Bayesian Nonparametric Inference for Stochastic Epidemic Models

Xiaoguang Xu

Thesis submitted to The University of Nottingham
for the degree of Doctor of Philosophy

May 2015

Corrections

The following errors were discovered after the binding of the printed edition. The corrections are incorporated into the text of the online version only.

p.29. the acceptance ratio is simplified to $\frac{T\times\lambda^*\times\sigma(g(l'))}{M+1}$ should read: the acceptance ratio is simplified to $\frac{T\times\lambda^*\times\sigma(-g(l'))}{M+1}$

p.30. the acceptance ratio is simplified to $\frac{M}{T\times\lambda^*\times\sigma(g(l))}$ should read: the acceptance ratio is simplified to $\frac{M}{T\times\lambda^*\times\sigma(-g(l))}$

p.31.

$$\times\frac{\prod_{k=1}^{K}\sigma(g(s_k))\prod_{m=1}^{M}\sigma(-g(\tilde{s}_m))\times\frac{\sigma(g(l'))}{\sigma(g(l))}\times\pi(\mathbf{g}'_{M+K}|\{s_k\}_{k=1}^{K},\{\tilde{s}_m\}_{m=1}^{M'},\theta)}{\prod_{k=1}^{K}\sigma(g(s_k))\prod_{m=1}^{M}\sigma(-g(\tilde{s}_m))\times\pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^{K},\{\tilde{s}_m\}_{m=1}^{M},\theta)}$$

should read:

$$\times\frac{\prod_{k=1}^{K}\sigma(g(s_k))\prod_{m=1}^{M}\sigma(-g(\tilde{s}_m))\times\frac{\sigma(-g(l'))}{\sigma(-g(l))}\times\pi(\mathbf{g}'_{M+K}|\{s_k\}_{k=1}^{K},\{\tilde{s}_m\}_{m=1}^{M'},\theta)}{\prod_{k=1}^{K}\sigma(g(s_k))\prod_{m=1}^{M}\sigma(-g(\tilde{s}_m))\times\pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^{K},\{\tilde{s}_m\}_{m=1}^{M},\theta)}$$

the acceptance ratio is simplified to $\frac{\sigma(g(l'))}{\sigma(g(l))}$ should read: the acceptance ratio is simplified to $\frac{\sigma(-g(l'))}{\sigma(-g(l))}$

p.38.

$$\pi(\gamma|\boldsymbol{\tau},\mathbf{I},I_1,h(t))\propto\Gamma\left(\nu_\gamma+K,\lambda_\gamma+\int_{I_1}^{T}\gamma Y_s\mathrm{d}s\right)$$

should read:

$$\pi(\gamma|\boldsymbol{\tau},\mathbf{I},I_1,h(t))\propto\Gamma\left(\nu_\gamma+K,\lambda_\gamma+\int_{I_1}^{T}Y_s\mathrm{d}s\right)$$

p.39.

$$\propto\prod_{j=2}^{K}\sigma(g(I_{j^-}))\times\prod_{i=1}^{K}Y_{\tau_{i^-}}\exp\left\{-\int_{I_1}^{T}Y_s\mathrm{d}s\right\}\times\pi(\mathbf{g}_{M+K}|M,\mathbf{I},\{\tilde{I}_s\}_{s=1}^{M},\theta)$$

should read:

$$\propto \prod_{j=2}^{K} \sigma(g(I_{j^-})) \times \prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\left\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\right\} \times \pi(\mathbf{g}_{M+K}|M, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)$$

$$= \frac{\frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I_{j^-})|I_j, \mathbf{g}'_{M+K}, \mathbf{I}', \{\tilde{I}_s\}_{s=1}^{M}, T, \theta)}{\frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I'_{j^-})|I'_j, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, T, \theta)} \times \frac{\prod_{j=2}^{K} \sigma(-g(I'_{j^-}))}{\prod_{j=2}^{K} \sigma(-g(I_{j^-}))}$$
$$\times \frac{\prod_{i=1}^{K} Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} Y'_s \mathrm{d}s\} \times \pi(\mathbf{g}'_{M+K}|\mathbf{I}', \{\tilde{I}_s\}_{s=1}^{M}, \theta)}{\prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.$$

should read:

$$= \frac{\frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I_{j^-})|I_j, \mathbf{g}'_{M+K}, \mathbf{I}', \{\tilde{I}_s\}_{s=1}^{M}, T, \theta)}{\frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I'_{j^-})|I'_j, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, T, \theta)} \times \frac{\prod_{j=2}^{K} \sigma(g(I'_{j^-}))}{\prod_{j=2}^{K} \sigma(g(I_{j^-}))}$$
$$\times \frac{\prod_{i=1}^{K} Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y'_s \mathrm{d}s\} \times \pi(\mathbf{g}'_{M+K}|\mathbf{I}', \{\tilde{I}_s\}_{s=1}^{M}, \theta)}{\prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.$$

p.40. the acceptance ratio is simplified to

$$\frac{\sigma(-g(I'_{j^-}))}{\sigma(-g(I_{j^-}))} \times \frac{\prod_{i=1}^{K} Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} Y'_s \mathrm{d}s\}}{\prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} Y_s \mathrm{d}s\}}.$$

should read: the acceptance ratio is simplified to

$$\frac{\sigma(g(I'_{j^-}))}{\sigma(g(I_{j^-}))} \times \frac{\prod_{i=1}^{K} Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y'_s \mathrm{d}s\}}{\prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\}}.$$

p.41.

$$\frac{(T - I_1) \times h^* \times \sigma(g(\tilde{I}'_{s^-}))}{M+1},$$

and

$$\frac{M}{(T - I_1) \times \beta^* \times \sigma(g(\tilde{I}_{s^-}))}.$$

should read:

$$\frac{(T - I_1) \times h^* \times \sigma(-g(\tilde{I}'_{s^-}))}{M+1},$$

and

$$\frac{M}{(T - I_1) \times \beta^* \times \sigma(-g(\tilde{I}_{s^-}))}.$$

The Metropolis-Hastings acceptance ratio is $\frac{\sigma(g(\tilde{I}'_{s^-}))}{\sigma(g(\tilde{I}_{s^-}))}$ should read: The Metropolis-Hastings acceptance ratio is $\frac{\sigma(-g(\tilde{I}'_{s^-}))}{\sigma(-g(\tilde{I}_{s^-}))}$

p.62. we have

$$\pi(\gamma|\boldsymbol{\tau}, \mathbf{I}, I_1, \beta(t)) \propto \Gamma\left(\nu_\gamma + K, \lambda_\gamma + \int_{I_1}^{T} \gamma Y_s \mathrm{d}s\right).$$

should read: we have

$$\pi(\gamma|\boldsymbol{\tau}, \mathbf{I}, I_1, \beta(t)) \propto \Gamma\left(\nu_\gamma + K, \lambda_\gamma + \int_{I_1}^{T} Y_s \mathrm{d}s\right).$$

p.64.

$$\begin{aligned}
&\pi(\mathbf{I}, \mathbf{g}_K|\boldsymbol{\tau}, M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M, \tilde{\beta}^*, \gamma, T, \theta) \\
&\propto \prod_{j=2}^{K} X_{I_{j^-}} Y_{I_{j^-}} \prod_{j=2}^{K} \sigma(g(I_{j^-})) \exp\{-\int_{I_1}^{T} X_s Y_s \mathrm{d}s\} \\
&\times \prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K}|M, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta).
\end{aligned}$$

should read:

$$\begin{aligned}
&\pi(\mathbf{I}, \mathbf{g}_K|\boldsymbol{\tau}, M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M, \tilde{\beta}^*, \gamma, T, \theta) \\
&\propto \prod_{j=2}^{K} X_{I_{j^-}} Y_{I_{j^-}} \prod_{j=2}^{K} \sigma(g(I_{j^-})) \exp\{-\int_{I_1}^{T} \tilde{\beta}^* X_s Y_s \mathrm{d}s\} \\
&\times \prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K}|M, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta).
\end{aligned}$$

p.65.

$$\begin{aligned}
&\times \frac{\prod_{j=2}^{K} X'_{I_{j^-}} Y'_{I_{j^-}} \times \prod_{j=2}^{K} \sigma(-g(I'_j)) \exp\{-\int_{I_1}^{T} X'_s Y'_s \mathrm{d}s\}}{\prod_{j=2}^{K} X_{I_{j^-}} Y_{I_{j^-}} \prod_{j=2}^{K} \sigma(-g(I_j)) \exp\{-\int_{I_1}^{T} X_s Y_s \mathrm{d}s\}} \\
&\times \frac{\prod_{i=1}^{K} Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} Y'_s \mathrm{d}s\} \times \pi(\mathbf{g}'_{M+K}|\{I'_j\}_{j=2}^{K}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}{\prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.
\end{aligned}$$

should read:

$$\begin{aligned}
&\times \frac{\prod_{j=2}^{K} X'_{I_{j^-}} Y'_{I_{j^-}} \times \prod_{j=2}^{K} \sigma(g(I'_j)) \exp\{-\int_{I_1}^{T} \tilde{\beta}^* X'_s Y'_s \mathrm{d}s\}}{\prod_{j=2}^{K} X_{I_{j^-}} Y_{I_{j^-}} \prod_{j=2}^{K} \sigma(g(I_j)) \exp\{-\int_{I_1}^{T} \tilde{\beta}^* X_s Y_s \mathrm{d}s\}} \\
&\times \frac{\prod_{i=1}^{K} Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y'_s \mathrm{d}s\} \times \pi(\mathbf{g}'_{M+K}|\{I'_j\}_{j=2}^{K}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}{\prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.
\end{aligned}$$

the acceptance ratio is simplified to

$$\frac{\prod_{j=2}^{K} X'_{I_{j^-}} Y'_{I_{j^-}} \times \sigma(-g(I'_j)) \exp\{-\int_{I_1}^{T} X'_s Y'_s \mathrm{d}s\}}{\prod_{j=2}^{K} X_{I_{j^-}} Y_{I_{j^-}} \sigma(-g(I_j)) \exp\{-\int_{I_1}^{T} X_s Y_s \mathrm{d}s\}} \times \frac{\prod_{i=1}^{K} Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} Y'_s \mathrm{d}s\}}{\prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} Y_s \mathrm{d}s\}}.$$

should read: the acceptance ratio is simplified to

$$\frac{\prod_{j=2}^{K} X'_{I_{j^-}} Y'_{I_{j^-}} \times \sigma(g(I'_j)) \exp\{-\int_{I_1}^{T} \tilde{\beta}^* X'_s Y'_s \mathrm{d}s\}}{\prod_{j=2}^{K} X_{I_{j^-}} Y_{I_{j^-}} \sigma(g(I_j)) \exp\{-\int_{I_1}^{T} \tilde{\beta}^* X_s Y_s \mathrm{d}s\}} \times \frac{\prod_{i=1}^{K} Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y'_s \mathrm{d}s\}}{\prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\}}.$$

p.66. The acceptance ratio is

$$\begin{aligned}
&\frac{q(M' \to M)}{q(M \to M')} \times \frac{\pi(M', \{\tilde{I}_s\}_{m=1}^{M'}, \mathbf{g}'_M | \mathbf{I}, \mathbf{g}_K, \theta, \tilde{\beta}^*)}{\pi(M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M | \mathbf{I}, \mathbf{g}_K, \theta, \tilde{\beta}^*)} \\
&= \frac{\frac{1}{2} \times \frac{1}{M'}}{\frac{1}{2} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(\tilde{I}'_{s^-}) | \tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)} \\
&\times \frac{(\tilde{\beta}^*)^{M'} \prod_{s=1}^{M'} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{m=1}^{M'} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}'_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{m=1}^{M'}, \theta)}{(\tilde{\beta}^*)^{M} \prod_{s=1}^{M} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{m=1}^{M} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.
\end{aligned}$$

should read: The acceptance ratio is

$$\begin{aligned}
&\frac{q(M' \to M)}{q(M \to M')} \times \frac{\pi(M', \{\tilde{I}_s\}_{s=1}^{M'}, \mathbf{g}'_M | \mathbf{I}, \mathbf{g}_K, \theta, \tilde{\beta}^*)}{\pi(M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M | \mathbf{I}, \mathbf{g}_K, \theta, \tilde{\beta}^*)} \\
&= \frac{\frac{1}{2} \times \frac{1}{M'}}{\frac{1}{2} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(\tilde{I}'_{s^-}) | \tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)} \\
&\times \frac{(\tilde{\beta}^*)^{M'} \prod_{s=1}^{M'} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^{M'} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}'_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{(\tilde{\beta}^*)^{M} \prod_{s=1}^{M} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^{M} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.
\end{aligned}$$

p.67.

$$\begin{aligned}
&\pi(\mathbf{g}'_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{m=1}^{M'}, \theta) \\
&= \pi(\mathbf{g}_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta) \times \pi(\mathbf{g}(\tilde{I}'_{s^-}) | \tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta),
\end{aligned}$$

the acceptance ratio is simplified to $\frac{(T-I_1) \times \tilde{\beta}^* \times \sigma(g(\tilde{I}'_{s^-})) \times X_{\tilde{I}'_{s^-}} Y_{\tilde{I}'_{s^-}}}{M+1}$ should read:

$$\begin{aligned}
&\pi(\mathbf{g}'_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta) \\
&= \pi(\mathbf{g}_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta) \times \pi(\mathbf{g}(\tilde{I}'_{s^-}) | \tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta),
\end{aligned}$$

the acceptance ratio is simplified to $\frac{(T-I_1) \times \tilde{\beta}^* \times \sigma(-g(\tilde{I}'_{s^-})) \times X_{\tilde{I}'_{s^-}} Y_{\tilde{I}'_{s^-}}}{M+1}$

$$q(M'\to M) = \frac{1}{2} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{m=1}^{M'}, \theta).$$

should read:

$$q(M'\to M) = \frac{1}{2} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta).$$

The acceptance ratio is

$$\begin{aligned}
&\frac{q(M'\to M)}{q(M\to M')} \times \frac{\pi(M', \{\tilde{I}_s\}_{m=1}^{M'}, \mathbf{g}'_M|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_K, \theta, \tilde{\beta}^*)}{\pi(M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \mathbf{g}_K, \theta, \tilde{\beta}^*)} \\
&= \frac{\frac{1}{2} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{m=1}^{M'}, \theta)}{\frac{1}{2} \times \frac{1}{M}} \\
&\times \frac{(\tilde{\beta}^*)^{M'} \prod_{s=1}^{M'} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{m=1}^{M'} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{m=1}^{M'}, \theta)}{(\tilde{\beta}^*)^{M} \prod_{s=1}^{M} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{m=1}^{M} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.
\end{aligned}$$

should read: The acceptance ratio is

$$\begin{aligned}
&\frac{q(M'\to M)}{q(M\to M')} \times \frac{\pi(M', \{\tilde{I}_s\}_{s=1}^{M'}, \mathbf{g}'_M|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_K, \theta, \tilde{\beta}^*)}{\pi(M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \mathbf{g}_K, \theta, \tilde{\beta}^*)} \\
&= \frac{\frac{1}{2} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{\frac{1}{2} \times \frac{1}{M}} \\
&\times \frac{(\tilde{\beta}^*)^{M'} \prod_{s=1}^{M'} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^{M'} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{(\tilde{\beta}^*)^{M} \prod_{s=1}^{M} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^{M} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.
\end{aligned}$$

$$= \pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{m=1}^{M'}, \theta) \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{m=1}^{M'}, \theta),$$

the acceptance ratio is simplified to $\frac{M}{(T-I_1)\times\tilde{\beta}^*\times\sigma(g(\tilde{I}_{s^-}))\times X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}}}$ should read:

$$= \pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta) \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta),$$

the acceptance ratio is simplified to $\frac{M}{(T-I_1)\times\tilde{\beta}^*\times\sigma(-g(\tilde{I}_{s^-}))\times X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}}}$

p.68.

$$\times \frac{\prod_{s=1}^{M-1} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \times \frac{X_{\tilde{I}'_{s^-}} Y_{\tilde{I}'_{s^-}}}{X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}}} \prod_{m=1}^{M} \sigma(-g(\tilde{I}_s)) \times \frac{\sigma(g(\tilde{I}'_s))}{\sigma(g(\tilde{I}_s))} \times \pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{\prod_{s=1}^{M-1} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{m=1}^{M} \sigma(-g(\tilde{I}_s)) \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.$$

should read:

$$\times \frac{\prod_{s=1}^{M-1} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \times \frac{X_{\tilde{I}'_{s^-}} Y_{\tilde{I}'_{s^-}}}{X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}}} \prod_{s=1}^{M} \sigma(-g(\tilde{I}_s)) \times \frac{\sigma(-g(\tilde{I}'_s))}{\sigma(-g(\tilde{I}_s))} \times \pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{\prod_{s=1}^{M-1} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^{M} \sigma(-g(\tilde{I}_s)) \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.$$

the acceptance ratio is simplified to $\frac{X_{\tilde{I}'_{s^-}} Y_{\tilde{I}'_{s^-}} \times \sigma(g(\tilde{I}'_{s^-}))}{X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \times \sigma(g(\tilde{I}_{s^-}))}$ should read: the acceptance ratio is simplified to $\frac{X_{\tilde{I}'_{s^-}} Y_{\tilde{I}'_{s^-}} \times \sigma(-g(\tilde{I}'_{s^-}))}{X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \times \sigma(-g(\tilde{I}_{s^-}))}$

p.89.

$$\pi(\gamma|\tilde{\boldsymbol{\beta}}(t), \boldsymbol{\tau}, \underset{\sim}{\mathbf{I}}) \propto \Gamma\left(\nu_\gamma + \sum_{i=1}^{k} n_i, \lambda_\gamma + \sum_{i=1}^{k} \int_{I_{i1}}^{T} \gamma Y_i(s) \mathrm{d}s\right).$$

should read:

$$\pi(\gamma|\tilde{\boldsymbol{\beta}}(t), \boldsymbol{\tau}, \underset{\sim}{\mathbf{I}}) \propto \Gamma\left(\nu_\gamma + \sum_{i=1}^{k} n_i, \lambda_\gamma + \sum_{i=1}^{k} \int_{I_{i1}}^{T} Y_i(s) \mathrm{d}s\right).$$

Abstract

Modelling of infectious diseases is a topic of great importance. Despite the enormous attention given to the development of methods for efficient parameter estimation, there has been relatively little activity in the area of nonparametric inference for epidemics. In this thesis, we develop new methodology which enables nonparametric estimation of the parameters which govern transmission within a Bayesian framework. Many standard modelling and data analysis methods use underlying assumptions (e.g. concerning the rate at which new cases of disease will occur) which are rarely challenged or tested in practice. we relax these assumptions and analyse data from disease outbreaks in a Bayesian nonparametric framework.

We first apply our Bayesian nonparametric methods to small-scale epidemics. In a standard SIR model, the overall force of infection is assumed to have a parametric form. We relax this assumption and treat it as a function which only depends on time. Then we place a Gaussian process prior on it and infer it using data-augmented Markov Chain Monte Carlo (MCMC) algorithms. Our methods are illustrated by applications to simulated data as well as Smallpox data. We also investigate the infection rate in the SIR model using our methods. More precisely, we assume the infection rate is time-varying and place a Gaussian process prior on it. Results are obtained using data augmentation methods and standard MCMC algorithms. We illustrate our methods using simulated data and respiratory disease data. We find our methods work fairly well for the stochastic SIR model.

We also investigate large-scaled epidemics in a Bayesian nonparametric framework. For large epidemics in large populations, we usually observe surveillance data which typically provide number of new infection cases occurring during observation periods. We infer the infection rate for each observation pe-

riod by placing Gaussian process priors on them. Our methods are illustrated by the real data, i.e. a time series of incidence of measles in London (1948-1957).

Acknowledgements

I feel very fortunate to have Philip O'Neill and Theodore Kypraios as my supervisors. They were always there for me and I am very grateful for all their guidance, advice and patience during my PhD studies.

Many thanks to my parents for all their support and patience.

I would also like to thank my friends in Nottingham for making my life more enjoyable.

Many thanks to the University of Nottingham School of Mathematical Sciences who fully funded my PhD study.

Finally, there is nothing I could do to sufficiently express my gratitude to Lin. She has been very supportive and understanding.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

Understanding the spread of communicable infectious diseases is of great importance in order to prevent major future outbreaks and therefore it remains high on the global scientific agenda. Mathematical and statistical modelling has become a valuable tool and has been widely used in the analysis of infectious disease dynamics. It is of interest to make statistical inference for the parameters of stochastic epidemic models given observed data. This is not a standard problem due to the fact that, in general, the underlying transmission process is partially observed (e.g. infection times are not observed). To date, almost all of the literature concerning statistical inference for epidemic models adopts an approach based on a parametric framework, in which a model is a family of distributions that can be described using a finite number of parameters. O'Neill & Roberts (1999) and Gibson & Renshaw (1998) present the first Bayesian approach using Markov Chain Monte Carlo (MCMC) methods for the so-called SIR (susceptible-infective-removed) stochastic epidemic model, given temporal data on case diagnosis times, in which infection times are not observed and are treated as unknown model parameters. O'Neill & Becker (2001) extend the techniques and develop an MCMC algorithm for performing Bayesian inference for a non-Markovian SIR model where the infectious period follows a Gamma distribution. O'Neill et al. (2000) develop MCMC algorithms to analyse both temporal and final size data from household outbreaks. Numerous other papers have utilised MCMC algorithms to analyse infectious disease data,

including Neal & Roberts, (2005), Lekone & Finkenstädt, (2006), Cauchemez & Ferguson (2008), Jewell et al., (2009) and McKinley et al., (2009). In general terms, inference for stochastic epidemic models has benefited considerably from the use of MCMC methods.

Despite the enormous attention given to the development of methods for efficient parameter estimation, there has been relatively little activity in the area of nonparametric inference for epidemics. Becker & Yip (1989) and Becker (1989) consider nonparametric estimation of the infection rate in an SIR model, specifically allowing the infection rate to depend on time. They applied martingale approaches to temporal epidemic data where the initial state and final state of the epidemic are assumed to be known. Chen et al. (2008) use classical nonparametric methods to estimate the infection rate of an epidemic model over time. They use local polynomial methods and martingale estimating equations to develop a closed-form estimator of the intensity function and its derivatives for multiplicative counting process models. They also apply the proposed estimators to analyse the infection rate of the 2003 SARS epidemic in Beijing, China. Cauchemez & Ferguson (2008) point out that the martingale approaches provided simple but efficient ways to estimate key quantities and approximate confidence regions for the parameters. However, it would be difficult when more complex situations are considered such as (i) the initial state of the epidemic is unknown, (ii) observed data are further aggregated temporally, and (iii) under-reporting, seasonality in transmission rates are what one must account for. Therefore, martingale approaches appear limited in terms of their range of epidemic applications. Lau & Yip (2008) adopt a nonparametric kernel estimator to reconstruct the infection process given that the removal process is observed. Hence, they estimate the basic reproduction number, R_0, with an unknown initial number of susceptibles and the methods are illustrated by an application to data from a smallpox epidemic. However, the person-to-person infection rate was assumed constant, which implied that the method may not be suitable if the infection rate varies throughout the epidemic.

Recently, Dureau et al. (2013) developed stochastic extensions to the deterministic SEIR epidemic model and assigned integrated Brownian motion for the time-varying effective contact rate (where an SEIR model, a variant of the standard SIR model, is characterised by four states: susceptible (S), exposed (E),

infected (I), removed (R) and it is assumed that there exists a latent period between time of infection and time of infectiousness during which individuals are in the 'exposed' state). Under a Bayesian framework, they adopted a particle Markov Chain Monte Carlo algorithm to implement inference (Persing, 2014). The performance of the proposed computational methods was validated on simulated data and applied to the 2009 A/H1N1 pandemic in England. However, the approach they used is not fully nonparametric as the time-varying infection rate is assumed to be governed by a diffusion process and is estimated with a parametric approach.

The motivation behind this thesis is to develop new methodology which enables nonparametric estimation of the parameters which govern transmission within a Bayesian framework. In the following sections we recall some key concepts that will be relevant.

1.2 Bayesian inference

In this section, we will review the fundamentals of Bayesian theory. For detailed discussions of the theory see Bernardo & Smith (1994).

1.2.1 Bayes' theorem

In a Bayesian framework, parameters are viewed as having probability distributions rather than fixed values, as is the case in Frequentist inference (Moyé, 2008). Bayesian inference is based around Bayes' theorem, which, for observed data $\mathbf{X}$ and model parameters $\boldsymbol{\theta}$, is

$$\pi(\boldsymbol{\theta}|\mathbf{X}) = \frac{\pi(\mathbf{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{X})} \propto \pi(\mathbf{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{1.1}$$

where $\pi(\mathbf{X}) = \sum_{\boldsymbol{\theta}} \pi(\mathbf{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ for the discrete case and the sum is all over possible values of $\boldsymbol{\theta}$ or $\pi(\mathbf{X}) = \int_{\boldsymbol{\theta}} \pi(\mathbf{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ in the case of continuous $\boldsymbol{\theta}$. In Bayes' theorem (1.1), $\pi(\boldsymbol{\theta}|\mathbf{X})$ is the posterior density of $\boldsymbol{\theta}$, $\pi(\mathbf{X}|\boldsymbol{\theta})$ is the likelihood of $\mathbf{X}$, and $\pi(\boldsymbol{\theta})$ is the prior density of $\boldsymbol{\theta}$.

In order to infer $\boldsymbol{\theta}$, we need to determine the posterior distribution, $\pi(\boldsymbol{\theta}|\mathbf{X})$,

which is the combination of the likelihood, $\pi(\mathbf{X}|\boldsymbol{\theta})$, and the prior, $\pi(\boldsymbol{\theta})$. In Bayesian inference, it is essential to specify prior distributions for model parameters, $\boldsymbol{\theta}$, as poorly chosen prior distributions can significantly affect the posterior estimates. We may choose certain prior distributions to reflect the existing knowledge about the parameters in question. Whereas, uninformative prior distributions may be chosen due to lack of existing evidence for parameter values, or a desire to derive estimates based solely on the observed data.

We aim to evaluate the posterior distribution of parameters, $\boldsymbol{\theta}$. However, the integral, $\pi(\mathbf{X}) = \int_{\boldsymbol{\theta}} \pi(\mathbf{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$, is often intractable (Gilks et al., 1996). With the development of computer technology towards the end of last century, Markov Chain Monte Carlo (MCMC) methods are widely used to tackle this problems.

1.2.2 Markov Chain Monte Carlo

MCMC methods are used to draw samples $\boldsymbol{\theta}$ while exploring the state space $\mathbf{X}$ using a Markov chain mechanism. This mechanism is constructed so that the samples $\boldsymbol{\theta}$ mimic samples drawn from the target distribution $\pi(\boldsymbol{\theta}|\mathbf{X})$. We use MCMC methods when we cannot draw samples from $\pi(\boldsymbol{\theta}|\mathbf{X})$ directly, but can evaluate $\pi(\boldsymbol{\theta}|\mathbf{X})$ up to a normalising constant (Andrieu et al., 2003).

In this section we will review some well studied MCMC algorithms. Detailed discussions about the theory and applications can be found in Gilks et al. (1996), Tanner (1996) and Robert & Casella (1999). MCMC had its origins in the 1950s (Metropolis et al., 1953) and was then generalised by Hastings (Hastings, 1970) to yield the so-called Metropolis-Hastings algorithm.

1.2.2.1 The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm manages to generate samples from a posterior density, $\pi(\boldsymbol{\theta}|\mathbf{X})$, in a way which does not require the computation of its normalisation constant. In each iteration t, the next state $\boldsymbol{\theta}^{(t+1)}$ is chosen as follows. First, a candidate point, $\boldsymbol{\theta}'$, is sampled from a proposal density, $q(\cdot|\boldsymbol{\theta}^{(t)})$. Then $\boldsymbol{\theta}'$ is accepted with probability $\phi(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}')$ where

$$\phi(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}') = \min\left(1, \frac{\pi(\boldsymbol{\theta}'|\mathbf{X})q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(t)}|\mathbf{X})q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(t)})}\right).$$

The next state becomes $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}'$ if the candidate point is accepted, otherwise $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$ which implies that the chain does not move. The algorithm is implemented as follows:

Algorithm The Metropolis Hastings Algorithm

1. Initialise $\boldsymbol{\theta}^{(0)}$
2. Set $t = 0$
3. Repeat the following steps:
   - Sample $\boldsymbol{\theta}' \propto q(\cdot|\boldsymbol{\theta}^{(t)})$
   - Sample $\mathbf{U} \propto U(0,1)$
   - If $\mathbf{U} \leq \phi(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}')$
     - set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}'$
   - Else
     - set $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$
   - $t = t + 1$

1.2.2.2 Gibbs sampler

The parameters of the vector, $\boldsymbol{\theta}$, in the Metropolis-Hastings algorithm can be updated simultaneously, or repeated for individual elements, or groups of elements. The Gibbs sampler, introduced by Geman & Geman (1984), is a special case of the Metropolis-Hastings algorithm where the proposed candidate is always accepted (Casella & George, 1992, Gelfand & Smith, 1990). Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_n)$ be the current parameter vector consisting of n parameters. Let $\boldsymbol{\theta}_{-i} = (\theta_1, \theta_2, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_n)$ be the vector, $\boldsymbol{\theta}$, with θ_i removed, where $i = 1, 2, \cdots, n$. In each iteration of the Gibbs sampler, each θ_i of $\boldsymbol{\theta}$ is drawn conditional on the value of all the others. The functions $\pi_i(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{X})$ are called the full conditional distributions of $\pi(\boldsymbol{\theta}|\mathbf{X})$. Therefore, for the Gibbs sampler, in each iteration t, we sample from the full conditional distributions, $\pi_i(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{X})$. Each of the parameters, θ_i, is conditional on the latest values of the elements of $\boldsymbol{\theta}$.

The algorithm is implemented as follows:

Algorithm Gibbs Sampler

1. Initialise $\boldsymbol{\theta}^{(0)}$
2. Set $t = 0$
3. Repeat the following steps:

   Draw $\theta_1^{(t+1)}$ from $\pi(\theta_1|\theta_2^{(t)}, \theta_3^{(t)}, \cdots, \theta_n^{(t)})$

   Draw $\theta_2^{(t+1)}$ from $\pi(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \cdots, \theta_n^{(t)})$

   Draw $\theta_3^{(t+1)}$ from $\pi(\theta_3|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_4^{(t)}, \cdots, \theta_n^{(t)})$

   .

   .

   .

   Draw $\theta_n^{(t+1)}$ from $\pi(\theta_n|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \cdots, \theta_{n-1}^{(t+1)})$

   $t = t + 1$

1.2.2.3 Burn-in

The term burn-in is referred to as the practice of discarding the early iterations of MCMC diminish the effect of the initial distribution (Gelman et al., 2004). The length of the burn-in usually depends on the initial values and how fast the MCMC chain converges to the desired distribution. We can determine burn-in from MCMC trace plots (Gilks et al., 1996).

1.2.2.4 Thinning

It is fairly common practice to "thin" the Markov chain when computer storage is a problem or if one wishes to reduce autocorrelation. We can keep only every k^{th} simulation draw from each sequence and discard the rest of them. This practice is often termed thinning. we can set k to some value high enough that successive draws of parameter $\boldsymbol{\theta}$ are approximately independent (Gilks et al., 1996).

1.2.2.5 Proposal distributions and convergence

In the Metropolis-Hastings algorithm, the choice of proposal density $q(\cdot)$ is of great importance and affects the rate at which the Markov chain converges to the desired distribution, as well as the rate at which the distribution is explored.

We commonly propose samples from a Gaussian distribution. This proposal distribution is a Normal distribution at the current parameter value and the only thing that needs to be specified is the variance σ^2. If σ^2 is too small, the proposal is likely to be very close to the current value. Then the acceptance rate is likely to be high but the Markov chain will take longer to converge to the desired distribution. Conversely, if σ^2 is too large, the proposal will be rejected almost every time as the acceptance rate is likely to be low. Therefore, it is implied that a careful choice of σ^2 is important for efficient sampling although the Metropolis-Hastings algorithm will converge for any proposal distribution. A general rule is to choose σ^2 in a way such that the acceptance rate for each proposal is 0.234 (Roberts et al., 1997).

It is important to ensure the MCMC algorithm has converged to the desired distribution. A simple test is to run several independent MCMC chains with different initial points, in order to ensure they appear to sample from the same distribution after a burn-in period. Gelman & Rubin (1992) and Geweke (1992) proposed more formal methods to test if the chain converges.

1.3 Epidemic modelling

Modelling of infectious diseases is a topic of enormous importance and many governments now routinely use modelling as a way of helping to formulate public health policy, e.g. how best to prepare for a possible influenza pandemic; how best to design a programme of vaccination; how best to react to an emerging disease like SARS, etc (Tildesley & Ryan, 2012, Balatif, 2014). In general, mathematical and statistical modelling is playing a fundamental role in the fight against the spread of disease.

1.3.1 Stochastic SIR model

We now describe the most well studied stochastic model for the transmission of infectious diseases (Bailey, 1975, Chapter 6, Andersson & Britton, 2000, Chapter 2). Consider a closed, homogeneous and homogeneously mixing population of individuals subdivided at any time $t \geq 0$ into three states: susceptible, infective and removed. Let X_t, Y_t and Z_t denote respectively the number of

individuals who are susceptible, infective, and removed at time t, and let N and a denote the initial number of susceptible and infective individuals respectively. We assume that there are no removed individual initially. An individual makes contacts with each of the N susceptibles at times given by the points of a homogeneous Poisson process with rate β. The infectious periods of different individuals are assumed to be independent and identically distributed according to the distribution of a random variable which can have any arbitrary but specified distribution. Susceptible individuals become infective immediately after they contact an infective. In other words, at time t, new infections occur according to an inhomogeneous Poisson process with rate $\beta X_t Y_t$. In epidemics, we refer to the rate, $\beta X_t Y_t$, as the overall force of infection. It is assumed that there is no latent period (infected but not infectious) for the newly infected individual. The infectious individual is removed from the infection process once its infectious period ends. A removed individual can be dead, or recovered and immune to further infections and then would not play any further part in the epidemic. The epidemic will die out as soon as there are no infective individuals present in the population.

Epidemic models like this are often called SIR epidemic models. The Markovian case where the infectious period follows an exponential distribution is the-so-called *general stochastic epidemic* (GSE). Lloyd (2001) pointed out that this assumption is equivalent to assuming that it is constant for the chance of recovery within a given time interval, regardless of the time since infection and it is unrealistic epidemiologically as the chance of recovery in a given time interval is initially small but increases over time, corresponding to the infectious period distribution. In this thesis, we always assume that the SIR model is Markovian for simplicity and comparison with other methods. However, our methods can also be applied to general infectious period distributions.

The epidemic can be defined according to the following transition probabilities, where the first two transitions correspond respectively to an infection and a removal:

$$P[X_{t+\delta t} = X_t - 1, Y_{t+\delta t} = Y_t + 1 | \mathcal{H}_t] = \beta X_t Y_t \delta t + o(\delta t),$$

$$P[X_{t+\delta t} = X_t, Y_{t+\delta t} = Y_t - 1 | \mathcal{H}_t] = \gamma Y_t \delta t + o(\delta t),$$

$$P[X_{t+\delta t} = X_t, Y_{t+\delta t} = Y_t | \mathcal{H}_t] = 1 - \beta X_t Y_t \delta t - \gamma Y_t \delta t + o(\delta t),$$

where β, a constant, is referred to as the infection rate, γ, a constant, is referred to as the removal rate and $\mathcal{H}_t$ is the sigma-algebra generated by the history of the process up to time t, i.e. $\mathcal{H}_t = \sigma\{(X_u, Y_u) : 0 \leq s \leq t\}$, with $\mathcal{H}_0 = \sigma\{(X_0 = N, Y_0 = a) : 0 \leq s \leq t\}$ specifying the initial conditions.

