Bayesian Non-Parametrics for Stochastic Infectious Disease Models

Rowland Seymour, School of Mathematical Sciences
Supervisors: Dr Theodore Kypraios and Prof Philip O'Neill

University of Nottingham
UK | CHINA | MALAYSIA

Epidemic Modelling

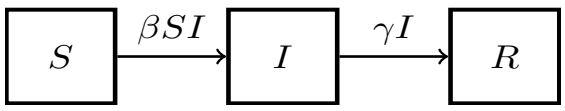
Epidemic modelling is used to understand, analyse, predict and prevent outbreaks of infectious diseases. Stochastic models are being used in a wide range of applications, for example:

- developing control measures for Ebola in west Africa,
- understanding the spread of Zika in South America,
- and protecting red squirrels from Squirrel Pox in Scottish islands.

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

SIR Models

Stochastic epidemic models are usually compartmental models and the most common model is an SIR model. Individuals are either in the susceptible, infected or removed category.



Non-Parametric Models

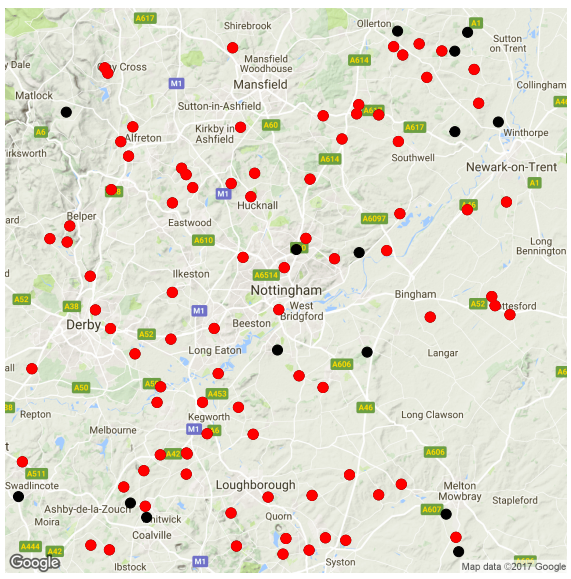
Using non-parametric models allow the data to speak for itself and reduces the chance of incorrect conclusions, which arise from assuming specific forms of the infection rate.

A common way to model the infection rate is to assume it has the form $\beta(t) = \beta_0 S_t I_t$. In this project, we will not assume such a parametric form. Instead, we will only assume the factors the infection rate depends on, such as time, position or information on how individuals are connected.

After choosing the factors, we will put a Gaussian process (GP) prior distribution on the infection rate, β. A GP is a collection a random variables, any finite subset of which has a joint Gaussian distribution.

We specify the mean and covariance functions of the GP, which are analogous to the Multivariate Gaussian distribution. They input our assumptions about the model into the GP, how quickly the output should vary.

Avian Influenza



Infected Not Infected

This is a hypothetical outbreak of Avian Influenza among poultry farms in the East Midlands. We simulated the outbreak using a distance dependent infection rate. Given the positions of the farms as well as their culling dates, we will now use our non-parametric methodology to infer the distance dependent infection rate.

Model

We compute the vector **d**, containing the distances between all of the farms sorted into ascending order, and use this as our input. We then put the following GP prior on $\beta_{i,j}$

$$\beta_{i,j} = \exp\{g(\mathbf{d})\}, \quad g \sim \mathcal{GP}\big(\mathbf{0},\ k(\mathbf{d},\mathbf{d})\big)$$

$$k(\mathbf{x},\mathbf{x}') = \alpha^2 \exp\Big\{-\frac{(\mathbf{x}-\mathbf{x}')^2}{l^2}\Big\}.$$

