Recent Developments in Bayesian Non-Parametric Inference for Epidemic Models

Theo Kypraios
University of Nottingham
http://www.maths.nott.ac.uk/∼tk

School of Mathematics and Statistics, University of Sheffield

Epidemic Modelling

- Mathematical and statistical modelling has played a fundamental role in the fight against disease spread (eg policy making).
- Enormous attention has been given to the development of:
  - realistic (parametric) model of varying complexity,
  - computationally efficient methods for efficient parameter estimation (MCMC, SMC, ABC etc)
- There has been relatively little activity in the area of non–parametric inference; see, for example, Becker and Yip (1989), Boys and Giles (2007).

Why Non-Parametric?

In general, adopting a non–parametric approach:

- helps to avoid erroneous conclusions ...
- ... and biased results arising from the use of parametric models with (perhaps) inappropriate assumptions.
- Offers great modelling flexibility.
- Allows the data to speak for themselves.

Why Non-Parametric?

In general, adopting a non–parametric approach:

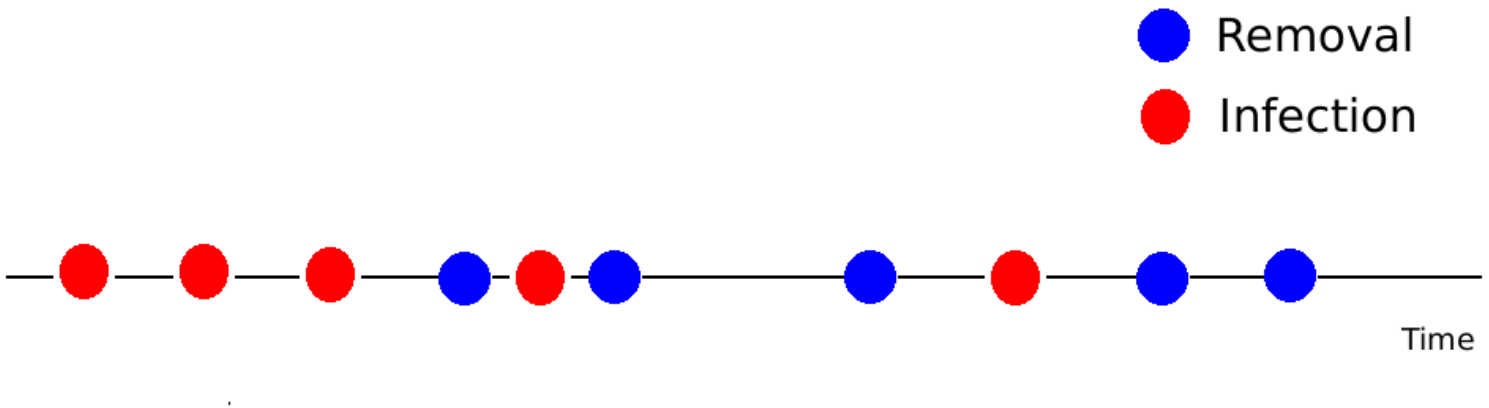
- helps to avoid erroneous conclusions ...
- ... and biased results arising from the use of parametric models with (perhaps) inappropriate assumptions.
- Offers great modelling flexibility.
- Allows the data to speak for themselves.

Disclaimer: Parametric models are of great value!

Roadmap

Susceptible-Infective-Removed (SIR) Model

- Consider an at-risk population of size N.
- At a given time, each individual of the population is susceptible, infective or removed.
- The epidemic is initiated by one infective in an otherwise fully-susceptible population.
- The epidemic ends when there are no more infectives.



Homogeneously-mixing SIR Model

- Underlying assumptions:
  - **S** $\rightarrow$ **I**: New infections occur at the points of a time non–homogeneous Poisson process with rate, for example,

$$\beta S_t I_t$$

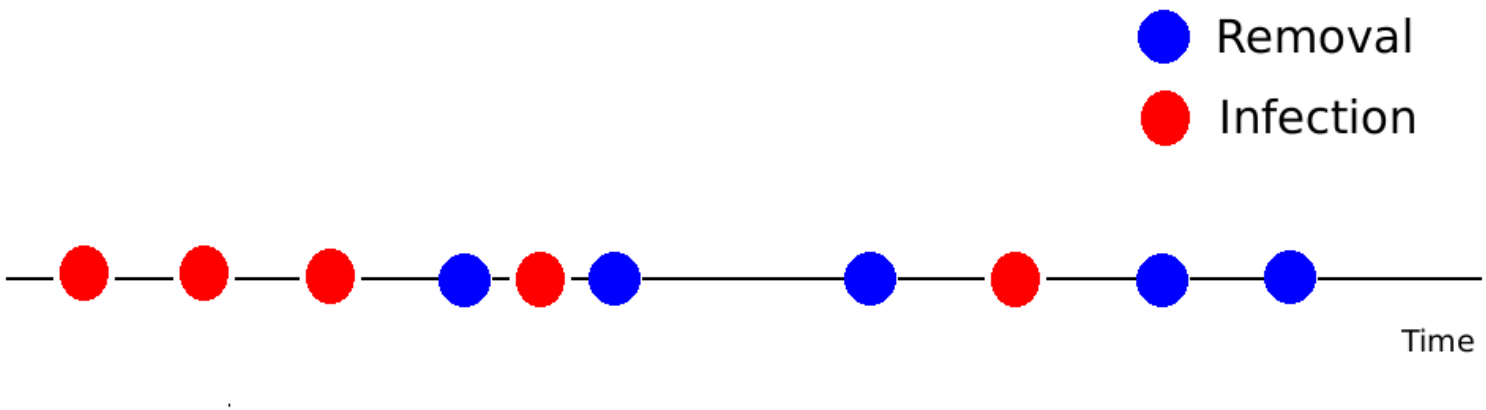
    where S_t and I_t are the number of susceptibles and infectives at time t respectively.

  - **I** $\rightarrow$ **R**: Infectives become removed after an infectious period which has an Exponential distribution with rate γ

$$R_i - I_i \sim \text{Exp}(\gamma)$$

Susceptible-Infective-Removed (SIR) Model: Inference

Suppose that some data are available:



For a statistical inference viewpoint, interest lies in estimating the infection rate β and the removal rate γ.

Relaxing the Usual Assumptions

- Underlying assumptions:
  - **S** $\rightarrow$ **I**: New infections occur at the points of a time non–homogeneous Poisson process with rate, for example,

$$\beta S_t I_t \quad \text{or} \quad \beta S_t I_t^{\delta} \quad \text{or} \quad \beta S_t^{\delta_1} I_t^{\delta_2} \quad \text{or} \quad \ldots$$

  - **I** $\rightarrow$ **R**: Infectives become removed after an infectious period which is distributed as:

$$\text{Exp}(\gamma) \quad \text{or} \quad \text{Gamma}(\mu, \nu) \quad \text{or} \quad \text{Weibull}(\mu, \nu) \quad \text{or} \quad \ldots$$

Non-Parametric

Main Idea

We relax the mass-action assumption i.e. new infections occur at rate $\beta S_t I_t$.

- Underlying Assumptions:
  - **S** $\rightarrow$ **I**: New infections occur at the points of a time non–homogeneous Poisson process with rate

$$h(t) > 0 \quad (t \in \mathbb{R})$$

  - **I** $\rightarrow$ **R**: Infectives become removed after an infectious period which has an arbitrary, but specified distribution, for example:

$$\text{Exp}(\gamma)$$

Hang on a Minute . . .

Criticism:

Aren't we moving from a model of the dynamics of a process to a descriptive model of events as a function of time?

Response:

That is indeed true. BUT there is still benefit to it (more details later).

One can also consider semi-parametric models, e.g. $h(t) =$ *function of* (S_t, I_t) *or* $h(t) =$ *function of* $(S_t I_t)$.

Bayesian Inference

We wish to infer $h(t)$ from data within a Bayesian framework:

- Surely, there are an uncountably infinite set of possible functions.
- How are we going to compute with this set in finite time?
- How do we place a prior distribution over a function?

Bayesian Inference

In this talk we discuss

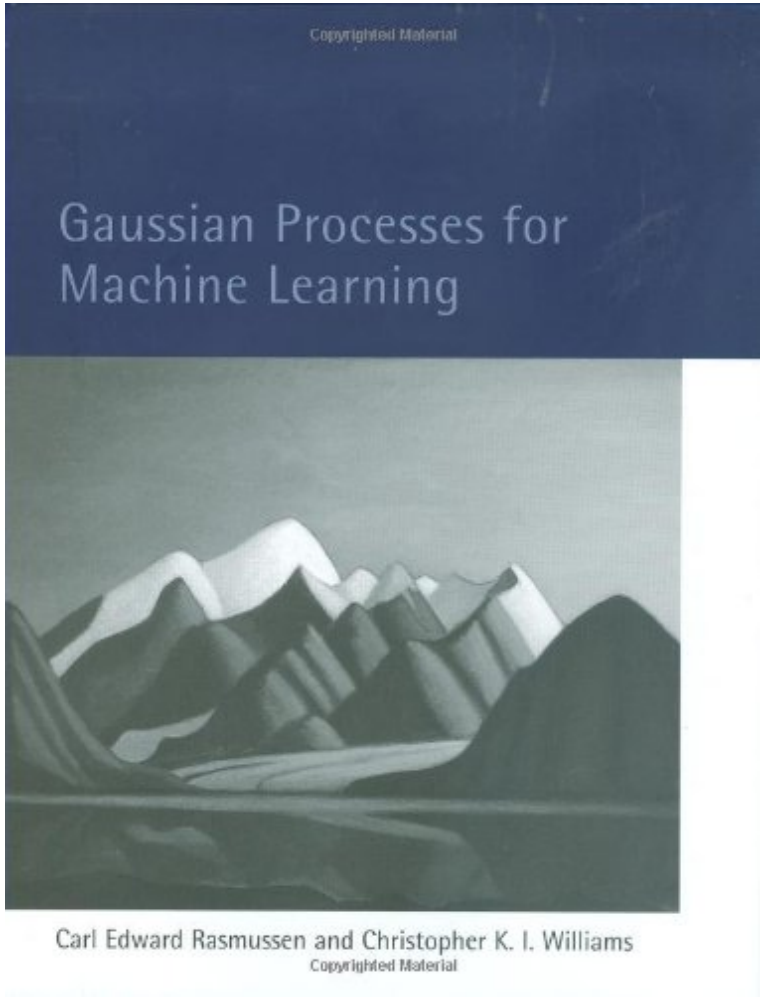
- three different prior choices /models for the unknown function $h(t)$
  - a Gaussian Process (GP);
  - a Step Function;
  - a B-Spline.
- how to deal with (very) large populations using Variational Bayes methods.
- extensions for models in discrete time / heterogeneously mixing populations.

