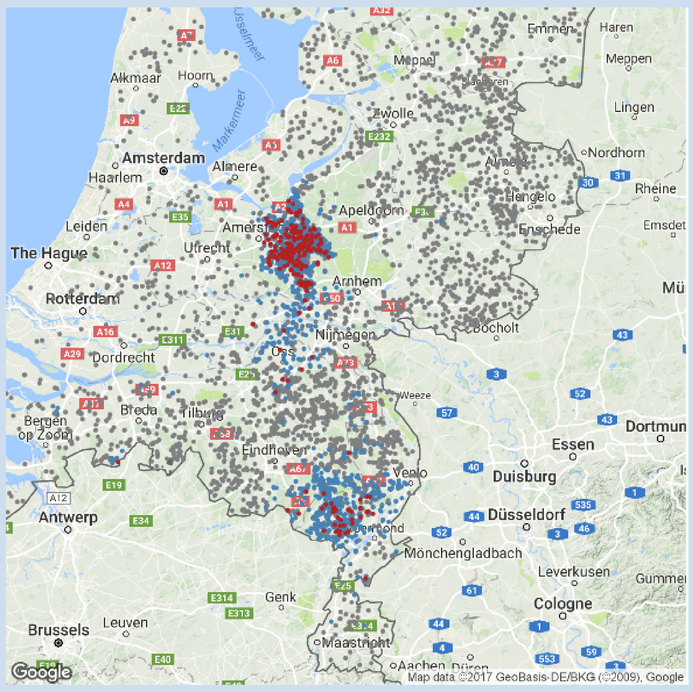

Figure 1. An outbreak of the Avian Influenza in the Netherlands. Red farms were infected and culled, blue farms were pre-emptively culled.

To model the outbreak we build a compartmental model, where individuals are either **S**usceptible, **I**nfected or **R**emoved.

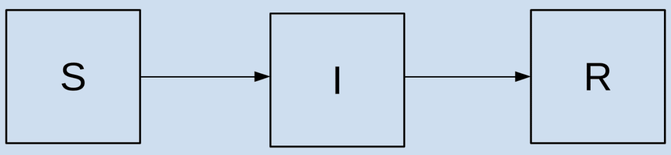

Figure 2. A typical SIR model.

We are interested in the rate at which individuals move between classes. Particularly, how the infection rate varies between individuals, or over distance for example. Standard modelling methods involve assuming a given form of the infection rate and inferring the model parameters. This can be restrictive as we need to write down the exact form of the function.

The inference is difficult because of the large amount of missing data. To evaluate the likelihood function we need to observe both the infection and removal times. We have developed a new method to infer the infection rate nonparametrically, and investigate how smooth this function is. We can implement our new method alongside current data augmentation methods for inferring the unobserved infection times.

Bayesian Nonparametric Methods and MCMC

The likelihood function is given by

$$\pi(\mathbf{i},\mathbf{r}|\boldsymbol{\beta},\gamma) \propto \underbrace{\exp\Big(-\sum_{j=1}^{n}\sum_{k=1}^{N}\beta_{j,k}\big((r_j \wedge i_k)-(i_j \wedge i_k)\big)\Big)}_{\text{Total Infectious Pressure}} \times \underbrace{\prod_{\substack{j=1\\ j\neq\kappa}}^{n}\Big(\sum_{k\in\mathcal{Y}_j}\beta_{k,j}\Big)}_{\substack{\text{Pressure infectives put}\\ \text{on each susceptible}}} \times \underbrace{\prod_{j=1}^{n} f(r_j-i_j|\gamma)}_{\text{Infectious period distribution}} .$$

We place a Gaussian Process prior distribution on the infection rate β, which allows us to nonparametrically estimate this parameter. To do this, we construct a covariance matrix using the distances between every pair of individuals. We put an exponential prior on the GP length scale parameter, which controls how smooth the inferred infection rate is.

$$\log\beta \sim \mathcal{GP}(\mathbf{0},\ \boldsymbol{\Sigma}_d), \quad \Sigma_d = k(\mathbf{d},\mathbf{d};\alpha,l).$$