1.3.2 Bayesian inference for the SIR model from partially observed data

As mentioned in Section 1.1, O'Neill & Roberts (1999) present the first Bayesian approach using MCMC methods for the SIR model given the removal times. We now give a brief description of the algorithms they developed for sampling the infection rate, β, removal rate, γ and other parameters from the desired posterior distribution. We assume that the epidemic is known to have ceased, i.e. the number of infection times equals to the number of removal times. We first give a list of notation that we use:

X_t and Y_t: the number of susceptibles and infectives respectively in the population at time $t \geq 0$

$\boldsymbol{\tau} = (\tau_1 = 0, \tau_2, \cdots, \tau_n)$: the observed (ordered) removal times where $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_n$

I_1: the first infection time

$\mathbf{I} = (I_2, I_3, \cdots, I_n)$: the unobserved (ordered) infection times excluding the first infection time during $(I_1, T]$, where $T \geq \tau_n$ and $I_2 \leq I_3 \leq \cdots \leq I_n$.

Suppose that, a priori, Gamma distributions with shape and rate parameters, ν_β and λ_β and ν_γ and λ_γ are put on β and γ respectively. Specifically, the prior densities of β and γ are given by

$$g(x; \nu_\beta, \lambda_\beta) = \frac{x^{\nu_\beta - 1} \exp(-x\lambda_\beta)}{\lambda_\beta^{-\nu_\beta} \Gamma(\nu_\beta)} \quad \text{for } x \geq 0 \text{ and } \nu_\beta, \lambda_\beta > 0,$$

$$g(x; \nu_\gamma, \lambda_\gamma) = \frac{x^{\nu_\gamma - 1} \exp(-x\lambda_\gamma)}{\lambda_\gamma^{-\nu_\gamma} \Gamma(\nu_\gamma)} \quad \text{for } x \geq 0 \text{ and } \nu_\gamma, \lambda_\gamma > 0.$$

We also suppose that I_1 has prior density given by $\delta \exp(\delta y) I(y \leq 0)$, where $\delta > 0$, and where $I(\cdot)$ is the indicator function. Conditionally on $\boldsymbol{\tau}$, $\mathbf{I}$, β and γ,

the density of I_1 is as follows:

$$\pi(y|\boldsymbol{\tau}, \mathbf{I}, \beta, \gamma) = \Lambda \exp\{-\Lambda(I_2 - y)\},\ y \in (-\infty, I_2),$$

where $\Lambda = \delta + \gamma + \beta * N$. According to O'Neill & Roberts (1999), the MCMC algorithm is implemented as follow:

Algorithm MCMC algorithm

1. Initialise β^0, γ^0, I_1^0 and $\mathbf{I}^0$
2. Sample β by using a Gibbs step and drawing from
   $\pi(\beta|\boldsymbol{\tau}, \mathbf{I}, I_1, \gamma) \equiv \Gamma\left(n - 1 + \nu_\beta, \lambda_\beta + \int_{I_1}^{T} X_t Y_t \mathrm{d}t\right)$
3. Sample γ by using a Gibbs step and drawing from
   $\pi(\gamma|\boldsymbol{\tau}, \mathbf{I}, I_1, \beta) \equiv \Gamma\left(n + \nu_\gamma, \lambda_\gamma + \int_{I_1}^{T} Y_t \mathrm{d}t\right)$
4. Sample I_1 by using a Gibbs step and drawing from
   $\pi(I_1|\boldsymbol{\tau}, \mathbf{I}, \beta, \gamma) \equiv \mathrm{Exp}(\beta N + \gamma + \delta)$
5. Move an infection time by using a Metropolis-Hastings algorithm with acceptance probability: $\min\left(1, \frac{\pi(\mathbf{I}-s+t|\boldsymbol{\tau}, I_1, \beta, \gamma)}{\pi(\mathbf{I}|\boldsymbol{\tau}, I_1, \beta, \gamma)}\right)$
6. repeat step 2-5

More precisely for step 5, we choose one of the existing infection times, s, uniformly at random and then obtain a replacement infection time, t, by sampling uniformly on (I_1, T).

It is shown that MCMC methods are used to carry out Bayesian inference for the stochastic SIR models given partially observed data. The approach is not only robust but very flexible and can be extended to more complex models. We will take advantage of this approach throughout the whole thesis.

1.4 Bayesian nonparametric model

Nonparametric models are increasingly used in preference to parametric models, due to the latter's lack of sufficient flexibility to represent a wide variety of data. Nonparametric models constitute an approach to model selection and adaptation, where the sizes of models are allowed to grow with data size (Orbanz & Teh, 2010). This is as opposed to parametric models which use a fixed number of parameters. Bayesian nonparametric methods provide a Bayesian

framework for model selection and adaptation using nonparametric models. A Bayesian formulation of nonparametric problems is nontrivial, since a Bayesian model defines prior and posterior distributions on a single fixed parameter space, but the dimension of the parameter space in a nonparametric approach should change with sample size. The Bayesian nonparametric solution to this problem is to use an infinite-dimensional parameter space, and to invoke only a finite subset of the available parameters on any given finite data set. This subset generally grows with the data set. In the context of Bayesian nonparametric models, "infinite-dimensional" can therefore be interpreted as "of finite but unbounded dimension". More precisely, a Bayesian nonparametric model is a model that (i) constitutes a Bayesian model on an infinite-dimensional parameter space and (ii) can be evaluated on a finite sample in a manner that uses only a finite subset of the available parameters to explain the sample. Therefore, the Bayesian nonparametric model has an infinite number of dimensions a priori. The word "nonparametric" is misleading. It is not that there are no parameters, it is in fact that there are an infinite number of parameters. One of the most popular tools for Bayesian nonparametric modelling of probability distributions is the Dirichlet process (DP) (Ferguson, 1973). However, the Dirichlet process is not suitable for modelling continuous random variables that require probability density functions since samples from the Dirichlet process are discrete with probability one. To tackle this problem, the Dirichlet process is often used to add a countably-infinite number of parameters into a continuous model. One of the most popular examples is the infinite mixture of parametric distributions (Escobar & West, 1995, Rasmussen, 2000, Gershman & Blei, 2011). Other approaches for using a Dirichlet process to construct a prior that assigns mass to continuous densities include kernel convolution (Lo, 1984) and the Dirichlet diffusion tree (Neal, 2001, 2003). For more discussion of topics related to Bayesian nonparametric models, see Walker et al. (1999) or Ghosh & Romamoorthi (2003). Another popular and well studied tool for specifying prior beliefs about probability densities is the Gaussian process (O'Hagan, 1978). We will use the Gaussian process as a main tool throughout the whole thesis.

Bayesian nonparametric inference, first developed in the statistics community, has become more attractive after the development of computational methods,

such as MCMC. It has led to the development of a variety of models and applications of these models in disciplines such as computer vision, computational biology, cognitive science, signal processing, etc. However, there is little literature concerning Bayesian nonparametric estimation for epidemics, which further motivates our work.

1.5 Gaussian process

1.5.1 Introduction

Given some observations of a dependent variable at certain values of the independent variable, x, one might want to find the best estimate of the dependent variable at a new value, x'. We may expect the relationship between the observations and the variable x, denoted by $f(x)$, to be linear, quadratic, cubic, or even non-polynomial. Then we can use the principles of model selection to choose among the various possibilities. Rather than assuming $f(x)$ relates to some specific models, e.g. $f(x) = ax + b$, a Gaussian process can represent $f(x)$ rigorously and precisely by letting the data describe the story for themselves (Ebden, 2008). In other words, a Gaussian process provides an approach to give a prior density to every possible function, where higher density values are given to functions that we consider to be more likely.

1.5.2 Definition

We now give the definition of a Gaussian process below (Rasmussen & Williams, 2006):

Definition. *A Gaussian process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution.*

In other words, a GP can be considered as a probability distribution over a (possibly infinite) number of elements of a set $\mathcal{X}$, such that the distribution over any finite subset of them is a multi-variate Gaussian. The set $\mathcal{X}$ is chosen based on what sort of elements the Gaussian process should support, but the range is always the real line. Denote by $\mu(\cdot)$ and $K(\cdot, \cdot)$ respectively the mean

function and the positive semidefinite covariance function of the GP, denoted by $f(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. For $n \in \mathbb{N}$ and any finite set of inputs $x_1, x_2, \cdots, x_n$:

$$f(\mathbf{x}) \sim N(\mu(\mathbf{x}), K(\cdot, \cdot)),$$

where $f(\mathbf{x}) = (f(x_1), f(x_2), \cdots, f(x_n))$, $\mu(\mathbf{x}) = (\mu(x_1), \mu(x_2), \cdots, \mu(x_n))$ and

$$K(\cdot, \cdot) = \begin{pmatrix} K(x_1, x_2) & K(x_1, x_3) & \cdots & K(x_1, x_n) \\ K(x_2, x_1) & K(x_2, x_2) & \cdots & K(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_n, x_1) & K(x_n, x_2) & \cdots & K(x_n, x_n) \end{pmatrix}.$$

Therefore, a GP specifies a distribution over functions and is completely specified by its mean and covariance function. More precisely, the Gaussian process provides a nonparametric distribution over functions of the form $g : \mathcal{X} \to \mathbb{R}$, where $\mathbb{R}$ refers to the extended real numbers. In this thesis, we will investigate the epidemic models with the Gaussian process in time space. Therefore, it can be assumed that the domain $\mathcal{X}$ is the 1-dimensional time space $\mathbb{R}$. Adams (2009) points out that from the function-space view, the main idea of the Gaussian process is that if we condition on a finite set of input values in the domain $\mathcal{X}$ with size N, denoted $\{\mathbf{x}_n \in \mathcal{X}\}_{n=1}^N$, we obtain a joint Gaussian distribution over the corresponding output values in the range $\mathbb{R}^N$. The dimensions of this Gaussian distribution can be considered as corresponding to the outputs $\{f(\mathbf{x}_n)\}_{n=1}^N$ of a random function $f(\mathbf{x})$. The N-dimensional Gaussian distribution is specified by an N- dimensional mean vector $\boldsymbol{\mu}_N$ and an $N \times N$ covariance matrix $\mathbf{K}_N$, where $[\boldsymbol{\mu}_N]_n = m(\mathbf{x}_n)$ and $[\mathbf{K}_N]_{n,n'} = K(\mathbf{x}_n, \mathbf{x}_{n'})$.

1.5.3 Covariance function

In this thesis, we will take the mean function to be zero, i.e. $m(\mathbf{x}) = 0, \forall \mathbf{x} \in \mathcal{X}$. Hence, to specify a Gaussian process, we only need to choose an appropriate covariance function which determines what kind of functions we can sample from the Gaussian process. The choice of an appropriate covariance function for a Gaussian process largely depends on the problem under consideration and sometimes, it can be difficult to choose one. We will place Gaussian process priors on the key quantities of interest (e.g. infection rate, in epidemics)

with different well-studied covariance functions and compare the results. For detailed discussions of covariance function selection see, Paciorek (2003) and Kronberger & Kommenda (2013). Below is a list of the covariance functions we will use (Rasmussen & Williams, 2006)

Squared exponential:

$$K(x_i, x_j) = \alpha^2 \exp\left(-\frac{(x_i - x_j)^2}{2\theta^2}\right), \tag{1.2}$$

Exponential:

$$K(x_i, x_j) = \alpha^2 \exp\left(-\frac{|x_i - x_j|}{\theta}\right), \tag{1.3}$$

Matérn ($\nu = \frac{3}{2}$):

$$K(x_i, x_j) = \alpha^2 \left(1 + \frac{\sqrt{3}|x_i - x_j|}{\theta}\right) \exp\left(-\frac{\sqrt{3}|x_i - x_j|}{\theta}\right), \tag{1.4}$$

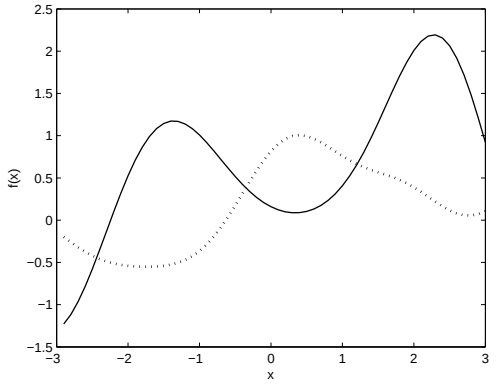
Periodic:

$$K(x_i, x_j) = \alpha^2 \exp\left(-\frac{2\sin^2(\frac{x_i - x_j}{2})}{\theta^2}\right). \tag{1.5}$$

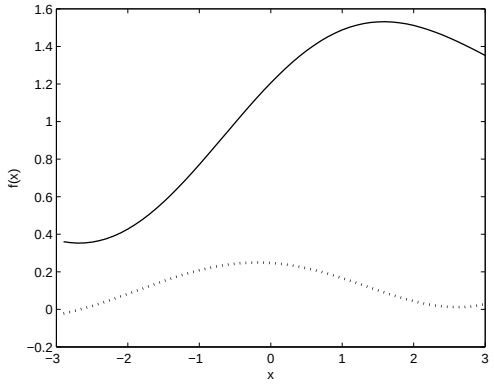
Here α and θ refer to the hyperparameters which respectively control the vertical scale and horizontal length-scale of the Gaussian process. More precisely, the larger α is, the larger vertical range the function will reach and the larger θ is, the more slowly varying the function generated will be. Figure 1.1 (a), (b) and (c) show samples drawn from the Gaussian process priors with the squared exponential covariance function with different hyperparameters settings. The increase in the horizontal length-scale from (a) to (b) produces more slowly fluctuating functions and the increase in the vertical scale from (a) to (c) produces functions that have larger deviation from zero. Figure 1.2 (a) and (b) show samples drawn from the Gaussian process priors with the exponential and Matérn covariance function respectively with same hyperparameter settings, i.e. $\alpha = 1$ and $\theta = 1$. Note that, the functions in Figure 1.1 (a) are also sampled with the same hyperparameter settings. One can see that the functions in Figure 1.2 (a) are more rough than the ones in Figure 1.2 (b) and the functions in Figure 1.1 (a) are the smoothest among the three figures. The squared exponential and exponential covariance functions both belong to the Matérn class of covariance functions shown below (Stein, 1999, Rasmussen & Williams, 2006):

$$K(x_i, x_j) = \alpha^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|x_i - x_j|}{\theta} \right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}|x_i - x_j|}{\theta} \right), \tag{1.6}$$

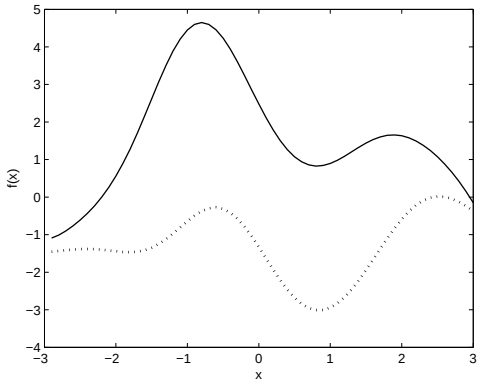
where ν is a positive parameter and K_ν is a modified Bessel function (Abramowitz & Stegun, 1965, Section, 9.6). As $\nu \to \infty$, we obtain the smoothest covariance function in the class, the squared exponential covariance function (1.2). By setting $\nu = 1/2$, we obtain the exponential covariance function (1.3) which generates a very rough process. Figure 1.2 (c) shows samples drawn from the Gaussian process priors with the periodic covariance function and the periodic properties of the covariance function in Figure 1.2 (c) can clearly be seen.



(a) Squared exponential covariance function with $\alpha = 1$ and $\theta = 1$: $\exp\left(-\frac{(x-x')^2}{2(1)^2}\right)$

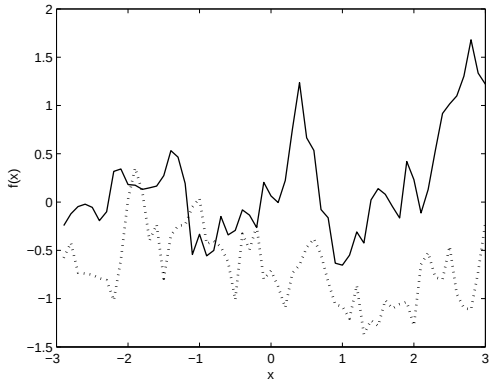


(b) Squared exponential covariance function with $\alpha = 1$ and $\theta = 3$: $\exp\left(-\frac{(x-x')^2}{2(3)^2}\right)$

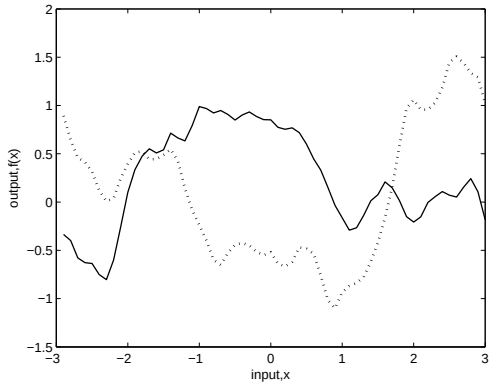


(c) Squared exponential covariance function with $\alpha = 3$ and $\theta = 1$: $3^2 \exp\left(-\frac{(x-x')^2}{2(1)^2}\right)$

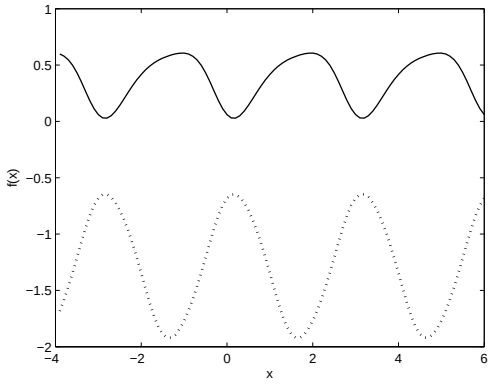
Figure 1.1: Samples drawn from Gaussian process priors. The figure shows two functions (solid and dotted) drawn from each of three Gaussian process priors with the squared exponential covariance function. The sample functions are obtained using a discretization of the x-axis of 60 equally spaced points. The corresponding covariance function is given below each plot.



(a) Exponential covariance function with $\alpha = 1$ and $\theta = 1$: $\exp\left(-\frac{|x-x'|}{1}\right)$



(b) Matérn ($\nu = \frac{3}{2}$) covariance function with $\alpha = 1$ and $\theta = 1$ $\left(1 + \frac{\sqrt{3}|x-x'|}{1}\right)\exp\left(-\frac{\sqrt{3}|x-x'|}{1}\right)$



(c) Periodic covariance function with $\alpha = 1$ and $\theta = 3$. Period is 3: $\exp\left(-\frac{2\sin^2(\pi(x-x')/3)}{3^2}\right)$

Figure 1.2: Samples drawn from Gaussian process priors. The figure shows two functions (solid and dotted) drawn from each of six Gaussian process priors with different covariance functions. The functions are obtained using a discretization of the x-axis of 60 and 100 equally spaced points for plots (a) and (b) and plot (f) respectively. The corresponding covariance function is given below each plot.

1.5.4 Gaussian process prediction

In this section, we describe how one could draw samples from the Gaussian process prior as well as predict function values at new points given observations.

Given a set of inputs $x_1, x_2, \cdots, x_n$, we can draw samples $f(x_1), f(x_2), \cdots, f(x_n)$ from the Gaussian process prior distribution: $\mathbf{f} \sim N(\mathbf{0}, K(\cdot,\cdot))$, where $\mathbf{f} = \{f(x_n)\}_{n=1}^{N}$. For example, Figure 1.3 shows 5 samples drawn from the GP prior distribution.

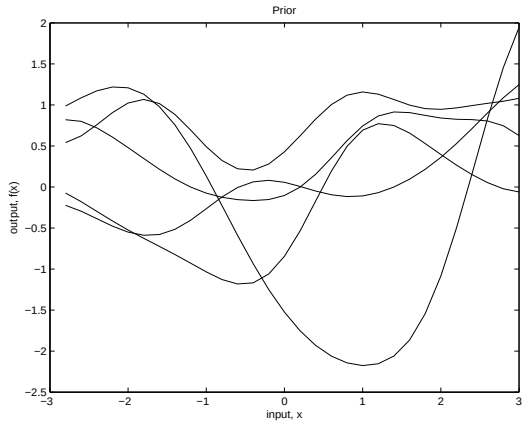


Figure 1.3: 5 random samples are generated from the GP with the same covariance function, the squared exponential covariance function, where the hyperparameters $\alpha = 1$ and $\theta = 1$. The samples are obtained using a discretization of the x-axis of 60 equally spaced points.

We are not only interested in drawing samples from the prior, but also we want to incorporate the knowledge that the new Gaussian inputs provide about the function. Given a set of observed data $D = \{x_n, f(x_n)|n = 1, 2, \cdots, N\} = \{\mathbf{X}, \mathbf{f}\}$, at N^* new points $\mathbf{X}^*$, the predictive distribution of $\mathbf{f}^*$ has the posterior density, $\pi(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}, \mathbf{f})$. According to the GP prior distribution, the joint distribution of $\mathbf{f}$ and $\mathbf{f}^*$ is:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}^* \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X},\mathbf{X}) & K(\mathbf{X},\mathbf{X}^*) \\ K(\mathbf{X}^*,\mathbf{X}) & K(\mathbf{X}^*,\mathbf{X}^*) \end{bmatrix}\right),$$

where $K(\mathbf{X}, \mathbf{X}^*)$ denotes the $N \times N^*$ matrix of the covariances calculated at all pairs of $\mathbf{X}$ and $\mathbf{X}^*$, and similarly for the other entries $K(\mathbf{X},\mathbf{X})$, $K(\mathbf{X}^*,\mathbf{X})$ and $K(\mathbf{X}^*,\mathbf{X}^*)$.

Then, conditioned on the GP prior distribution on the observations, the GP posterior density $\pi(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}, \mathbf{f})$ can be obtained from the following expression:

$$\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}, \mathbf{f} \sim N\left(K(\mathbf{X}^*,\mathbf{X})K(\mathbf{X},\mathbf{X})^{-1}\mathbf{f}, K(\mathbf{X}^*,\mathbf{X}^*) - K(\mathbf{X}^*,\mathbf{X})K(\mathbf{X},\mathbf{X})^{-1}K(\mathbf{X},\mathbf{X}^*)\right).$$

The derivation of the mean and covariance function is in Rasmussen and Williams (2006), section A.2. Note that conditional distribution uses standard results on multi-variate Normal distributions. Figure 1.4 shows 5 samples drawn from the GP posterior distribution.

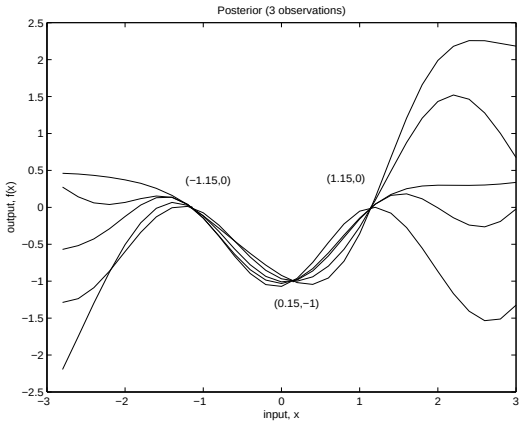


Figure 1.4: Given three observations (-1.15,0), (0.15,-1), (1.15,0), 5 random samples are generated from the Gaussian process with the same covariance function, the squared exponential covariance function, where the hyperparameters $\alpha = 1$ and $\theta = 1$. The samples are obtained using a discretization of the x-axis of 60 equally spaced points.

Figure 1.3 and Figure 1.4 show that given the observations, the function adjusts itself to pass through the given points and the posterior uncertainty would reduce close to the observations.

1.6 Overview

In this chapter, we presented some background information about parametric and nonparametric inference for stochastic epidemic models. We also gave the fundamentals of Bayesian inference and MCMC methods. We then discussed Bayesian nonaprametric models and gave a description of Gaussian processes that we will use in the following chapters.

The rest of this thesis develops Bayesian nonparametric methods for the SIR

epidemic models in small populations and large populations respectively.

In Chapter 2, we explore the behaviour of the overall force of infection in a Bayesian nonparametric framework. We draw inference for the overall force of infection without making specific assumptions on its parametric functional form and place a Gaussian process prior on the overall force of infection. Our methods are illustrated with simulated datasets as well as the real outbreak data. We also compare our findings with results obtained from other parametric methods.

In Chapter 3, we develop Bayesian nonparametric methods for the infection rate in small-scale epidemics. We relax the usual assumption that the infection rate is constant throughout the whole epidemic and treat it as a function which only depends on time. We place a Gaussian process prior on the infection rate and apply the methods to simulated datasets and real outbreak data. Finally, we compare our results with ones obtained from other parametric methods.

In Chapter 4, we apply our methods to large-scale epidemics. Epidemic time-series data are used which consist of number of observed infection cases each day or weeks, etc. For large epidemics in large populations, there is no option but to approximate the epidemic model since the number of infectives is usually so large that it is not possible to augment the data with the times of infection/removal of each case. We adapt an approximation method due to Cauchemez & Ferguson (2008) to develop a Bayesian nonparametric method of inference for epidemics in large populations. We place a Gaussian process prior on the infection rates for each observation period and estimate them on the log scale. Our methods are illustrated with a simulated dataset and real outbreak data. Finally, we compare the results with ones obtained from Cauchemez & Ferguson.

In Chapter 5, we review the thesis and any model limitations.

CHAPTER 2

Bayesian nonparametric estimation for the overall force of infection in small-scale epidemics

2.1 Introduction

It has been widely recognised that mathematical and statistical modelling has become a valuable tool in the analysis of infectious disease dynamics by supporting the development of control strategies, informing policy-making at the highest levels, and in general playing a fundamental role in the fight against the spread of disease. However, many standard modelling and data analysis methods use underlying assumptions (e.g. concerning the rate at which new cases of disease will occur) which are rarely challenged or tested in practice. Our aim is to relax these assumptions by using ideas from Bayesian Nonparametric Statistics. This allows us to analyse data from disease outbreaks without making underlying assumptions which may not be appropriate, and in turn it means that the conclusions taken from the analysis are more reliable.

There has been considerable research activity devoted to the development of Bayesian nonparametric methods. Particularly, the methods have been frequently applied in survival analysis. Nieto-Barajas & Walker (2004) investigate a new class of nonparametric priors for modelling a cumulative distribution function and hence model an unknown survival function in a Bayesian nonparametric framework. Kottas (2006) develops a computational method to ob-

tain the posterior distribution of general functionals of a Dirichlet process mixture and model the survival distribution via a flexible Dirichlet process mixture, with a Weibull kernel. Lorio et al. (2009) develop a Dependent Dirichlet Process model for survival analysis data without conducting survival curve estimates to satisfy the ubiquitous proportional hazards assumption.

Recently, Palacios & Minin (2013) extended advances in Gaussian process-based nonparametric inference for Poisson processes (Adams et al., (2009)) and developed a method for estimating population size dynamics under the coalescent in a Bayesian nonparametric framework.

However, to the best of our knowledge, there is no literature concerning Bayesian nonparametric estimation for epidemics.

In the SIR model, during its infectious period, an individual makes contacts with each of the N susceptibles at times given by the points of a homogeneous Poisson process with rate β, a constant value. This implies that at time t, infection times are generated by an inhomogeneous Poisson process with rate $\beta X_t Y_t$, where X_t and Y_t denote respectively the number of susceptibles and infectives at time t. Such a rate is usually termed the overall force of infection in epidemic modelling.

In this chapter, we will address the question of estimating the overall force of infection in a Bayesian nonparametric framework given either complete observation of an epidemic (i.e. both infection and removal times) or partial observation, with just removals observed. Specifically, we relax the assumption of the overall force of infection being of the form, $\beta X_t Y_t$, and treat it as a function which only depends on time, denoted by $h(t)$. Inferring $h(t)$ is equivalent to inferring a function in a Bayesian framework, i.e. a prior is needed. As discussed in Chapter 1, Section 1.5.2, Gaussian processes are used to describe a distribution over functions, we will take advantage of this and place Gaussian process priors on $h(t)$.

We will first introduce Bayesian nonparametric methods for inhomogeneous Poisson processes introduced by Adams et al. (2009) in the first half of Section 2.2. In the second half of Section 2.2, we will adapt the methods from Adams et al. to simulated Poisson datasets by using a Gaussian process to place a prior on the inhomogeneous Poisson process intensity. We will then develop methods

for the epidemic model estimation problem in section 3 and illustrate them as applied to both artificial and real data. We will also carry out sensitivity analysis by placing different Gaussian process priors on the overall force of infection of an epidemic in Section 2.4. Finally, we finish with conclusions in Section 2.5.

2.2 Estimation for inhomogeneous Poisson process

2.2.1 Introduction

The Poisson process is one of the most widely studied and used models for point data in time and space. The inhomogeneous Poisson process, a variant of the Poisson process, allows the intensity to vary across its domain, e.g. time. In inference problems, the functional form for the variation is usually unknown in advance. In this setting, another stochastic process like a Gaussian process (Rathbun & Cressie, 1994, Møller et al., 1998), Dirichlet process (Kottas & Sansó, 2007), etc. is usually introduced to describe nonparametrically the variation in the Poisson intensity function. This construction is called a Cox process (Cox, 1955). A stochastic process is called a Gaussian Cox process if the intensity function is a transformation of a random realization of a Gaussian process. By placing a Gaussian process prior distribution on the intensity function, it is not necessary to have a preconceived idea of the particular functional form. Unfortunately, however, for likelihood inference, we will have a problem when integrating an infinite-dimensional random function. The problem will be discussed in detail in Section 2.2.4.1.

2.2.2 Sigmoidal Gaussian Cox process model

We aim to use a Gaussian process prior for the inhomogeneous Poisson process intensity which is positive all the time. However, the output space for the Gaussian process is the real line. Therefore, we need a transformation. Adams et al. (2009) modified the Gaussian Cox process model in order to carry out fully nonparametric Bayesian inference via MCMC. The model is called a Sigmoidal Gaussian Cox Process (SGCP) since a sigmoid function is introduced to transform random function values into a random intensity function $\lambda(t)$ via the

relationship

$$\lambda(t) = \lambda^* \sigma(g(t)), \tag{2.1}$$

where λ^* is an upper bound on $\lambda(t)$, $\sigma(\cdot)$ is the logistic function, $\sigma(z) = (1 + e^{-z})^{-1}$, and $g(\cdot)$ is a random function which has a Gaussian process prior. Given a mean function $m(\cdot)$ and a positive definite covariance function $K(\cdot)$ parameterised by hyperparameters, α and θ, the function values $g(t)$ can be drawn from a multivariate Gaussian distribution.

2.2.3 Generating exact Poisson data

Lewis & Shedler (1979) present an algorithm for generating an inhomogeneous Poisson process with intensity, $\lambda(t)$, via thinning of a homogeneous Poisson process. The main idea is to consider a realisation of points from a homogeneous Poisson process with intensity, λ, and accept or reject each point with acceptance probability, $\lambda(t)/\lambda$, where $\lambda(t) \leq \lambda$. Then the collection of accepted points constructs a realisation of the inhomogeneous Poisson process with intensity, $\lambda(t)$.

We now describe the algorithm in detail for generating exact data from a time-inhomogeneous Poisson process with intensity given by (2.1) via a thinning method (Lewis & Shedler 1979, Adams et al. 2009). The "exact" means samples generated are not biased, for instance, by the initial state of a Markov chain.

1. We simulate a set of event times $\{\hat{s}_j\}_{j=1}^J$ from a homogeneous Poisson process with rate λ^* within a time region $[0, T]$. To achieve this, we sample J, the number of events by drawing it from a Poisson distribution with parameter $\lambda^* T$. Then we distribute the J events $\{\hat{s}_j\}_{j=1}^J$ uniformly into $[0, T]$ (Pasupathy, 2011).

2. We consider these events $\{\hat{s}_j\}_{j=1}^J$ as input points for a Gaussian process and draw samples $\{g(\hat{s}_j)\}_{j=1}^J$ at the locations $\{\hat{s}_j\}_{j=1}^J$. Specifically, we sample from the distribution, $\pi(\{g(\hat{s}_j)\}_{j=1}^J | \{\hat{s}_j\}_{j=1}^J, \theta)$, where θ denotes the hyperparameter in the covariance function.

3. We generate J uniform random variates on $(0, 1)$, denoted by $\{r_j\}_{j=1}^J$ and only accept the events for which $r_j < \sigma(g(\hat{s}_j))$. Then the accepted events form the exact Poisson data which are drawn from an inhomogeneous Poisson pro-

cess with an intensity function λ.

Note that λ is the result of applying (2.1) to a random function g drawn from a Gaussian process. The algorithm is also applicable if the intensity λ is known. Given a particular intensity function λ, when simulating data, the probability of accepting the events is $\lambda(\hat{s}_j)/\lambda^*$ instead of $\sigma(g(\hat{s}_j))$. Examples can be found in Adams et al. (2009). Figure 2.1 shows the procedure for generating exact Poisson data via a thinning method.

2.2.4 Inference

We now investigate the posterior distribution of λ given a set of observed data, denoted by $\{s_k\}_{k=1}^K$, using the SGCP model as the prior. Here, s_k is assumed to be a time within a region $[0, T]$.

2.2.4.1 Doubly intractable problem

In terms of Bayes' Theorem, the posterior density for λ given the data $\{s_k\}_{k=1}^K$ is

$$\begin{aligned}\pi(\lambda|\{s_k\}_{k=1}^K) &= \pi(\mathbf{g}|\lambda^*, \{s_k\}_{k=1}^K) \\ &= \frac{\mathcal{GP}(\mathbf{g}) \exp\{-\int_0^T \lambda^* \sigma(g(s))\mathrm{d}s\} \prod_k \lambda^* \sigma(g(s_k))}{\int \mathcal{GP}(\mathbf{g}) \exp\{-\int_0^T \lambda^* \sigma(g(s))\mathrm{d}s\} \prod_k \lambda^* \sigma(g(s_k))\mathrm{d}\mathbf{g}},\end{aligned} \tag{2.2}$$

where $\mathbf{g}$ denotes the infinite-dimensional object corresponding to $g(t)$ and $\mathcal{GP}$ denotes a Gaussian process prior on $\mathbf{g}$. It is notable that (2.2) is intractable due to the presence of an intractable integral over $[0, T]$ in the numerator and an intractable integral over possible values of $\mathbf{g}$ in the denominator. Particularly, the value of the integral over $\mathbf{g}$ is a function of the parameters, $\mathbf{g}$, the variables being sampled. Such problem makes the posterior (2.2) doubly intractable (Murray, 2007). Therefore, performing MCMC inference directly via the posterior density for λ is an intractable problem.