GPs and Epidemics

Gaussian Processes

A Gaussian process (GP) is a generalization of the Gaussian probability distribution.

- The Gaussian distribution is over vectors, whereas the Gaussian process is over functions.
- GPs are used to describe a distribution over functions.



Definition

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

A GP is completely specified by its mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$.

We shall write

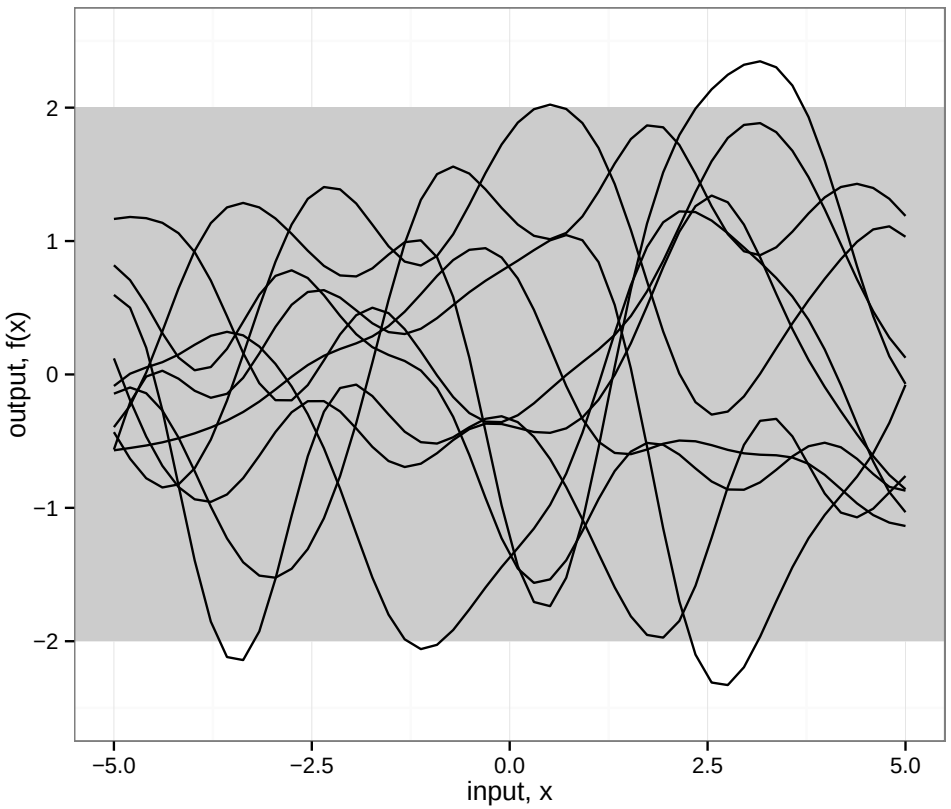
$$f \sim \mathcal{GP}\left(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right)$$

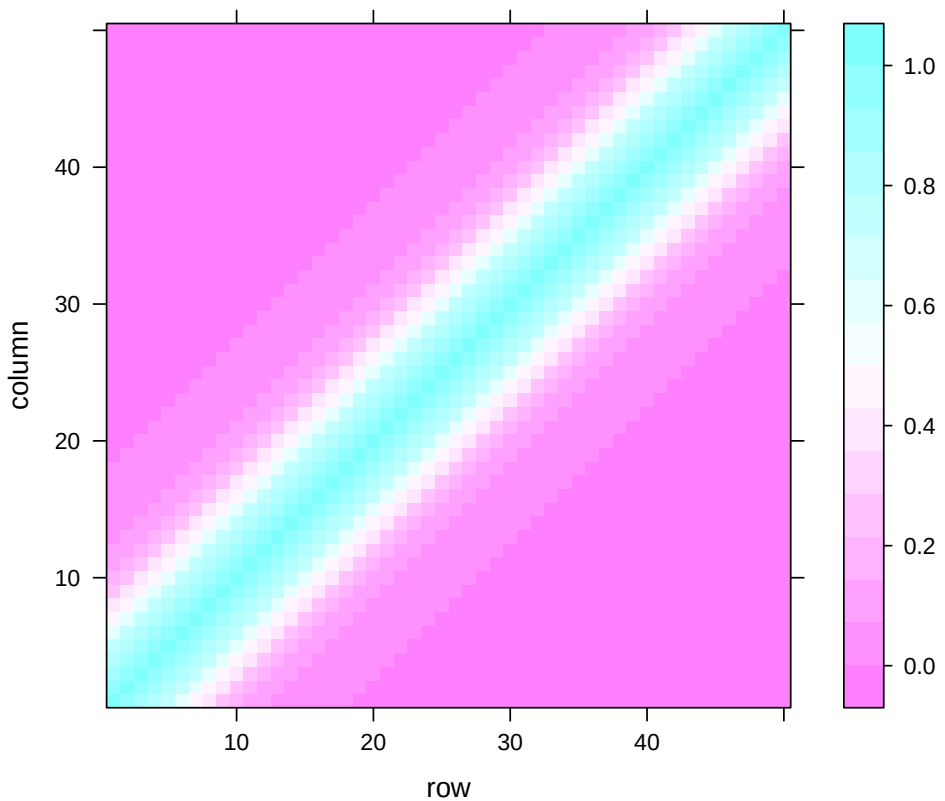
The covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$:

- is a crucial ingredient in GPs.
- It encodes our assumptions about the function which we wish to learn.

Gaussian Processes

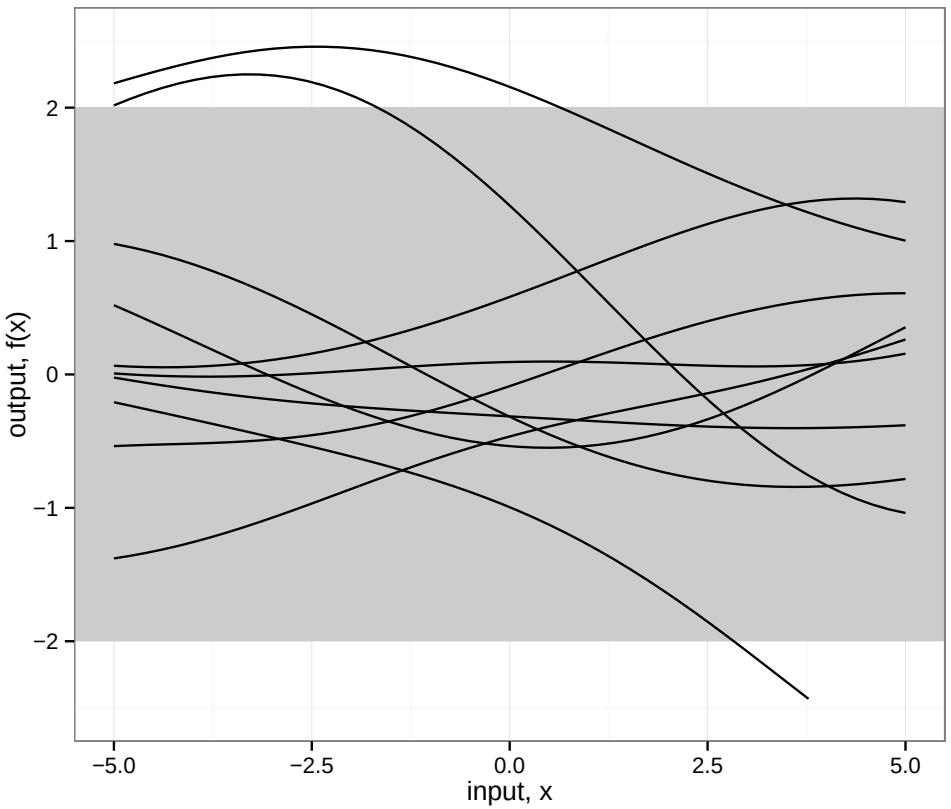
Square Exponential ($l = 1$)