We exponentiate the GP to ensure the infection rate is always positive. We assume the farms remain infected for a period drawn from a $\Gamma(\lambda, \gamma)$ distribution, and that we wish to infer the rate parameter γ. We do this by putting an exponential prior distribution on this parameter. We use these prior distributions to derive the following posterior distribution

$$\pi(\mathbf{g},\gamma|\mathbf{i},\mathbf{r},\lambda) \propto \mathcal{GP}(\mathbf{g}) \exp\Big(-\sum_{j=1}^{n}\sum_{k=1}^{N} \exp\{g(||j-k||)\}\big((r_j \wedge i_k) - (i_j \wedge i_k)\big)\Big)$$

$$\times \prod_{j=1}^{n}\Big(\sum_{k\in\mathcal{Y}_j} \exp\{g(||j-k||)\}\Big)\gamma^{n\lambda}\exp\Big\{-\gamma\sum_{j=1}^{n}(r_j-i_j)\Big\}\gamma\exp\{-\nu\gamma\},$$

where **i** and **r** are the infection and culling dates respectively, g is the unknown infection rate, and there were N individuals in the population and n were infected.

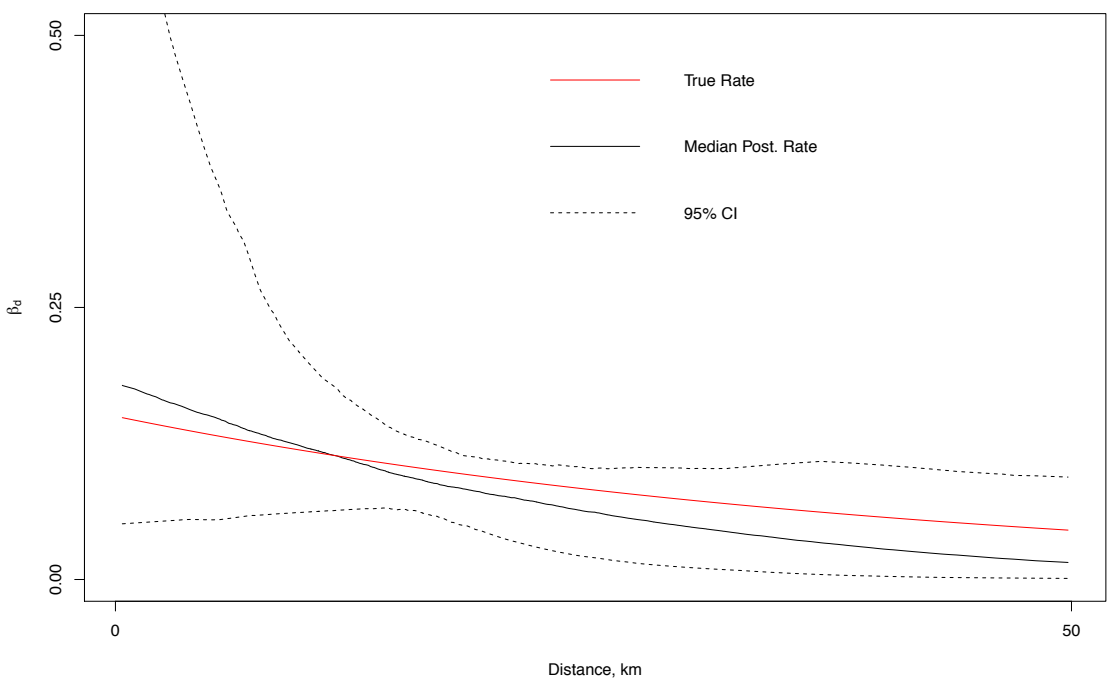
Method

We use MCMC to explore the posterior distribution. We do this by using **data augmentation** to infer the infection times by treating them as random variables, a **Gibbs sampler** to infer γ, and **under relaxed MCMC** to update the GP. In under relaxed MCMC, we propose new values of the infection rate as follows

$$g' = \exp\big\{\sqrt{1-\delta^2}\log g + \delta\nu\big\}, \quad \nu \sim N(\mathbf{0},\ k(\mathbf{d},\mathbf{d})).$$

Results

Using our non-parametric method, we correctly estimate the shape and scale of the infection rate for the Avian Influenza dataset, with a maximum absolute error of 0.03.



This work is supported by EPSRC (EP/N50970X/1)

rowland.seymour@nottingham.ac.uk