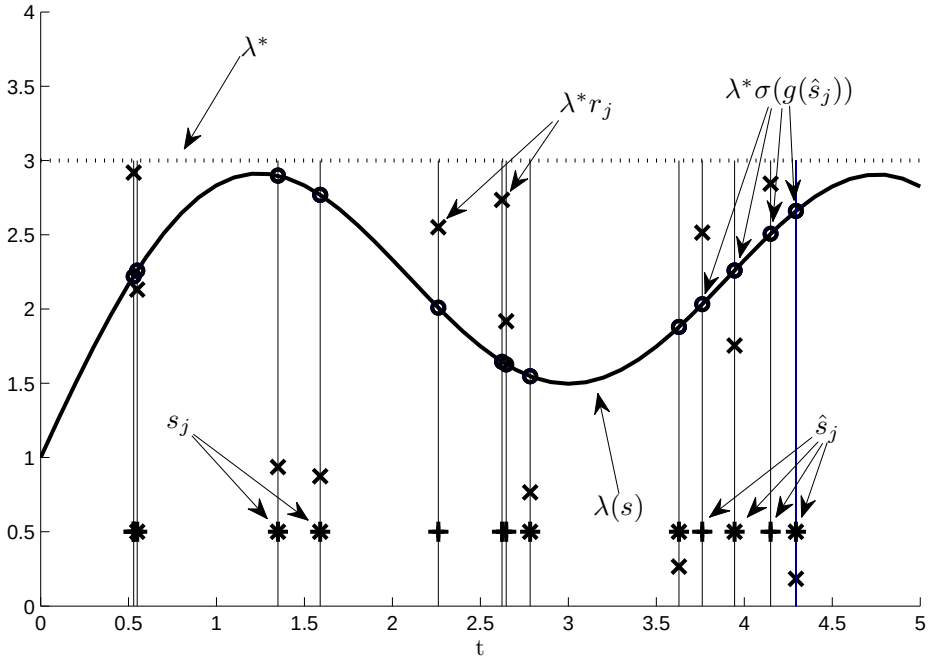


Figure 2.1: Procedure for generating exact Poisson data via a thinning method. A Poisson time series, $\{\hat{s}_j\}_{j=1}^J$ ($+$ and $*$ marks), is generated from a homogeneous Poisson process with rate, λ^*. At each point of the time series, a sample, $g(\hat{s}_j)$, is drawn from the Gaussian process. Then the function value is transformed through the sigmoid function so that it is in form of $\lambda^* \sigma(g(\hat{s}_j))$ ($\circ$ marks) and is positive and bounded above by λ^*. Variates, $\{\lambda^* r_j\}_{j=1}^J$ ($\times$ marks), are drawn uniformly on $(0, \lambda)$ in the vertical coordinate. If the variates are greater than the transformed function values, the corresponding events are discarded ($+$ marks). The remaining events, $\{s_j\}_{j=1}^J$ ($*$ marks), are the exact Poisson events generated from the inhomogeneous Poisson process corresponding to the random function, $\lambda(s)$.

2.2.5 Data augmentation

To solve the problem above, Adams et al. (2009) adopt a data-augmentation approach to make the Markov chain tractable, as follows. Specifically, consider augmenting the observed data with the following variables: the number of thinned events, M, the locations of thinned events, $\{\tilde{s}_m\}_{m=1}^M$, the function values at the observed data, $\mathbf{g}_K = (g(s_1), g(s_2), \cdots, g(s_K))$ and the function values at the thinned events, $\mathbf{g}_M = (g(\tilde{s}_1), g(\tilde{s}_2), \cdots, g(\tilde{s}_M))$. We can sample the above parameters from the joint distribution with density

$$\begin{aligned}&\pi(\{s_k\}_{k=1}^K, M, \{\tilde{s}_m\}_{m=1}^M, \mathbf{g}_{M+K}|\lambda^*, T, \theta)\\ &= (\lambda^*)^{M+K}\exp\{-\lambda^* T\}\prod_{k=1}^K \sigma(g(s_k))\prod_{m=1}^M \sigma(-g(\tilde{s}_m)) \times \pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta),\end{aligned} \tag{2.3}$$

where $\mathbf{g}_{M+K}$ denotes a vector concatenating $\mathbf{g}_M$ and $\mathbf{g}_K$. More specifically, we can sample λ^*, M, $\{\tilde{s}_m\}_{m=1}^M$, $\mathbf{g}_{M+K}$ as well as θ respectively from the full conditional distributions (2.4), (2.5), (2.6), (2.7) and (2.8). We have that

$$\pi(\lambda^*|M, T, \{s_k\}_{k=1}^K) \propto (\lambda^*)^{K+M-1}\exp\{-\lambda^* T\} \times \pi(\lambda^*), \tag{2.4}$$

$$\begin{aligned}&\pi(M, \{\tilde{s}_m\}_{m=1}^M, \mathbf{g}_M|\lambda^*, \mathbf{g}_K, \{s_k\}_{k=1}^K, \theta, T)\\ &\propto (\lambda^*)^M \prod_{m=1}^M \sigma(-g(\tilde{s}_m)) \times \pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta),\end{aligned} \tag{2.5}$$

$$\begin{aligned}&\pi(\{\tilde{s}_m\}_{m=1}^M, \mathbf{g}_M|\lambda^*, \mathbf{g}_K, \{s_k\}_{k=1}^K, M, \theta, T)\\ &\propto \prod_{m=1}^M \sigma(-g(\tilde{s}_m)) \times \pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta),\end{aligned} \tag{2.6}$$

$$\begin{aligned}&\pi(\mathbf{g}_{M+K}|M, \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta)\\ &\propto \prod_{k=1}^K \sigma(g(s_k))\prod_{m=1}^M \sigma(-g(\tilde{s}_m)) \times \pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta),\end{aligned} \tag{2.7}$$

$$\pi(\theta|M, \mathbf{g}_{M+K}, \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M)$$
$$\propto \pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta) \times \pi(\theta). \quad (2.8)$$

2.2.6 MCMC sampling

We here describe in detail a procedure of updating parameters via an MCMC algorithm for inference of the inhomogeneous Poisson process intensity. The algorithm for updating the upper bound parameter, λ^*, the number of thinned events, M, and the locations of thinned events, $\{\tilde{s}_m\}_{m=1}^M$, is the same in Adams et al. (2009). However, the algorithm for updating function values, $\mathbf{g}_{M+K}$, and the hyperparameters, α and θ, is different. Adams et al. adopted Hamiltonian Monte Carlo methods for inference of the function values and the hyperparameters and obtained good mixing of the Markov chain. We used a standard MCMC algorithm which is much easier to implement.

1. Updating the upper bound parameter, λ^*

We use a Gibbs step to update λ^*. As the locations of the observed data and the thinned events are drawn from a homogeneous Poisson process on time region $[0, T]$ with rate λ^*, the mean of λ^* is $\frac{K+M}{T}$, where K is the number of observed data and M is the number of thinned events. According to (2.4), with the gamma prior $\Gamma(\alpha_\lambda, \beta_\lambda)$ put on λ^*, we can sample λ^* from the full conditional distribution, $\Gamma(\alpha_\lambda + K + M, \beta_\lambda + T)$.

2. Updating the number of thinned events, M

We use a Metropolis-Hastings step to sample from the full conditional distribution of the number of thinned events, see (2.5). We choose an insertion move or a deletion move each with probability $\frac{1}{2}$. For the insertion move, we propose a new location l' which is uniformly sampled from $[0, T]$ and then draw a corresponding function value $g(l')$ from the conditioned multivariate Gaussian distribution at l'. Specifically, we sample $g(l')$ from the distribution given by $\pi(\mathbf{g}(l')|l', \mathbf{g}_{M+K}, \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta)$, where $K(l', \mathbf{S})K(\mathbf{S}, \mathbf{S})^{-1}\mathbf{g}_{M+K}$ and $K(l', l') - K(l', \mathbf{S})K(\mathbf{S}, \mathbf{S})^{-1}K(\mathbf{S}, l')$ are the mean and variance of the distribution, where $\mathbf{S} = (s_1, s_2, \cdots, s_K, \tilde{s}'_1, \tilde{s}'_2, \cdots, \tilde{s}'_M)$ denotes locations of observed data and thinned events. The new value can be seen as a prediction of the Gaus-

sian process. For the deletion move, a thinned event l, is uniformly selected and removed from the current M events. We note that the set of $\{\tilde{s}_m\}_{m=1}^{M}$ and $\mathbf{g}_M$ are updated correspondingly when updating the number of thinned events, M. Below is the derivation of the probability of acceptance for the insertion and deletion move.

- Insertion move:

The forward and backward densities are:

$$q(M \to M') = \frac{1}{2} \times \frac{1}{T} \times \pi(\mathbf{g}(l')|l', \mathbf{g}_{M+K}, \{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta),$$

$$q(M' \to M) = \frac{1}{2} \times \frac{1}{M'}.$$

The acceptance ratio is

$$\begin{aligned}
&\frac{q(M' \to M)}{q(M \to M')} \times \frac{\pi(M', \{\tilde{s}_m\}_{m=1}^{M'}, \mathbf{g}'_{M+K}|\{s_k\}_{k=1}^{K}, \theta, \lambda^*)}{\pi(M, \{\tilde{s}_m\}_{m=1}^{M}, \mathbf{g}_{M+K}|\{s_k\}_{k=1}^{K}, \theta, \lambda^*)} \\
&= \frac{\frac{1}{2} \times \frac{1}{M'}}{\frac{1}{2} \times \frac{1}{T} \times \pi(\mathbf{g}(l')|l', \mathbf{g}_{M+K}, \{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta)} \\
&\times \frac{(\lambda^*)^{M'+K} \prod_{k=1}^{K} \sigma(g(s_k)) \prod_{m=1}^{M'} \sigma(-g(\tilde{s}_m)) \times \pi(\mathbf{g}'_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M'}, \theta)}{(\lambda^*)^{M+K} \prod_{k=1}^{K} \sigma(g(s_k)) \prod_{m=1}^{M} \sigma(-g(\tilde{s}_m)) \times \pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta)}.
\end{aligned}$$

As $M' = M + 1$ and

$$\begin{aligned}
&\pi(\mathbf{g}'_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M'}, \theta) \\
&= \pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta) \times \pi(\mathbf{g}(l')|l', \mathbf{g}_{M+K}, \{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta),
\end{aligned}$$

the acceptance ratio is simplified to $\frac{T \times \lambda^* \times \sigma(-g(l'))}{M+1}$.

If the move is accepted, the proposal l' and the corresponding function value $\mathbf{g}(l')$ will be added to the set of $\{\tilde{s}_m\}_{m=1}^{M}$ and $\mathbf{g}_M$ respectively.

- Deletion Move:

The forward and backward densities are:

$$q(M \to M') = \frac{1}{2} \times \frac{1}{M},$$

$$q(M' \to M) = \frac{1}{2} \times \frac{1}{T} \times \pi(\mathbf{g}(l)|l, \mathbf{g}'_{M+K}, \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^{M'}, \theta).$$

The acceptance ratio is

$$\begin{aligned}
&\frac{q(M' \to M)}{q(M \to M')} \times \frac{\pi(M', \{\tilde{s}_m\}_{m=1}^{M'}, \mathbf{g}'_{M+K} | \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \mathbf{g}_{M+K}, \theta, \lambda^*)}{\pi(M, \{\tilde{s}_m\}_{m=1}^M, \mathbf{g}_{M+K} | \{s_k\}_{k=1}^K, \theta, \lambda^*)} \\
&= \frac{\frac{1}{2} \times \frac{1}{T} \times \pi(\mathbf{g}(l)|l, \mathbf{g}'_{M+K}, \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^{M'}, \theta)}{\frac{1}{2} \times \frac{1}{M}} \\
&\times \frac{(\lambda^*)^{M'+K} \prod_{k=1}^K \sigma(g(s_k)) \prod_{m=1}^{M'} \sigma(-g(\tilde{s}_m)) \times \pi(\mathbf{g}'_{M+K} | \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^{M'}, \theta)}{(\lambda^*)^{M+K} \prod_{k=1}^K \sigma(g(s_k)) \prod_{m=1}^{M} \sigma(-g(\tilde{s}_m)) \times \pi(\mathbf{g}_{M+K} | \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^{M}, \theta)}.
\end{aligned}$$

As $M' = M - 1$ and

$$\begin{aligned}
&\pi(\mathbf{g}_{M+K} | \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta) \\
&= \pi(\mathbf{g}'_{M+K} | \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^{M'}, \theta) \times \pi(\mathbf{g}(l)|l, \mathbf{g}'_{M+K}, \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^{M'}, \theta),
\end{aligned}$$

the acceptance ratio is simplified to $\frac{M}{T \times \lambda^* \times \sigma(-g(l))}$.

If the move is accepted, the chosen thinned event l and the corresponding function value $\mathbf{g}(l)$ will be discarded from the set of $\{\tilde{s}_m\}_{m=1}^M$ and $\mathbf{g}_M$ respectively.

3. Updating locations of thinned events, $\{\tilde{s}_m\}_{m=1}^M$

We use a Metropolis-Hastings step to sample from the full conditional distribution, (2.6). We first uniformly draw a new location l' from $[0, T]$ and then draw a new function value $\mathbf{g}(l')$ from the Gaussian process, conditioned on the current $\mathbf{g}_{M+K}$. Specifically, we sample $\mathbf{g}(l')$ from the conditional distribution, $\pi(\mathbf{g}(l')|l', \mathbf{g}_{M+K}, \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta)$, and the mean and variance of this distribution are $K(l', \mathbf{S})K(\mathbf{S}, \mathbf{S})^{-1}\mathbf{g}_{M+K}$ and $K(l', l') - K(l', \mathbf{S})K(\mathbf{S}, \mathbf{S})^{-1}K(\mathbf{S}, l')$ respectively. Then one of the existing thinned events is selected uniformly at random and replaced by the proposed new values. The forward and backward densities for this move are:

$$q(l \to l') = \frac{1}{M} \times \frac{1}{T} \times \pi(\mathbf{g}(l')|l', \mathbf{g}_{M+K}, \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^M, \theta),$$

$$q(l' \to l) = \frac{1}{M'} \times \frac{1}{T} \times \pi(\mathbf{g}(l)|l, \mathbf{g}'_{M+K}, \{s_k\}_{k=1}^K, \{\tilde{s}_m\}_{m=1}^{M'}, \theta).$$

The acceptance ratio is

$$\begin{aligned}
&\frac{q(l'\to l)}{q(l\to l')} \times \frac{\pi(\{\tilde{s}_m\}_{m=1}^{M'}, \mathbf{g}'_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \mathbf{g}_{M+K}, \theta)}{\pi(\{\tilde{s}_m\}_{m=1}^{M}, \mathbf{g}_{M+K}|\{s_k\}_{k=1}^{K}, \theta)} \\
&= \frac{\frac{1}{M'} \times \frac{1}{T} \times \pi(\mathbf{g}(l)|l, \mathbf{g}'_{M+K}, \{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M'}, \theta)}{\frac{1}{M} \times \frac{1}{T} \times \pi(\mathbf{g}(l')|l', \mathbf{g}_{M+K}, \{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta)} \\
&\times \frac{\prod_{k=1}^{K} \sigma(g(s_k)) \prod_{m=1}^{M} \sigma(-g(\tilde{s}_m)) \times \frac{\sigma(-g(l'))}{\sigma(-g(l))} \times \pi(\mathbf{g}'_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M'}, \theta)}{\prod_{k=1}^{K} \sigma(g(s_k)) \prod_{m=1}^{M} \sigma(-g(\tilde{s}_m)) \times \pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta)}.
\end{aligned}$$

As $M' = M$ and

$$\begin{aligned}
&\pi(\mathbf{g}'_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M'}, \theta) \times \pi(\mathbf{g}(l)|l, \mathbf{g}'_{M+K}, \{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M'}, \theta) \\
&= \pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta) \times \pi(\mathbf{g}(l')|l', \mathbf{g}_{M+K}, \{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta),
\end{aligned}$$

the acceptance ratio is simplified to $\frac{\sigma(-g(l'))}{\sigma(-g(l))}$.

If the move is accepted, the old value l and $\mathbf{g}(l)$ will be replaced by the proposal l' and the corresponding value $g(l')$, respectively.

4. Updating function values, $\mathbf{g}_{M+K}$

Given the observed data $\{s_k\}_{k=1}^{K}$, the thinned events $\{\tilde{s}_m\}_{m=1}^{M}$ and the current hyperparameter θ, we wish to sample from the distribution on the function values $\mathbf{g}_{M+K}$, see (2.7). We propose new function values $\mathbf{g}'_{M+K}$ by a proposal of the form $\mathbf{g}'_{M+K} = \delta\mathbf{g}_{M+K} + \sqrt{1-\delta^2}\mathbf{h}_{M+K}$, where $\mathbf{h}_{M+K}$ is drawn from the Gaussian process at $\{s_k\}_{k=1}^{K}$ and $\{\tilde{s}_m\}_{m=1}^{M}$ and δ is in $[0, 1)$. Specifically, we sample $\mathbf{h}_{M+K}$ from the distribution $\pi(\mathbf{h}_{M+K}|\{s_k\}_{k=1}^{K}, \{\tilde{s}_m\}_{m=1}^{M}, \theta)$. This mechanism is an underrelaxed MCMC method discussed by Adams et al.(2009). Such a method is used to keep the contribution of the current function values on some level δ when making proposals. One needs to choose a large δ (which is near one) to obtain a good acceptance ratio when the number of the parameters is large. If we obtain the candidate function values $\mathbf{g}'_{M+K}$ by simply drawing samples from the Gaussian process, we may have a very low acceptance ratio of the proposal as the Gaussian process typically defines a narrow mass (Adams et al., 2009) and it will become worse with increase of the number of the parameters.

To calculate the acceptance probability, first note that

$$\begin{aligned}&\pi(\mathbf{g}_{M+K}|M,\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta,\lambda^*,T)\\&\propto\prod_{k=1}^K\sigma(g(s_k))\prod_{m=1}^M\sigma(-g(\tilde{s}_m))\times\pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta).\end{aligned}$$

We shall abbreviate $\pi(\mathbf{g}_{M+K}|M,\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta,\lambda^*,T)$ by $\pi(\mathbf{g}_{M+K})$ and $\pi(\mathbf{g}'_{M+K}|M,\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta,\lambda^*,T)$ by $\pi(\mathbf{g}'_{M+K})$. Then the acceptance ratio is

$$\begin{aligned}&\frac{\pi(\mathbf{g}'_{M+K})}{\pi(\mathbf{g}_{M+K})}\times\frac{q(\mathbf{g}'_{M+K}\to\mathbf{g}_{M+K})}{q(\mathbf{g}_{M+K}\to\mathbf{g}'_{M+K})}\\&=\frac{\prod_{k=1}^K\sigma(g'(s_k))\prod_{m=1}^M\sigma(-g'(\tilde{s}_m))\times\pi(\mathbf{g}'_{M+K}|\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta)}{\prod_{k=1}^K\sigma(g(s_k))\prod_{m=1}^M\sigma(-g(\tilde{s}_m))\times\pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta)}\\&\times\frac{\pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta)}{\pi(\mathbf{g}'_{M+K}|\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta)}\\&=\frac{\prod_{k=1}^K\sigma(g'(s_k))\prod_{m=1}^M\sigma(-g'(\tilde{s}_m))}{\prod_{k=1}^K\sigma(g(s_k))\prod_{m=1}^M\sigma(-g(\tilde{s}_m))}.\end{aligned}$$

Acceptance ratios generally simplify as we propose from the GP conditional density which appears in the likelihood so that cancellation occurs.

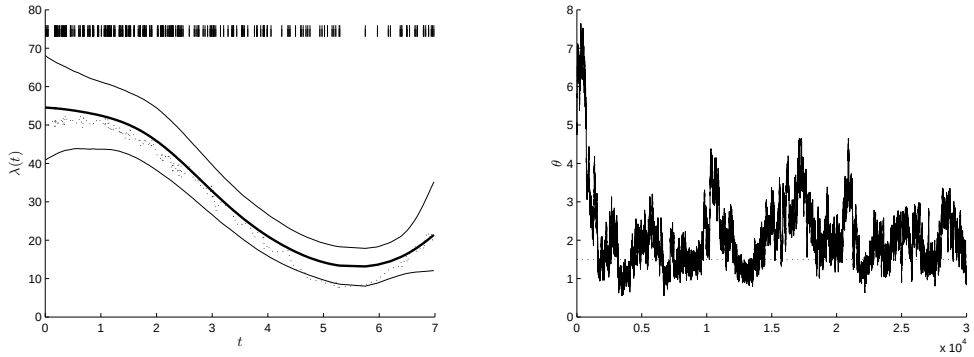
5. Updating the hyperparameter, θ

As we set the hyperparameter α to a certain value for our case, we only need to sample from the posterior distribution on the hyperparameter θ in the covariance function. The full conditional distribution is given in (2.8). We assign an exponential prior on θ, i.e. $\theta\sim\text{Exp}(\lambda_\theta)$. Then we propose a new θ' from a normal distribution with mean, the current value of θ and variance, σ^2. Let Σ_θ denotes the covariance matrix with the hyperparameter θ. The Metropolis-Hastings acceptance ratio for this proposal is

$$\begin{aligned}&\frac{\pi(\theta'|M,\mathbf{g}_{M+K},\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\lambda^*,T)}{\pi(\theta|M,\mathbf{g}_{M+K},\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\lambda^*,T)}\times\frac{q(\theta'\to\theta)}{q(\theta\to\theta')}\\&=\frac{\pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta')\times\pi(\theta')}{\pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta)\times\pi(\theta)}\times\frac{\exp(-(\theta'-\theta)^2/2\sigma^2)}{\exp(-(\theta-\theta')^2/2\sigma^2)}\\&=\frac{\pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta')\times\lambda_\theta\exp(-\lambda_\theta\theta')}{\pi(\mathbf{g}_{M+K}|\{s_k\}_{k=1}^K,\{\tilde{s}_m\}_{m=1}^M,\theta)\times\lambda_\theta\exp(-\lambda_\theta\theta)}\\&=\frac{|\Sigma_{\theta'}|^{-\frac{1}{2}}\times\exp(-\mathbf{g}_{M+K}{}^T{\Sigma_{\theta'}}^{-1}\mathbf{g}_{M+K}/2-\lambda_\theta\theta')}{|\Sigma_\theta|^{-\frac{1}{2}}\times\exp(-\mathbf{g}_{M+K}{}^T{\Sigma_\theta}^{-1}\mathbf{g}_{M+K}/2-\lambda_\theta\theta)}.\end{aligned}$$

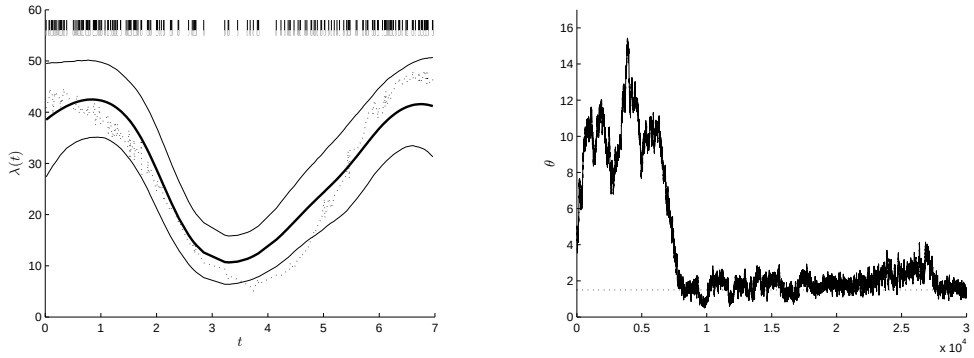
2.2.7 Simulated Poisson data study

We first apply our methods to simulated Poisson datasets generated from an inhomogeneous Poisson process with random intensity $\lambda(t)$. We choose the squared exponential covariance function for the Gaussian process prior and fix the hyperparameter, α, to the truth, $\alpha = 1$, i.e. we only update the hyperparameter, θ, in the covariance function. We have applied the methods to several different data sets. We here choose 3 of our results to show that our approach works fairly well. Figures 2.2 (a), 2.3 (a) and 2.4 (a) give the posterior mean of $\lambda(t)$ for different data sets. It is notable that the true $\lambda(t)$ values in 2.2 (a) and 2.3 (a) are generated from the transformed Gaussian process with the same parameter setting. However, they are different due to the random Gaussian process. For the computer implementation of the algorithm, the run times of 30000 iterations were around 3 hours and 1.2 hours for the first 2 results and the last one respectively. The MCMC trace plot of the hyperparameter θ is shown in Figures 2.2 (b), 2.3 (b) and 2.4 (b). We also apply our methods to simulated Poisson datasets generated from a constant intensity function $\lambda(t) = b$, where b is a constant. Such Poisson data are homogeneous as there are no rejections during the procedure of generating the inhomogeneous Poisson data from the initial homogeneous Poisson data. We shall expect that the estimated intensity is quite flat. Figure 2.5 and 2.6 show posterior means of the estimated intensity $\lambda(t)$ with 95% credible intervals. The " | " marks give locations of the Poisson events. From the results obtained, we can see that our methods work fairly well. Specifically, the estimated intensities appear to recover the truth and are covered by the 95% credible intervals.



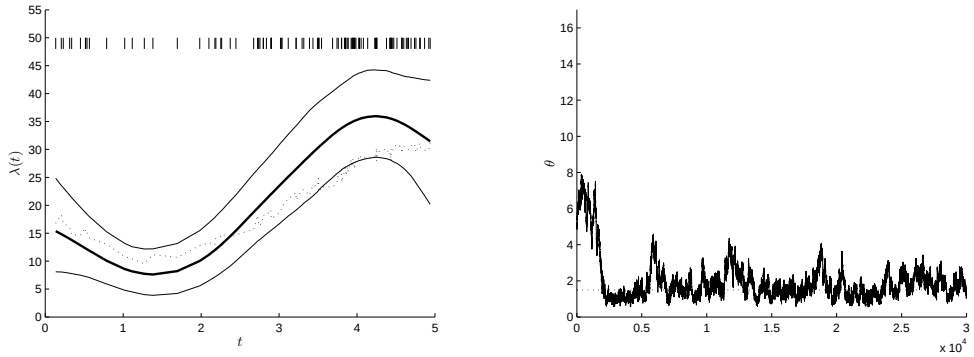
(a) (b)

Figure 2.2: Plot (a) shows posterior mean of the inhomogeneous Poisson process intensity $\lambda(t)$ at each Poisson point compared with the true $\lambda(t)$ generated from a transformed Gaussian process (see (2.1)) given by $\lambda^* = 60$, $T = 7$ and a zero mean and a value of hyperparameter, $\theta = 1.5$, for the Gaussian process. The Poisson data are shown by " | " marks. The 95% credible intervals are shown as well. Plot (b) shows the MCMC trace plot of the hyperparameter θ compared with the true θ (dotted line) for the first Poisson data set.



(a) (b)

Figure 2.3: Plot (a) shows posterior mean of the inhomogeneous Poisson process intensity $\lambda(t)$ at each Poisson point compared with the true $\lambda(t)$ generated from a transformed Gaussian process given by $\lambda^* = 60$, $T = 7$ and a zero mean and a value of hyperparameter, $\theta = 1.5$, for the Gaussian process. The Poisson data are shown by " | " marks. The 95% credible intervals are shown as well. Plot 2.3 (b) shows the MCMC trace plot of the hyperparameter θ compared with the true θ (dotted line) for the second Poisson data set.



(a) (b)

Figure 2.4: Plot (a) shows posterior mean of the inhomogeneous Poisson process intensity $\lambda(t)$ at each Poisson point compared with the true $\lambda(t)$ generated from a transformed Gaussian process given by $\lambda^* = 40$, $T = 5$ and a zero mean and a value of hyperparameter, $\theta = 1.5$, for the Gaussian process. The Poisson data are shown by " | " marks. The 95% credible intervals are shown as well. Plot 2.4 (b) shows the MCMC trace plot the hyperparameter θ compared with the true θ (dotted line) for the third Poisson data set.

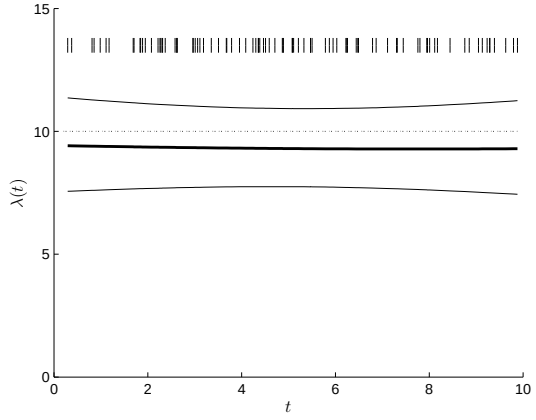


Figure 2.5: Posterior mean of the inhomogeneous Poisson process intensity $\lambda(t)$ at each Poisson point compared with the truth $\lambda(t) = 10$ (dotted line). The Poisson data (" | " marks) are generated from the intensity $\lambda(t) = 10$ within a time region $[0, 10]$.

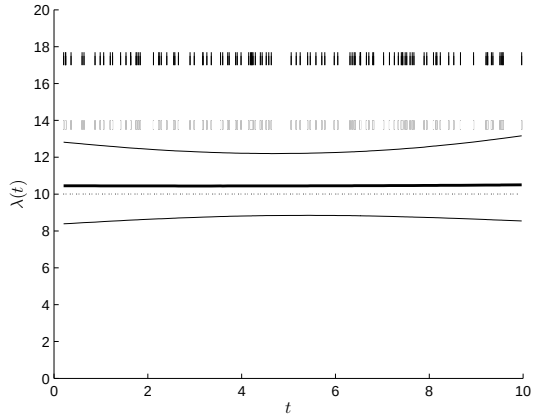


Figure 2.6: Posterior mean of the inhomogeneous Poisson process intensity $\lambda(t)$ at each Poisson point compared with the truth $\lambda(t) = 10$ (dotted line). The Poisson data, different from the one in Figure 2.5, (" | " marks) are generated with the same parameter setting, i.e. from the intensity $\lambda(t) = 10$ within a time region $[0, 10]$.

2.3 Estimation for epidemic models

For real life epidemics, data concerning the process of infection are seldom available, i.e. only removal times are observed. We have discussed Bayesian inference for the SIR model in Chapter 1 by assuming infections occur according to an inhomogeneous Poisson process with rate $\beta X_t Y_t$, at time t, where X_t and Y_t refer to the number of susceptibles and infectives respectively at time t. Recall that $\beta X_t Y_t$ is termed the overall force of infection in epidemic modelling. We now investigate the overall force of infection of the SIR model nonparametrically in a Bayesian framework. We relax the assumption of the overall force of infection being $\beta X_t Y_t$ and treat it as a function which only depends on time, denoted by $h(t)$. Then the epidemic can be described according to the following transition probabilities:

$$P[X_{t+\delta t} = X_t - 1, Y_{t+\delta t} = X_t + 1 | \mathcal{H}_t] = h(t)\delta t + o(\delta t),$$

$$P[X_{t+\delta t} = X_t, Y_{t+\delta t} = Y_t - 1 | \mathcal{H}_t] = \gamma Y_t \delta t + o(\delta t),$$

$$P[X_{t+\delta t} = X_t, Y_{t+\delta t} = Y_t | \mathcal{H}_t] = 1 - h(t)\delta t - \gamma Y_t \delta t + o(\delta t).$$

Let the observed (ordered) successive removal times be $\boldsymbol{\tau} = (\tau_1 = 0, \tau_2, \cdots, \tau_K)$, these being the times of all the removals during $[0, T]$ where $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_K$ and $T > 0$. Let I_1 be the unobserved time of the first infection and let $\mathbf{I} = (I_2, I_3, \cdots, I_K)$ denote the remaining unobserved successive infection times during $[I_1, T]$, where $I_1 \leq I_2 \leq \cdots \leq I_K$. As we assume the epidemic is known to have ceased, the number of infection times and removal times are equal. If the initial infective does not manage to infect any other susceptible by the first removal time, τ_1, the epidemic is ceased. Therefore, in order to obtain epidemics with $n \leq 2$, the following constrains are imposed: $I_i < I_{i+1} < \tau_i$, for $i = 2, 3, \cdots, K-1$. Conditionally on the overall force of infection, $h(t)$, the removal rate, γ, and the first infection time, I_1, we can write the density of $(\boldsymbol{\tau}, \mathbf{I})$ as

$$\begin{aligned}&\pi(\boldsymbol{\tau}, \mathbf{I} | h(t), \gamma, I_1) \\ &= \prod_{j=2}^{K} h(I_{j^-}) \exp\left\{-\int_{I_1}^{T} h(s)\mathrm{d}s\right\} \times \prod_{i=1}^{K} \gamma Y_{\tau_{i^-}} \exp\left\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\right\}. \end{aligned} \tag{2.9}$$

We now place a Gaussian process prior on $h(t)$ via the SGCP model, i.e.

$$h(t) = h^* \sigma(g(t)),$$

where h^* refers to an upper bound on $h(t)$, $\sigma(\cdot)$ refers to the logistic function, $\sigma(z) = (1 + e^{-z})^{-1}$ and $g(\cdot)$ refers to a random function which has a Gaussian process prior with mean function $\mathbf{0}$ and covariance function with hyperparameter, θ, and we assume the hyperparameter, α, is fixed to 2 for this case. According to Bayes' Theorem, we write the posterior distribution over $h(t)$ as

$$\begin{aligned}&\pi(\mathbf{h} | \boldsymbol{\tau}, \mathbf{I}, I_1, T, \gamma) = \pi(\mathbf{g} | h^*, \boldsymbol{\tau}, \mathbf{I}, I_1, T, \gamma) \\ &= \frac{\mathcal{GP}(\mathbf{g}) \prod_{j=2}^{K} h^* \sigma(g(I_{j^-})) \exp\{-\int_{I_1}^{T} h^* \sigma(g(s)) \mathrm{d}s\} \prod_{i=1}^{K} \gamma Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\}}{\int \mathcal{GP}(\mathbf{g}) \prod_{j=2}^{K} h^* \sigma(g(I_{j^-})) \exp\{-\int_{I_1}^{T} h^* \sigma(g(s)) \mathrm{d}s\} \prod_{i=1}^{K} \gamma Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\} \mathrm{d}\mathbf{g}},\end{aligned} \tag{2.10}$$

where $\mathbf{h}$ and $\mathbf{g}$ denote the infinite-dimensional objects corresponding to $h(t)$ and $g(t)$. We now have a similar problem, i.e. doubly-intractable, as we had for the case discussed in Section 2.2. To tackle this, we augment the posterior distribution shown in (2.10) considering the overall force of infection, $h(t)$, as

an inhomogeneous Poisson process intensity. More precisely, we consider augmenting the observed data with the following variables: the number of thinned events, M, the locations of thinned events, $\{\tilde{I}_s\}_{s=1}^{M}$, the function values at the infection times, $\mathbf{g}_K = (g(I_{2^-}), g(I_{3^-}), \cdots, g(I_{K^-}))$ and the function values at the thinned events, $\mathbf{g}_M = (g(\tilde{I}_{1^-}), g(\tilde{I}_{2^-}), \cdots, g(\tilde{I}_{M^-}))$. We then can sample the parameters as well as the additional latent variables from the joint distribution below.

$$\begin{aligned}
&\pi(\boldsymbol{\tau}, \mathbf{I}, M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_{M+K} | h^*, \gamma, I_1, T, \theta) \\
&\propto (h^*)^{K+M-1} \exp(-h^*(T - I_1)) \prod_{j=2}^{K} \sigma(g(I_{j^-})) \prod_{s=1}^{M} \sigma(-g(\tilde{I}_{s^-})) \\
&\times \prod_{i=1}^{K} \gamma Y_{\tau_{i^-}} \exp\left\{ -\int_{I_1}^{T} \gamma Y_s \mathrm{d}s \right\} \times \pi(\mathbf{g}_{M+K} | M, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta),
\end{aligned} \tag{2.11}$$

where $\mathbf{g}_{M+K}$ denotes a vector concatenating of $\mathbf{g}_M$ and $\mathbf{g}_K$. It is notable that we only have $K-1$ elements in the vector $\mathbf{g}_K$ since we do not model the function value at the first infection time.