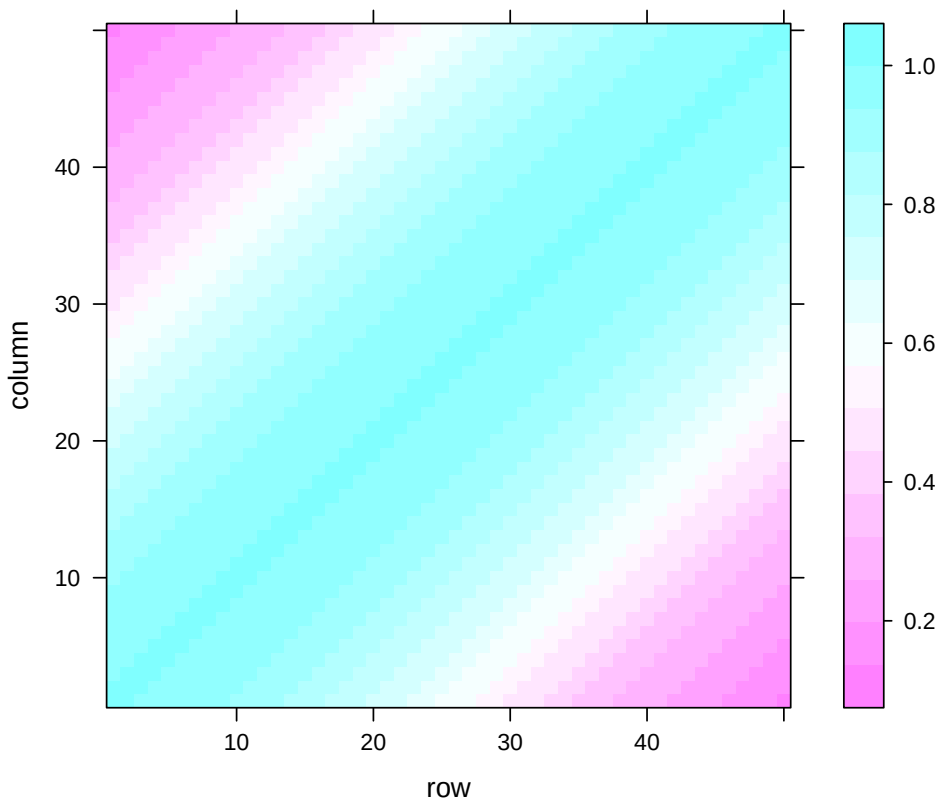




Gaussian Processes

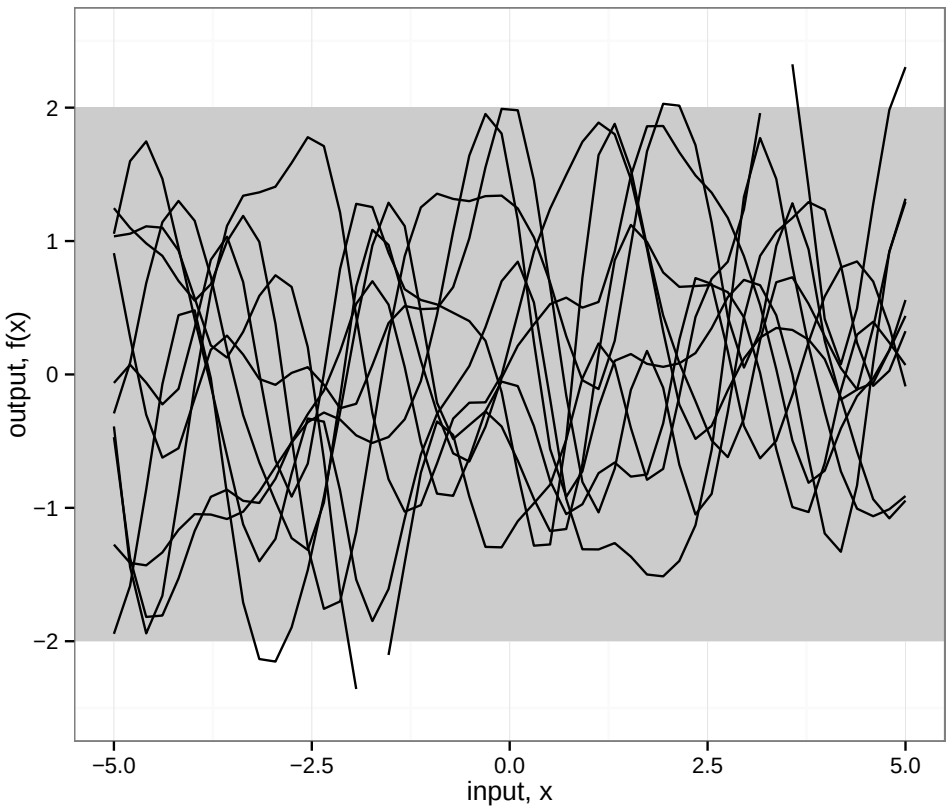
Square Exponential ($l = 5$)

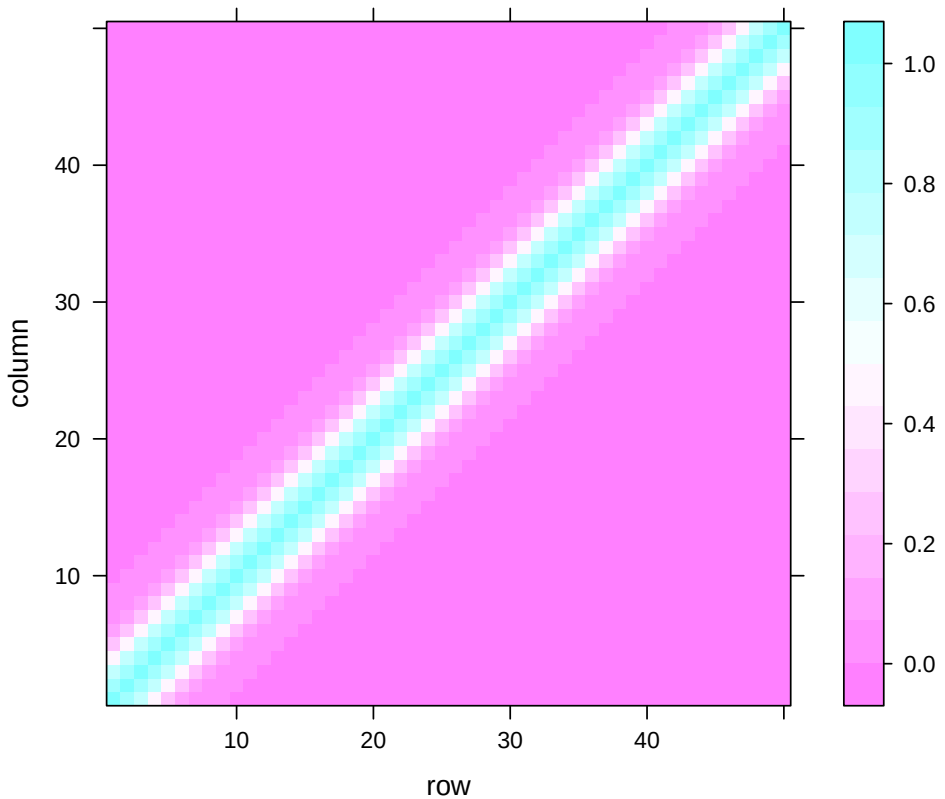




Gaussian Processes

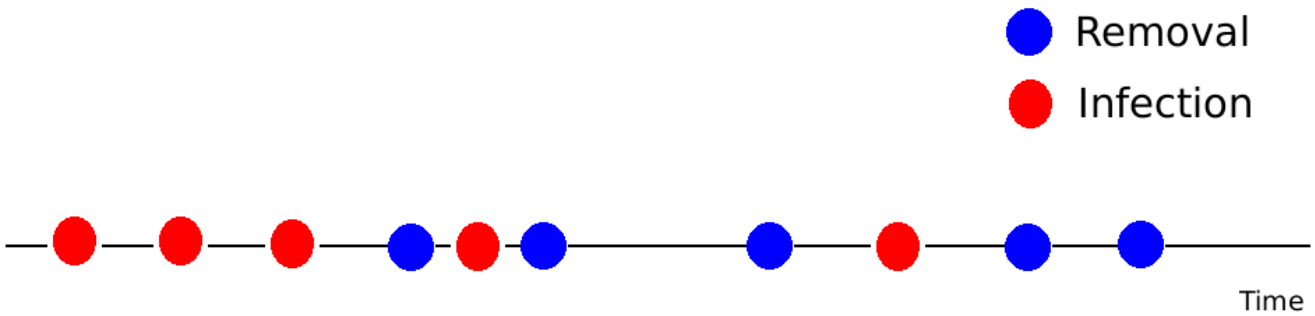
Square Exponential ($l = 0.5$)





Non Parametric Estimation of a Poisson's Process Intensity

- Suppose (for now) that we have a fully observed epidemic (i.e. both infection and removal times).



- Estimating $h(t)$ is equivalent to estimating the intensity function of a time−inhomogeneous Poisson process.
- Likelihood−based inference is generally intractable due to the need to integrate an infinite-dimensional random function.
- Various approximations have been introduced to deal with this intractability (e.g. Diggle, 1985).

GPs on Stochastic Epidemic Models

Our approach is to

1. adopt a Bayesian framework;
2. assign a GP prior on $h(t)$;
3. overcome the intractability by incorporating a data−augmentation framework ...
4. ... and develop efficient Markov Chain Monte Carlo algorithms to explore the posterior distribution of interest.

.

GPs on Stochastic Epidemic Models

Our approach is to

1. adopt a Bayesian framework;
2. assign a GP prior on $h(t)$;
3. overcome the intractability by incorporating a data−augmentation framework ...
4. ... and develop efficient Markov Chain Monte Carlo algorithms to explore the posterior distribution of interest.

Such an approach avoids any approximations.

Log Gaussian Cox Process (LGCP)

GP ingredients

- Input space $\mathcal{X}$ (e.g. $\mathbb{R}^D$)
- Output space $\mathcal{Y} = \mathbb{R}$
- Positive-definite covariance function $k(\boldsymbol{x}, \boldsymbol{x}'; \theta) \to \mathbb{R}$
- mean function $m(\boldsymbol{x}; \theta) : \mathcal{X} \to \mathcal{Y}$

The idea is to use a GP as a prior for the Poisson process intensity $h(t) \quad (t > 0)$. But the output space is $\mathbb{R}$ ...

Do the natural thing, i.e. transform $f(\cdot)$:

$$\begin{aligned} g(t) &\sim \mathcal{GP}(t, \theta) \\ h(t) &= \exp(g(t)) \end{aligned}$$

Doubly Intractable Inference

The likelihood of events $\{t_p\}_{p=1}^P$ between 0 and T

$$P\left(\{t_p\}_{p=1}^P \mid h(t) = \boldsymbol{h}\right) = \exp\left\{-\int_0^T \exp\left(g(t)\right)\,\mathrm{d}t + \sum_{p=1}^P g(t_p)\right\}$$

Problems

- $h(t)$ is infinite dimensional;
- the posterior distribution is doubly intractable in the sense that the likelihood is only known up to a constant which depends on the parameters of interest.
- Inference is hard and routine MCMC algorithms do not work.
- Recent advances in MCMC enable inference for such problems [Möller et al., 2004, Murray et al., 2006] given that you are able to generate *exact realisations* from the model.

Exact Simulation of an Inhomgeneous Poisson Process

Aim: Simulate points from a PP with intensity $\lambda\phi(t)$ in [0, T]

1. `Draw` $N \sim Po(\lambda T)$
2. `Distribute the` N `points, say` $\{t_p\}_{p=1}^{P}$`, independently and uniformly on` $[0, T]$
3. `Remove` t_p `with probability` $1 - \phi(t_p)$

The remaining events are points from the desired Poisson process.

- We did not have to discover the function at more than a finite number of locations.
- We did not have to integrate the function.
- Data are exactly drawn from a Poisson process with the desired intensity.

Inference via Latent History

Adams, Murray and MacKay, 2009

Given the P events on $[0, T]$ and the GP prior, the posterior is still intractable.

