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 the 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. Nonparametric inference is made more challenging as the likelihood function is intractable. This is because the times at which individuals were infected are unobserved, and in some cases we do not observe which individuals were infected. We develop a novel method for inferring $f(\cdot)$ nonparametrically, removing the need to make questionable parametric assumptions. 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.

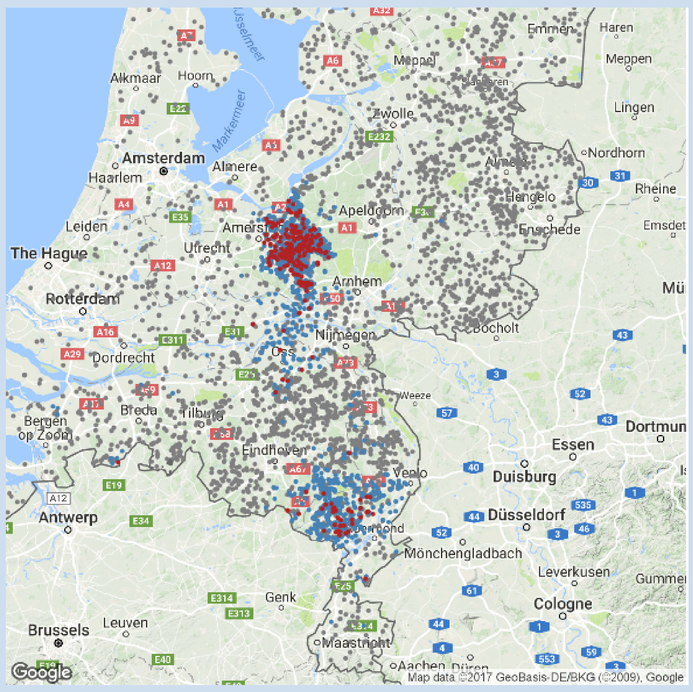

Figure 1. An outbreak of the Avian Influenza in the Netherlands. Red farms were infected and culled, blue farms were pre-emptively culled.

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

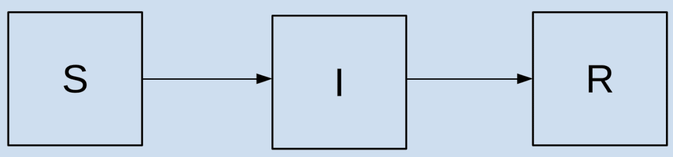

Figure 2. A typical SIR model.

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 made more difficult by a 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.

The Likelihood Function

To construct the likelihood function, we use the infectious pressure paradigm, where each infectious individual exerts infectious pressure onto the susceptible individuals. The likelihood function is given by

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

The likelihood function is intractable as we do not observe the infection times, $\mathbf{i}$. This is made more difficult in the Avian Influenza data set, as many farms were pre-emptively culled without their infection status being known. To infer the infection times and statuses, as well as the infection rate, we develop an efficient data augmentation MCMC algorithm.

Bayesian Nonparametric Methods

We put 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 underrelaxed MCMC to infer the infection rate. We propose new infection rate functions β' by

$$\log\beta' = \sqrt{1-\delta^2}\log\beta + \delta\nu, \quad \nu \sim \mathcal{GP}(\mathbf{0},\ \boldsymbol{\Sigma}_d).$$

To infer the unobserved infection times, at each iteration of the algorithm we propose

- moving an infection time,
- infecting a susceptible individual and creating their infection time,
- or making an infected individual susceptible and removing their infection time.

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.

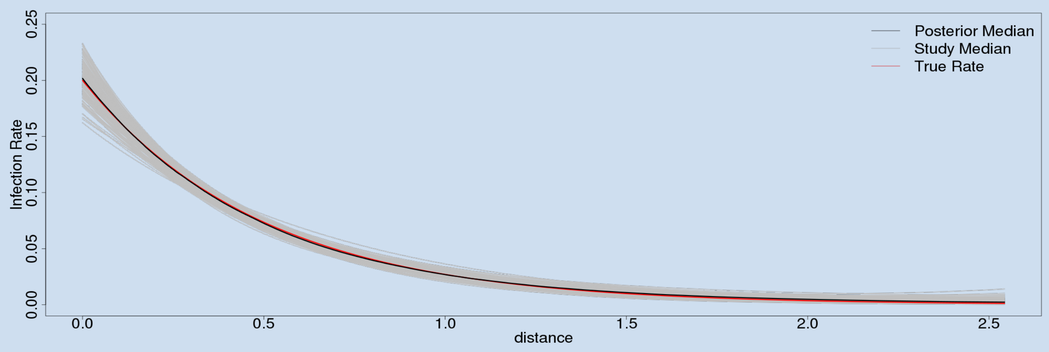

Figure 4. The results of the simulation study compared against the true infection rate

We now run our algorithm on the Avian Influenza data set. These are preliminary results assuming a fixed infectious period. We can compare this to our best parametric model. The results of the parametric and nonparametric results are similar but differ slightly suggesting that our new nonparametric method allows greater flexibility.

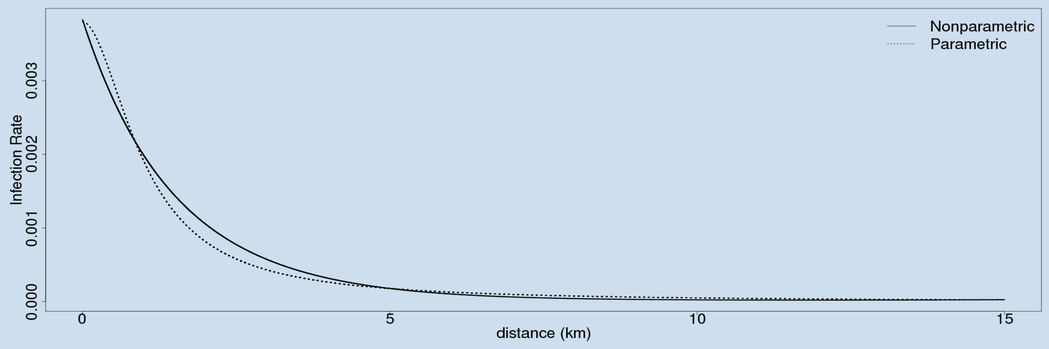

Figure 5. Nonparametric and parametric estimates for the Avian Influenza Infection rate.