2.3.1 MCMC algorithms

1. Sampling γ

Considering (2.11), we have the full conditional density of γ

$$\pi(\gamma | \boldsymbol{\tau}, \mathbf{I}, I_1, h(t)) \propto \prod_{i=1}^{K} \gamma Y_{\tau_{i^-}} \exp\left\{ -\int_{I_1}^{T} \gamma Y_s \mathrm{d}s \right\} \pi(\gamma), \tag{2.12}$$

where $\pi(\gamma)$ denotes a prior on γ. According to the algorithms developed by O'Neill & Roberts (1999), we use a Gibbs step to sample from the distribution (2.12). With a conjugate gamma prior $(\nu_\gamma, \lambda_\gamma)$ put on γ, we have

$$\pi(\gamma | \boldsymbol{\tau}, \mathbf{I}, I_1, h(t)) \propto \Gamma\left(\nu_\gamma + K, \lambda_\gamma + \int_{I_1}^{T} Y_s \mathrm{d}s \right).$$

2. Sampling I_1

We use a Gibbs step to sample from the desired distribution. Suppose that I_1 has prior density given by $\eta \exp(\eta y) I(y < 0)$, where $\eta > 0$, and $I(\cdot)$ is the indicator function. Following from (2.11) and the fact that $I_1 < I_2$, conditionally

on $\boldsymbol{\tau}$, $\mathbf{I}$, $h(t)$ and γ, the density of I_1 is as follows:

$$\pi(y|\boldsymbol{\tau}, \mathbf{I}, h(t), \gamma) = \Lambda \exp\{-\Lambda(I_2 - y)\},\ y \in (-\infty, I_2),$$

where $\Lambda = \eta + \gamma + h(I_{2^-}) * N$.

3. Sampling I

Considering (2.11), we write the full conditional density of $\mathbf{I}$

$$\begin{aligned}&\pi(\mathbf{I}, \mathbf{g}_K|\boldsymbol{\tau}, M, \{\tilde{I}_s\}_{s=1}^M, \mathbf{g}_M, h^*, \gamma, I_1, T, \theta)\\ &\propto \prod_{j=2}^{K} \sigma(g(I_{j^-})) \times \prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\left\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\right\} \times \pi(\mathbf{g}_{M+K}|M, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta).\end{aligned} \tag{2.13}$$

According to the method presented by O'Neill & Roberts (1999), we sample from the distribution (2.13) using a Metropolis-Hastings algorithm. We first uniformly draw a new location I_j' from $[I_1, T]$ and then draw a new function value $\mathbf{g}(I_{j^-}')$ from the Gaussian distribution, conditioned on the current $\mathbf{g}_{M+K}$. Specifically, the distribution is $\pi(\mathbf{g}(I_{j^-}')|I_j', \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta)$, with the mean and variance $K(I_j', \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{g}_{M+K}$ and $K(I_j', I_j') - K(I_j', \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}(\mathbf{X}, I_j')$ respectively, where $\mathbf{X} = (I_2, I_3, \cdots, I_K, \tilde{I}_1, \tilde{I}_2, \cdots, \tilde{I}_M)$ denotes locations of infection times and thinned events. Then one of the existing infection times, I_j, is selected uniformly at random and replaced by the proposed new values. We shall write $\mathbf{I}'$ to denote the set of infection times $\mathbf{I}$ with the selected I_j excluded and with the proposed I_j' included, i.e. $\mathbf{I}' = \mathbf{I} - \{I_j\} + \{I_j'\}$. The forward and backward densities are:

$$q(I_j \to I_j') = \frac{1}{K-1} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(I_{j^-}')|I_j', \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta),$$

$$q(I_j' \to I_j) = \frac{1}{K-1} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(I_{j^-})|I_j, \mathbf{g}_{M+K}', \mathbf{I}', \{\tilde{I}_s\}_{s=1}^M, \theta).$$

We abbreviate $\pi(\mathbf{I}, \mathbf{g}_K|\{\tau_i\}_{i=1}^K, M, \{\tilde{I}_s\}_{s=1}^M, \mathbf{g}_M, \beta^*, \gamma, I_1, T, \theta)$ by $\pi(\mathbf{I}, \mathbf{g}_K)$. Then

the acceptance ratio is

$$\begin{aligned}
&\frac{q(I'_j \to I_j)}{q(I_j \to I'_j)} \times \frac{\pi(\{I'_j\}_{j=2}^K, \mathbf{g}'_K)}{\pi(\{I_j\}_{j=2}^K, \mathbf{g}_K)} \\
&= \frac{\frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I_{j^-})|I_j, \mathbf{g}'_{M+K}, \mathbf{I}', \{\tilde{I}_s\}_{s=1}^M, T, \theta)}{\frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I'_{j^-})|I'_j, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, T, \theta)} \times \frac{\prod_{j=2}^K \sigma(g(I'_{j^-}))}{\prod_{j=2}^K \sigma(g(I_{j^-}))} \\
&\times \frac{\prod_{i=1}^K Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^T \gamma Y'_s \mathrm{d}s\} \times \pi(\mathbf{g}'_{M+K}|\mathbf{I}', \{\tilde{I}_s\}_{s=1}^M, \theta)}{\prod_{i=1}^K Y_{\tau_{i^-}} \exp\{-\int_{I_1}^T \gamma Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta)}.
\end{aligned}$$

As

$$\begin{aligned}
&\pi(\mathbf{g}'_{M+K}|\mathbf{I}', \{\tilde{I}_s\}_{s=1}^M, \theta) \times \pi(\mathbf{g}(I_{j^-})|I_j, \mathbf{g}'_{M+K}, \mathbf{I}', \{\tilde{I}_s\}_{s=1}^M, \theta) \\
&= \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta) \times \pi(\mathbf{g}(I'_{j^-})|I'_j, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta),
\end{aligned}$$

the acceptance ratio is simplified to

$$\frac{\sigma(g(I'_{j^-}))}{\sigma(g(I_{j^-}))} \times \frac{\prod_{i=1}^K Y'_{\tau_{i^-}} \exp\{-\int_{I_1}^T \gamma Y'_s \mathrm{d}s\}}{\prod_{i=1}^K Y_{\tau_{i^-}} \exp\{-\int_{I_1}^T \gamma Y_s \mathrm{d}s\}}.$$

4. Sampling h^*

According to (2.11), we have the full conditional density of h^*

$$\pi(h^*|M, I_1, T, \mathbf{I}) \propto (h^*)^{K+M-1} \exp\{-h^*(T - I_1)\}\pi(h^*),$$

where $\pi(h^*)$ is a prior on h^*. We use a Gibbs step to sample h^*. With a conditionally-conjugate gamma prior $\Gamma(\nu_h, \lambda_h)$ put on h^*, we can sample from the full conditional distribution, $\Gamma(\nu_h + K + M - 1, \lambda_h + T - I_1)$.

5. Sampling M

Similar to the algorithm discussed in Section 2.2.5, we use a Metropolis-Hastings step to sample from the full conditional distribution of the number of thinned events below

$$\begin{aligned}
&\pi(M, \{\tilde{I}_s\}_{s=1}^M, \mathbf{g}_M|h^*, \mathbf{g}_K, \mathbf{I}, \theta, I_1, T) \\
&\propto (h^*)^M \prod_{s=1}^M \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta).
\end{aligned}$$

There are insertion move and deletion move each with probability $\frac{1}{2}$. For the insertion move, we propose a new location $\tilde{I}'_s$ which is uniformly sampled from

$[I_1, T]$ and then draw a corresponding function value $\mathbf{g}(\tilde{I}'_{s^-})$ from the conditioned multivariate Gaussian distribution at $\tilde{I}'_s$. Specifically, the distribution is $\pi(\mathbf{g}(\tilde{I}'_{s^-})|\tilde{I}'_s, \mathbf{g}_{M+K}, \{I_j\}_{j=2}^K, \{\tilde{I}_s\}_{s=1}^M, \theta)$, and the mean and variance of this distribution are $K(\tilde{I}'_s, \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{g}_{M+K}$ and $K(\tilde{I}'_s, \tilde{I}'_s) - K(\tilde{I}'_s, \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}K(\mathbf{X}, \tilde{I}'_s)$ respectively. The new value can be seen as a prediction of Gaussian process. For the deletion move, a thinned event $\tilde{I}_s$, is uniformly selected and removed from the current M events. The Metropolis-Hastings acceptance ratio for the insertion and deletion move is respectively:

$$\frac{(T - I_1) \times h^* \times \sigma(-g(\tilde{I}'_{s^-}))}{M + 1},$$

and

$$\frac{M}{(T - I_1) \times \beta^* \times \sigma(-g(\tilde{I}_{s^-}))}.$$

The derivation of the acceptance ratio for both moves is similar to the case discussed in Section 2.2.5 and will not be given here.

6. Sampling $\{\tilde{I}_s\}_{s=1}^M$

We use a Metropolis-Hastings step to perform sampling from the full conditional distribution of $\{\tilde{I}_s\}_{s=1}^M$ below.

$$\begin{aligned}&\pi(\{\tilde{I}_s\}_{s=1}^M, \mathbf{g}_M|h^*, \mathbf{g}_K, \mathbf{I}, M, \theta, I_1, T)\\ &\propto \prod_{s=1}^M \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta).\end{aligned}$$

We first uniformly draw a new location $\tilde{I}'_s$ from $[I_1, T]$. Conditioned on the current $\mathbf{g}_{M+K}$, we then draw a new function value $\mathbf{g}(\tilde{I}'_{s^-})$ from the multivariate Gaussian distribution, given by $\pi(\mathbf{g}(\tilde{I}'_{s^-})|\tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta)$, with the mean and variance $K(\tilde{I}'_s, \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{g}_{M+K}$ and $K(\tilde{I}'_s, \tilde{I}'_s) - K(\tilde{I}'_s, \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}K(\mathbf{X}, \tilde{I}'_s)$ respectively. Then one of the existing thinned events is selected uniformly at random and replaced by the proposed new values. The Metropolis-Hastings acceptance ratio is $\frac{\sigma(-g(\tilde{I}'_{s^-}))}{\sigma(-g(\tilde{I}_{s^-}))}$.

7. Sampling $\mathbf{g}_{M+K}$

Given the infection times, $\mathbf{I}$, the locations of thinned events, $\{\tilde{I}_s\}_{s=1}^M$, and the current hyperparameter, θ, we wish to sample from the full conditional distri-

bution of function values, $\mathbf{g}_{M+K}$, below.

$$\pi(\mathbf{g}_{M+K}|M, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)$$
$$\propto \prod_{j=2}^{K} \sigma(g(I_{j^-})) \prod_{s=1}^{M} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta).$$

We use a Metropolis-Hastings step to sample $\mathbf{g}_{M+K}$. By using the underrelaxed MCMC method discussed in Section 2.2.5, we propose new function values $\mathbf{g}'_{M+K}$ by a proposal of the form $\mathbf{g}'_{M+K} = \delta\mathbf{g}_{M+K} + \sqrt{1-\delta^2}\mathbf{h}_{M+K}$, where $\mathbf{h}_{M+K}$ is drawn from the Gaussian process at $\mathbf{I}$ and $\{\tilde{I}_s\}_{s=1}^{M}$ and δ is in $[0, 1)$. Specifically, we sample $\mathbf{h}_{M+K}$ from the distribution $\pi(\mathbf{h}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)$. The new function values are accepted with probability

$$\min\left(1, \frac{\prod_{j=2}^{K} \sigma(g'(I_{j^-})) \prod_{s=1}^{M} \sigma(-g'(\tilde{I}_{s^-}))}{\prod_{j=2}^{K} \sigma(g(I_{j^-})) \prod_{s=1}^{M} \sigma(-g(\tilde{I}_{s^-}))}\right).$$

8. Sampling θ

As we fix the hyperparameter α for our case, we only need to sample from the full conditional distribution of the hyperparameter, θ, given below.

$$\pi(\theta|M, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}) \propto \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta) \times \pi(\theta),$$

where $\pi(\theta)$ is a prior on θ. We assign an exponential prior on θ, i.e. $\theta \sim \text{Exp}(\lambda_\theta)$. Then we propose a new θ' from a normal distribution with mean, the current value of θ and variance, σ^2. The Metropolis-Hastings acceptance ratio for this proposal is

$$= \frac{|\Sigma_{\theta'}|^{-\frac{1}{2}} \times \exp(-\mathbf{g}_{M+K}{}^T \Sigma_{\theta'}{}^{-1} \mathbf{g}_{M+K}/2 - \lambda_\theta \theta')}{|\Sigma_\theta|^{-\frac{1}{2}} \times \exp(-\mathbf{g}_{M+K}{}^T \Sigma_\theta{}^{-1} \mathbf{g}_{M+K}/2 - \lambda_\theta \theta)}.$$

For the MCMC algorithms discussed above, the standard deviations of the proposals were tuned to obtain an acceptance rate of 20-30%.

2.3.2 Simulated complete data

We now simulate a general stochastic epidemic with parameters infection rate, removal rate (in the SIR model, removals occur at rate γX_t, where γ is removal rate), initial number of susceptibles and initial number of infective individuals.

We first assume that we obtain a set of infection times. Then we can derive the duration of the process, V, i.e. $V =$ last removal time - first infection time. We shall use these data to consider a scenario that the epidemic is known to be completed, i.e. infection times are known. We choose the squared exponential covariance function for the Gaussian process prior and the hyperparameter α is fixed to 2 in order to obtain efficient mixing of the MCMC algorithms and the topic of the choice of α was not explored. The prior of the hyperparameter, θ, was set to be $Exp(0.1)$. We also used Gamma distribution with mean 10^4 and variance 10^8 providing conjugate priors for the upper bound, $h(t)$. We applied our approach to 3 different simulated epidemic datasets which are generated from the SIR model with 3 different parameter settings. Table 2.1 gives infection rates, removal rates, initial number of susceptibles and initial number of infectives for generating the data. Figures 2.7, 2.8 and 2.9 show the posterior means of force of infection at each infection time for SE-Data 1, SE-Data 2 and SE-Data 3 respectively. From the figures, we can see that the model fits SE-Data 1 not as well as SE-Data 2 and SE-Data 3, i.e. for SE-Data 1, the estimated intensity struggles to capture the feature of the data $\beta X_t Y_t$ in the middle period of the epidemic. One explanation is that there is a total of 87 infections within a time region $[0, 11.6]$ for the first data set whereas there are 130 and 189 infections within time regions $[0, 8.1]$ and $[0, 10.3]$ respectively for the second and third data sets. One might think there will be a relationship between the size of the population and the covariance structure of the Gaussian process, i.e. weaker correlation in small populations. For the computer implementation of the algorithm, the run times of 200000 iterations were around 4.5 hours, 10 hours and 13 hours respectively.

Table 2.1: Infection rates, β, removal rates, γ, initial number of susceptibles, N, and initial number of infectives, a, for 3 of the data sets. Such parameter settings are used to generate 3 different epidemic process from the SIR model, i.e. 3 simulated epidemic data set named as SE-Data 1, SE-Data 2 and SE-Data 3.

	β	γ	N	a
parameter setting 1	0.025	1	100	1
parameter setting 2	0.015	1	150	1
parameter setting 3	0.015	1	200	1

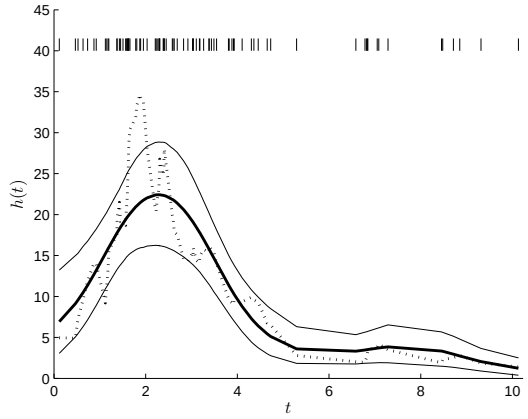


Figure 2.7: Posterior mean of the overall force of infection $h(t)$ (solid line) at each infection time compared with the original data $\beta X_t Y_t$ (dotted line) generated from the general stochastic epidemic with parameters infection rate $\beta = 0.025$, removal rate $\gamma = 1$, initial number of susceptibles $N = 100$ and initial number of infective individuals $a = 1$. The 95% credible intervals are shown as well. There is a total of 87 infections during the whole epidemic and all the infection times are assumed to be known. The " | " marks in the plot represent the observed data, i.e. the infection times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2.

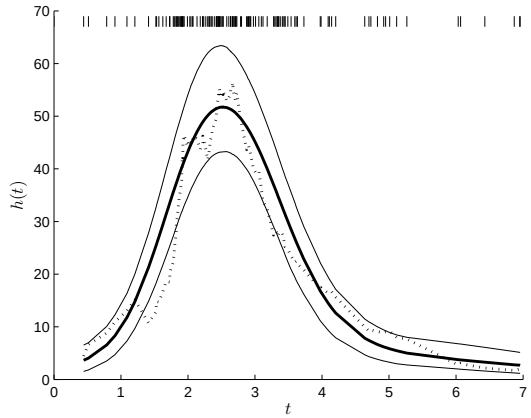


Figure 2.8: Posterior mean of the overall force of infection $h(t)$ (solid line) at each infection time compared with the original data $\beta X_t Y_t$ (dotted line) generated from the general stochastic epidemic with parameters infection rate $\beta = 0.015$, removal rate $\gamma = 1$, initial number of susceptibles $N = 150$ and initial number of infective individuals $a = 1$. The 95% credible intervals are shown as well. There is a total of 130 infections during the whole epidemic and all the infection times are assumed to be known. The " | " marks in the plot represent the observed data, i.e. the infection times. The squared exponential covariance function is used for the Gaussian process

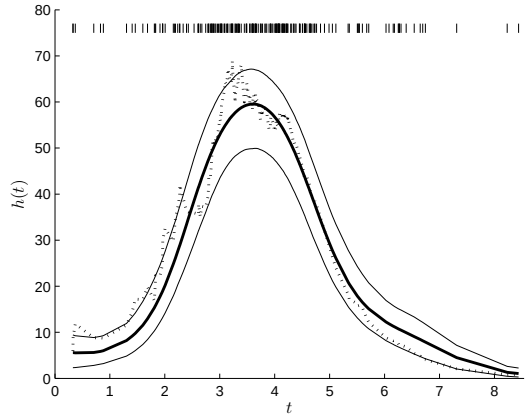


Figure 2.9: Posterior mean of the overall force of infection $h(t)$ (solid line) at each infection time compared with the original data $\beta X_t Y_t$ (dotted line) generated from the general stochastic epidemic with parameters infection rate $\beta = 0.015$, removal rate $\gamma = 1$, initial number of susceptibles $N = 200$ and initial number of infective individuals $a = 1$. The 95% credible intervals are shown as well. There is a total of 189 infections during the whole epidemic and all the infection times are assumed to be known. The " | " marks in the plot represent the observed data, i.e. the infection times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2.

2.3.3 Simulated partial data

We now assume the epidemic data are partially observed, i.e. only removal times are known. Based on the results given in Section 2.3.2 where complete data are assumed to be observed, we apply our methods to SE-Data 2 and SE-Data 3 to perform inference for the overall force of infection, $h(t)$. We first assume the removal rate, γ, is known and then we relax this assumption and use the Gibbs step method (discussed in Chapter 1, Section 1.2.2.2) to perform inference for γ. We used Gamma distribution with mean 10^4 and variance 10^8 providing conjugate priors for the removal rate, γ. In Figure 2.10, we show the estimation results for SE-Data 2 under different assumptions. For the case where we only observe incomplete data, we plot the posterior mean of $h(t)$ at each estimated infection times. Figure 2.10 (b) is under the assumption that we

observe removal times and the removal rate is also known and Figure 2.10 (c) is under the assumption that we only observe removal times. For comparison, we give the estimation results (shown in Figure 2.10 (a)) where we observe complete data. Compared to the case where we observe the complete data, the estimated intensities in Figure 2.10 (b) and (c) recover the truth but with larger uncertainties. The estimated intensities appear to struggle to recover the truth between the time interval, $[2.5, 3]$, but covered by the 95% credible intervals. The larger uncertainties in the case where only removal times are observed can be considered as a consequence of having the inferred infection times. Between Figure 2.10 (b) and (c), we find that the estimated intensity in (c) has much larger uncertainty than in (b) where the removal rate is known, although it appears to roughly capture the feature of the data, $\beta X_t Y_t$, at each infection time. One explanation could be that the estimation of the removal rate makes the uncertainty even larger. Figure 2.11 shows MCMC trace plot of the hyperparameter, θ, the overall force of infection at time $t = 1.52$ and $t = 2.03$ respectively. In Figure 2.12, we have the estimation results for SE-Data 3 under the same assumption settings as for SE-Data 2 and the results recover the truth with different uncertainties.

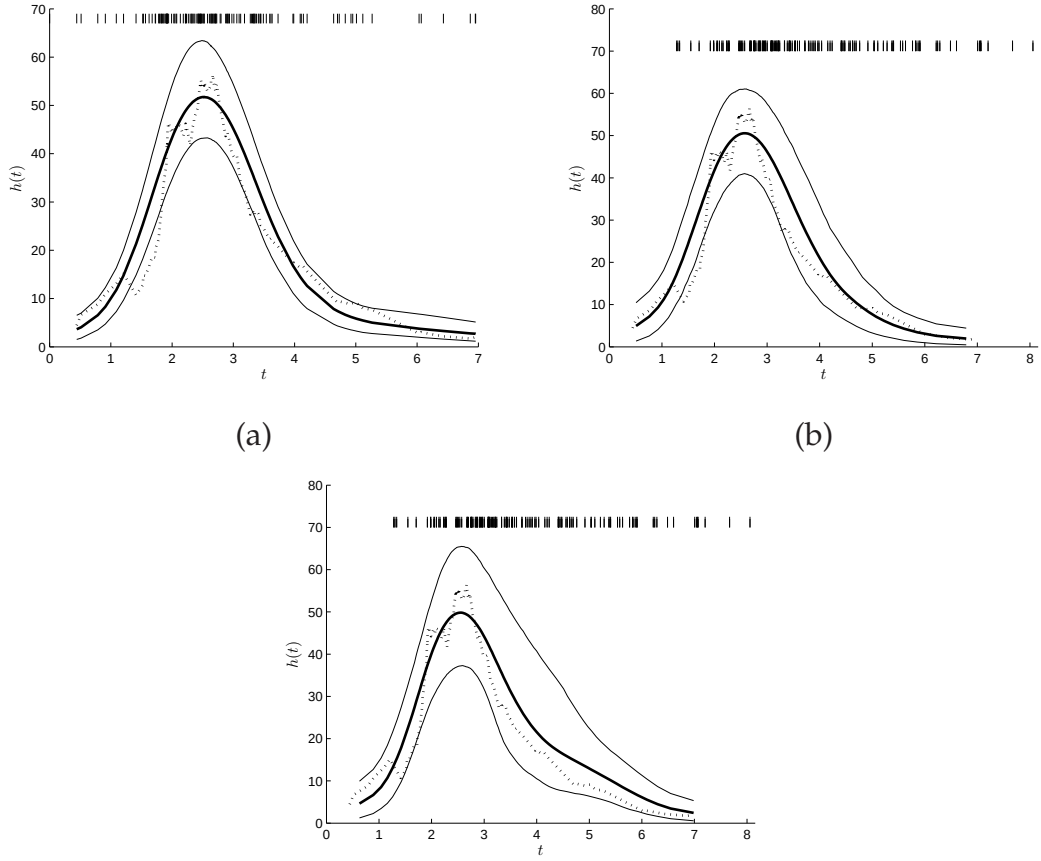


Figure 2.10: Dataset SE-Data 2 recovered by placing the Gaussian process prior. Three plots show posterior mean of the overall force of infection $h(t)$ (solid line) at each infection time compared with the original data $\beta X_t Y_t$ (dotted line) generated from the general stochastic epidemic with parameters $\beta = 0.015$, $\gamma = 1$, $N = 150$ and $a = 1$. Plot (a) corresponds to the case where complete data are observed. Plot (b) corresponds to the case where removal times are known as well as the removal rate. Plot (c) corresponds to the case where only removal times are known. The 95% credible intervals are shown for each of the plot. The " | " marks in each plot represent the observed data, i.e. the infection times for plot (a) and the removal times for plot (b) and (c). There is a total of 130 infections during the whole epidemic. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2.

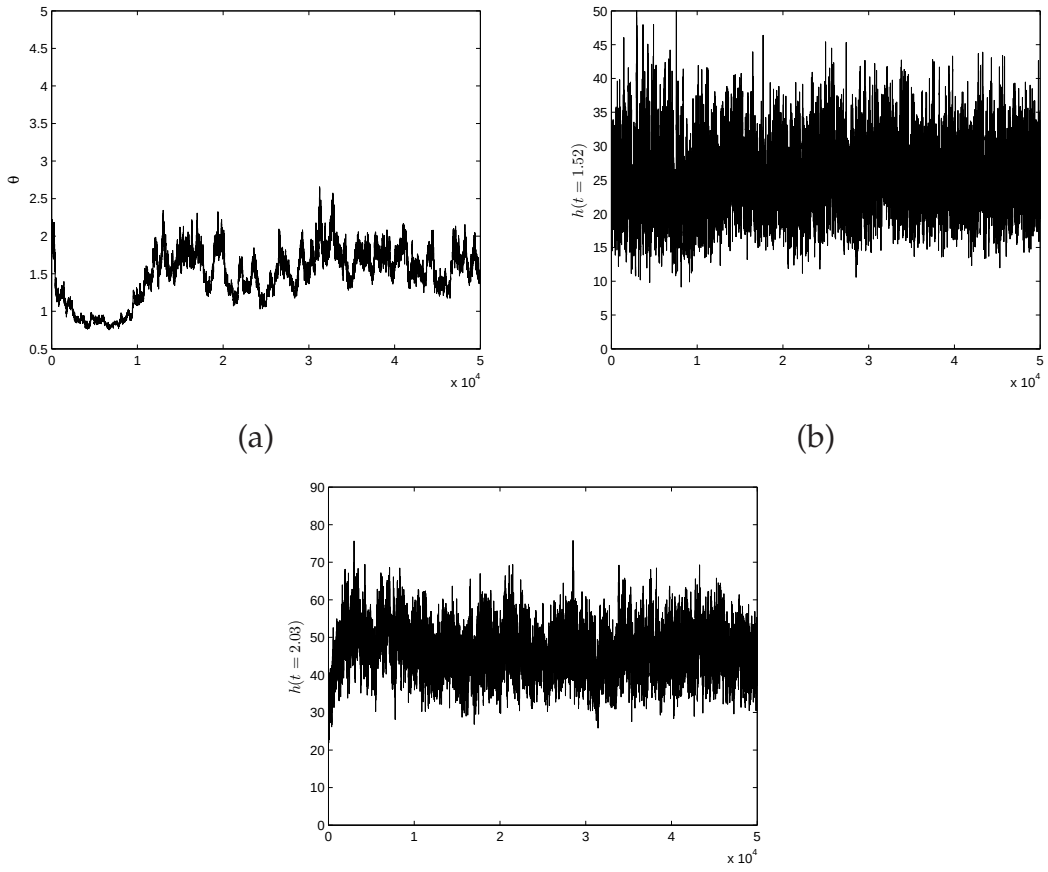


Figure 2.11: Dataset SE-Data 2 recovered by placing the Gaussian process prior where only removal times are known. Plot (a) shows MCMC trace plot of the hyperparameter, θ. Plot (b) and (c) show MCMC trace plot of the overall force of infection at time $t = 1.52$ and $t = 2.03$ respectively. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2.

2.3.4 Smallpox data

We now apply our methods to real life data, Smallpox data obtained from Bailey (1975, p.125). There is a total of 30 cases in a community of 120 individuals at risk. The data are summarised by the 29 time intervals between the detection of cases. O'Neill & Roberts (1999), Eichner & Dietz (2003) and Becker & Yip (1989) have analysed the data and Becker & Yip used a martingale approach un-

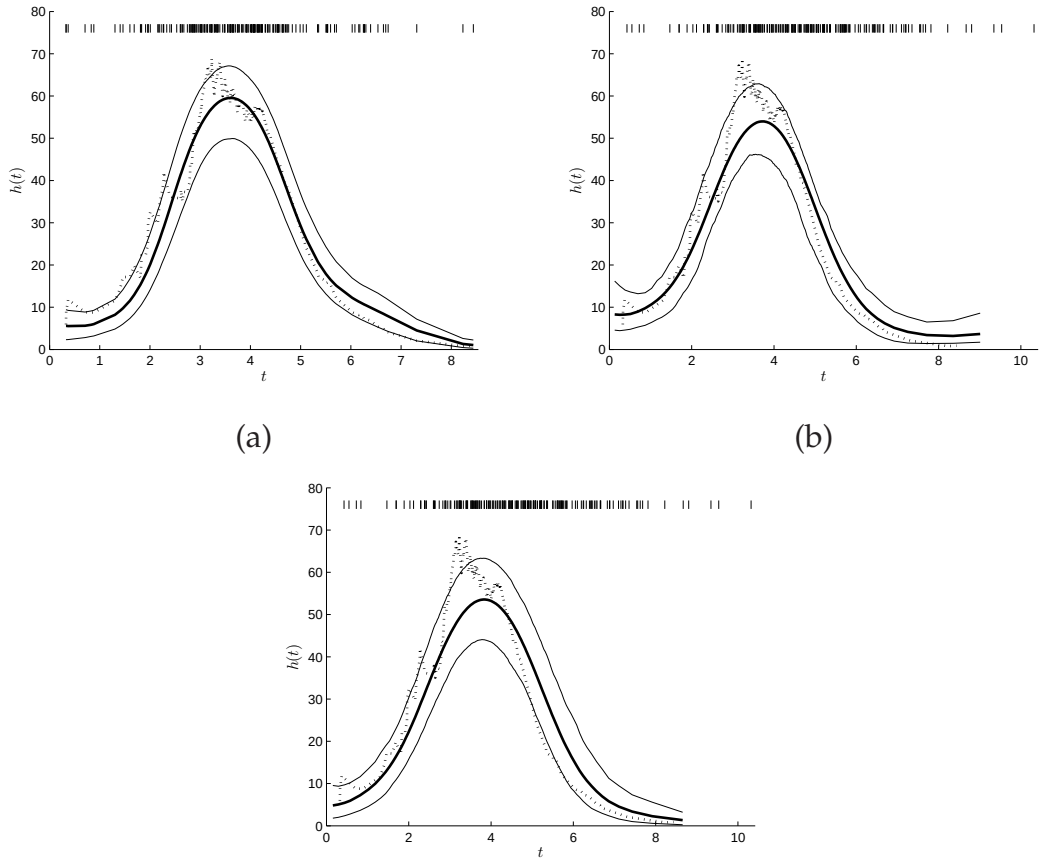


Figure 2.12: Dataset SE-Data 3 recovered by placing the Gaussian process prior. Three plots show posterior mean of the overall force of infection $h(t)$ (solid line) at each infection time compared with the original data $\beta X_t Y_t$ (dotted line) generated from the general stochastic epidemic with parameters $\beta = 0.015$, $\gamma = 1$, $N = 200$ and $a = 1$. Plot (a) corresponds to the case where infection times are known. Plot (b) corresponds to the case where removal times are known as well as the removal rate. Plot (c) corresponds to the case where only removal times are known. The 95% credible intervals are shown for each of the plot. The " | " marks in each plot represent the observed data, i.e. the infection times for plot (a) and the removal times for plot (b) and (c). There is a total of 189 infections during the whole epidemic. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2.

der a nonparametric framework where infection times are assumed to be fixed. However, no one has applied Bayesian nonparametric methods to the Smallpox data. We assume only removal times are known. All the infection times for each individual and the removal rate are inferred by the MCMC methods. We then apply our methods to the data and place a Gaussian process prior using the squared exponential covariance function on the overall force of infection, $h(t)$, at each infection time. Figure 2.15 gives the estimated intensity with 95% credible intervals. In Figure 2.15, we show densities of 4 key quantities of interest including the removal rate, γ, the hyperparameter, θ, the upper bound of the intensity, h^*, and the number of thinned events, M. Table 2.2 gives statistical summaries for these parameters. O'Neill & Roberts (1999) estimated the infection rate, β, and the removal rate, γ in a Bayesian framework under the assumption that only removal times are observed and β is constant. They obtained posterior summaries as follows: (1) β: mean 0.0009, standard deviation 0.00019; (2) γ: mean 0.098, standard deviation 0.021. From Table 2.2, we find that for the removal rate, the mean values are similar, and the standard deviations a little larger. In order to compare the estimates of β, we estimate $\hat{\beta}(t) = \frac{h(t)}{X_t Y_t}$ using our Bayesian nonparametric methods. Figure 2.15 shows the posterior mean of the infection rate, $\hat{\beta}(t)$. Compared to the estimates of β from O'Neill & Roberts, we find a larger mean of the infection rates at each infection time and the estimated functions are not smooth. The latter presumably is due to that the estimator $\hat{\beta}(t)$ is constructed by smooth functions divided by very rough functions which give rough functions. Specifically, h is expected to be a smooth function as we place a Gaussian process prior on it, even if a rough covariance function is used. Whereas, $X_t Y_t$ is a very rough function. Therefore, $\hat{\beta}$ is not a function as smooth as a constant.

From Table 2.2 and Figure 2.13 and 2.15, we conclude that the variability shows that the standard SIR model, with infection rate constant over time, is not appropriate for the data.

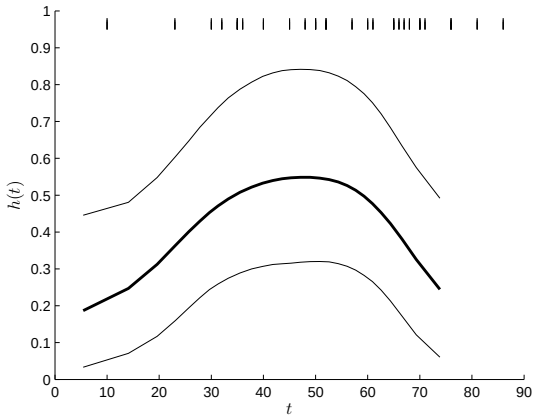


Figure 2.13: Posterior mean of force of infection $h(t)$ (solid line) at each estimated infection time. The 95% credible intervals are shown as well. There is a total of 30 infections during the whole epidemic and all the infection times and the removal rate are assumed to be unknown. Only removal times are known. The " | " marks in the plot represent the observed data, i.e. the removal times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2.