However, if we augment the state with the "latent history" of the generative procedure ... and assume there were K thinned events, $\{s_k\}_{k=1}^K$ we can write down the full joint distribution:

$$\pi\left(\{t_p\}_{p=1}^P, \{s_k\}_{k=1}^K, \boldsymbol{g} | \lambda, \theta\right) =$$

$$\lambda^{P+K} \exp\{-\lambda T\} \quad \text{[homogeneous Poisson process]}$$

$$\times \prod_{p=1}^{P} g(t_p) \times \prod_{k=1}^{K} (1 - g(t_p)) \quad \text{[probability of unthinned/thinned events]}$$

$$\times \mathcal{GP}\{g(t_p)_{p=1}^P\}, \{g(s_k)_{k=1}^K\} | \theta \quad \text{[GP prior]}$$

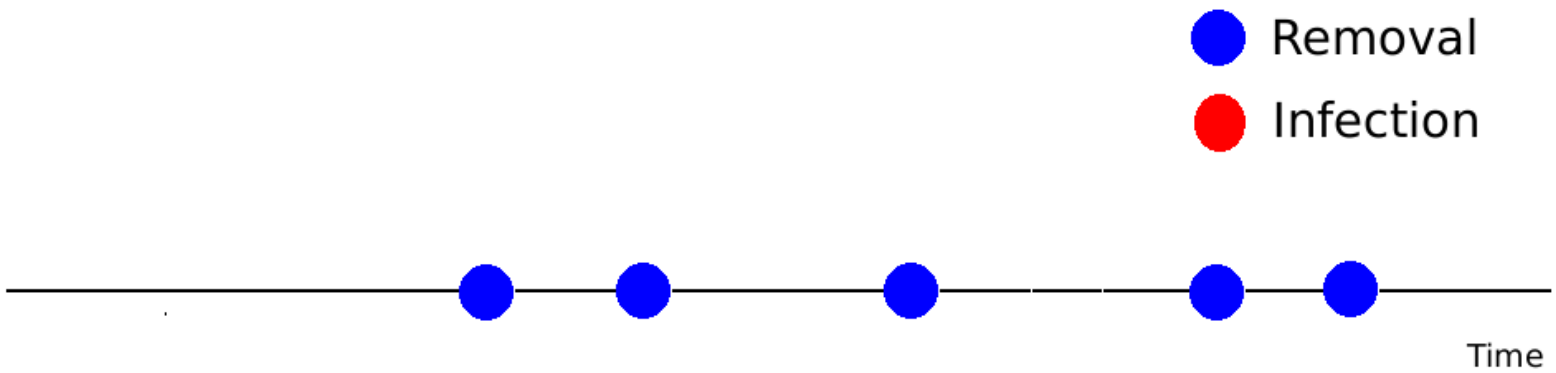
Exploring the Posterior Distribution

$\pi\left(\{t_p\}_{p=1}^{P}, \{s_k\}_{k=1}^{K}, \boldsymbol{g}|\lambda, \theta\right)$ is not pleasant but tractable and can sample from it using MCMC:

- We update each part of the latent state separately, conditioned on the others using a Gibbs−like sampler.
- Insert and remove latent thinned events via Metropolis−Hastings.
- Move latent thinned events around via Metropolis−Hastings.
- Sample the latent function Metropolis−Hastings (or Hamiltonian Monte Carlo for more efficiency)
- The hyperparameters of the GP can also be updated.

Accounting for Unobserved Infection Times

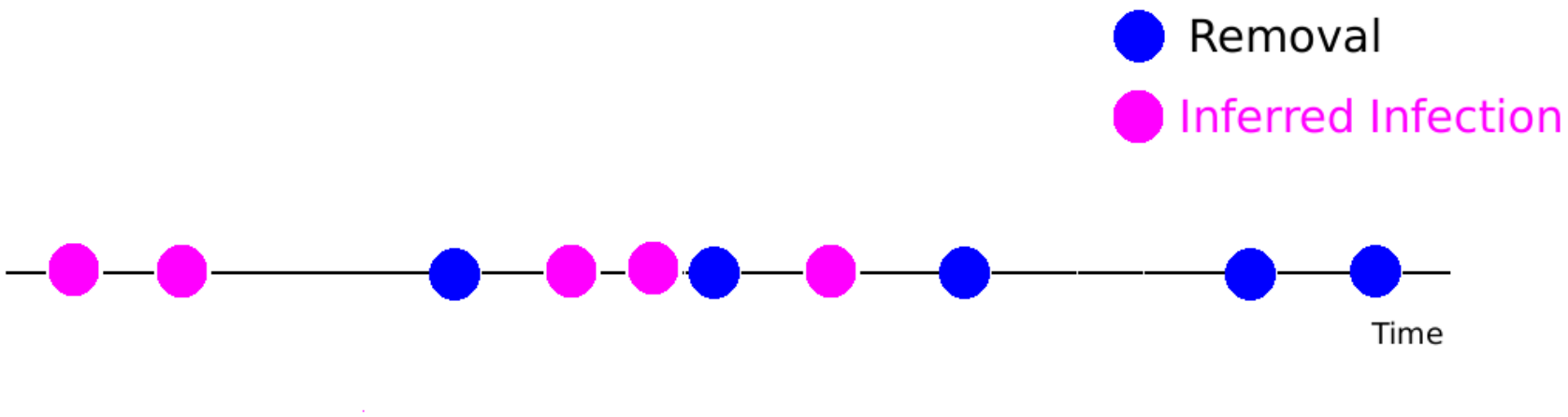
So far we have assumed that we observe both the infection and removal times, but infection times are rarely observed.



Accounting for Unobserved Infection Times

So far we have assumed that we observe both the infection and removal times, but infection times are rarely observed.

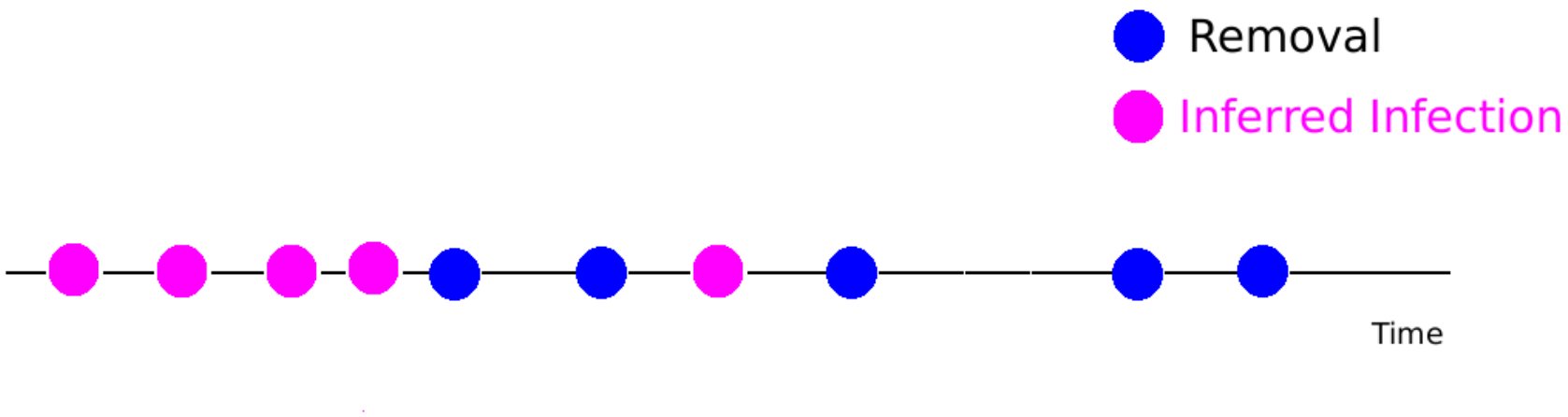
Augment the space (even further) with them and update them as well via an MCMC scheme.



Accounting for Unobserved Infection Times

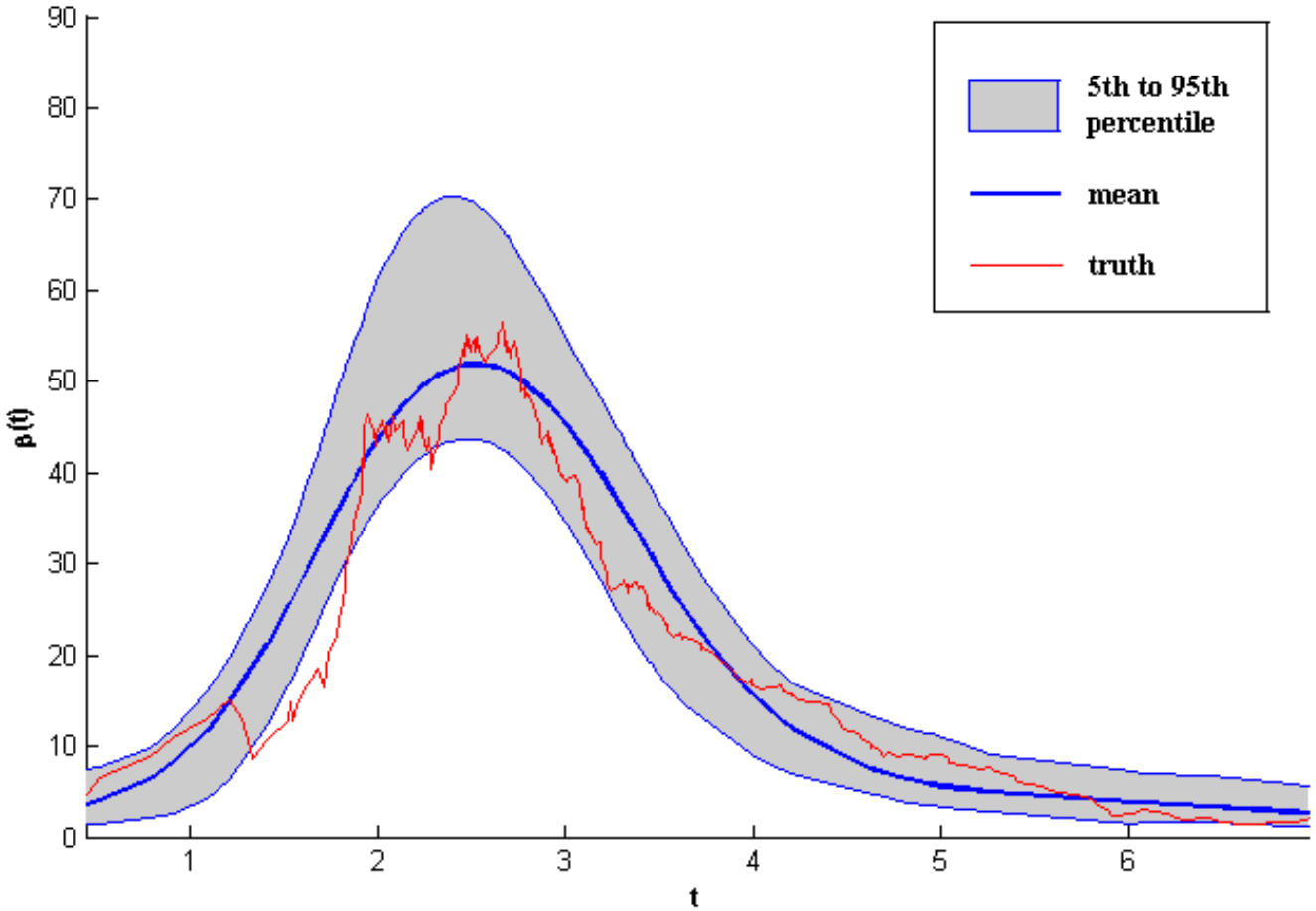
So far we have assumed that we observe both the infection and removal times, but infection times are rarely observed.

Augment the space (even further) with them and update them as well via an MCMC scheme.