We use a Monte Carlo Markov Chain (MCMC) algorithm to infer the model parameters. We use underrelaxed MCMC to propose updating the infection rate β to β', using the form

$$\log\beta' = \sqrt{1-\delta^2}\log\beta + \delta\nu, \qquad \nu \sim \mathcal{GP}(0,\Sigma_d), \qquad \delta\in(0,1].$$

We infer the infectious period distribution rate parameter γ using a Gibbs sampler, and data augmentation to infer the unobserved infection times. The full algorithm is shown below.

Algorithm: Nonparametric Inference for Spatially Heterogeneous Epidemic Models

Input : removal times $\mathbf{r}$, pair-wise distances $\mathbf{d}$

1. Start the Markov chain with initial estimates $\gamma^{(0)}$, $\beta^{(0)}$, and $\mathbf{i}^{(0)}$.

Repeat the following steps

2. Update γ using a Gibbs sampler and the density $\pi(\gamma|\beta,\mathbf{x},\mathbf{i},\mathbf{r})$
3. Update β using an underrelaxed proposal in a Metropolis Hastings algorithm
4. Choose one of the following three events with equal probability:
   - Move an infection time
   - Add the infection time for a susceptible, pre-emptively removed individual
   - Remove the infection time for an infected, pre-emptively removed individual

Results

We first run a simulation study to asses our new method. We simulate 250 outbreaks of a disease with a known infection rate across 1,000 individuals. We then run our algorithm on each outbreak to infer the infection rate and infection times. The results are shown in figure 3, and we can see the median infection rate for the study estimates the true rate well across the entire domain.

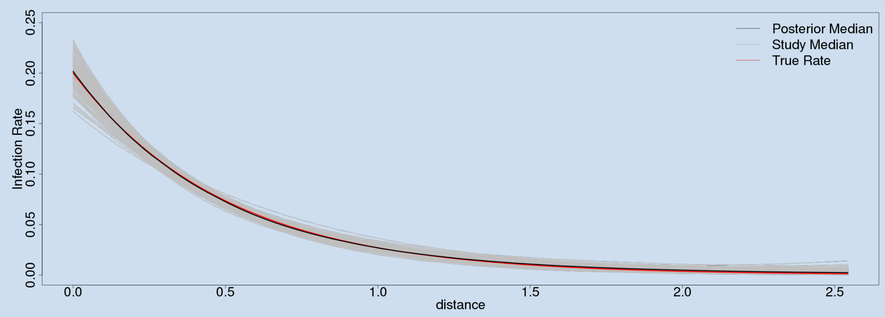

Figure 3. The results of the simulation study compared against the true infection rate.

We now run our algorithm on the Avian Influenza data set. These are preliminary results assuming pre-emptively culled farms were susceptible. We infer the infection times for the infected farms, and estimate that 33 farms were infected before the biosecurity measures were put in place.

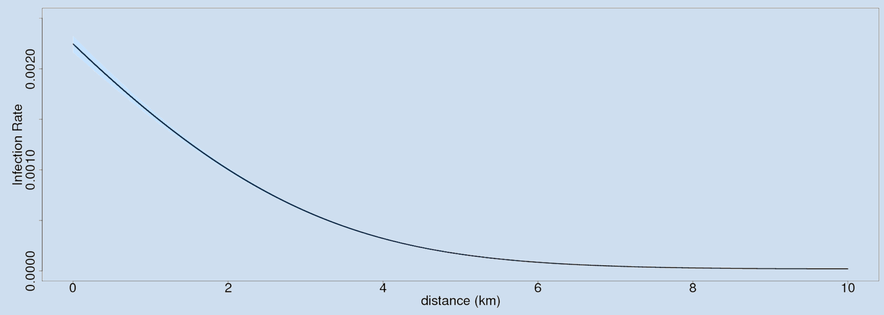

Figure 4. The preliminary analysis of the Avian Influenza Data Set.