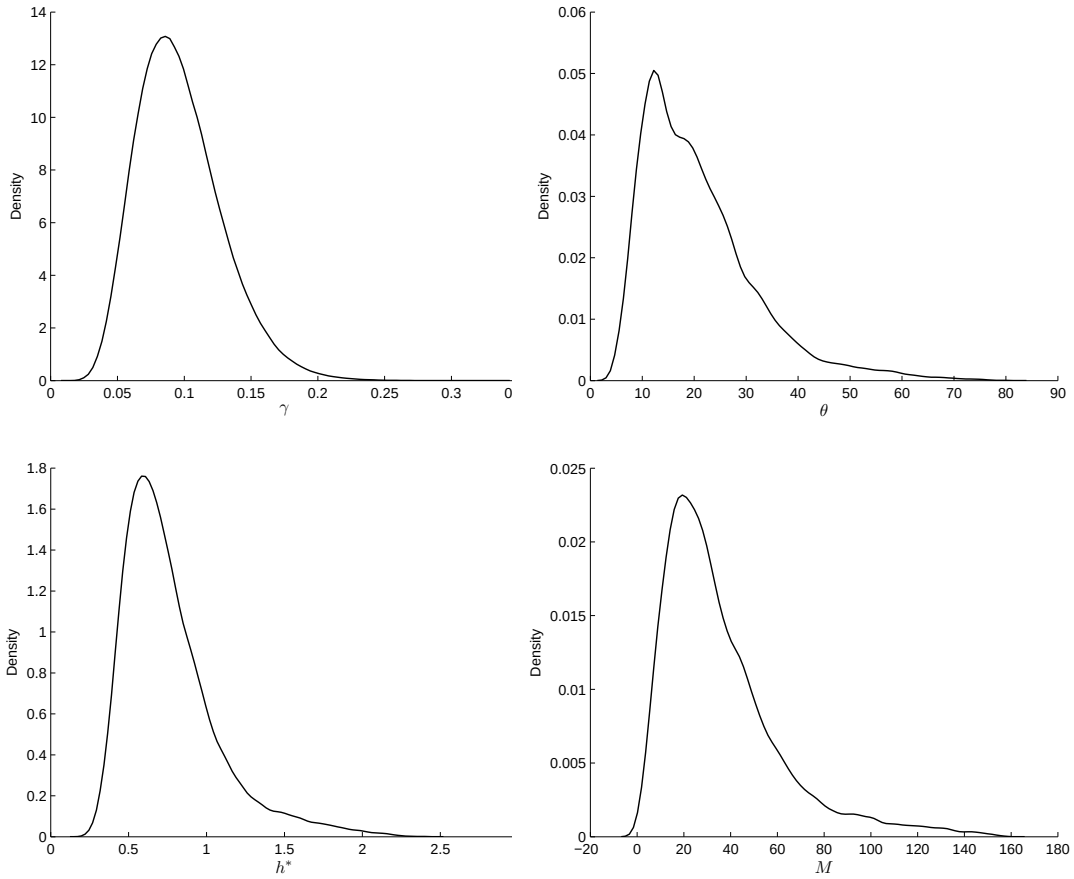


Figure 2.14: Densities of removal rate, γ, hyperparameter, θ, upper bound h^* and number of thinned events, M, for the Smallpox data analysis.

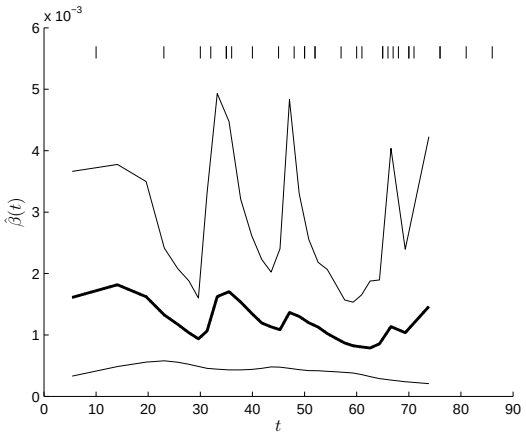


Figure 2.15: Posterior mean of the infection rate, $\hat{\beta}(t)$ at each estimated infection time for the Smallpox data. The 95% credible intervals are shown as well. There is a total of 30 infections during the whole epidemic and all the infection times and the removal rate are assumed to be unknown. Only removal times are known. The " | " marks in the plot represent the observed data, i.e. the removal times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2.

Table 2.2: Mean and standard deviation of the removal rate, γ, the hyperparameter, θ, the upper bound h^* and the number of thinned events, M.

	γ	θ	h^*	M
mean (std)	0.10 (0.03)	20.90 (11.04)	0.77 (0.31)	35.80 (25.23)

2.4 Sensitivity to the Gaussian process prior for the SIR epidemic model

In this section, we are concerned with investigation of how different covariance functions for the Gaussian process prior affect the performance of the inference of $h(t)$. We choose 3 different covariance functions, i.e. squared exponential covariance function, exponential covariance function and Matérn class covariance function with ν set to $3/2$. As discussed in Chapter 1, the squared exponential

covariance function ($\nu \to \infty$) and exponential covariance ($\nu = 1/2$) can be considered as two special cases in the Matérn class covariance functions and the bigger the value of ν is the smoother the Gaussian process will be.

We now apply our methods to SE-Data 2 using the covariance functions discussed above for the Gaussian process prior and perform inference for the overall force of infection, $h(t)$, by assuming we observe the complete data. Figure 2.16 shows the estimated intensities using squared exponential covariance function, exponential covariance function and Matérn class covariance function for the Gaussian process prior respectively for SE-Data 2. Compared to the estimated intensity in Figure 2.16 (a) where the squared exponential covariance function is used, the estimated intensity in Figure 2.16 (b) is quite rough and behaves less smoothly than the one in (a). Compared to the estimated intensity in Figure 2.16 (b), the one in Figure 2.16 (c) behaves much more smoothly. In terms of smoothness, the plots in (a) and (c) are similar. From all of the 3 different estimated intensities, we can conclude that they all recover the truth although they seem not able to capture the dramatic changes in the data, for instance, the change at around time point 1.5 and the bimodal feature in time interval $[2, 3]$. It is notable that the posterior means naturally look smoother than any posterior sample from the Gaussian process. Figure 2.17 (a), (b) and (c) show posterior samples of the overall force of infection on the actual process in the simulation.

We also evaluate our methods for the three different Gaussian process priors for the Smallpox. Figure 2.18 (a), (b) and (c) show the recovered trajectories for the squared exponential covariance function, exponential covariance function and Matérn class covariance function respectively. The plot (a) is much smoother than (b) and (c). We find that different Gaussian process priors produce estimates with different level of smoothness. Apart from this, they do not make much difference.

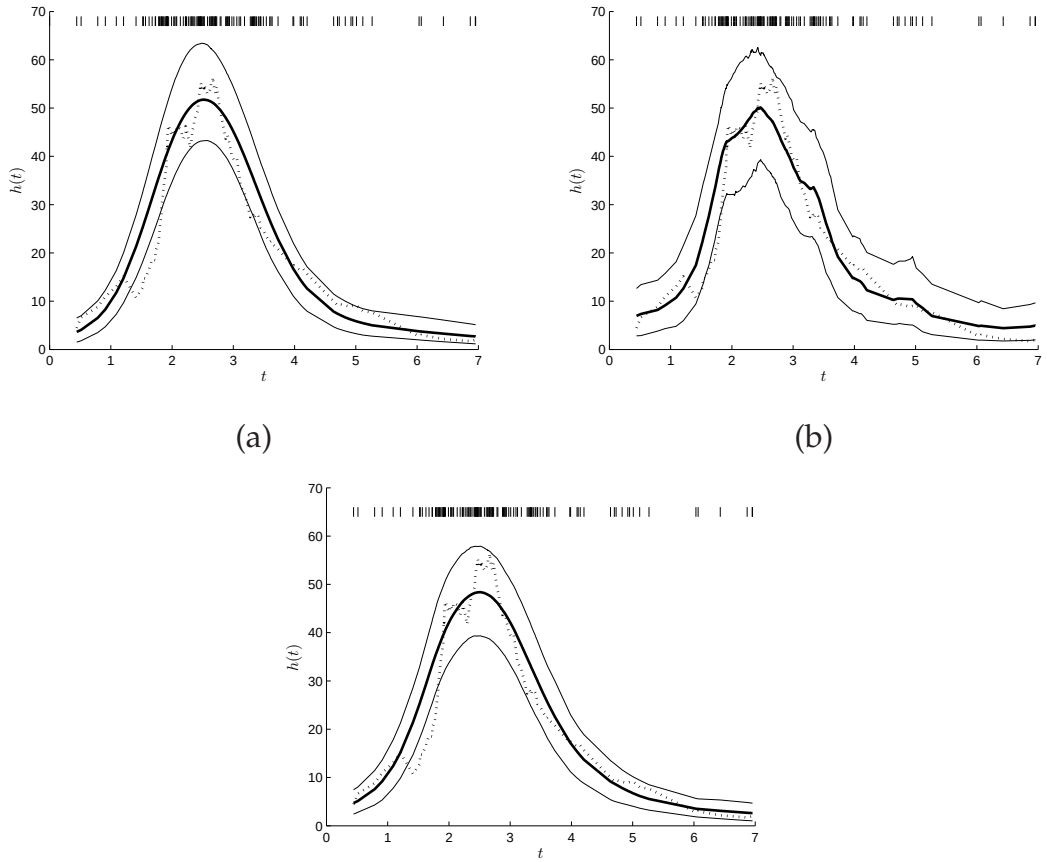


Figure 2.16: Dataset SE-Data 2 recovered by placing three different Gaussian process priors. Three plots show posterior mean of the overall force of infection $h(t)$ (solid line) at each infection time compared with the original data $\beta X_t Y_t$ (dotted line) generated from the general stochastic epidemic with parameters infection rate $\beta = 0.015$, removal rate $\gamma = 1$, initial number of susceptibles $N = 200$ and initial number of infective individuals $a = 1$. Plot (a) corresponds to the squared exponential covariance function see, (1.2). Plot (b) corresponds to the exponential covariance function see, (1.3). Plot (c) corresponds to the Matérn covariance function with ν set to $3/2$ see, (1.4). The 95% credible intervals are shown for each of the plot. The " | " marks in each plot represent the observed data, i.e. the infection times. There is a total of 130 infections during the whole epidemic and all the infection times are assumed to be known. The hyperparameter of the three covariance functions, α, is set to 2.

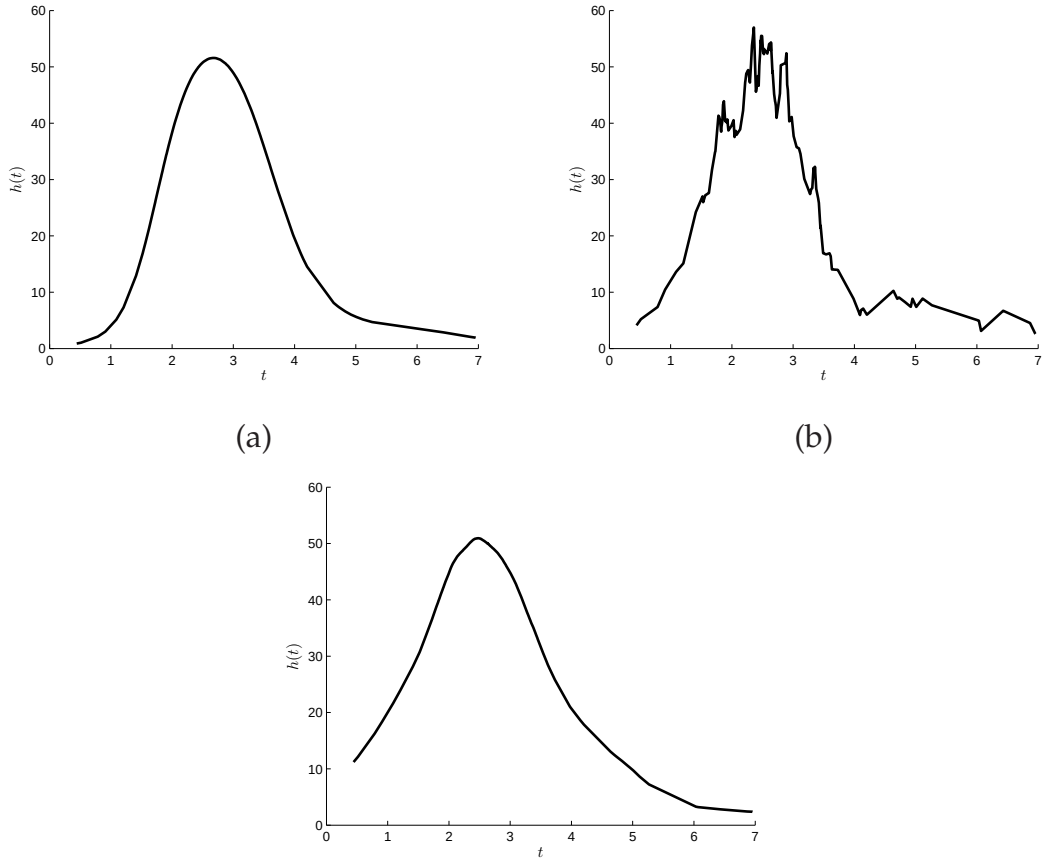


Figure 2.17: Dataset SE-Data 2 recovered by placing three different Gaussian process priors. Three plots show posterior samples of the overall force of infection $h(t)$ (solid line) at each infection time. Plot (a) corresponds to the squared exponential covariance function see, (1.2). Plot (b) corresponds to the exponential covariance function see, (1.3). Plot (c) corresponds to the Matérn covariance function with ν set to $3/2$ see, (1.4). The hyperparameter of the three covariance functions, α, is set to 2.

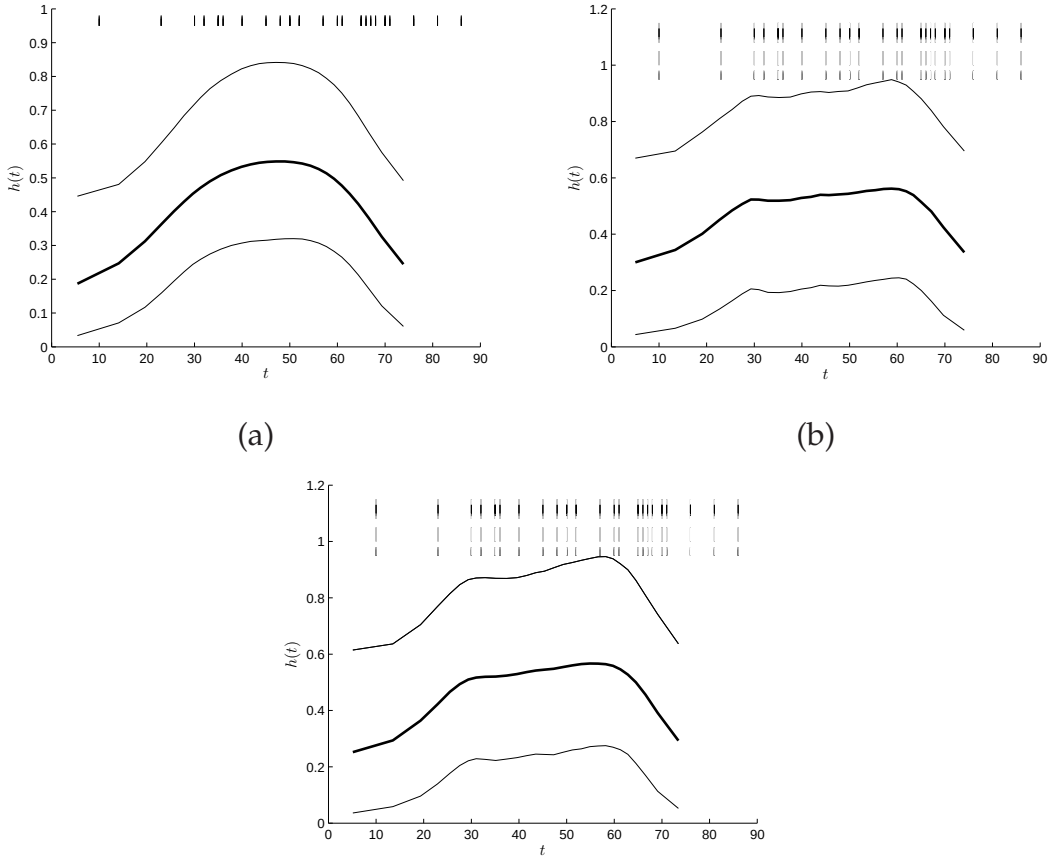


Figure 2.18: Smallpox recovered by placing three different Gaussian process priors. Plot (a) corresponds to the squared exponential covariance function see, (1.2). Plot (b) corresponds to the exponential covariance function see, (1.3). Plot (c) corresponds to the Matérn covariance function with ν set to $3/2$ see, (1.4). The 95% credible intervals are shown for each of the plot. The " | " marks in each plot represent the observed data, i.e. the removal times. There is a total of 30 infections during the whole epidemic and only removal times are known. The hyperparameter of the three covariance functions, α, is set to 2.

2.5 Conclusion

This Chapter firstly described Bayesian nonparametric methods for inhomogeneous Poisson processes introduced by Adams et al. (2009). Then we adapted

the methods and applied our methods to the epidemic data including the simulated data generated from the SIR epidemic model as well as the real outbreak data, the Smallpox data. We found that our methods work fairly well for epidemic models in small populations from the simulated epidemic data although some dramatic changes within a small time period in the data may not be captured since the Gaussian process prior generally produces smooth estimate. For the Smallpox data, we compared our results with the ones obtained from O'Neill & Roberts (1999). We have similar mean and a slightly larger standard deviation in terms of the estimate of the removal rate. In order to compare the estimate of the infection rate, we introduced an estimator, $\hat{\beta}(t) = \frac{h(t)}{X_t Y_t}$. We found that the posterior mean of $\hat{\beta}(t)$ is larger than O'Neill & Roberts' estimate at each estimated infection time although the estimated infection rate from O'Neill & Roberts is fully covered in our 95% credible intervals. The large uncertainty of the estimated infection rate and the large standard deviation of the estimated parameters such as the estimated removal rate etc. indicate that the standard SIR model is not appropriate for the data. Finally, we explored the effects of placing a Gaussian process prior with different covariance functions on the overall force of infection, $h(t)$ at each infection time.

We have successfully developed Bayesian nonparametric methods for the overall force of infection in small-scale epidemics where we assume the overall force of infection has the form, $h(t)$ which does not consider the information of the number of susceptible and infectives. In the standard SIR model, the overall force of infection is often assumed to be of the form $\beta X_t Y_t$. Simulated data are generated based on such parametric assumptions. In the near future, one could further investigate our methods with simulated data which are generated with a modified overall force of infection, e.g. which has a form, $\beta X_t {Y_t}^v$, where $v > 0$.

CHAPTER 3

Bayesian nonparametric estimation for the infection rate in small-scale epidemics

3.1 Introduction

In Chapter 2, we investigated the behaviour of the overall force of infection in the SIR epidemic model by assuming it is of the form $h(t)$, i.e. the overall force of infection is assumed to be a function of time without any particular form on it. We adopted a Bayesian framework to estimate $h(t)$ and therefore we placed a Gaussian process prior on it. The promising estimation results give us motivation to further explore the behaviour of the overall force of infection using Bayesian nonparametric methods.

It is known that in the SIR model, during its infectious period, an individual makes contacts with each of the N susceptibles at times given by the points of a homogeneous Poisson process with rate β, a constant value. Although the assumption that the infection rate, β, is constant over time is a common one in epidemic modelling, this is not always realistic (Fang et al., 2004). Specifically, β could vary over time as a result of factors such as behaviour change in response to the epidemic, the introduction of control or mitigation measures, greater public awareness of the epidemic, and so on. From a modelling perspective, one could also introduce a time-dependent infection rate as a proxy for population homogeneity. For instance the idea that individuals who avoid

infection by a given time t may be less likely to become infected thereafter, could be modelled by letting $\beta = \tilde{\beta}(t)$ be decreasing over time. There has been many studies on estimating time-dependent infection rate of various infectious diseases where epidemic models are fitted using parametric functions for $\tilde{\beta}(t)$ see, (Becker, 1989, p.137, Pollicott et al., 2009, Pollicott et al., 2012, Smirnova & Tuncer, 2014). However, no literature is concerned with the estimation of the infection rate in a Bayesian nonparametric work which further motivates us.

Previously, we used $h(t)$ to represent the overall force of infection and ignored the number of susceptibles and infectives in the population. In this Chapter, we will address the question of estimating $\beta = \tilde{\beta}(t)$ in a Bayesian nonparametric framework and we will consider both the case of complete observations (i.e. both infection and removal times) and partial observations (i.e. just removals observed). In other words, the overall force of infection will have a form, $\tilde{\beta}(t)X_tY_t$. We will also introduce extended multi-group SIR models and estimate key quantities of interest such as infection rate of the epidemic for the models from multi-group epidemic data via Bayesian nonparametric methods and then illustrate our methods with simulated and real life data. Compared to the multi-group model, the single group SIR model is in fact an extreme case that there is only one type of susceptible in the whole population. The datasets SE-Data 1, SE-Data 2, SE-Data 3 and the Smallpox data we studied in Chapter 2 are all single group data. In Section 3.2, we will apply our methods to single group epidemic data. We will then investigate effect of different Gaussian process priors placed on the infection rate in Section 3.3. In Section 3.4, we will describe the multi-group SIR model and apply our methods to the multi-group data. In Section 3.5, we will compare the approaches discussed in Chapter 2 and Chapter 3. Finally, we will give conclusions in Section 3.6.

3.2 Estimation for epidemic models from single group data

In this section, we apply our Bayesian nonparametric methods to single group epidemic data and place a Gaussian process prior on the infection rate, $\tilde{\beta}(t)$. In terms of the observed data, we again perform inference from the complete data

first (i.e. infection and removal times) and then from the partially observed data (i.e. only removal times). We assume that the initial number of susceptibles, N, and the initial number of infectives, a, are available. Before we discuss the inference, let us first define the model where the infection rate is assumed to be a function of time, i.e. $\tilde{\beta}(t)$. According to Chapter 1, Section 1.3.1, The epidemic can be described according to the following transition probabilities:

$$\begin{aligned}
P[X_{t+\delta t} = X_t - 1, Y_{t+\delta t} = X_t + 1 | \mathcal{H}_t] &= \tilde{\beta}(t) X_t Y_t \delta t + o(\delta t), \\
P[X_{t+\delta t} = X_t, Y_{t+\delta t} = Y_t - 1 | \mathcal{H}_t] &= \gamma Y_t \delta t + o(\delta t), \\
P[X_{t+\delta t} = X_t, Y_{t+\delta t} = Y_t | \mathcal{H}_t] &= 1 - \tilde{\beta}(t) X_t Y_t \delta t - \gamma Y_t \delta t + o(\delta t).
\end{aligned}$$

3.2.1 Inference

We now describe a Bayesian nonparametric method to infer the infection rate, $\tilde{\beta}(t)$, given the observed data which consist of a set of removal times. The unobserved infection times are treated as unknown parameters of the model. We assume the epidemic is known to have ceased. Therefore, the number of infections equals to the number of removals. We denote by $\{I_j\}_{j=1}^K$ and $\{\tau_i\}_{i=1}^K$ the ordered infection and removal times of the epidemic respectively. Conditionally on the infection rate, $\tilde{\beta}(t)$, the removal rate, γ, and the first infection time, I_1, we have the density of $(\boldsymbol{\tau} = \{\tau_i\}_{i=1}^K, \mathbf{I} = \{I_j\}_{j=2}^K)$ to be

$$\begin{aligned}
&\pi(\boldsymbol{\tau}, \mathbf{I} | \tilde{\beta}(t), \gamma, I_1) \\
&= \prod_{j=2}^{K} \tilde{\beta}(I_{j^-}) X_{I_{j^-}} Y_{I_{j^-}} \exp\left\{ -\int_{I_1}^{T} \tilde{\beta}(s) X_s Y_s \mathrm{d}s \right\} \times \prod_{i=1}^{K} \gamma Y_{\tau_{i^-}} \exp\left\{ -\int_{I_1}^{T} \gamma Y_s \mathrm{d}s \right\},
\end{aligned} \tag{3.1}$$

where I_{j^-} gives the left limit, so for example $X_{I_{j^-}} = \lim_{s \uparrow\uparrow I_{j^-}} (X_s)$.

We now sample the removal rate, γ, from the desired posterior distribution. According to the algorithms developed by O'Neill & Roberts (1999), we use a Gibbs step to sample from the distribution $\pi(\gamma | \boldsymbol{\tau}, \mathbf{I}, I_1, \tilde{\beta}(t))$. With a gamma prior $(\nu_\gamma, \lambda_\gamma)$ put on γ, we have

$$\pi(\gamma | \boldsymbol{\tau}, \mathbf{I}, I_1, \beta(t)) \propto \Gamma\left(\nu_\gamma + K, \lambda_\gamma + \int_{I_1}^{T} Y_s \mathrm{d}s \right).$$

We can also use a Gibbs step to sample I_1 from the distribution with an Exponential prior, i.e. I_1 has prior density given by $\eta \exp(\eta y) I(y < 0)$, where $\eta > 0$, and $I(\cdot)$ is the indicator function.

$$\pi(y|\boldsymbol{\tau}, \mathbf{I}, \tilde{\beta}(t), \gamma) = \Lambda \exp\{-\Lambda(I_2 - y)\},\ y \in (-\infty, I_2),$$

where $\Lambda = \eta + \gamma + \tilde{\beta}(I_{2^-}) * N$.

We now define an algorithm to sample $\tilde{\beta}(t)$ from a tractable posterior distribution.

We use the SGCP model to describe the infection rate, $\tilde{\beta}(t)$. Specifically, let

$$\tilde{\beta}(t) = \tilde{\beta}^* \sigma(g(t)),$$

$\tilde{\beta}^*$: an upper bound on $\tilde{\beta}(t)$,
$\sigma(\cdot)$: the logistic function, $\sigma(z) = (1 + e^{-z})^{-1}$,
$g(\cdot)$: a random function which has a Gaussian process prior.

For random infinite-dimensional $\tilde{\beta}(t)$ such as the SGCP model, the integral $\int_{I_1}^{T} \tilde{\beta}(s) X_s Y_s \mathrm{d}s$ is not tractable. According to Bayes' Theorem, we write the posterior distribution over $\tilde{\beta}(t)$ using the SGCP model as a prior:

$$\begin{aligned}
&\pi(\tilde{\boldsymbol{\beta}}|\boldsymbol{\tau}, \mathbf{I}, I_1, \gamma) = \pi(\mathbf{g}|\tilde{\beta}^*, \boldsymbol{\tau}, \mathbf{I}, I_1, \gamma) \\
&= \frac{\mathcal{GP}(\mathbf{g}) \prod_{j=2}^{K} \tilde{\beta}^* \sigma(g(I_{j^-})) X_{I_{j^-}} Y_{I_{j^-}} e^{-\int_{I_1}^{T} \tilde{\beta}^* \sigma(g(s)) X_s Y_s \mathrm{d}s} \times \prod_{i=1}^{K} \gamma Y_{\tau_{i^-}} e^{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s}}{\int \mathcal{GP}(\mathbf{g}) \prod_{j=2}^{K} \tilde{\beta}^* \sigma(g(I_{j^-})) X_{I_{j^-}} Y_{I_{j^-}} e^{-\int_{I_1}^{T} \tilde{\beta}^* \sigma(g(s)) X_s Y_s \mathrm{d}s} \times \prod_{i=1}^{K} \gamma Y_{\tau_{i^-}} e^{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s} \mathrm{d}\mathbf{g}},
\end{aligned} \tag{3.2}$$

where $\mathbf{g}$ denotes the infinite-dimensional object corresponding to $g(t)$. The integral over $[I_1, T]$ in the numerator and the integral over $\mathbf{g}$ in the denominator are both intractable. In particular, the value of the integral over $\mathbf{g}$ in fact depends on $\mathbf{g}$ and is no longer a constant. In other words, standard MCMC methods are unable to tackle this problem.

We now augment the posterior distribution (3.2) considering the infection rate, $\tilde{\beta}(t)$, as an inhomogeneous Poisson process intensity. Specifically, consider-

ing augmenting the observed data with the following variables: the number of thinned events, M, the locations of thinned events, $\{\tilde{I}_s\}_{s=1}^{M}$, the function values at the infection times, $\mathbf{g}_K = (g(I_{2^-}), g(I_{3^-}), \cdots, g(I_{K^-}))$ and the function values at the thinned events, $\mathbf{g}_M = (g(\tilde{I}_{1^-}), g(\tilde{I}_{2^-}), \cdots, g(\tilde{I}_{M^-}))$, we can sample the above from the joint distribution below.

$$
\begin{aligned}
&\pi(\boldsymbol{\tau}, \mathbf{I}, M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_{M+K} | \tilde{\beta}^*, \gamma, T, \theta) \\
&\propto \prod_{j=2}^{K} \tilde{\beta}^* X_{I_{j^-}} Y_{I_{j^-}} \prod_{j=2}^{K} \sigma(g(I_{j^-})) \prod_{s=1}^{M} \tilde{\beta}^* X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^{M} \sigma(-g(\tilde{I}_{s^-})) \exp\{-\int_{I_1}^{T} \tilde{\beta}^* X_s Y_s \mathrm{d}s\} \\
&\times \prod_{i=1}^{K} \gamma Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K} | M, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta),
\end{aligned}
$$

where $\mathbf{g}_{M+K}$ denotes a vector consisting of $\mathbf{g}_M$ and $\mathbf{g}_K$. As the first infection time, I_1, is not modelled in the SIR epidemic model, we have in fact $K-1$ function values at the infection times.

3.2.1.1 MCMC algorithms

1. Sampling I

We now impute the rest of the infection times, $\mathbf{I}$, using a Metropolis-Hastings algorithm and sample from the distribution below:

$$
\begin{aligned}
&\pi(\mathbf{I}, \mathbf{g}_K | \boldsymbol{\tau}, M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M, \tilde{\beta}^*, \gamma, T, \theta) \\
&\propto \prod_{j=2}^{K} X_{I_{j^-}} Y_{I_{j^-}} \prod_{j=2}^{K} \sigma(g(I_{j^-})) \exp\{-\int_{I_1}^{T} \tilde{\beta}^* X_s Y_s \mathrm{d}s\} \\
&\times \prod_{i=1}^{K} Y_{\tau_{i^-}} \exp\{-\int_{I_1}^{T} \gamma Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K} | M, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta).
\end{aligned}
$$

We first uniformly draw a new location I_j' from $[I_1, T]$ and then draw a new function value $\mathbf{g}(I_{j^-}')$ from the distribution, $\pi(\mathbf{g}(I_{j^-}') | I_j', \mathbf{g}_{M+K}, \mathbf{X}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)$, and $K(I_j', \mathbf{X}) K(\mathbf{X}, \mathbf{X})^{-1} \mathbf{g}_{M+K}$ and $K(I_j', I_j') - K(I_j', \mathbf{X}) K(\mathbf{X}, \mathbf{I})^{-1} (\mathbf{X}, I_j')$ are the mean and

variance of the distribution respectively, where $\mathbf{X} = (I_2, I_3, \cdots, I_K, \tilde{I}_1, \tilde{I}_2, \cdots, \tilde{I}_M)$ denotes locations of infection times and thinned events. Then one of the existing infection times is selected uniformly at random and replaced by the proposed new values. The forward and backward densities are:

$$q(I_j \to I_j') = \frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I_{j^-}')|I_j', \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta),$$

$$q(I_j' \to I_j) = \frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I_{j^-})|I_j, \mathbf{g}_{M+K}', \{I_j'\}_{j=2}^K, \{\tilde{I}_s\}_{s=1}^M, \theta).$$

We abbreviate $\pi(\mathbf{I}, \mathbf{g}_K|\boldsymbol{\tau}, M, \{\tilde{I}_s\}_{s=1}^M, \mathbf{g}_M, \tilde{\beta}^*, \gamma, T, \theta)$ by $\pi(\mathbf{I}, \mathbf{g}_K)$. Then the acceptance ratio is

$$\begin{aligned}
&\frac{q(I_j' \to I_j)}{q(I_j \to I_j')} \times \frac{\pi(\{I_j'\}_{j=2}^K, \mathbf{g}_K')}{\pi(\{I_j\}_{j=2}^K, \mathbf{g}_K)} \\
&= \frac{\frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I_{j^-})|I_j, \mathbf{g}_{M+K}', \{I_j'\}_{j=2}^K, \{\tilde{I}_s\}_{s=1}^M, T, \theta)}{\frac{1}{K-1} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(I_{j^-}')|I_j', \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, T, \theta)} \\
&\times \frac{\prod_{j=2}^K X_{I_{j^-}}' Y_{I_{j^-}}' \times \prod_{j=2}^K \sigma(g(I_j')) \exp\{-\int_{I_1}^T \tilde{\beta}^* X_s' Y_s' \mathrm{d}s\}}{\prod_{j=2}^K X_{I_{j^-}} Y_{I_{j^-}} \prod_{j=2}^K \sigma(g(I_j)) \exp\{-\int_{I_1}^T \tilde{\beta}^* X_s Y_s \mathrm{d}s\}} \\
&\times \frac{\prod_{i=1}^K Y_{\tau_{i^-}}' \exp\{-\int_{I_1}^T \gamma Y_s' \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K}'|\{I_j'\}_{j=2}^K, \{\tilde{I}_s\}_{s=1}^M, \theta)}{\prod_{i=1}^K Y_{\tau_{i^-}} \exp\{-\int_{I_1}^T \gamma Y_s \mathrm{d}s\} \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta)}.
\end{aligned}$$

As

$$\begin{aligned}
&\pi(\mathbf{g}_{M+K}'|\{I_j'\}_{j=2}^K, \{\tilde{I}_s\}_{s=1}^M, \theta) \times \pi(\mathbf{g}(I_{j^-})|I_j, \mathbf{g}_{M+K}', \{I_j'\}_{j=2}^K, \{\tilde{I}_s\}_{s=1}^M, \theta) \\
&= \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta) \times \pi(\mathbf{g}(I_{j^-}')|I_j', \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta),
\end{aligned}$$

the acceptance ratio is simplified to

$$\frac{\prod_{j=2}^K X_{I_{j^-}}' Y_{I_{j^-}}' \times \sigma(g(I_j')) \exp\{-\int_{I_1}^T \tilde{\beta}^* X_s' Y_s' \mathrm{d}s\}}{\prod_{j=2}^K X_{I_{j^-}} Y_{I_{j^-}} \sigma(g(I_j)) \exp\{-\int_{I_1}^T \tilde{\beta}^* X_s Y_s \mathrm{d}s\}} \times \frac{\prod_{i=1}^K Y_{\tau_{i^-}}' \exp\{-\int_{I_1}^T \gamma Y_s' \mathrm{d}s\}}{\prod_{i=1}^K Y_{\tau_{i^-}} \exp\{-\int_{I_1}^T \gamma Y_s \mathrm{d}s\}}.$$

We now sample $\tilde{\beta}^*$, M, $\{\tilde{I}_s\}_{s=1}^M$, $\mathbf{g}_{M+K}$ and the hyperparameter, θ, respectively from their conditional posterior distributions.

2. Sampling $\tilde{\beta}^*$

According to the full conditional density for $\tilde{\beta}^*$

$$\pi(\tilde{\beta}^*|M, T, \{I_j\}_{j=1}^K) \propto (\tilde{\beta}^*)^{K+M-1} \exp\{-\int_{I_1}^T \tilde{\beta}^* X_s Y_s \mathrm{d}s\} \pi(\tilde{\beta}^*),$$

where $\pi(\tilde{\beta}^*)$ is a prior on $\tilde{\beta}^*$. With a conditionally-conjugate gamma prior, $\Gamma(\nu_{\tilde{\beta}}, \lambda_{\tilde{\beta}})$, put on $\tilde{\beta}^*$, we use a Gibbs step to update $\tilde{\beta}^*$. We can sample from the full conditional distribution, $\Gamma(\nu_{\tilde{\beta}} + K + M - 1, \lambda_{\tilde{\beta}} + \int_{I_1}^{T} X_s Y_s \mathrm{d}s)$.