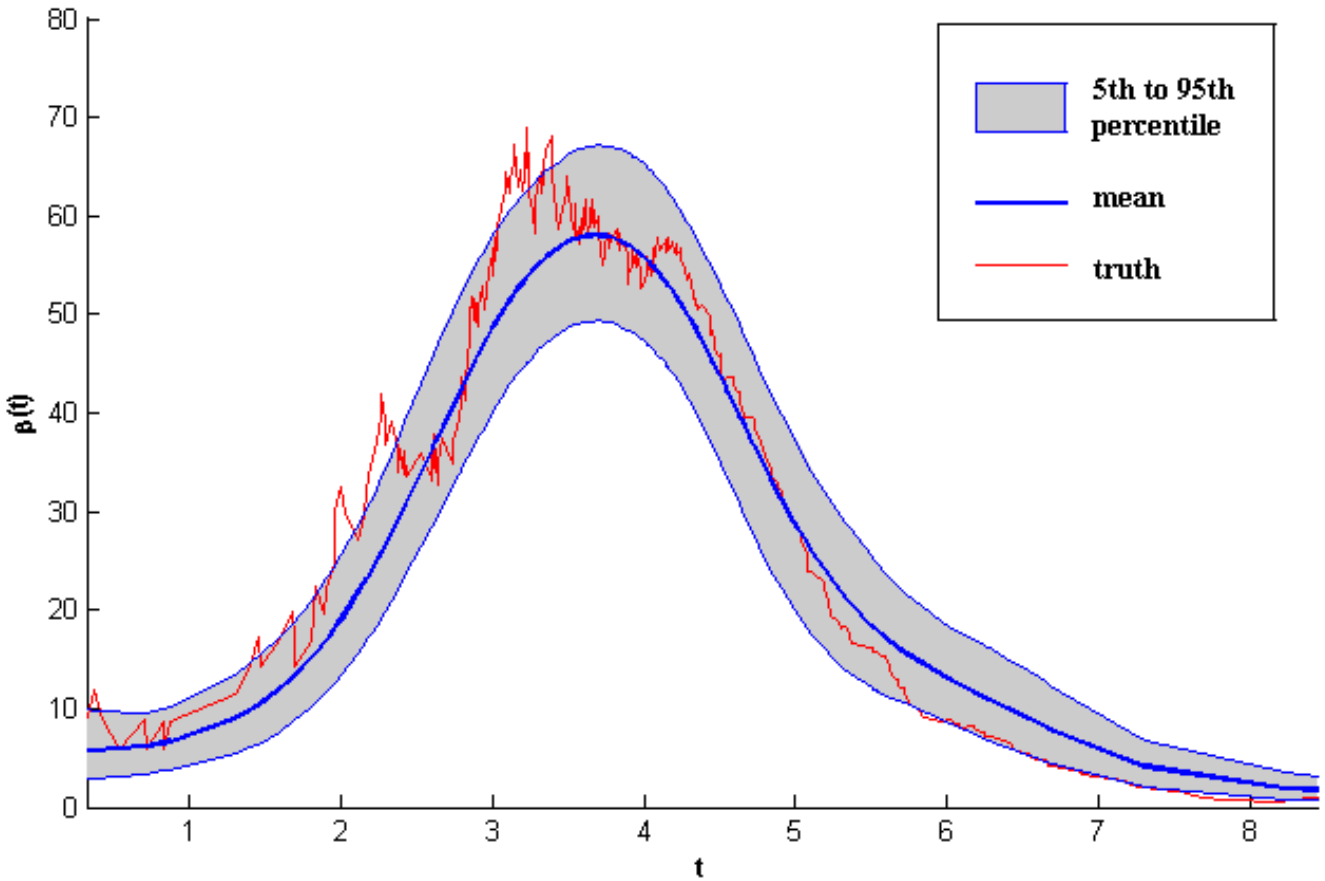
Does this Really Work?

SIR model, $N = 150$ true intensity: $\beta S_t I_t$, synthetic data



Does this Really Work?

SIR model, $N = 200$ true intensity: $\beta S_t I_t$, synthetic data



GPs for Epidemic Models

Reflections

- Flexible models
- Covariance functions.
- Need efficient MCMC samplers.
- Computational cost can be quite high.

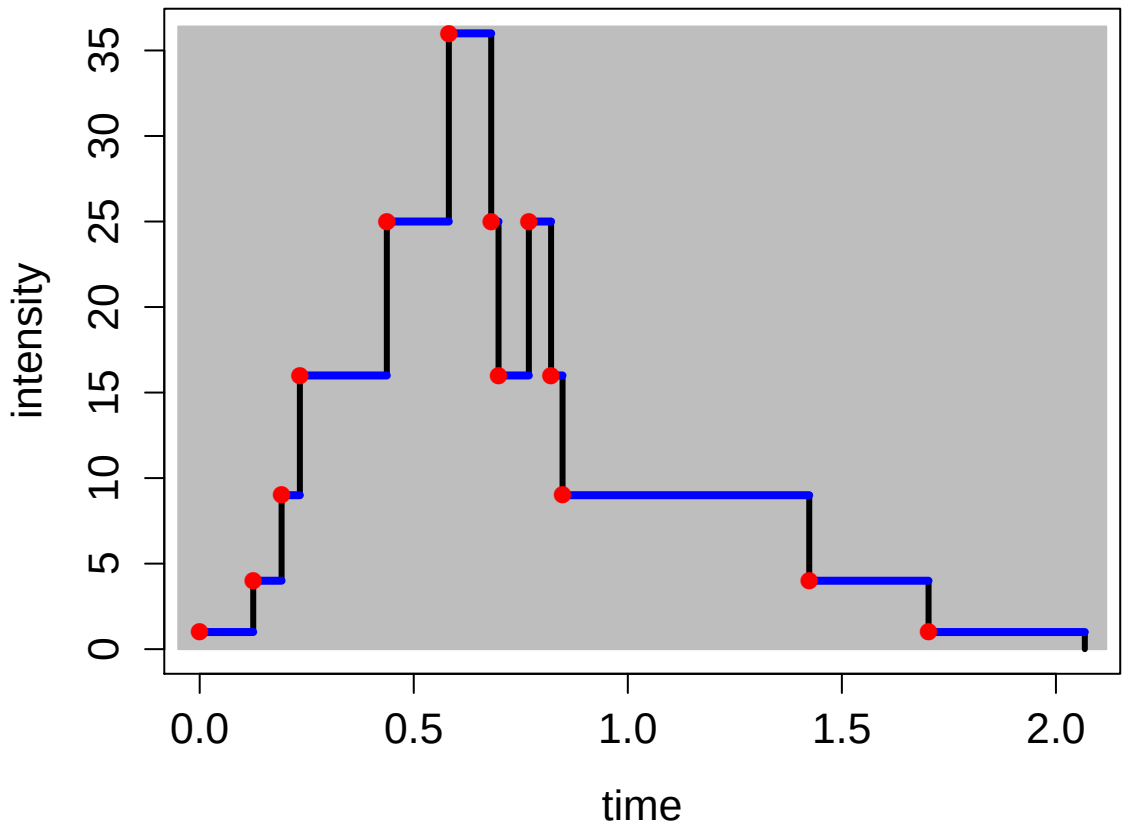
Step Function and B-Splines

Model the infection rate $h(t)$ via Step−Functions

An alternative approach (model) for $h(t)$ is to treat it as a *step function*.

Ingredients:

- changepoints (number + locations)
- the heights.



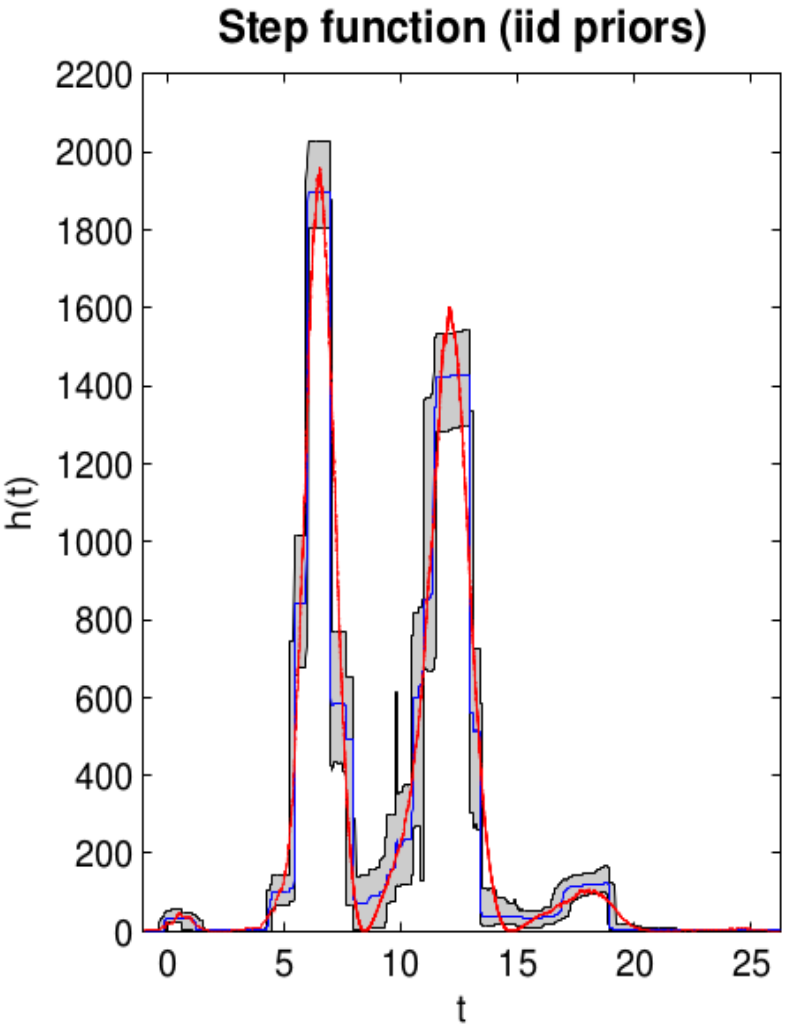
Model the infection rate $h(t)$ via Step−Functions

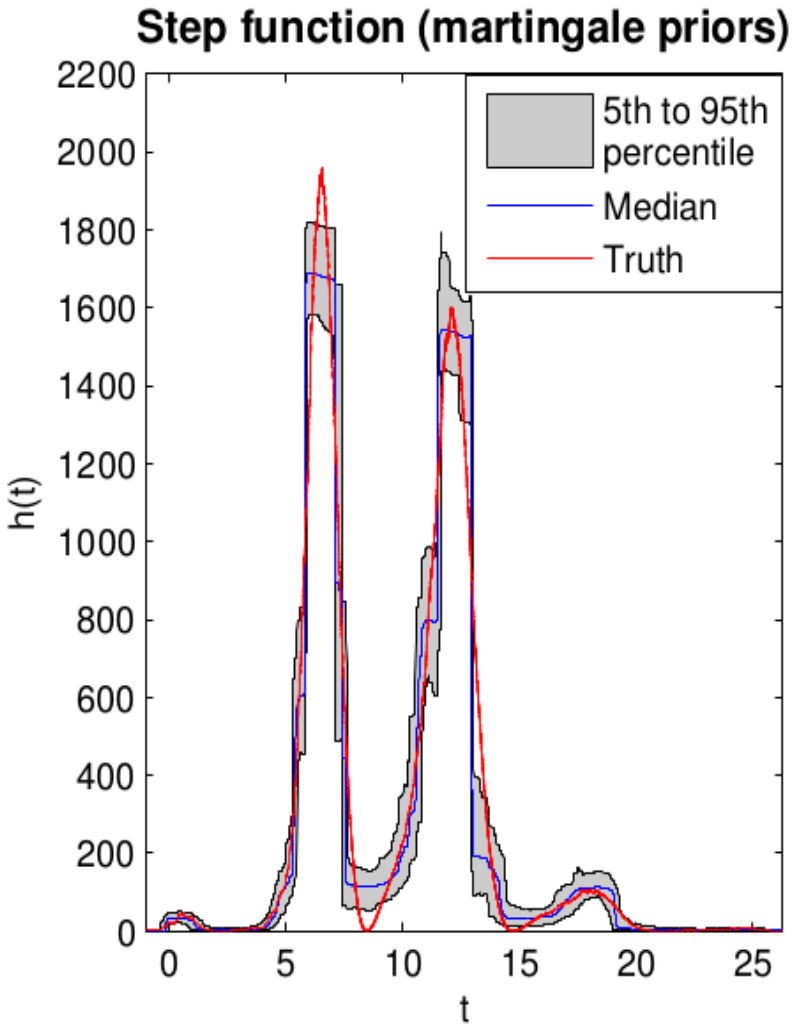
- Modelling $h(t)$ enables straightforward calculation of $\int_0^T h(t)\,\mathrm{d}t$ required for the likelihood function.
- However:
  - the number of change points, say k,
  - their locations, $s_1, \ldots, s_k$
  - and the height of the function at these points $(h_0, h_1, \ldots, h_k)$

  are unknown and needed to be estimate from the observed data (i.e. removal times).
- Can be done using RJ-MCMC (birth/death of changepoints etc).
- Can incorporate smoothness by using *Martingale priors*. [Similar to Arjas and Gasbarra, 1994]

Illustration

True model: $h(t) = \beta(1 + \cos(t - I_{(1)}))S_t I_t$, $N = 10,000$, 6971 infectives





Introducing (Even) More Smoothness

An alternative to step functions we consider is a 2nd-order B-spline, which is a continuous, piecewise quadratic function.

We assume given $k + 6$ knots that $h(t)$ is a linear combination of B-spline basis functions:

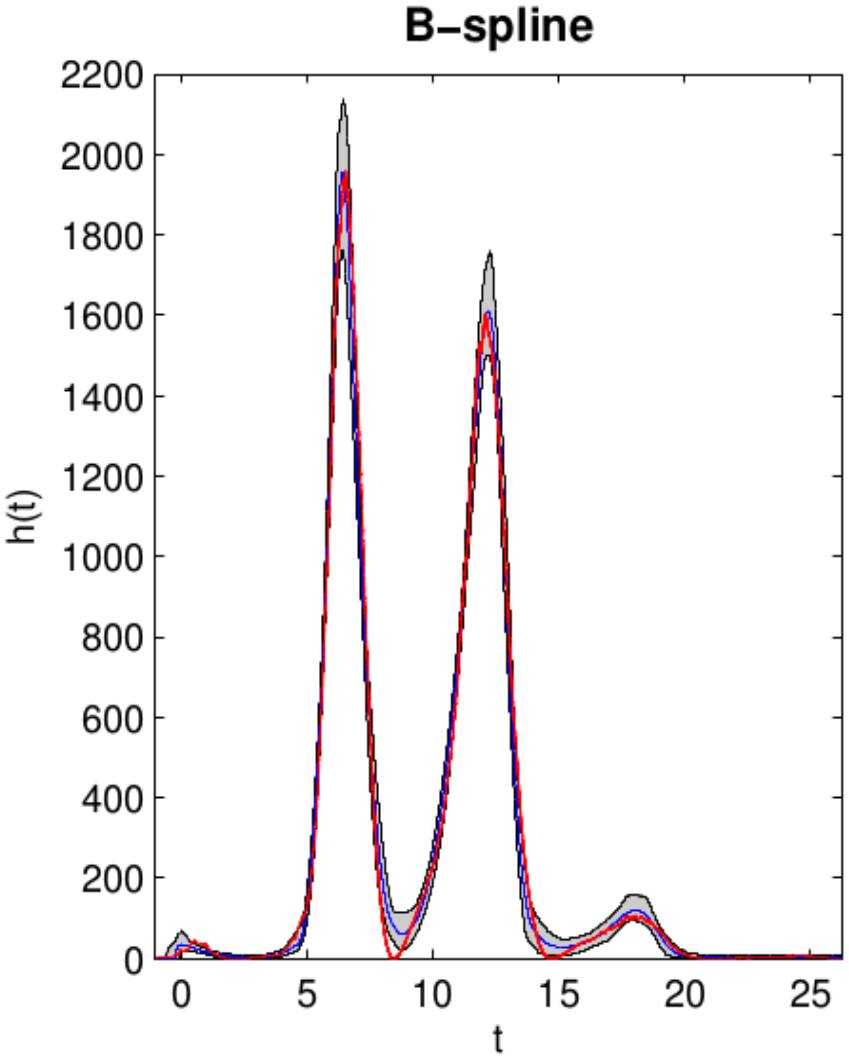
$$h(t) = \sum_{i=0}^{k+2} P_{i+1} b_{i,2}(t),$$

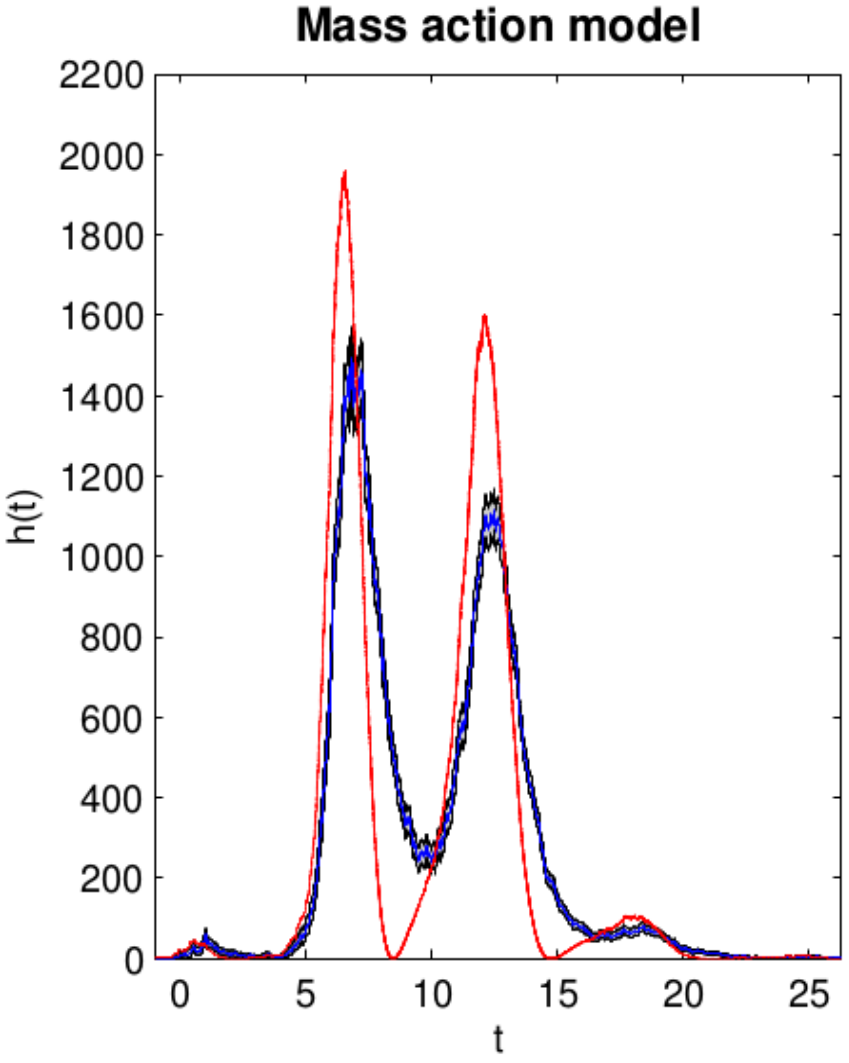
where $b_{i,j}(t)$ is the ith B-spline basis function of order j.

Again, one can employ a RJ-MCMC algorithm to sample from the posterior distribution.