3. Sampling M

We use a Metropolis-Hastings step to sample from the full conditional distribution of the number of thinned events below

$$\begin{aligned}
&\pi(M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M | \tilde{\beta}^*, \mathbf{g}_K, \mathbf{I}, \theta, T, I_1) \\
&\propto (\tilde{\beta}^*)^M \prod_{s=1}^{M} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^{M} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta).
\end{aligned}$$

First, select either an insertion move or a deletion move, each with probability $\frac{1}{2}$. For the insertion move, we propose a new location $\tilde{I}'_s$ which is uniformly sampled from $[I_1, T]$ and then draw a corresponding function value $\mathbf{g}(\tilde{I}'_{s^-})$ from the conditioned multivariate Gaussian distribution at $\tilde{I}'_s$. Specifically, the distribution is $\pi(\mathbf{g}(\tilde{I}'_{s^-}) | \tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)$, and the mean and variance of this distribution are $K(\tilde{I}'_s, \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{g}_{M+K}$ and $K(\tilde{I}'_s, \tilde{I}'_s) - K(\tilde{I}'_s, \mathbf{I})K(\mathbf{I}, \mathbf{X})^{-1}K(\mathbf{X}, \tilde{I}'_s)$ respectively. The new value can be seen as a prediction of the Gaussian process. For the deletion move, a thinned event $\tilde{I}_s$, is uniformly selected and removed from the current M events. Below is the derivation of the probability of acceptance for the insertion and deletion move.

- Insertion move:

The forward and backward densities are:

$$q(M \to M') = \frac{1}{2} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(\tilde{I}'_{s^-}) | \tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta),$$

$$q(M' \to M) = \frac{1}{2} \times \frac{1}{M'}.$$

The acceptance ratio is

$$\begin{aligned}
&\frac{q(M' \to M)}{q(M \to M')} \times \frac{\pi(M', \{\tilde{I}_s\}_{s=1}^{M'}, \mathbf{g}'_M | \mathbf{I}, \mathbf{g}_K, \theta, \tilde{\beta}^*)}{\pi(M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M | \mathbf{I}, \mathbf{g}_K, \theta, \tilde{\beta}^*)} \\
&= \frac{\frac{1}{2} \times \frac{1}{M'}}{\frac{1}{2} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(\tilde{I}'_{s^-}) | \tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)} \\
&\times \frac{(\tilde{\beta}^*)^{M'} \prod_{s=1}^{M'} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^{M'} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}'_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{(\tilde{\beta}^*)^{M} \prod_{s=1}^{M} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^{M} \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K} | \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.
\end{aligned}$$

As $M' = M + 1$ and

$$\begin{aligned}&\pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)\\ &= \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta) \times \pi(\mathbf{g}(\tilde{I}'_{s^-})|\tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta),\end{aligned}$$

the acceptance ratio is simplified to $\frac{(T-I_1)\times\tilde{\beta}^*\times\sigma(-g(\tilde{I}'_{s^-}))\times X_{\tilde{I}'_{s^-}} Y_{\tilde{I}'_{s^-}}}{M+1}$.

- Deletion Move:

The forward and backward densities are:

$$q(M\to M') = \frac{1}{2}\times\frac{1}{M},$$

$$q(M'\to M) = \frac{1}{2}\times\frac{1}{T-I_1}\times\pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta).$$

The acceptance ratio is

$$\begin{aligned}&\frac{q(M'\to M)}{q(M\to M')}\times\frac{\pi(M', \{\tilde{I}_s\}_{s=1}^{M'}, \mathbf{g}'_M|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_K, \theta, \tilde{\beta}^*)}{\pi(M, \{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \mathbf{g}_K, \theta, \tilde{\beta}^*)}\\ &= \frac{\frac{1}{2}\times\frac{1}{T-I_1}\times\pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{\frac{1}{2}\times\frac{1}{M}}\\ &\times\frac{(\tilde{\beta}^*)^{M'}\prod_{s=1}^{M'} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}}\prod_{s=1}^{M'}\sigma(-g(\tilde{I}_{s^-}))\times\pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{(\tilde{\beta}^*)^{M}\prod_{s=1}^{M} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}}\prod_{s=1}^{M}\sigma(-g(\tilde{I}_{s^-}))\times\pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)}.\end{aligned}$$

As $M' = M - 1$ and

$$\begin{aligned}&\pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta)\\ &= \pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta) \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta),\end{aligned}$$

the acceptance ratio is simplified to $\frac{M}{(T-I_1)\times\tilde{\beta}^*\times\sigma(-g(\tilde{I}_{s^-}))\times X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}}}$.

4. Sampling $\{\tilde{I}_s\}_{s=1}^{M}$

We use a Metropolis-Hastings step to perform sampling from the full conditional distribution of locations of the thinned events below.

$$\begin{aligned}&\pi(\{\tilde{I}_s\}_{s=1}^{M}, \mathbf{g}_M|\tilde{\beta}^*, \mathbf{g}_K, \mathbf{I}, M, \theta, T, I_1)\\ &\propto\prod_{s=1}^{M} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}}\prod_{s=1}^{M}\sigma(-g(\tilde{I}_{s^-}))\times\pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M}, \theta).\end{aligned}$$

We first uniformly draw a new location $\tilde{I}'_s$ from $[I_1, T]$ and then draw a new function value $\mathbf{g}(\tilde{I}'_{s^-})$ from the Gaussian distribution, conditioned on the current $\mathbf{g}_{M+K}$. More precisely, the distribution is $\pi(\mathbf{g}(\tilde{I}'_{s^-})|\tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta)$, and the mean and variance of the distribution are $K(\tilde{I}'_s, \mathbf{X})K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{g}_{M+K}$ and $K(\tilde{I}'_s, \tilde{I}'_s) - K(\tilde{I}'_s, \mathbf{I})K(\mathbf{X}, \mathbf{X})^{-1}K(\mathbf{X}, \tilde{I}'_s)$ respectively. One of the existing thinned events is then selected uniformly at random and replaced by the proposed new values. The forward and backward densities are:

$$q(\tilde{I}_s \to \tilde{I}'_s) = \frac{1}{M} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(\tilde{I}'_{s^-})|\tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta),$$

$$q(\tilde{I}'_s \to \tilde{I}_s) = \frac{1}{M'} \times \frac{1}{T - I_1} \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta).$$

The acceptance ratio is

$$\begin{aligned}
&\frac{q(\tilde{I}'_s \to \tilde{I}_s)}{q(\tilde{I}_s \to \tilde{I}'_s)} \times \frac{\pi(\{\tilde{I}_s\}_{s=1}^{M'}, \mathbf{g}'_M|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \mathbf{g}_K, \theta)}{\pi(\{\tilde{I}_s\}_{s=1}^M, \mathbf{g}_M|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \mathbf{g}_K, \theta)} \\
&= \frac{\frac{1}{M'} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{\frac{1}{M} \times \frac{1}{T-I_1} \times \pi(\mathbf{g}(\tilde{I}'_{s^-})|\tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta)} \\
&\times \frac{\prod_{s=1}^{M-1} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \times \frac{X_{\tilde{I}'_{s^-}} Y_{\tilde{I}'_{s^-}}}{X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}}} \prod_{s=1}^M \sigma(-g(\tilde{I}_s)) \times \frac{\sigma(-g(\tilde{I}'_s))}{\sigma(-g(\tilde{I}_s))} \times \pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta)}{\prod_{s=1}^{M-1} X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \prod_{s=1}^M \sigma(-g(\tilde{I}_s)) \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta)}.
\end{aligned}$$

As $M' = M$ and

$$\begin{aligned}
&\pi(\mathbf{g}'_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta) \times \pi(\mathbf{g}(\tilde{I}_{s^-})|\tilde{I}_s, \mathbf{g}'_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^{M'}, \theta) \\
&= \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta) \times \pi(\mathbf{g}(\tilde{I}'_{s^-})|\tilde{I}'_s, \mathbf{g}_{M+K}, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta),
\end{aligned}$$

the acceptance ratio is simplified to $\frac{X_{\tilde{I}'_{s^-}} Y_{\tilde{I}'_{s^-}} \times \sigma(-g(\tilde{I}'_{s^-}))}{X_{\tilde{I}_{s^-}} Y_{\tilde{I}_{s^-}} \times \sigma(-g(\tilde{I}_{s^-}))}$.

5. Sampling $\mathbf{g}_{M+K}$

Given the infection times $\mathbf{I}$, the locations of thinned events $\{\tilde{I}_s\}_{s=1}^M$ and the current hyperparameter θ, we wish to sample from the full conditional distribution of function values, $\mathbf{g}_{M+K}$, below.

$$\begin{aligned}
&\pi(\mathbf{g}_{M+K}|M, \mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta) \\
&\propto \prod_{j=2}^K \sigma(g(I_{j^-})) \prod_{s=1}^M \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K}|\mathbf{I}, \{\tilde{I}_s\}_{s=1}^M, \theta).
\end{aligned}$$

By using the underrelaxed MCMC method discussed in Chapter 2, Section 2.2.5, we propose new function values $\mathbf{g}'_{M+K}$ by a proposal of the form $\mathbf{g}'_{M+K} = \delta\mathbf{g}_{M+K} + \sqrt{1-\delta^2}\mathbf{h}_{M+K}$, where $\mathbf{h}_{M+K}$ is drawn from the Gaussian process at $\mathbf{I}$ and $\{\tilde{I}_s\}_{s=1}^M$ and δ is in $[0,1)$.

We shall abbreviate

$$\pi(\mathbf{g}_{M+K}|M,\mathbf{I},\{\tilde{I}_s\}_{s=1}^M,\theta)$$

by $\pi(\mathbf{g}_{M+K})$ and

$$\pi(\mathbf{g}'_{M+K}|M,\mathbf{I},\{\tilde{I}_s\}_{s=1}^M,\theta)$$

by $\pi(\mathbf{g}'_{M+K})$. Then the acceptance ratio is

$$\begin{aligned}
&\frac{\pi(\mathbf{g}'_{M+K})}{\pi(\mathbf{g}_{M+K})} \times \frac{q(\mathbf{g}'_{M+K}\to\mathbf{g}_{M+K})}{q(\mathbf{g}_{M+K}\to\mathbf{g}'_{M+K})} \\
&= \frac{\prod_{j=2}^K \sigma(g'(I_{j^-}))\prod_{s=1}^M \sigma(-g'(\tilde{I}_{s^-})) \times \pi(\mathbf{g}'_{M+K}|\mathbf{I},\{\tilde{I}_s\}_{s=1}^M,\theta)}{\prod_{j=2}^K \sigma(g(I_{j^-}))\prod_{s=1}^M \sigma(-g(\tilde{I}_{s^-})) \times \pi(\mathbf{g}_{M+K}|\mathbf{I},\{\tilde{I}_s\}_{s=1}^M,\theta)} \\
&\times \frac{\pi(\mathbf{g}_{M+K}|\mathbf{I},\{\tilde{I}_s\}_{s=1}^M,\theta)}{\pi(\mathbf{g}'_{M+K}|\mathbf{I},\{\tilde{I}_s\}_{s=1}^M,\theta)} \\
&= \frac{\prod_{j=2}^K \sigma(g'(I_{j^-}))\prod_{s=1}^M \sigma(-g'(\tilde{I}_{s^-}))}{\prod_{j=2}^K \sigma(g(I_{j^-}))\prod_{s=1}^M \sigma(-g(\tilde{I}_{s^-}))}.
\end{aligned}$$

6. Sampling θ

As we fix the hyperparameter α to 1 for our case, we only need to sample from the full conditional distribution of the hyperparameter θ shown below.

$$\begin{aligned}
&\pi(\theta|M,\mathbf{g}_{M+K},\mathbf{I},\{\tilde{I}_s\}_{s=1}^M) \\
&\propto \pi(\mathbf{g}_{M+K}|\mathbf{I},\{\tilde{I}_s\}_{s=1}^M,\theta) \times \pi(\theta)
\end{aligned}$$

We assign an exponential prior on θ, i.e. $\theta\sim\text{Exp}(\lambda_\theta)$. Then we propose a new θ' from a normal distribution with mean, the current value of θ and variance, σ^2.

The Metropolis-Hastings acceptance ratio for this proposal is

$$\begin{aligned}
&\frac{\pi(\theta'|M,\mathbf{g}_{M+K},\mathbf{I},\{\tilde{I}_s\}_{s=1}^{M})}{\pi(\theta|M,\mathbf{g}_{M+K},\mathbf{I},\{\tilde{I}_s\}_{s=1}^{M})}\times\frac{q(\theta'\to\theta)}{q(\theta\to\theta')}\\
&=\frac{\pi(\mathbf{g}_{M+K}|\mathbf{I},\{\tilde{I}_s\}_{s=1}^{M},\theta')\times\pi(\theta')}{\pi(\mathbf{g}_{M+K}|\mathbf{I},\{\tilde{I}_s\}_{s=1}^{M},\theta)\times\pi(\theta)}\times\frac{\exp(-(\theta'-\theta)^2/2\sigma^2)}{\exp(-(\theta-\theta')^2/2\sigma^2)}\\
&=\frac{\pi(\mathbf{g}_{M+K}|\mathbf{I},\{\tilde{I}_s\}_{s=1}^{M},\theta')\times\lambda_\theta\exp(-\lambda_\theta\theta')}{\pi(\mathbf{g}_{M+K}|\mathbf{I},\{\tilde{I}_s\}_{s=1}^{M},\theta)\times\lambda_\theta\exp(-\lambda_\theta\theta)}\\
&=\frac{|\Sigma_{\theta'}|^{-\frac{1}{2}}\times\exp(-\mathbf{g}_{M+K}{}^T\Sigma_{\theta'}{}^{-1}\mathbf{g}_{M+K}/2-\lambda_\theta\theta')}{|\Sigma_\theta|^{-\frac{1}{2}}\times\exp(-\mathbf{g}_{M+K}{}^T\Sigma_\theta{}^{-1}\mathbf{g}_{M+K}/2-\lambda_\theta\theta)}.
\end{aligned}$$

For the MCMC algorithms discussed above, the standard deviations of the proposals were tuned to obtain an acceptance rate of 20-30%.

3.2.2 Simulated complete single group data

3.2.2.1 Constant infection rate

We first apply our methods to SE-Data 1, SE-Data 2 and SE-Data 3 which we simulated in Chapter 2 and assume that a set of infection times are available. We choose the squared exponential covariance function for the Gaussian process prior and assume that the hyperparameter, α, is fixed to 1. It is known that datasets SE-Data 1, SE-Data 2 and SE-Data 3 are all generated with a constant infection rate. We shall expect that the posterior means of $\tilde{\beta}(t)$ are fairly flat. Figure 3.1, 3.2 and 3.3 show the posterior means of infection rate at each infection time for SE-Data 1, SE-Data 2 and SE-Data 3 respectively. As expected, the estimated infection rates, $\tilde{\beta}(t)$, at each infection time are close to the truth and appear to be quite smooth and flat. It is notable that in Figure 3.2, the 95% credible intervals become wider from time point 4. One explanation could be that we have far fewer observations between $[4,7]$ than the one between $[2,4]$. The same reason also applies to the estimate in Figure 3.3.

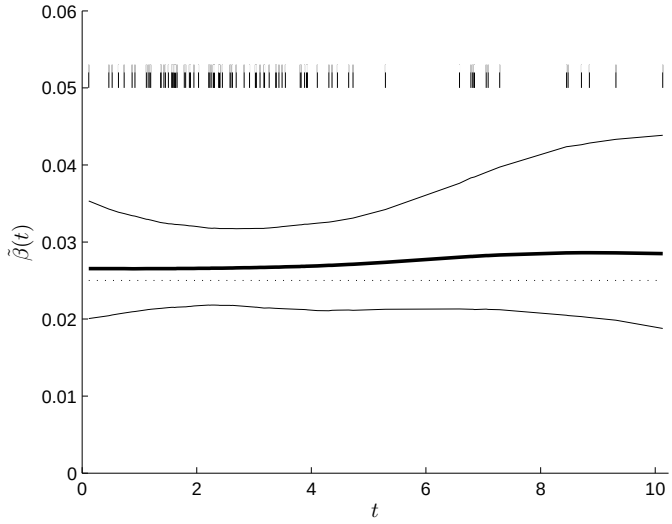


Figure 3.1: Posterior mean of the infection rate $\tilde{\beta}(t)$ (solid line) at each infection time for SE-Data 1. The 95% credible intervals are shown as well. True infection rate (dotted line), $\beta = 0.025$. There is a total of 87 infections during the whole epidemic and all the infection times are assumed to be known. The " | " marks in the plot represent the observed data, i.e. the infection times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

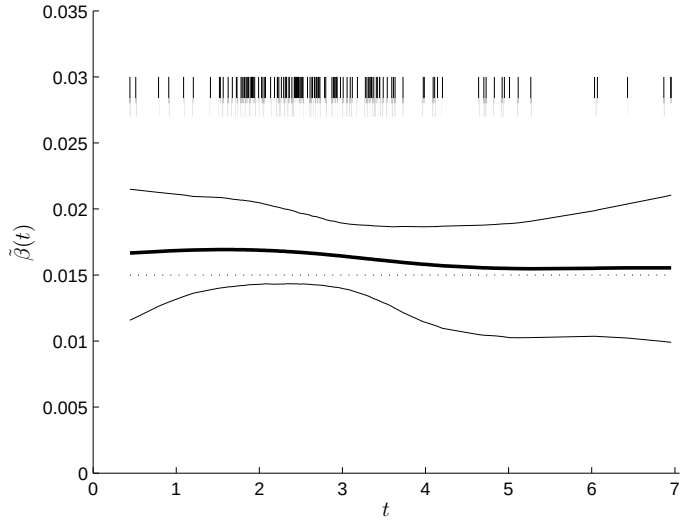


Figure 3.2: Posterior mean of the infection rate $\tilde{\beta}(t)$ (solid line) at each infection time for SE-Data 2. The 95% credible intervals are shown as well. True infection rate (dotted line), $\beta = 0.015$. There is a total of 130 infections during the whole epidemic and all the infection times are assumed to be known. The " | " marks in the plot represent the observed data, i.e. the infection times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

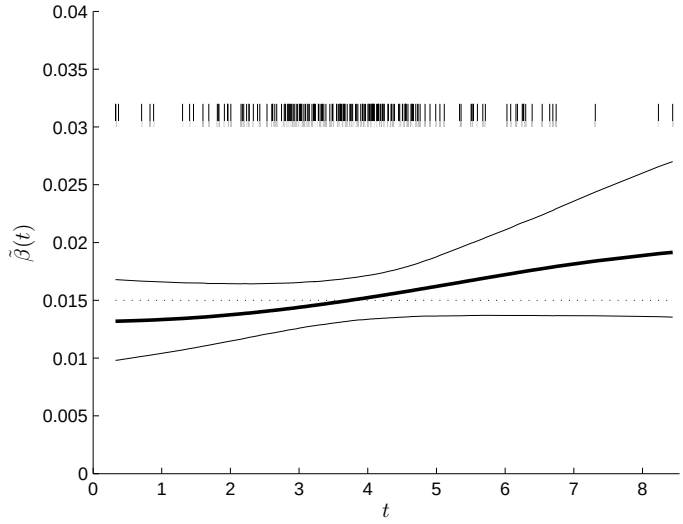


Figure 3.3: Posterior mean of the infection rate $\tilde{\beta}(t)$ (solid line) at each infection time for SE-Data 3. The 95% credible intervals are shown as well. True infection rate (dotted line), $\beta = 0.015$. There is a total of 189 infections during the whole epidemic and all the infection times are assumed to be known. The " | " marks in the plot represent the observed data, i.e. the infection times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

3.2.2.2 Infection rate varies throughout the epidemic

We now simulate a new epidemic data set (SE-Data 4) from the SIR model assuming that the infection rate varies throughout the whole epidemic. Table 3.1 gives infection rate, removal rate, initial number of susceptibles and initial number of infectives for generating the data. We then apply our methods to SE-Data 4 and assume that a set of infection times. We again choose the squared exponential covariance function for the Gaussian process prior and the hyperparameter, α, is fixed to 1. Figure 3.4 shows the posterior mean of the infection rate at each infection time for SE-Data 4. From Figure 3.4, we see the estimated infection rate recovers the truth fairly well which implies that our methods appear to be working reasonably well not only for the case where the simulated data are generated with a constant infection rate, but also for the case where the simulated data are generated with varying infection rate throughout the

epidemic.

Table 3.1: Infection rate, $\tilde{\beta}(t)$, removal rate, γ, initial number of susceptibles, N, and initial number of infectives, a, for the new data set. The parameter setting is used to generate an epidemic process from the SIR model, i.e. the simulated epidemic data set named as SE-Data 4.

	$\tilde{\beta}(t)$	γ	N	a
parameter setting	$\exp(-t/10)/150$	0.7	300	1

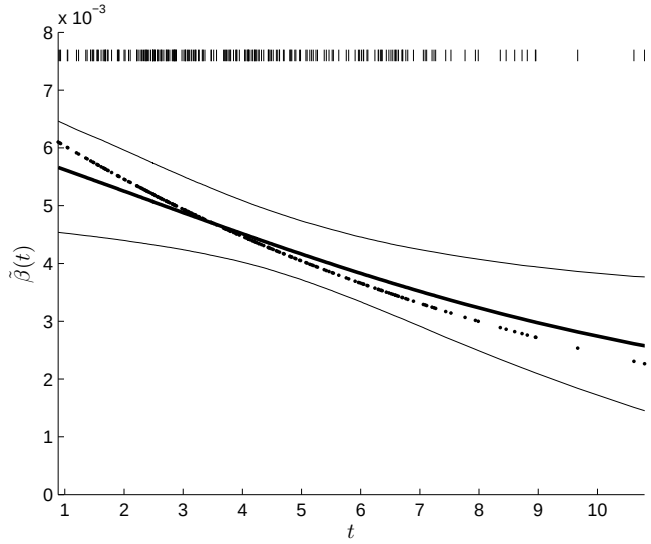


Figure 3.4: Posterior mean of the infection rate $\tilde{\beta}(t)$ (solid line) at each infection time compared with true infection rate (dotted line). The 95% credible intervals are shown as well. There is a total of 222 infections during the whole epidemic and all the infection times are assumed to be known. The " | " marks in the plot represent the observed data, i.e. the infection times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

3.2.3 Simulated partially observed single group data

3.2.3.1 Constant infection rate

We now assume we only observe removal times, i.e. we do not know infection times and removal rate of the epidemic. We first apply our methods to SE-Data 1, SE-Data 2 and SE-Data 3 which are generated with a constant infection rate. The squared exponential covariance function for the Gaussian process prior is adopted and the hyperparameter α is fixed to 1. Figure 3.5, 3.6 and 3.8 show the estimation results for SE-Data 1, SE-Data 2 and SE-Data 3 under different assumptions, i.e. complete data observed and only removal times observed. Each Figure gives the posterior mean of the infection rate. Density of the removal rate is also given in Figure 3.5 (c), 3.6 (c) and 3.8 (c) for the case where data are partially observed. From Figure 3.5 (b), 3.6 (b) and 3.8 (b), we can see that the estimated results of $\tilde{\beta}(t)$ recover the truth fairly well and it turns out that there is little difference between the case where the complete data are observed (shown in Figure 3.5 (a), 3.6 (a) and 3.8 (a)) and the case where we only observe incomplete data. From the density plots, we have good estimations of γ for all of the three dataset and it also appears that we have the least deviation from the posterior mean of the removal rate, γ for SE-Data 3. One explanation could be that there are larger numbers of people infected in SE-Data 3 than the other two data sets which leads to more accurate estimation of the removal rate. However, this is just one dataset only so it could be just due to chance. Figure 3.7 shows MCMC trace plot of the hyperparameter, θ, the infection rate at time $t = 2.42$ and $t = 3.42$ respectively for SE-Data 2.

3.2.3.2 Infection rate varies throughout the epidemic

We now apply our methods to SE-Data 4 which are generated with varying infection rate throughout the whole epidemic. We also assume we only observe removal times. The squared exponential covariance function for the Gaussian process prior is adopted and the hyperparameter, α, is fixed to 1. Figure 3.9 shows the posterior mean of the infection rate, $\tilde{\beta}(t)$, and density of the removal rate, γ. Compared to the results for the completed epidemic case shown in Figure 3.9 (a), the estimation shown in Figure 3.9 (b) is not as good. However,

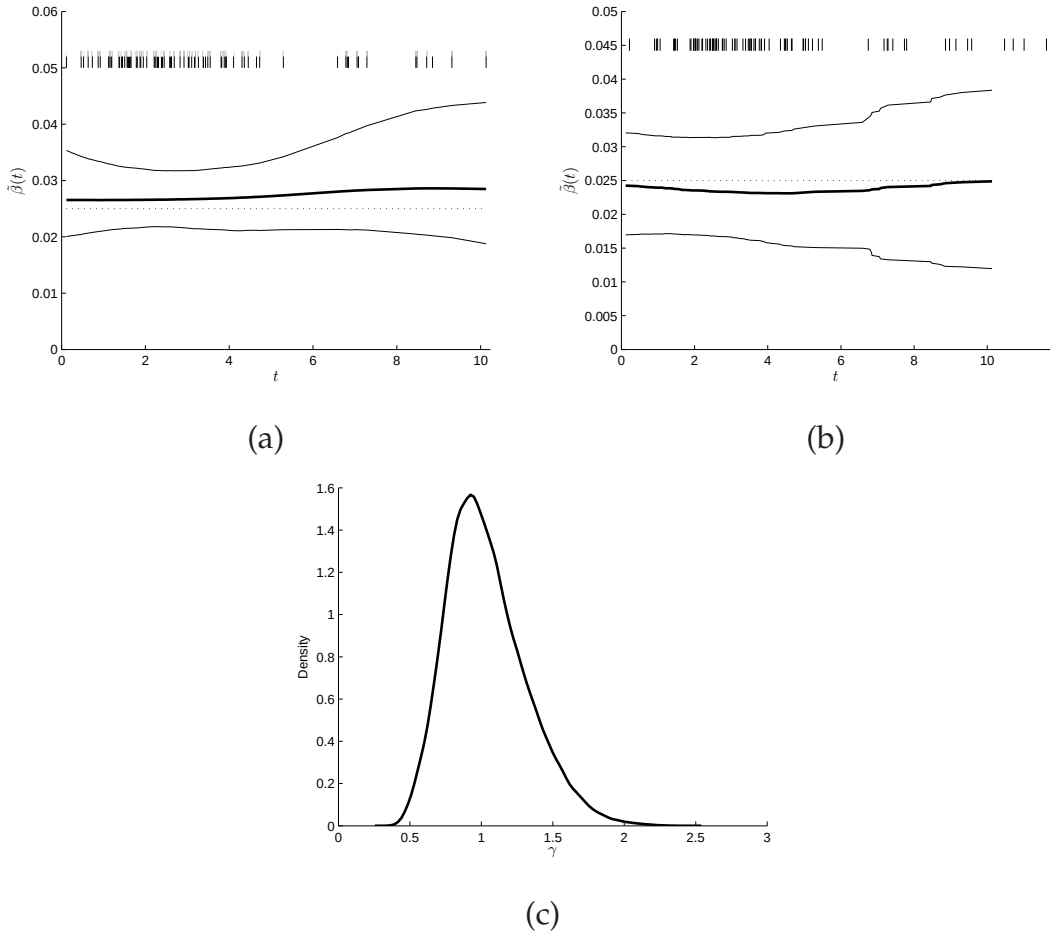


Figure 3.5: Posterior mean of the infection rate $\tilde{\beta}(t)$ (solid line) at each estimated infection time for SE-Data 1 (plot (a) and plot(b)). True infection rate (dotted line), $\beta = 0.025$ and true removal rate, $\gamma = 1$. The 95% credible intervals are shown in (a) and (b). Plot (a) is the case where complete data are observed. Plot (b) is the case where only removal times are observed. The " | " marks in each plot represent the observed data, i.e. the infection times in (a) and removal times in (b). Plot (c) shows density of the removal rate for the second case. There is a total of 87 infections during the whole epidemic. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

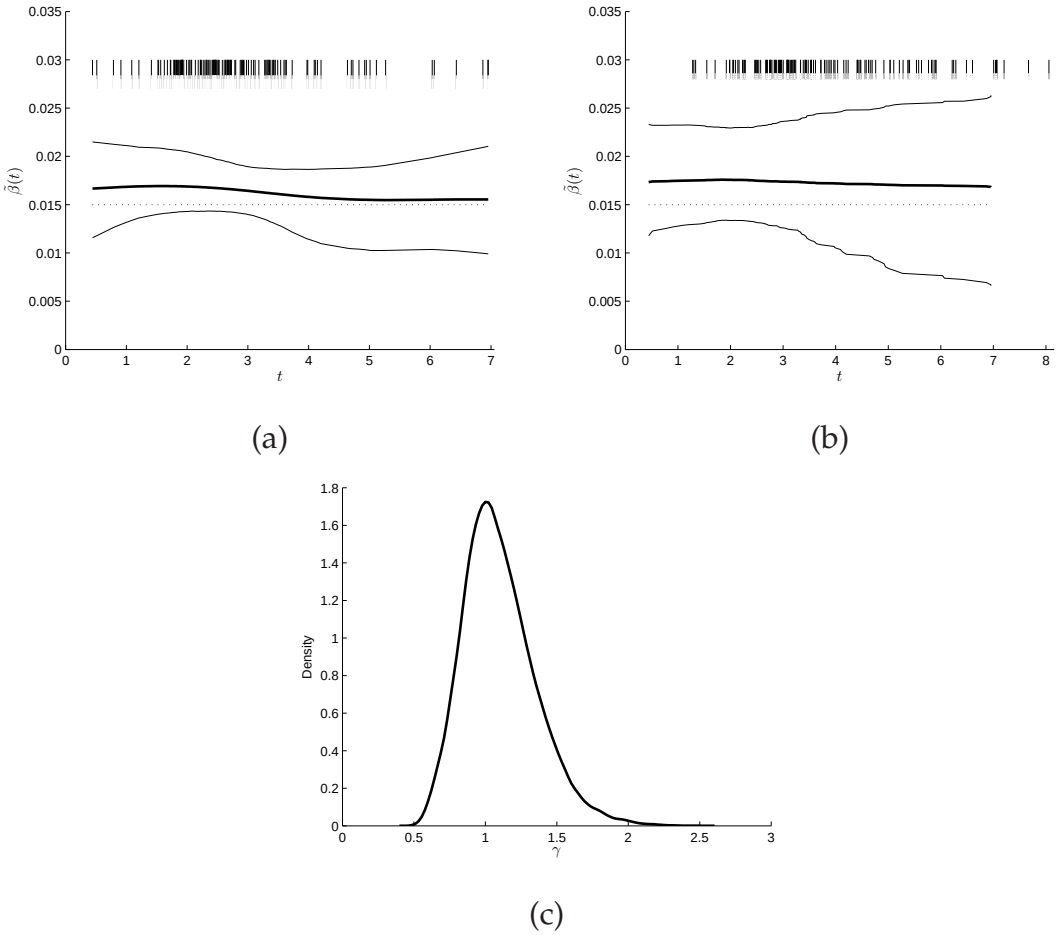


Figure 3.6: Posterior mean of the infection rate $\tilde{\beta}(t)$ (solid line) at each estimated infection time for SE-Data 2 (plot (a) and plot(b)). True infection rate (dotted line), $\beta = 0.015$ and true removal rate, $\gamma = 1$. The 95% credible intervals are shown in (a) and (b). Plot (a) is the case where complete data are observed. Plot (b) is the case where only removal times are observed. The " | " marks in each plot represent the observed data, i.e. the infection times in (a) and removal times in (b). Plot (c) shows density of the removal rate for the second case. There is a total of 130 infections during the whole epidemic. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

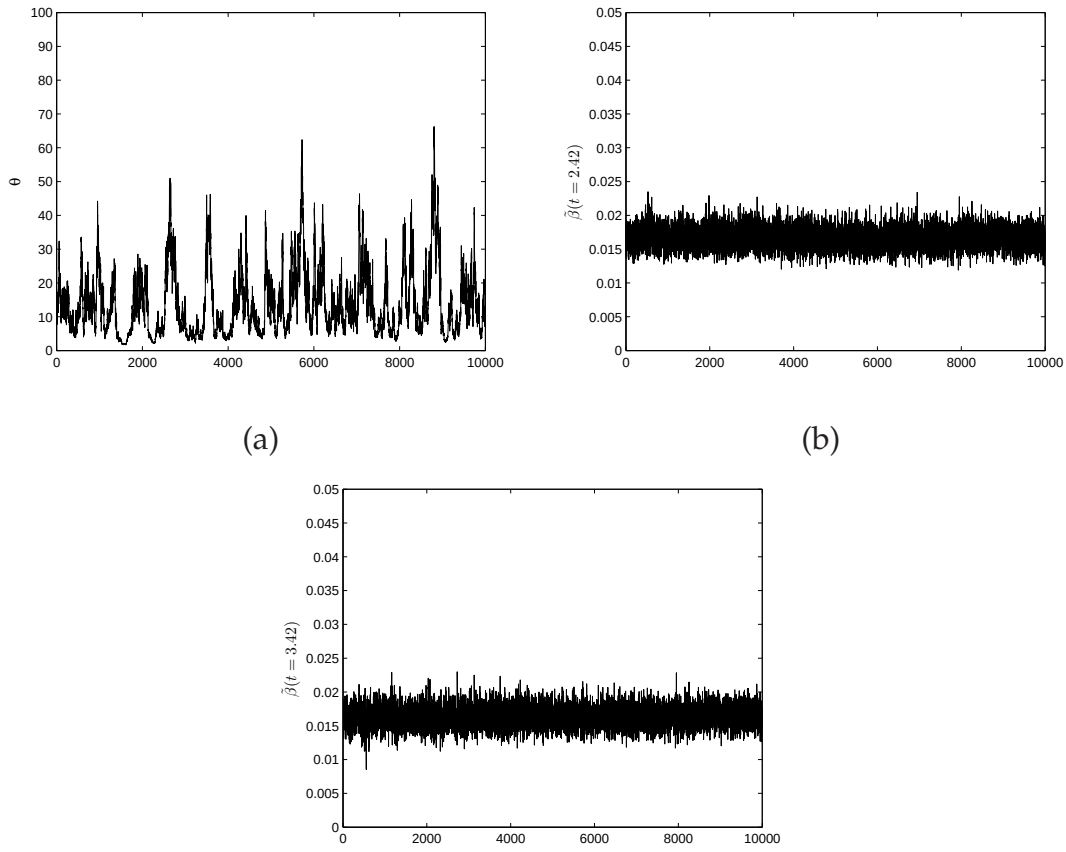


Figure 3.7: Dataset SE-Data 2 recovered by placing the Gaussian process prior where only removal times are known. Plot (a) shows MCMC trace plot of the hyperparameter, θ. Plot (b) and (c) show MCMC trace plot of the infection rate at time $t = 2.42$ and $t = 3.42$ respectively. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

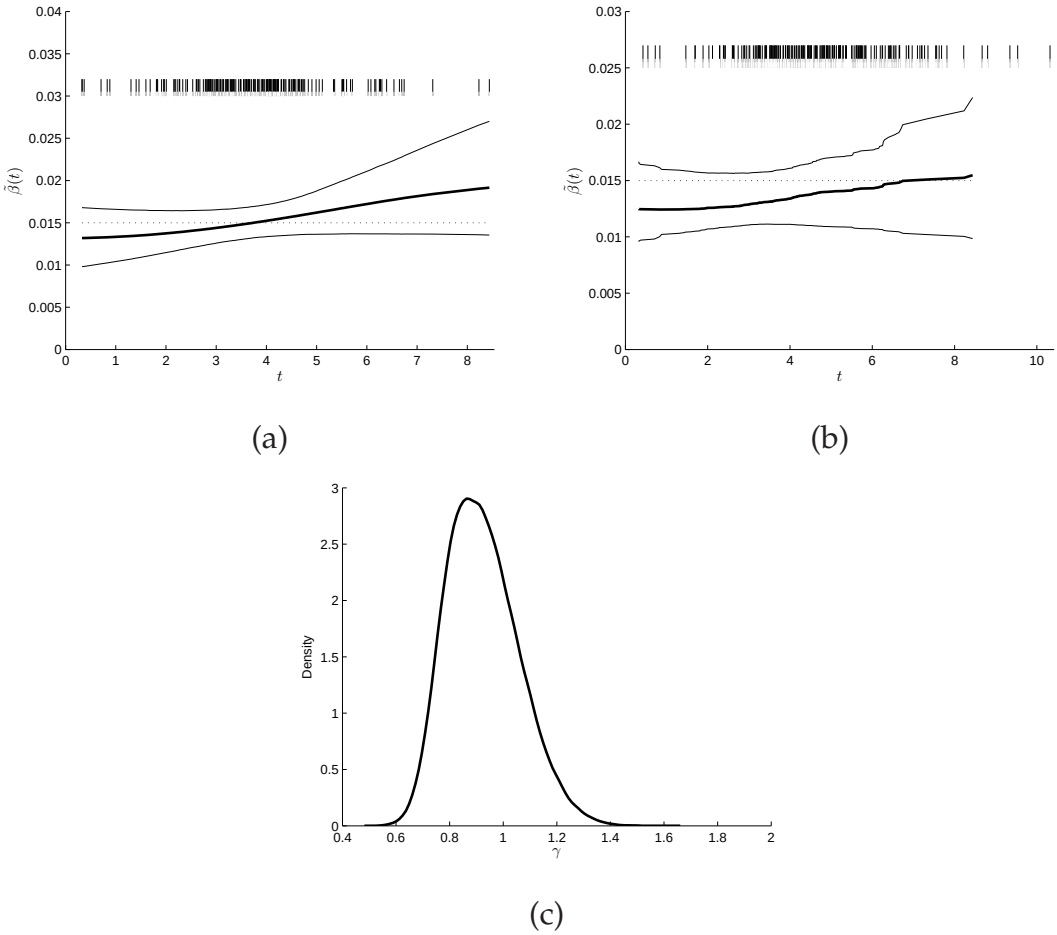


Figure 3.8: Posterior mean of the infection rate $\tilde{\beta}(t)$ (solid line) at each estimated infection time for SE-Data 3 (plot (a) and plot(b)). True infection rate (dotted line), $\beta = 0.015$ and true removal rate, $\gamma = 1$. The 95% credible intervals are shown in (a) and (b). Plot (a) is the case where complete data are observed. Plot (b) is the case where only removal times are observed. The " | " marks in each plot represent the observed data, i.e. the infection times in (a) and removal times in (b). Plot (c) shows density of the removal rate for the second case. There is a total of 189 infections during the whole epidemic. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

the truth totally lies within the 95% credible intervals. The density plot shows we have fairly good estimation of γ.