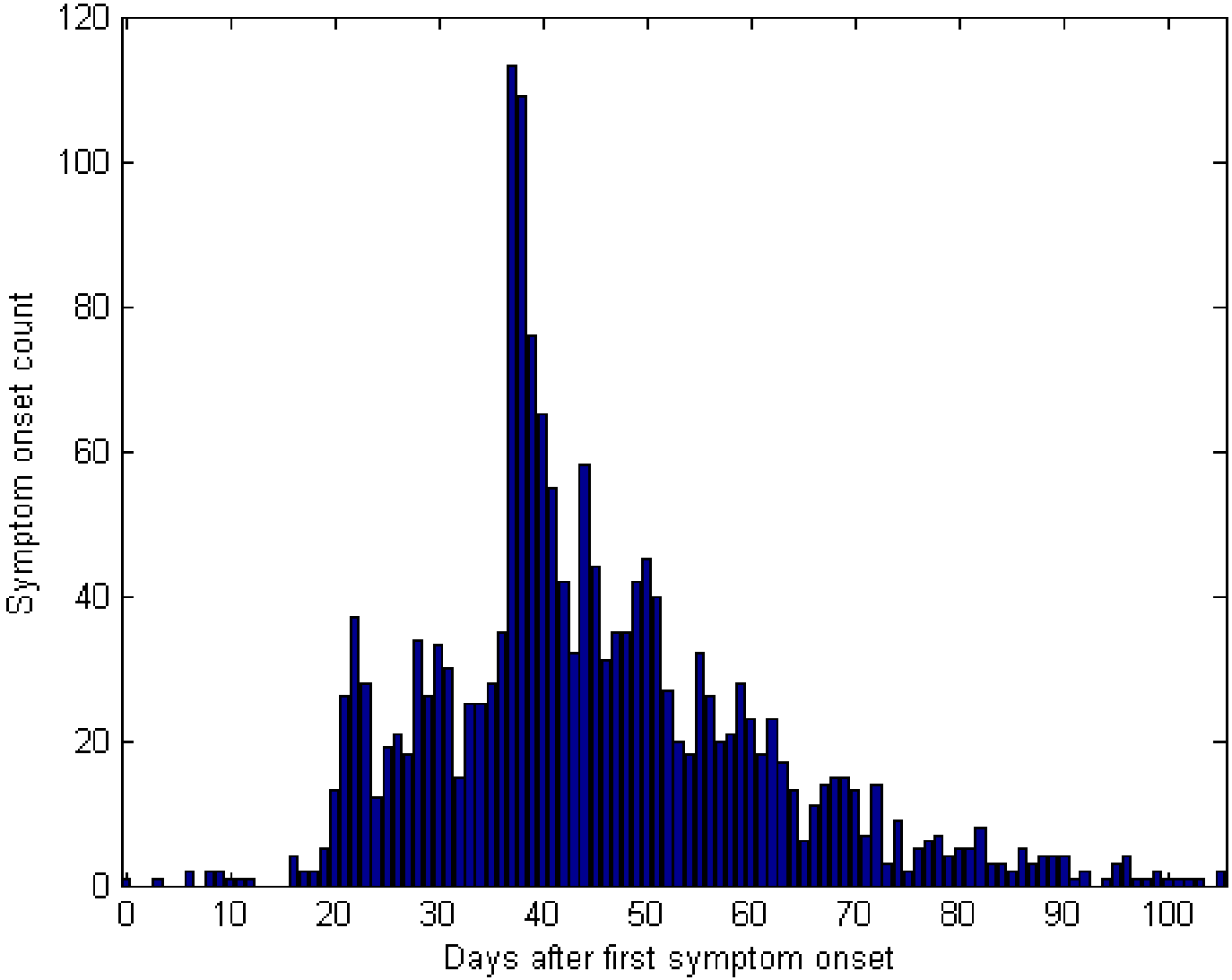
Illustration

True model: $h(t) = \beta(1 + \cos(t - I_{(1)}))S_t I_t$, $N = 10,000$, 6971 infectives



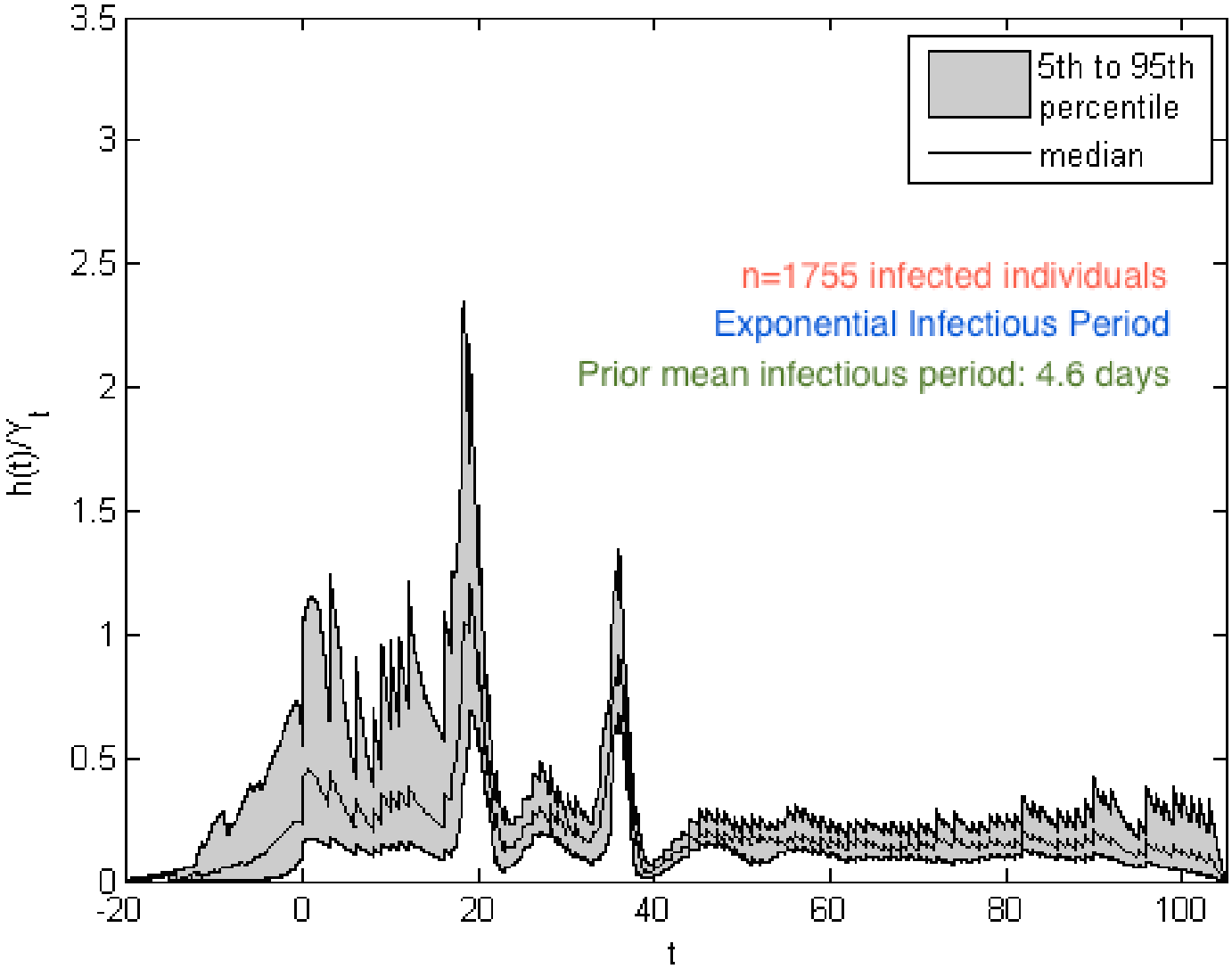


Application to 2003 Hong Kong SARS symptom onset data



Application to 2003 Hong Kong SARS symptom onset data

Bayesian non−parametric estimation of the rate per infective ($h(t)/I_t$)



Heterogeneously Mixing Populations

- There is often quite a lot of structure in the population:
  - households
  - workplaces
  - ...
- Hence, assuming homogeneously mixing often is reasonable.
- For example, a suitable parametric model for the infection rate β_{ij} which takes into account the distance between i and j (ρij).

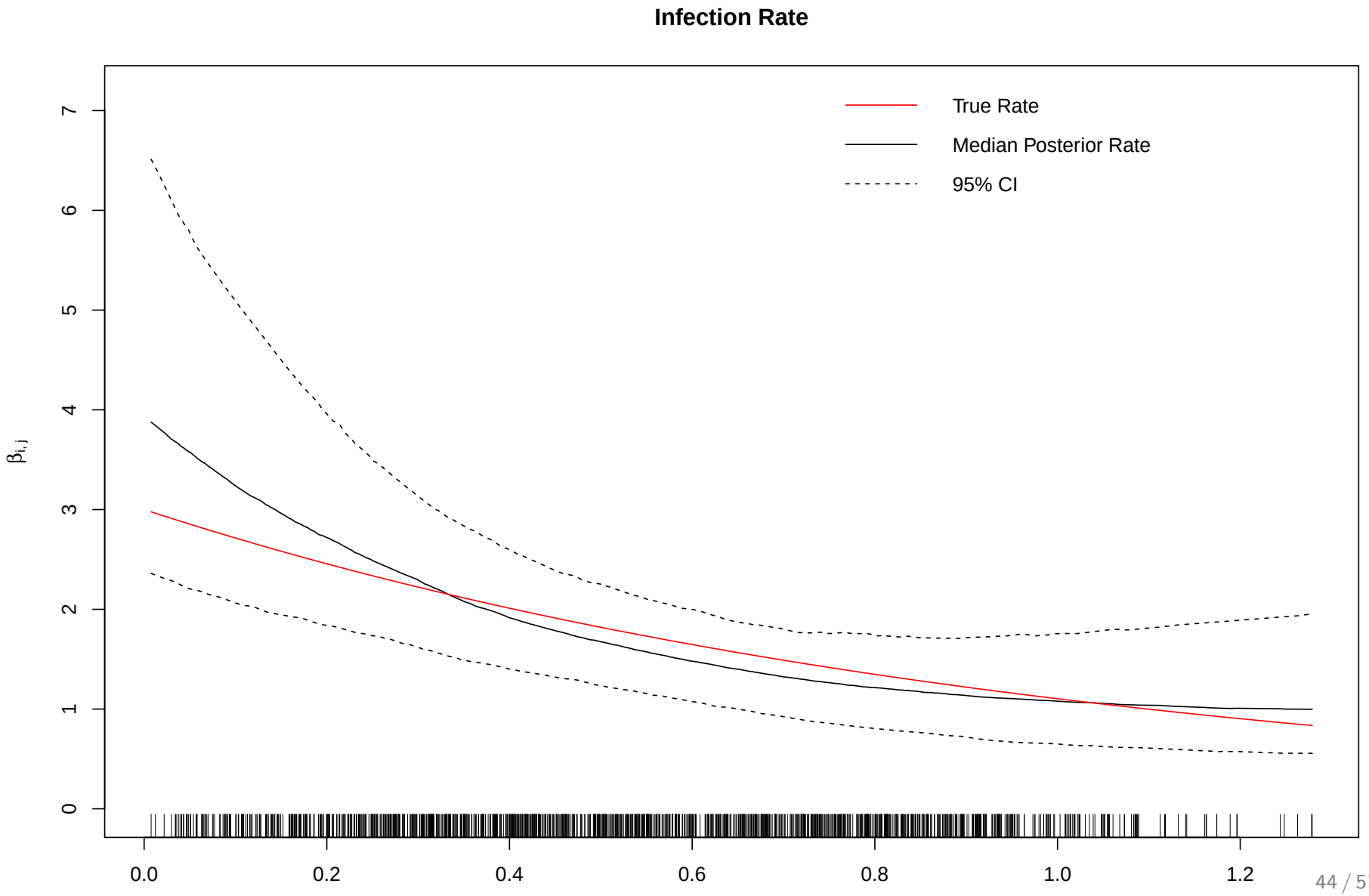
$$\beta_{ij} = \beta_0 \exp\{-\theta \rho_{ij}\}$$

- However, the choice the distance kernel is arbitrary.
- A non-parametric approach instead will assume:

$$\beta_{ij} = h(\rho_{ij})$$

- No free lunch!