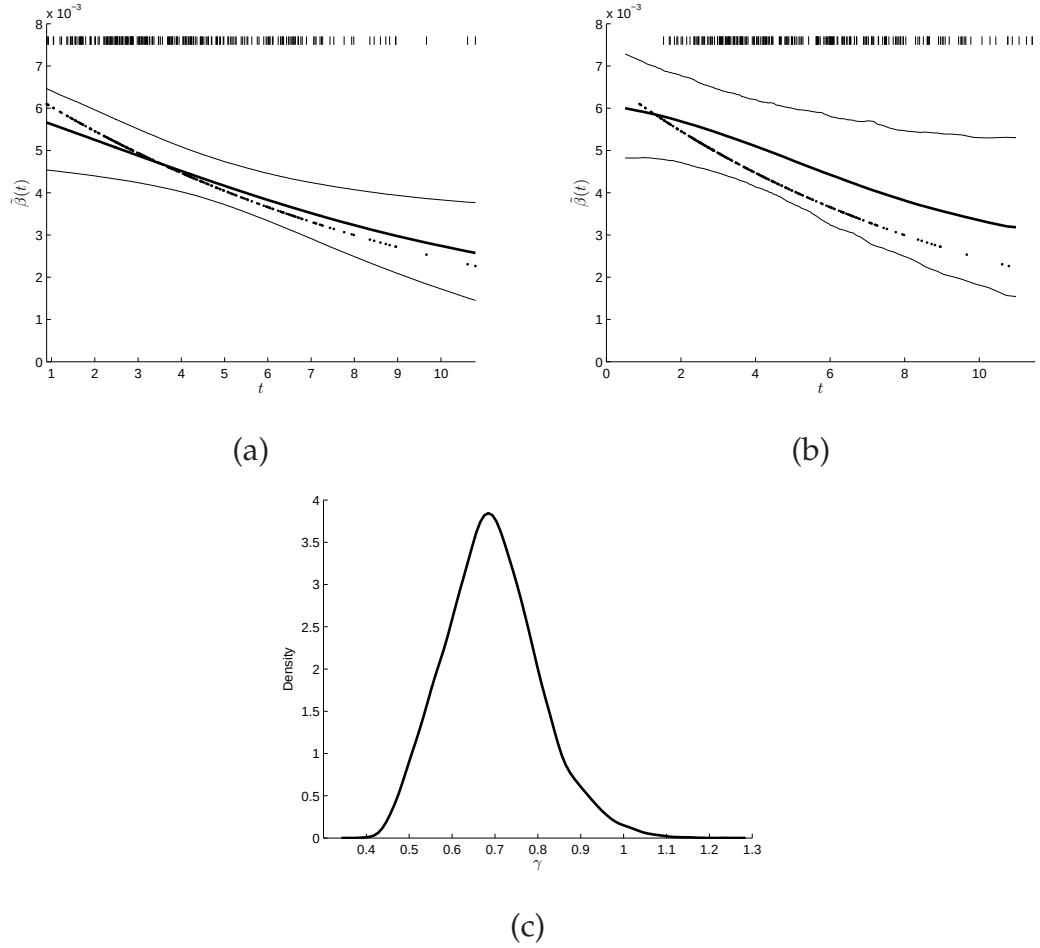
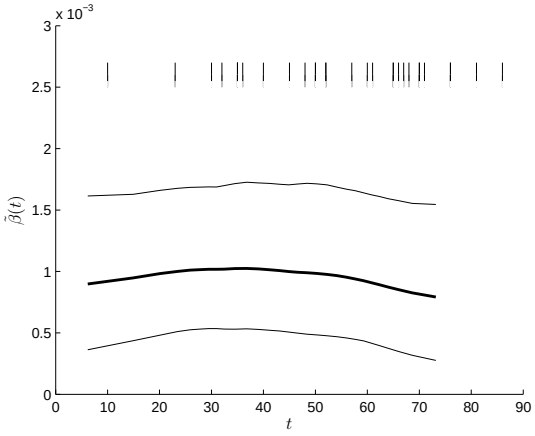


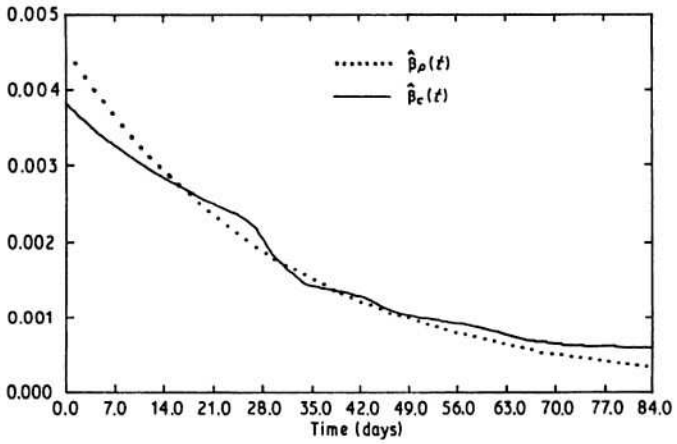
Figure 3.9: Posterior mean of the infection rate $\tilde{\beta}(t)$ (solid line) at each estimated infection time for SE-Data 4 (plot (a) and plot (b)). True removal rate, $\gamma = 0.7$. The 95% credible intervals are shown in (a) and (b). Plot (a) is the case where complete data are observed. Plot (b) is the case where only removal times are observed. The " | " marks in each plot represent the observed data, i.e. the infection times in (a) and removal times in (b). Plot (c) shows density of the removal rate for the second case. There is a total of 222 infections during the whole epidemic. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

3.2.4 Smallpox data

We now apply our methods to real life data, namely Smallpox data as described in Chapter 2, Section 2.3.4. We assume only removal times are known. All the infection times for each individual and the removal rate are inferred by the MCMC methods. We then apply our methods to the data and place a Gaussian process prior using the squared exponential covariance function on the infection rate, $\tilde{\beta}(t)$. The hyperparameter, α, is fixed to 1. Figure 3.10 (a) shows the posterior mean of the infection rate. For comparison, we give two estimation results of the infection rate from Becker (1989), p.137, shown in Figure 3.10 (b). The author assumed the infection rate is decreasing over time, hence an exponential function was used to model the infection rate. Although our results show some differences that the infection rate keeps increasing for nearly 30 days from the beginning of the epidemic, the decreasing part of our results is similar to Becker's results. Figure 3.11 shows the density of the removal rate, γ, the hyperparameter, θ, the upper bound, $\tilde{\beta}^*$ and the number of thinned events, M respectively. For the estimation of γ, our results appear to be close to the ones obtained by O'Neill & Roberts (1999).



(a)



(b)

Figure 3.10: Smallpox data recovered by placing the Gaussian process prior. The plot (a) shows posterior mean of the infection rate, $\tilde{\beta}(t)$, at each estimated infection time using our Bayesian nonparametric methods where the squared exponential covariance function is adopted. The " | " marks in the plot represent the observed data, i.e. the removal times. The plot (b), obtained from Becker (1989), p.137, shows two estimation results of the infection rate for smallpox data from Becker's methods. The infection rate is assumed to be a decreasing function of time in the two models.

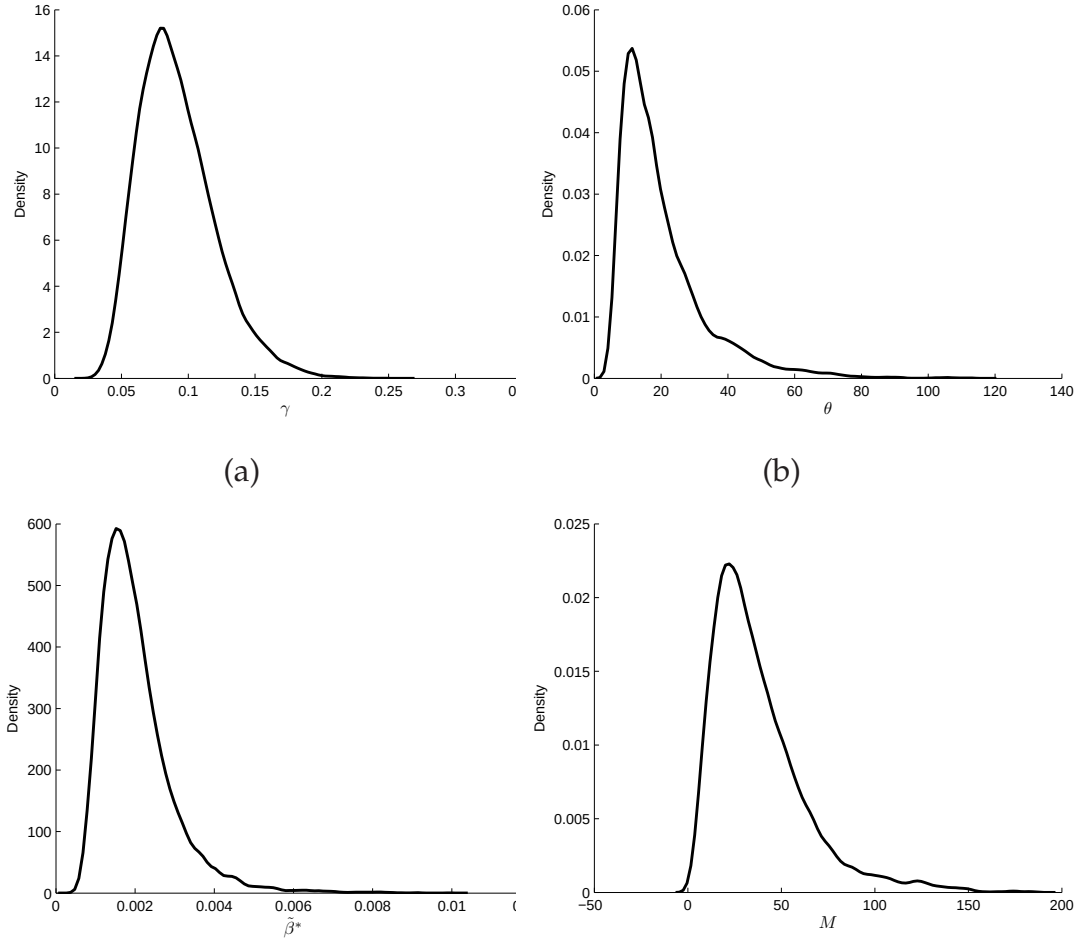


Figure 3.11: Smallpox data recovered by placing the Gaussian process prior on $\tilde{\beta}(t)$. Four plots show densities of the removal rate, γ, the hyperparameter, θ, the upper bound $\tilde{\beta}^*$ and the number of thinned events, M respectively. The squared exponential covariance function for the Gaussian process prior is adopted.

3.3 Sensitivity to the Gaussian process prior for the SIR epidemic model

Similar to Section 2.4, we investigate the effect of the choice of covariance function on the inferences in this section. We choose 3 different covariance functions, i.e. squared exponential covariance function, exponential covariance function and Matérn class covariance function with ν set to $3/2$.

We now apply our methods to SE-Data 2 and perform inference for the infection rate, $\tilde{\beta}(t)$, by assuming we observe the complete data. Figure 3.12 shows the estimated intensities using squared exponential covariance function, exponential covariance function and Matérn class covariance function for the Gaussian process prior respectively for SE-Data 2. Compared to the estimated intensity in Figure 3.12 (a) where the squared exponential covariance function is used, the estimated intensity in Figure 3.12 (b) is quite rough and has larger uncertainty. One explanation could be that a rough covariance function is chosen for the Gaussian process prior which is expected to produce a rough process. The estimated intensity in Figure 3.12 (c), has similar mean and uncertainty to the one in (a). In terms of smoothness, the plots in (a) and (c) are similar.

We also evaluate our methods for the three different Gaussian process priors for the Smallpox. Figure 3.13 shows the recovered trajectories for the squared exponential covariance function, exponential covariance function and Matérn class covariance function respectively. The plot (a) is slightly smoother than (b) and (c) and the results in (b) are very similar to the ones in (c). In general, we find that inferences are insensitive to the choice of Gaussian process priors for the Smallpox data.

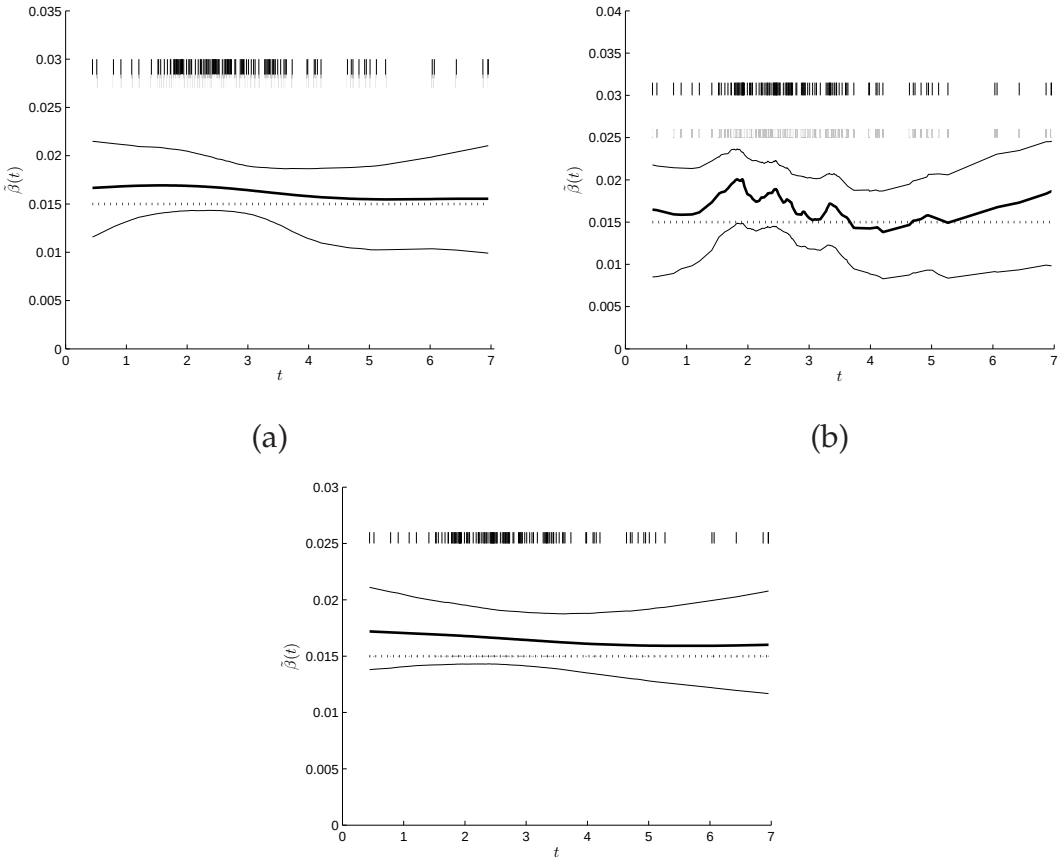


Figure 3.12: Dataset SE-Data 2 recovered by placing three different Gaussian process priors. Three plots show posterior mean of the infection rate, $\tilde{\beta}(t)$ (solid line) at each infection time compared with the true β (dotted line). The data are generated from the general stochastic epidemic with parameters infection rate $\beta = 0.015$, removal rate $\gamma = 1$, initial number of susceptibles $N = 150$ and initial number of infective individuals $a = 1$. Plot (a) corresponds to the squared exponential covariance function see, (1.2). Plot (b) corresponds to the exponential covariance function see, (1.3). Plot (c) corresponds to the Matérn covariance function with ν set to $3/2$ see, (1.4). The 95% credible intervals are shown for each of the plot. The " | " marks in each plot represent the observed data, i.e. the infection times. There is a total of 130 infections during the whole epidemic and all the infection times are assumed to be known. The hyper-parameter of the three covariance functions, α, is set to 1.

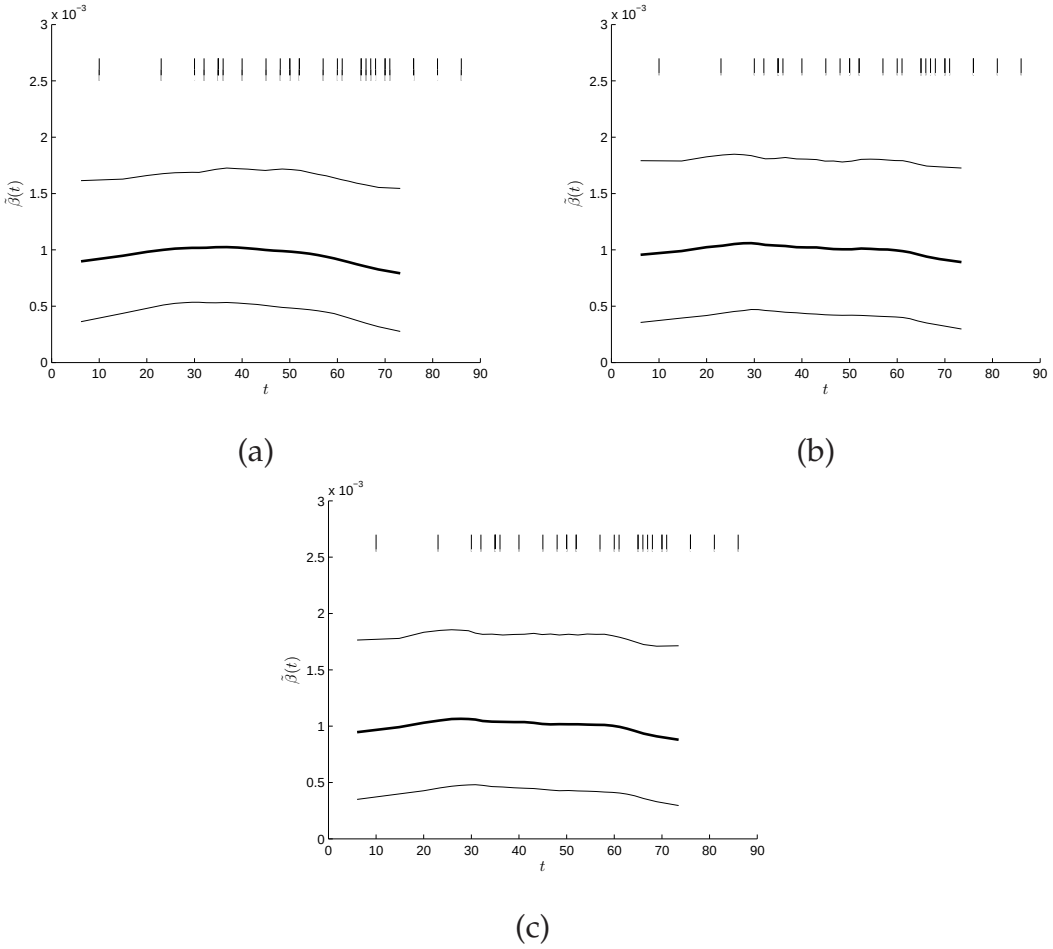


Figure 3.13: Smallpox recovered by placing three different Gaussian process priors. Three plots show posterior mean of the infection rate, $\tilde{\beta}(t)$ (solid line) at each estimated infection time. Plot (a) corresponds to the squared exponential covariance function see, (1.2). Plot (b) corresponds to the exponential covariance function see, (1.3). Plot (c) corresponds to the Matérn covariance function with ν set to $3/2$ see, (1.4). The 95% credible intervals are shown for each of the plot. The " | " marks in each plot represent the observed data, i.e. the removal times. There is a total of 30 infections during the whole epidemic and only removal times are known. The hyper-parameter of the three covariance functions, α, is set to 1.

3.4 Estimation for epidemic models from multi-group epidemic data

In Section 3.2, we demonstrated that our Bayesian nonparametric methods appear to be working fairly well for epidemic models from single group epidemic data which motivated us to investigate more complex cases with our methods. For example, the population can be no longer homogeneous, but partitioned into groups according to level of susceptibility. Hayakawa et al. (2003) developed Bayesian methods for SIR epidemic models with several kinds of susceptibles. In this Section, we will first introduce the multi-group model discussed in Hayakawa et al. (2003) and then apply our Bayesian nonparametric methods to multi-group epidemic data (simulated data and real life data).

3.4.1 Multi-group epidemic model

We now describe the extended multi-group SIR model discussed in Hayakawa et al. (2003). Consider a population consisting initially of k groups of susceptibles, where the groups are labelled $1, \cdots, k$, and group i contains N_i susceptibles, $i = 1, \cdots, k$. An epidemic is initiated in the population by one of the susceptibles becoming infected and this infection is assumed to occur via some process external to the population. For $t \geq 0$ and $i = 1, \cdots, k$, denote by $X_i(t)$ and $Y_i(t)$ the numbers of susceptibles and infectives, respectively, in group i at time t. Let $Y(t) = \sum_{i=1}^{k} Y_i(t)$ denote the total number of infectives in the population at time t. The epidemic is then defined according to the following transition probabilities, the transitions themselves corresponding respectively to an infection and a removal:

$$P\left((X_i(t+\delta t), Y(t+\delta t)) = (x-1, y+1) \middle| (X_i(t), Y(t)) = (x, y)\right) = \beta_i x y \delta t + o(\delta t),$$

$$P\left((X_i(t+\delta t), Y(t+\delta t)) = (x, y-1) \middle| (X_i(t), Y(t)) = (x, y)\right) = \gamma y \delta t + o(\delta t),$$

all other transitions having probability $o(\delta t)$. A removed individual does not take part in the epidemic any longer. The epidemic ceases if there are no more infectives left in the population. The assumptions imply that each susceptible

of type i is affected equally by any infectives irrespective of the infective's type, and that the durations of the infectious periods are independent and identically distributed Exponential random variables with rate γ.

Under the assumption that the infection rate for each group is no longer a constant but a function of time the transition probabilities now become:

$$P\left((X_i(t+\delta t), Y(t+\delta t)) = (x-1, y+1) \middle| (X_i(t), Y(t)) = (x,y)\right) = \tilde{\beta}_i(t)xy\delta t + o(\delta t),$$

$$P\left((X_i(t+\delta t), Y(t+\delta t)) = (x, y-1) \middle| (X_i(t), Y(t)) = (x,y)\right) = \gamma y \delta t + o(\delta t).$$

3.4.2 Inference

We first give a list of notation that we use.

$\boldsymbol{\tau}_i = (\tau_{i1}, \tau_{i1}, \cdots, \tau_{in_i})$: the vector of ordered removal times of type i

$\mathbf{I}_i = (I_{i1}, I_{i1}, \cdots, I_{im_i})$: the vector of ordered infection times of type i

$\mathbf{I}$=$(\mathbf{I}_1, \mathbf{I}_2, \cdots, \mathbf{I}_k)$

$\boldsymbol{\tau}$=$(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_k)$

$\tilde{\boldsymbol{\beta}}(t)$=$(\tilde{\beta}_1(t), \tilde{\beta}_2(t), \cdots, \tilde{\beta}_k(t))$

$\mathbf{N}$=$(N_1, N_2, \cdots, N_k)$

I_{i1}: the time of the first infection in group i

$I_{\min} = \min(I_{11}, I_{21}, \cdots, I_{k1})$: the time of the first infection in the total population

$i_{\min}$: type i for which $I_{i1} = I_{\min}$

n_i: the observed total number of removals of type i

m_i: the unobserved total number of infections of type i

$N = \sum_{i=1}^{k} N_i$: the initial number of susceptibles in the total population

$\underset{\sim}{\mathbf{I}}$: $\{\mathbf{I} \setminus I_{\min}\}$

As our interest focuses on the infection rate in each group, it is assumed that we know the initial number of susceptibles in each group, $\mathbf{N}$, and the type of

the time of the first infection in the total population, $i_{\min}$. According to the augmentation methods in Hayakawa et al. (2003), we can write the joint distribution over the infection rate in each group, $\tilde{\boldsymbol{\beta}}(t)$, the removal rate, γ, the time of the first infection in the total population. $I_{\min}$ by Bayes' Theorem below.

$$\pi(\tilde{\boldsymbol{\beta}}(t), \gamma, I_{\min}|\underset{\sim}{\mathbf{I}}, \boldsymbol{\tau}) \propto \pi(\underset{\sim}{\mathbf{I}}, \boldsymbol{\tau}|\tilde{\boldsymbol{\beta}}(t), \gamma, I_{\min})\pi(\tilde{\boldsymbol{\beta}}(t), \gamma, I_{\min}),$$

where

$$\begin{aligned}
&\pi(\underset{\sim}{\mathbf{I}}, \boldsymbol{\tau}|\tilde{\boldsymbol{\beta}}(t), \gamma, I_{\min}) \\
&= \left(\prod_{i=1}^{k}\left(\prod_{j=1}^{n_i}\gamma Y_i(\tau_{ij^-})\right)\left(\prod_{l=2}^{m_i}\tilde{\beta}_i(I_{il^-})X_i(I_{il^-})Y_i(I_{il^-})\right)\right)\left(\prod_{i=1, i\neq i_{\min}}^{k}\tilde{\beta}_i(I_{i1^-})X_i(I_{i1^-})Y_i(I_{i1^-})\right) \\
&\quad \times \exp\left(-\sum_{i=1}^{k}\left(\int_{I_{\min}}^{T}\tilde{\beta}_i(s)X_i(s)Y_i(s)\mathrm{d}s + \int_{I_{i1}}^{T}\gamma Y_i(s)\mathrm{d}s\right)\right), \qquad (3.3)
\end{aligned}$$

and $\pi(\tilde{\boldsymbol{\beta}}(t), \gamma, I_{\min})$ denotes the prior density of $\tilde{\boldsymbol{\beta}}(t)$, γ and $I_{\min}$.

Considering the removal part in (3.3), we can write the full conditional distribution of γ below.

$$\pi(\gamma|\tilde{\boldsymbol{\beta}}(t), \boldsymbol{\tau}, \underset{\sim}{\mathbf{I}}) \propto \left(\prod_{i=1}^{k}\left(\prod_{j=1}^{n_i}\gamma Y_i(\tau_{ij^-})\right)\right) \times \exp\left(-\sum_{i=1}^{k}\left(\int_{I_{i1}}^{T}\gamma Y_i(s)\mathrm{d}s\right)\right).$$

According to the algorithms described in Hayakawa et al. (2003), with a gamma prior $(\nu_\gamma, \lambda_\gamma)$ put on γ, we can sample from the distribution below.

$$\pi(\gamma|\tilde{\boldsymbol{\beta}}(t), \boldsymbol{\tau}, \underset{\sim}{\mathbf{I}}) \propto \Gamma\left(\nu_\gamma + \sum_{i=1}^{k} n_i, \lambda_\gamma + \sum_{i=1}^{k}\int_{I_{i1}}^{T} Y_i(s)\mathrm{d}s\right).$$

As we know $i_{\min}$, we use the similar method discussed in Section 3.2.1 to update $I_{\min}$. By using a Gibbs sampling approach, with a non-informative exponential prior put on, we can sample from the distribution if the first infection time is in group i

$$\pi(y|\boldsymbol{\tau}_i, \underset{\sim}{\mathbf{I}}, \tilde{\beta}_i, \gamma) = \Lambda \exp\{-\Lambda(I_2 - y)\},\ y \in (-\infty, I_2),$$

where $I_2 = \min(I_{i2}, \{I_{j1}\}_{j=1, j\neq i}^{k})$ and $\Lambda = \gamma + \tilde{\beta}(I_{2^-}) * N$.

It now only remains to find a way of sampling $\tilde{\boldsymbol{\beta}}(t)$ and $\underset{\sim}{\mathbf{I}}$. The main problem here is to make the infection part in (3.3) tractable. Similar to the methods used

for the single group case, we use the SGCP model to describe the infection rate for group i, $\tilde{\beta}_i(t)$.

$$\tilde{\beta}_i(t) = \tilde{\beta}_i^* \sigma(g(t)),$$

where $\tilde{\beta}_i^*$: an upper bound on $\tilde{\beta}_i(t)$.

To tackle the intractable problem, we again augment the posterior distribution like we did for the single group case. For group i and if $I_{i1} = I_{\min}$, the additional latent variables are: the number of thinned events, M_i, the locations of thinned events, $\tilde{\mathbf{I}}_i = (\tilde{I}_{i1}, \tilde{I}_{i2}, \cdots, \tilde{I}_{iM_i})$, the function values at the infection times, $\mathbf{g}_{m_i} = (g(I_{i2^-}), g(I_{i3^-}), \cdots, g(I_{im_i^-}))$ and the function values at the locations of thinned events, $\mathbf{g}_{M_i} = (g(\tilde{I}_{i2^-}), g(\tilde{I}_{i3^-}), \cdots, g(\tilde{I}_{iM_i^-}))$. If $I_{i1} \neq I_{\min}$, we need to consider the first infection time, I_{i1}, in group i. Therefore, $g(I_{i1^-})$ should be added to $\mathbf{g}_{m_i}$. We can now write the augmented likelihood we require below.

$$\begin{aligned}
&\pi(\underset{\sim}{\mathbf{I}}, \boldsymbol{\tau}, \{M_i, \tilde{\mathbf{I}}_i, \mathbf{g}_{M_i+m_i}\}_{i=1}^k | \{\tilde{\beta}_i^*\}_{i=1}^k, \gamma, I_{\min}, T, \theta_i) \\
&= \left(\prod_{i=1}^k \left(\prod_{j=1}^{n_i} \gamma Y_i(\tau_{ij^-}) \right) \left(\prod_{l=2}^{m_i} \tilde{\beta}_i^* X_i(I_{il^-}) Y_i(I_{il^-}) \sigma(g(I_{il^-})) \right) \right) \\
&\quad \times \left(\prod_{i=1}^k \left(\prod_{s=1}^{M_i} \tilde{\beta}_i^* X_i(\tilde{I}_{is^-}) Y_i(\tilde{I}_{is^-}) \sigma(-g(\tilde{I}_{is^-})) \right) \right) \\
&\quad \times \left(\prod_{i=1, i\neq i_{\min}}^k \tilde{\beta}_i^* X_i(I_{i1^-}) Y_i(I_{i1^-}) \sigma(g(I_{i1^-})) \right) \\
&\quad \times \exp\left(-\sum_{i=1}^k \left(\int_{I_{\min}}^T \tilde{\beta}_i^* X_i(s) Y(s) \mathrm{d}s + \int_{I_{i1}}^T \gamma Y_i(s) \mathrm{d}s \right) \right) \\
&\quad \times \left(\prod_{i=1}^k \pi(\mathbf{g}_{M_i+m_i} | M_i, \{I_{il}\}_{l=1+\delta_i}^{m_i}, \{\tilde{I}_{is}\}_{s=1}^{M_i}, \theta_i) \right), \qquad (3.4)
\end{aligned}$$

where $\delta_i = 1$ if $i = i_{\min}$ and 0 otherwise. Considering (3.4), we can now write the full conditional density of each parameter. For group i, we have

$$\pi(\tilde{\beta}_i^* | M_i, T, \underset{\sim}{\mathbf{I}}, I_{\min}) \propto \tilde{\beta}_i^{*\, m_i - \delta_i + M_i} \times \exp\left(-\int_{I_{\min}}^T \tilde{\beta}_i^* X_i(s) Y(s) \mathrm{d}s \right),$$

$$\pi(M_i, \tilde{\mathbf{I}}_i, \mathbf{g}_{M_i} | \tilde{\beta}_i^*, \mathbf{g}_{m_i}, \{I_{il}\}_{l=1+\delta_i}^{m_i}, \theta_i, T, I_{\min})$$
$$\propto (\tilde{\beta}_i^*)^{M_i} \prod_{s=1}^{M_i} X_i(\tilde{I}_{is^-}) Y_i(\tilde{I}_{is^-}) \sigma(-g(\tilde{I}_{is^-})) \times \pi(\mathbf{g}_{M_i+m_i} | \{I_{il}\}_{l=1+\delta_i}^{m_i}, \{\tilde{I}_{is}\}_{s=1}^{M_i}, \theta_i),$$

$$\pi(\tilde{\mathbf{I}}_i, \mathbf{g}_{M_i} | \tilde{\beta}_i^*, \mathbf{g}_{m_i}, \{I_{il}\}_{l=1+\delta_i}^{m_i}, M_i, \theta_i, T, I_{\min})$$
$$\propto \prod_{s=1}^{M_i} X_i(\tilde{I}_{is^-}) Y_i(\tilde{I}_{is^-}) \sigma(-g(\tilde{I}_{is^-})) \times \pi(\mathbf{g}_{M_i+m_i} | \{I_{il}\}_{l=1+\delta_i}^{m_i}, \{\tilde{I}_{is}\}_{s=1}^{M_i}, \theta_i),$$

$$\pi(\mathbf{g}_{M_i+m_i} | M_i, \{I_{il}\}_{l=1+\delta_i}^{m_i}, \{\tilde{I}_{is}\}_{s=1}^{M_i}, \theta_i)$$
$$\propto \prod_{l=1+\delta_i}^{m_i} \sigma(g(I_{il^-})) \prod_{s=1}^{M_i} \sigma(-g(\tilde{I}_{is^-})) \times \pi(\mathbf{g}_{M_i+m_i} | \{I_{il}\}_{l=1+\delta_i}^{m_i}, \{\tilde{I}_{is}\}_{s=1}^{M_i}, \theta_i).$$

$$\pi(\theta_i | M_i, \mathbf{g}_{M_i+m_i}, \{I_{il}\}_{l=1+\delta_i}^{m_i}, \{\tilde{I}_{is}\}_{s=1}^{M_i})$$
$$\propto \pi(\mathbf{g}_{M_i+m_i} | \{I_{il}\}_{l=1+\delta_i}^{m_i}, \{\tilde{I}_{is}\}_{s=1}^{M_i}, \theta_i)) \times \pi(\theta_i)$$

The algorithm of sampling the parameters above for each group is similar to the single group case in Section 3.2.1 and will not be discussed here.