Heterogeneously Mixing Populations



Semi-Parametric Models

- Are we completely ignoring key quantities that have been used in epidemic modelling? $(\beta S_t I_t \rightarrow h(t)?)$
  - $\beta S_t I_t \rightarrow \beta\, h_1(S_t) \cdot h_2\,(I_t)$
  - $\beta S_t I_t \rightarrow \beta\, h\,(S_t I_t)$
  - $\beta S_t I_t \rightarrow \beta(t) S_t I_t$; see Xu's thesis (2014).
  - ...
- In principle, the presented methodology could be adapted to fit such models.
- Another option is to use Variational Bayes.
- Connection between the SIR model and the log Gaussian Cox Process (LGCP).

Inference for LGCP

Approximate inference for LGCP can be performed by discretising time into bins and assuming that the width of the bin is small enough that the intensity may be considered constant across each bin.

- $y_i{}_{i=1}^{N} \sim Po(\lambda(t_i)\Delta t)$ where Δt is the width of each bin, and $\lambda(t_i)$ denotes the intensity in bin i.
- $f(t_i)_{i=1}^{N} \sim GP$

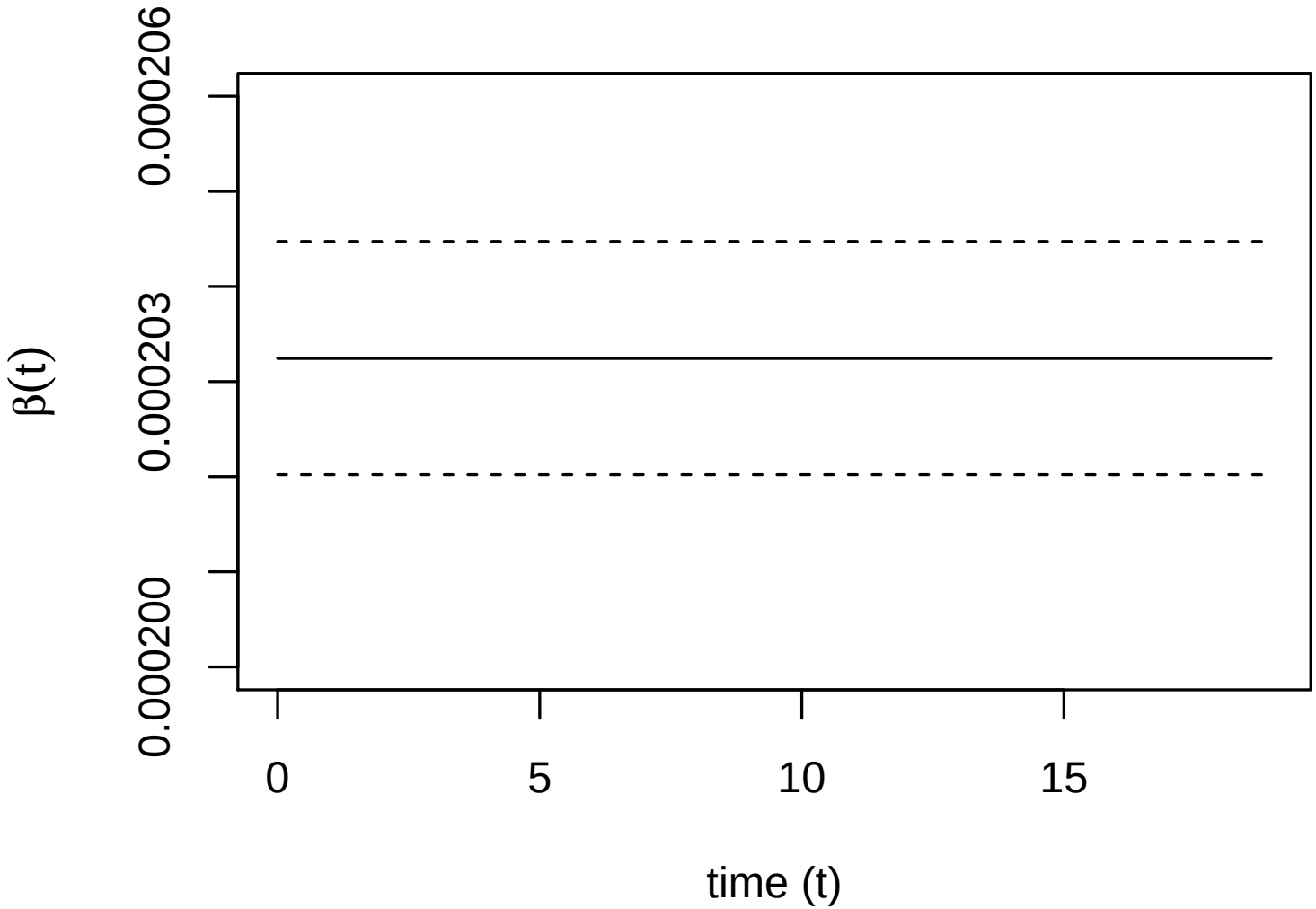
Any MCMC method will be computationally intensive when the dimension of f is large since we have to solve the system of equations which costs $O(N^3)$ operations.

The LGCP exacerbates this problem since we wish to discretise the time into as many bins as possible to improve the accuracy of a (discrete) approximation to the continuous model.

For that reason, we employ a variational Bayes framework to infer the posterior over f.

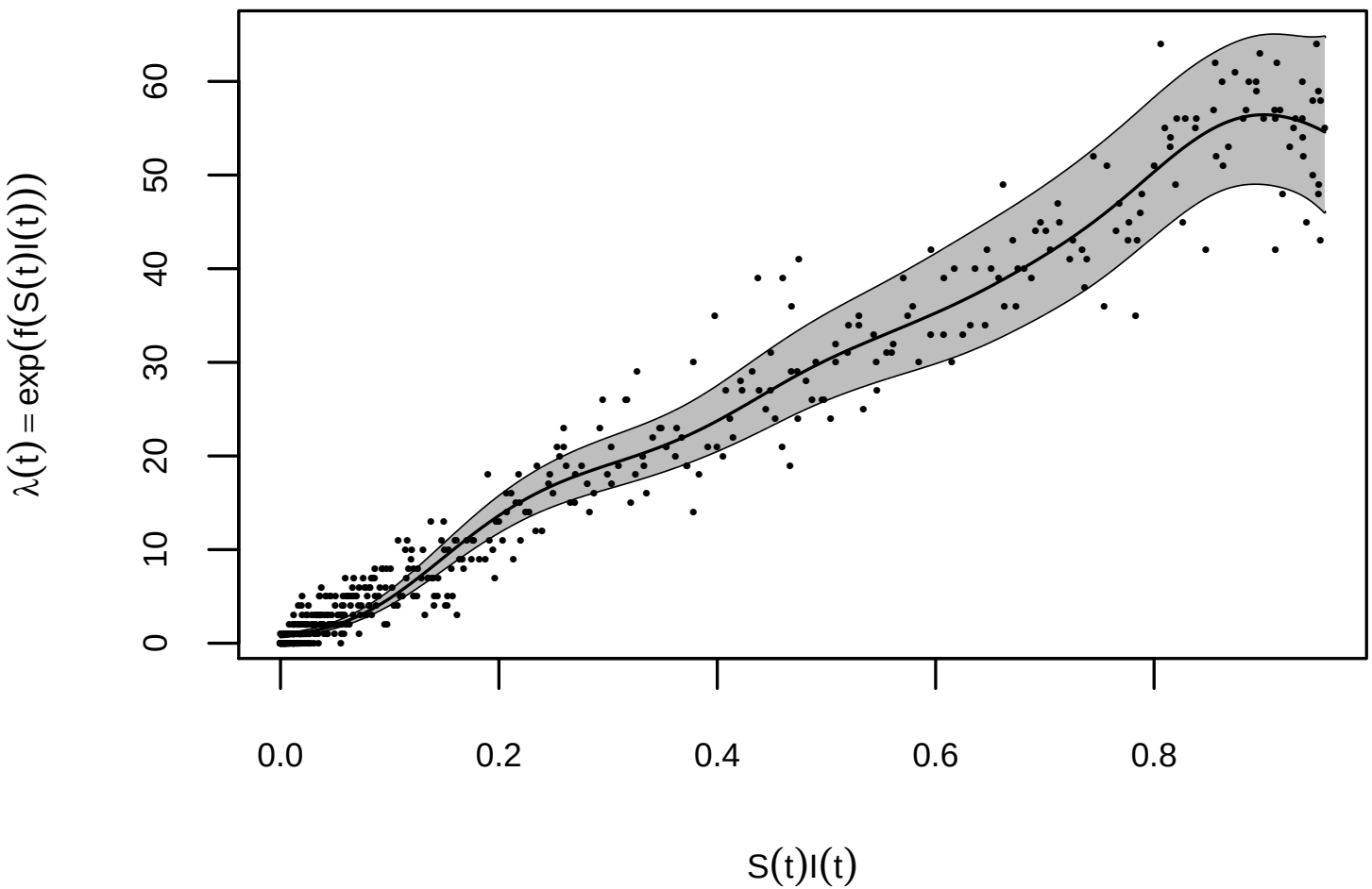
True Model: $\beta S(t)I(t)$

We fit the model: $\lambda(t) = \beta(t)S(t)I(t)$ with $\beta(t) = \exp(f(t))$ and a priori $f(t) \sim GP$.



True Model: $\beta S(t)I(t)$

We fit the model: $\lambda(t) = \exp(f(S(t) \cdot I(t))$ with $f(\cdot) \sim GP$ a priori.



Conclusions & Further Work

- Flexible Models
- Need to be careful with the choice of priors!
- Computational cost can be an issue!
- Model the distribution of the infectious period non parametrically (e.g Dirichlet Mixture Models)?
- Discrete time models (e.g. Reed−Frost type of models, infectiousness); some promising results already.
- Model the population structure non−parametrically (e.g. a prior distribution over networks).

Acknowledgements

- **The team**
  - GPs: Xiaoguang (Allen) Xu and Phil O'Neill @ University of Nottingham
  - GPs & β_{ij}: Rowland Seymour and Phil O'Neill University of Nottingham
  - Splines & Step Functions: Edward Knock @ University of Nottingham
  - VB: James Hensman @ Lancaster University
- **Funding**

References

- Kypraios, T. O'Neill, P.D. (2016) Bayesian Nonparametric Inference for stochastic epidemic models.
- Xu, X., Kypraios, T. and O'Neill, P.D. (2016) Bayesian nonparametric inference for stochastic epidemic models using Gaussian Processes. Biostatistics, doi: 10.1093/biostatistics/kxw011.
- Knock, E. and Kypraios, T. (2016) Bayesian non-parametric inference for infectious disease data. arXiv:1411.2624
- Hensman, J. and Kypraios, T. (2016) Variational Bayesian Non-Parametric Inference for Infectious Disease Models in *Machine Learning for Healthcare*, IET.