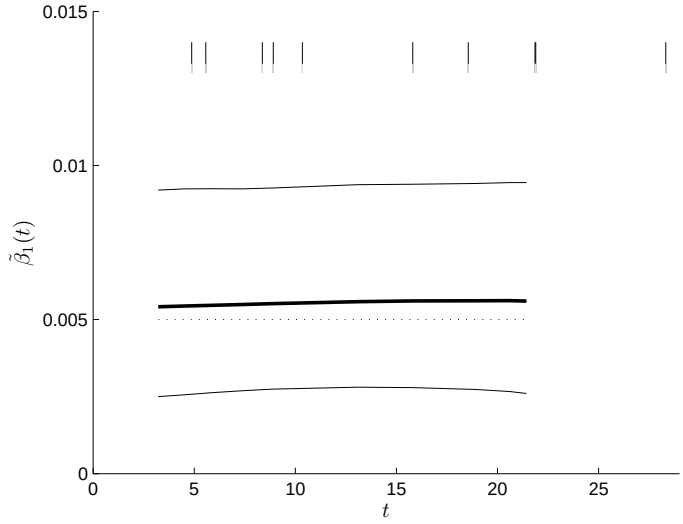
3.4.3 Simulated partially observed multi-group data

We now simulate a set of multi-group data with a parameter setting given in Table 3.2. The total population of 254 is partitioned into three groups. The groups' initial population size are reported as $N_1 = 25$, $N_2 = 36$ and $N_3 = 192$. The total number of cases in each of the three groups are 10, 12 and 30 respectively. We assume that only removal times of each group are observed. We apply our methods to the data and place the Gaussian process prior on $\tilde{\beta}_1$, $\tilde{\beta}_2$ and $\tilde{\beta}_3$ respectively. The prior distributions for the Gaussian process in each group are assumed to be independent. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1. Figure 3.14 (a), (b) and Figure 3.15 (a) show

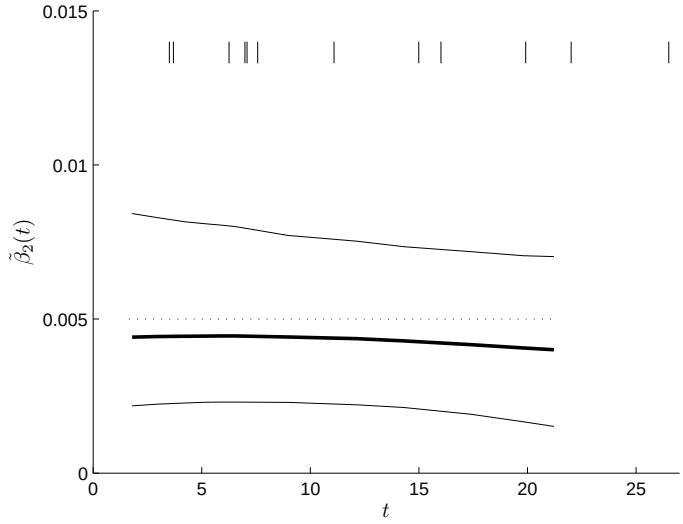
estimation results for the infection rate of each group. Compared to the truth, the posterior means of the infection rate of all three groups are smooth and flat and all close to the true values. In addition, the true values are all lie in the 95% credible intervals which although show large uncertainties. Figure 3.15 (b) shows the density of the estimated removal rate which implies that we have a good estimation for the removal rate.

Table 3.2: Infection rate, β, removal rate, γ, initial number of susceptibles, N, and initial number of infectives, a, for the new 3-group data set. The parameter setting is used to generate an epidemic process from the multi-group SIR model, i.e. the simulated epidemic data set named as SE-Data 5.

	β	N	a	γ
group 1	0.005	25	0	0.5
group 2	0.005	36	0	
group 3	0.002	192	1	



(a)

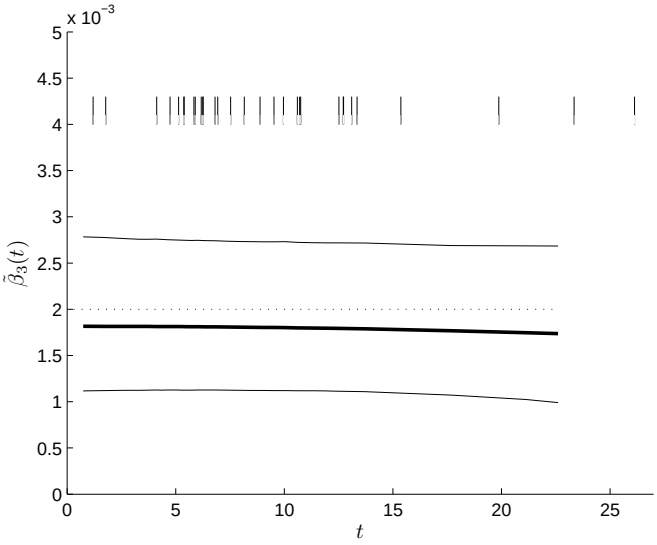


(b)

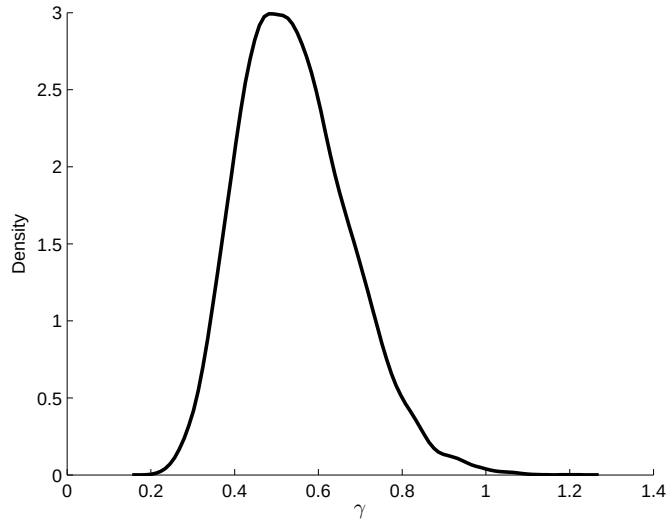
Figure 3.14: (a), posterior mean of the infection rate $\tilde{\beta}_1(t)$ (solid line) at each estimated infection time for SE-Data 5. True infection rate, $\beta_1 = 0.005$. The 95% credible intervals are shown as well. (b), posterior mean of the infection rate $\tilde{\beta}_2(t)$ (solid line) at each estimated infection time for SE-Data 5. True infection rate, $\beta_2 = 0.005$. The 95% credible intervals are shown as well. Only removal times of each group are known. The " | " marks in each plot represent the observed data, i.e. the removal times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter, α, is fixed to 1.

3.4.4 Respiratory disease multi-group data

We also apply our methods to the real multi-group data discussed in Becker & Hopper (1983) and Hayakawa et al. (2003). The dataset corresponds to removal times of individuals with a respiratory disease which occurred between October and November of 1967 on the island of Tristan da Cunha in the South Atlantic. The total population of the island of 255 was partitioned into three groups by age: infants, children and adults. As there was one unidentified case, we suppose $N = 254$. The groups' initial number of susceptibles are $N_1 = 25$, $N_2 = 36$ and $N_3 = 192$. It is notable that we assume the initial number of susceptibles for adults group is 192 instead of 193 given in (Hayakawa et al., 2003) since it is assumed that the group type for the first infection time in the total population is known, i.e. we know that $a_{\text{infants}=0}$, $a_{\text{children}=0}$ and $a_{\text{adults}=1}$, where a_i represents the initial number of infectives in group i. The total number of cases in each group was 9, 6 and 25 respectively. We apply our methods to the data and place the Gaussian process prior on $\tilde{\beta}_{\text{infants}=0}$, $\tilde{\beta}_{\text{children}=0}$ and $\tilde{\beta}_{\text{adults}=1}$ respectively. Figure 3.16 (a), (b) and Figure 3.17 (a) show the estimation results for the infection rate of each group. Table 3.3 gives estimation results of the infection rate and removal rate from Hayakawa et al. (2003) for the real multi-group data, i.e. mean and standard deviation of the infection rate and removal rate are given in Table 3.3. For the groups of infants and children, the results are fairly close to the ones given in Hayakawa et al. (2003). However, the result for the adults group starts from 0.001 and then keeps increasing until nearly 0.0015. One explanation could be that there is significantly more information given from the adults group, i.e. total population of 192 and 25 cases from adults group compared to total population of 25 and 9 cases from infants group and total population of 36 and 6 cases, the Gaussian process prior put on the infection rate may capture features that causes an increasing change on the infection rate from the observed data and tell a different story on the infection rate of the epidemic in the adults group. Figure 3.17 (b) gives the density of the removal rate which implies that we have a good estimation for the removal rate.

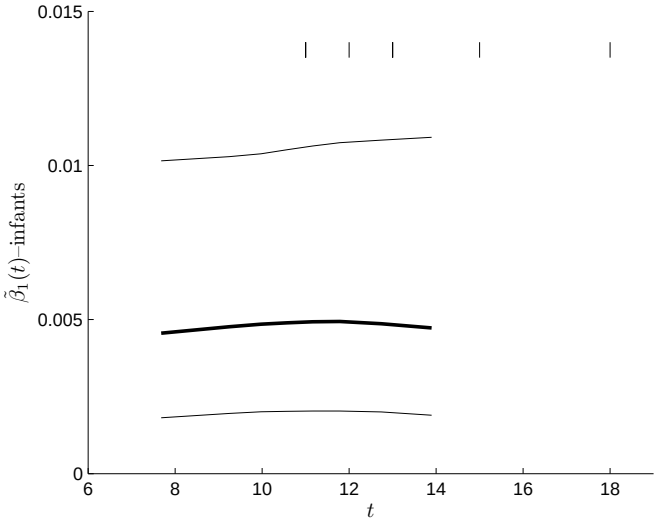


(a)

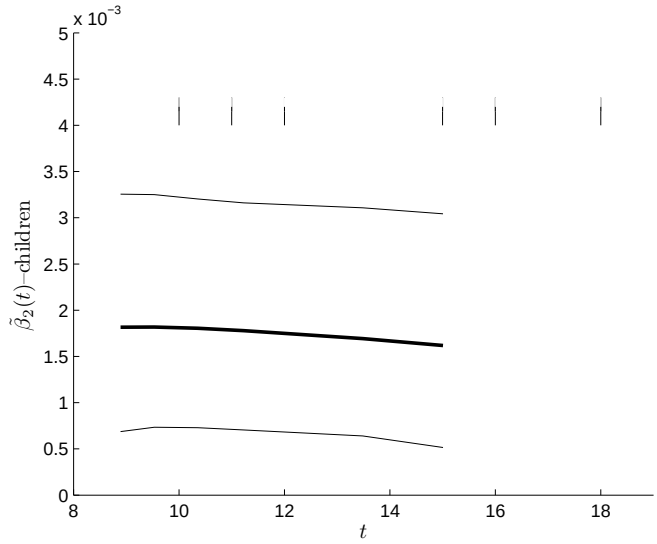


(b)

Figure 3.15: (a), posterior mean of the infection rate $\tilde{\beta}_3(t)$ (solid line) at each estimated infection time for SE-Data 5. True infection rate, $\beta = 0.002$ and true removal rate, $\gamma = 0.5$. The 95% credible intervals are shown as well. (b), density of the removal rate. Only removal times of each group are known. The " | " marks in the plot represent the observed data, i.e. the removal times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

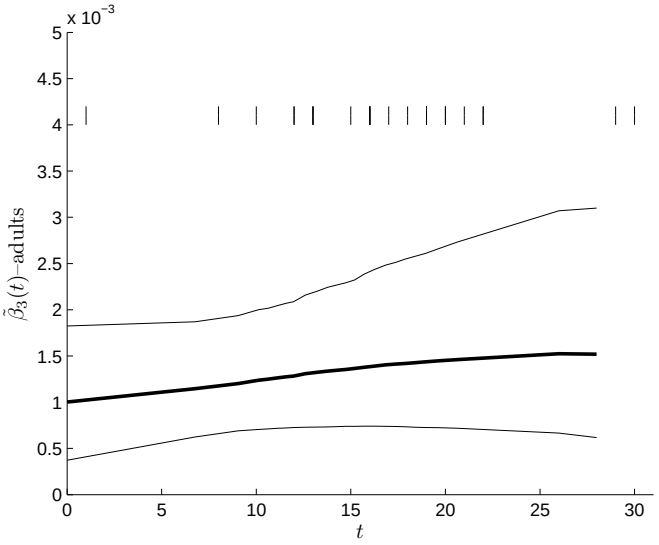


(a)

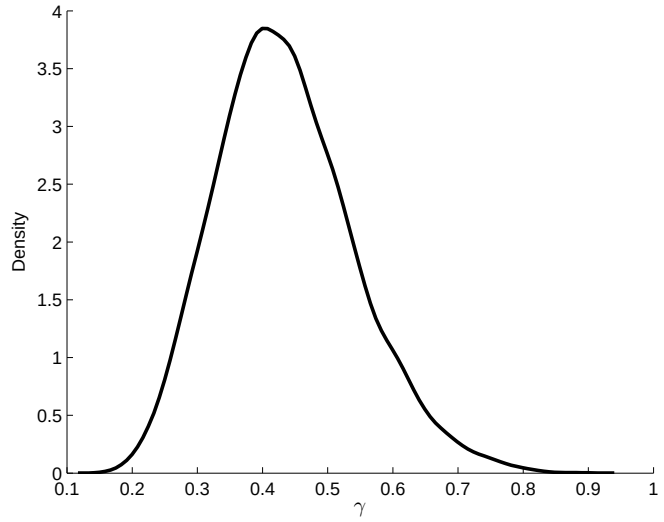


(b)

Figure 3.16: (a), posterior mean of the infection rate $\tilde{\beta}_{\text{infants}}(t)$ (solid line) at each estimated infection time for the respiratory disease data (Hayakawa et al., 2003). The 95% credible intervals are shown as well. (b), posterior mean of the infection rate $\tilde{\beta}_{\text{children}}(t)$ (solid line) at each estimated infection time for the respiratory disease data. The 95% credible intervals are shown as well. Only removal times of each group are known. The " | " marks in each plot represent the observed data, i.e. the removal times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.



(a)



(b)

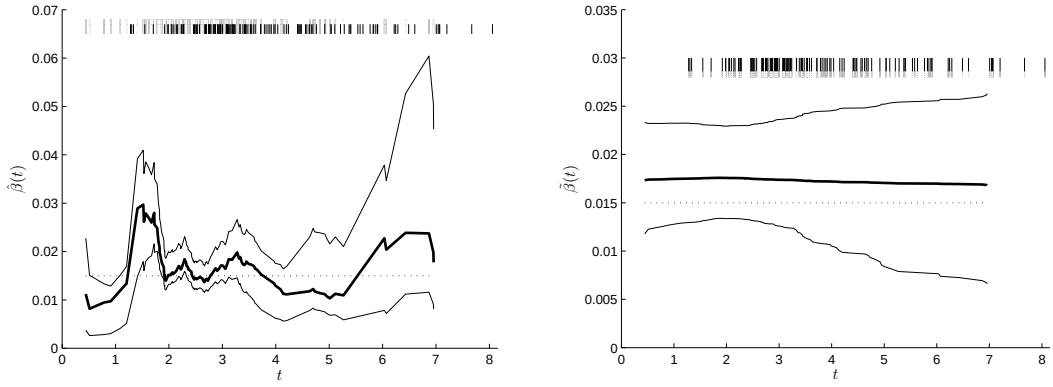
Figure 3.17: (a), posterior mean of the infection rate $\tilde{\beta}_{\text{adults}}(t)$ (solid line) at each estimated infection time for the respiratory disease data (Hayakawa et al., 2003). The 95% credible intervals are shown as well. (b), density of the removal rate. Only removal times in each group are known. The " | " marks in the plot represent the observed data, i.e. the removal times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is fixed to 1.

Table 3.3: Mean and standard deviation of the infection rate in each group for the real multi-group data (Hayakawa et al., 2003). The results are obtained from Hayakawa et al. (2003) where the initial number of susceptibles in each group is assumed to be known.

	β_1	β_2	β_3	γ
mean	0.00451	0.00181	0.00131	0.371
sd	0.00176	0.000824	0.000377	0.0952

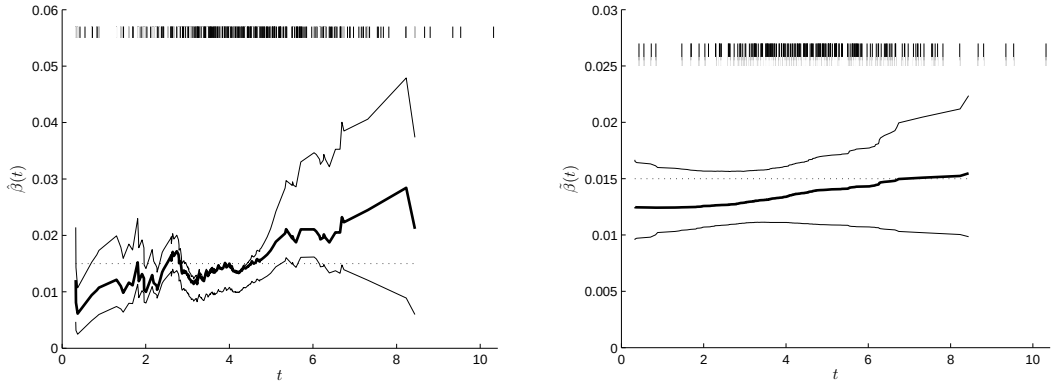
3.5 Comparison with $h(t)$ approach

In this section we compare the approaches discussed in this chapter and in Chapter 2. Recall that in Chapter 2, it is assumed that the overall force of infection is of the form, $h(t)$. Conversely in this chapter, we assume it is of the form, $\tilde{\beta}(t)X_tY_t$. In Chapter 2, Section, 2.3.4, we introduced an estimator of the infection rate which is $\hat{\beta}(t) = \frac{h(t)}{X_tY_t}$. We now compare estimates of $\hat{\beta}(t)$ with $\tilde{\beta}(t)$ using the simulated datasets SE-Data 2 and SE-Data 3 as well as the real outbreak data, the Smallpox data. We assume only removal times are observed. Figure 3.18 (a) and (b) show posterior mean of $\hat{\beta}(t)$ and $\tilde{\beta}(t)$ respectively for the dataset SE-Data 2. From Figure 3.18, the result in plot (a) struggles to recover the truth, although most part of the truth is covered in the 95% credible intervals. Compared to the estimate of $\tilde{\beta}(t)$ in plot (a), the estimate of $\hat{\beta}(t)$ is much more rough. The reason could be that the function, $\hat{\beta}$, is expected to produce a very rough process as discussed in Chapter 2, Section 2.3.4. Figure 3.19, for another dataset SE-Data 3, also illustrates this point. Figure 3.20 shows posterior mean of $\hat{\beta}(t)$ and $\tilde{\beta}(t)$ respectively for the Smallpox data. From Figure 3.20, we find that the estimate of $\hat{\beta}(t)$ at each infection time is over 0.001, whereas the estimate of $\tilde{\beta}(t)$ is around 0.001 all the time. Compared to the results in O'Neill & Roberts (1999), (i.e. the mean of β is 0.0009 and the standard deviation is 0.00019), we conclude that our approached discussed in Chapter 3 is a better way to estimate the infection rate for the SIR epidemic model.



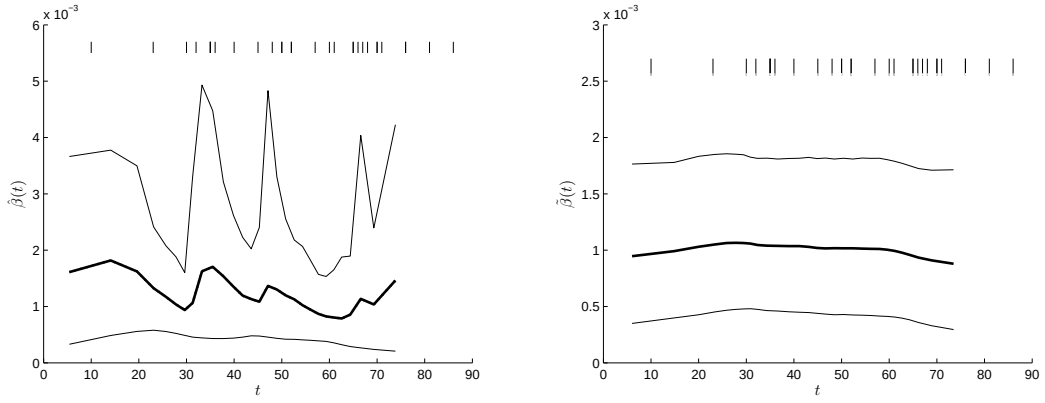
(a) (b)

Figure 3.18: Dataset SE-Data 2 recovered by placing the Gaussian process prior with different Bayesian nonparametric methods. Two plots show posterior mean of the infection rate, $\hat{\beta}(t)$ and $\tilde{\beta}(t)$ (solid line) at each infection time respectively. The original data $\beta X_t Y_t$ (dotted line) are generated from the general stochastic epidemic with parameters infection rate $\beta = 0.015$, removal rate $\gamma = 1$, initial number of susceptibles $N = 150$ and initial number of infective individuals $a = 1$. Only removal times are observed for both cases. Plot (a) corresponds to the case where the approach discussed in Chapter 2 is used. Plot (b) corresponds to the case where the approach discussed in Chapter 3 is used. The 95% credible intervals are shown for each of the plot. The " | " marks in each plot represent the observed data, i.e. the removal times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2 and 1 for $\hat{\beta}(t)$ and $\tilde{\beta}(t)$ respectively.



(a) (b)

Figure 3.19: Dataset SE-Data 3 recovered by placing the Gaussian process prior with different Bayesian nonparametric methods. Two plots show posterior mean of the infection rate, $\hat{\beta}(t)$ and $\tilde{\beta}(t)$ (solid line) at each infection time respectively. The original data $\beta X_t Y_t$ (dotted line) are generated from the general stochastic epidemic with parameters infection rate $\beta = 0.015$, removal rate $\gamma = 1$, initial number of susceptibles $N = 200$ and initial number of infective individuals $a = 1$. Only removal times are observed for both cases. Plot (a) corresponds to the case where the approach discussed in Chapter 2 is used. Plot (b) corresponds to the case where the approach discussed in Chapter 3 is used. The 95% credible intervals are shown for each of the plot. The " | " marks in each plot represent the observed data, i.e. the removal times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2 and 1 for $\hat{\beta}(t)$ and $\tilde{\beta}(t)$ respectively.



(a) (b)

Figure 3.20: Smallpox data recovered by placing the Gaussian process prior with different Bayesian nonparametric methods. Two plots show posterior mean of the infection rate, $\hat{\beta}(t)$ and $\tilde{\beta}(t)$ (solid line) at each infection time respectively. Only removal times are observed for both cases. Plot (a) corresponds to the case where the approach discussed in Chapter 2 is used. Plot (b) corresponds to the case where the approach discussed in Chapter 3 is used. The 95% credible intervals are shown for each of the plot. The " | " marks in each plot represent the observed data, i.e. the removal times. The squared exponential covariance function is used for the Gaussian process prior and the hyperparameter of the covariance function, α, is set to 2 and 1 for $\hat{\beta}(t)$ and $\tilde{\beta}(t)$ respectively.

3.6 Conclusion

In this Chapter, we investigated the behaviour of the overall force of infection in the SIR epidemic model by assuming it has the form $\tilde{\beta}(t)X_tY_t$, i.e. the infection rate is assumed to be a function of time. We developed a Bayesian nonparametric method for estimating such SIR epidemic model and obtained reasonably well estimation results from analysing the single group data including the simulated and the real life data. We also investigated the effect of different Gaussian process priors placed on $\tilde{\beta}(t)$ for the simulated and the real life data. We found that they do not make much difference apart from smoothness. We also described an extended SIR epidemic model with several kinds of susceptibles. We applied our methods to the simulated and real life multi-group data

and good estimation results indicated that our methods work fairly well for the multi-group model as well. Finally, we compared the approach with the one developed in Chapter 2 where we assume the overall force of infection is of the form, $h(t)$, which does not depend on the number of susceptibles and infectives but time. Specifically, we compared the estimate of the infection rate in two different Bayesian nonparametric methods, i.e. $\hat{\beta}(t)$ and $\tilde{\beta}(t)$. We showed that the approach which considering the information of the number of susceptibles and infectives is a better way to estimate the infection rate, although it took longer to run the MCMC algorithms due to the extra computation of the number of susceptibles and infectives.

The approaches developed in Chapter 2 and Chapter 3 both have a limitation, i.e. the computational complexity. It is known that Gaussian processes have heavy computational demands, i.e. $O(n^3)$ time complexity for n input points. As both our approaches are concerned with Gaussian process priors which involve heavy matrix computation, e.g. inverse of the matrix, computation times would become prohibitive when dealing with large datasets, for example, the number of cases exceeds a thousand. An approximation of the model may be considered to tackle this problem.

We have successfully applied Bayesian nonparametric methods to estimate the standard SIR model in small populations where we assume the infectious period follows an exponential distribution. A natural extension of the methods is the relax of the assumption of the infectious period. Specifically, one could apply our methods for the SIR model with non-exponential infectious period in the small-scale epidemics. Another extension could be that one may estimate the SEIR model using our methods where a latent period is considered.

CHAPTER 4

Bayesian nonparametric estimation for epidemic models in large populations from time-series data

4.1 Introduction

In previous chapters, we have developed methods for Bayesian nonparametric inference which involve data augmentation MCMC methods. Although such methods are powerful, they struggle to perform adequately as the dimensionality increases, i.e. as the number of unknown infection times becomes large. In this chapter, we adapt an approximation method due to Cauchemez & Ferguson (2008) to produce a Bayesian nonparametric method of inference for epidemics in large populations. Specifically, this chapter is concerned with methods to parameterise epidemic models from time-series data, e.g. number of observed cases each day or week, etc.

In practice, an epidemic process is only partially observed, and observations are often aggregated in time. Surveillance data typically provide counts of new infection cases occurring during observation periods of length T on a local or national basis and in some situations, the number of new infection cases is likely to be under-reported. We refer to such data as epidemic time-series data.

Standard MCMC uses data augmentation to tackle the missing data problem. However, for large epidemics in large populations, the number of infectives is normally so large that it is infeasible to augment the data with the times of in-

fection/removal of each case. In other words, if the infection times for each individual are all imputed and considered as additional parameters as well as the number of susceptibles and infectives at each infection time, it will take an unacceptably long time for the MCMC algorithm to converge with current computing techniques. Therefore, tractable approximations of the epidemic model, such as the SIR model, need to be adopted. Cauchemez & Ferguson (2008) proposed statistical methods in a Bayesian framework to estimate the continuous-time SIR model from time-series data, when large populations are considered. We first introduce the methods designed by Cauchemez & Ferguson (2008) and then apply our methods, Bayesian nonparametrics, to estimate key quantities of the SIR model, infection rates, from time-series data.

In this chapter, we will first introduce a method to approximate the SIR model when dealing with large-scale epidemics in Section 4.2 and then we will describe a method for inference using an MCMC algorithm in Section 4.3. The two sections together are a brief explanation of Cauchemez & Ferguson's method (2008). We will then describe a simulation study using Cauchemez & Ferguson's method in Section 4.4. In Section 4.5, we will describe our method using a Bayesian nonparametric framework and apply the method to simulated data. We will also explore the method further under different assumptions and parameter settings in this section. Finally, we apply Cauchemez & Ferguson's method and our method to the real data, i.e. a time series of incidence of measles in London (1948-1957) and compare the results of estimation using the two methods.

4.2 Approximation to the SIR epidemic model

4.2.1 Introduction

As described in the previous chapter, the SIR epidemic model is a continuous-time Markovian model which describes the spread of an infectious disease in a population. Below is a mathematical definition of the model with the addition of births into the susceptible population considered in terms of Markovian transition rates:

$$\begin{cases} P(\mathrm{d}S_t = 1|\mathcal{H}_t) = \nu(t)\mathrm{d}t + o(\mathrm{d}t), \\ P(\mathrm{d}S_t = -1, \mathrm{d}I_t = 1|\mathcal{H}_t) = \beta S_t I_t \mathrm{d}t + o(\mathrm{d}t), \\ P(\mathrm{d}I_t = -1, \mathrm{d}R_t = 1|\mathcal{H}_t) = \gamma I_t \mathrm{d}t + o(\mathrm{d}t), \end{cases}$$

where $\nu(t)$ is the birth rate and $\mathcal{H}_t$ is the σ-algebra generated by the history $\{S_u, I_u, R_u; 0 \leq u \leq t\}$. In the equations, the mortality due to disease is assumed to be neglected and the number of individuals who leave the susceptible population due to death or immigration is assumed to be neglected as well.

In reality, the epidemic process for infectious diseases like measles is partially observed. Consider data consisting of the number of new infections, $\{U_k\}_{k=0}^K$, and the number of births, $\{B_k\}_{k=0}^K$, during periods of length T labelled $k = 0, 1, 2, \cdots, K$, where T could be a number of weeks or months. We consider U_k as the number of times in interval $[kT, (k+1)T]$ when I_t increases by $+1$. The relationship among U_k, I_{kT}, $I_{(k+1)T}$ and $\tilde{R}_k$ is given by

$$U_k = I_{(k+1)T} - I_{kT} + \tilde{R}_k,$$

where $\tilde{R}_k$ is number of removals during the $(k+1)^{\text{th}}$ period, and I_{kT} and $I_{(k+1)T}$ are respectively the number of infectives at the beginning of the $(k+1)^{\text{th}}$ and $(k+2)^{\text{th}}$ period. As $\tilde{R}_k$ is always non-negative, then there is a constraint that $U_k \geq I_{(k+1)T} - I_{kT}$. Figures 4.1 and 4.2 show two possible trajectories for the number of infectives I_t with five new infections occurring during the $(k+1)^{\text{th}}$ period, i.e. $U_k = 5$. It is shown that, although the same number of new infections is observed in Figures 4.1 and 4.2, the infection and removal processes between the two cases are very different. As a result of stochastic fluctuations, important differences could be observed between trajectories, even if the infection rate and removal rate were the same.

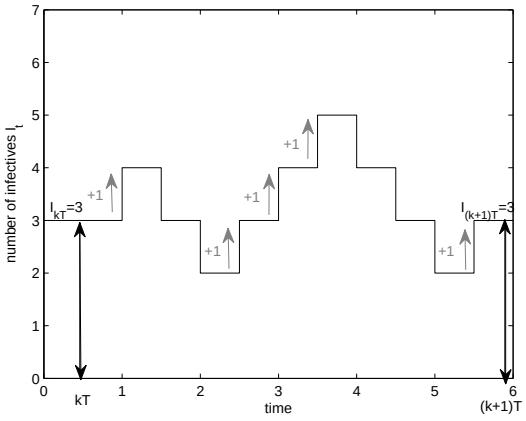


Figure 4.1: An example of the trajectory of the number of infectives I_t with $U_k = 5$, where U_k is the number of new infections occurring during the $(k+1)^{\text{th}}$ period from kT to $(k+1)T$ and $I_{kT} = 3$, $I_{(k+1)T} = 3$.

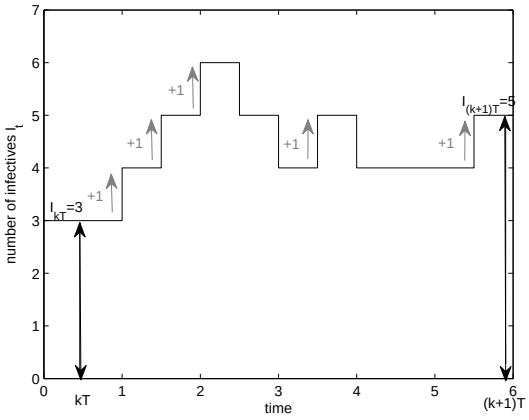


Figure 4.2: An example of the trajectory for the number of infectives I_t with $U_k = 5$, where U_k is the number of new infections occurring during the $(k+1)^{\text{th}}$ period from kT to $(k+1)T$ and $I_{kT} = 3$, $I_{(k+1)T} = 5$.

Suppose we observe I_{kT} and S_{kT}, the number of infectives and susceptibles at the beginning of each observation period (these are considered as parameters of the inference framework since they are unknown in practice). Then, the main issue for inference is to determine the joint probability $P(I_{(k+1)T}, U_k | I_{kT}, S_{kT})$ which also can be expressed as

$$P(I_{(k+1)T}, U_k | I_{kT}, S_{kT}) = P(I_{(k+1)T} | I_{kT}, S_{kT}) P(U_k | I_{(k+1)T}, I_{kT}, S_{kT})$$

(Cauchemez & Ferguson 2008). Without loss of generality, since the epidemic

process is time-homogeneous, we only look at the first observation period, i.e. set $k = 0$. Therefore, the equation above can be written as

$$P(I_T, U_0|I_0, S_0) = P(I_T|I_0, S_0)P(U_0|I_T, I_0, S_0). \tag{4.1}$$

The probability of I_T given I_0, S_0 and the probability of U_0 given I_T, I_0, S_0 in (4.1) are intractable in a large population, hence we use an approximation to tackle this problem. Specifically, the Cox-Ingersoll-Ross diffusion process is adopted to work out the aforementioned probabilities.

4.2.2 The Cox-Ingersoll-Ross diffusion process

The Cox-Ingersoll-Ross process, a diffusion process, is commonly used in financial markets to model interest rates (Cox et al. 1985). Here, we use this diffusion process to approximate the Markovian SIR process under the following assumptions:

(i) Changes in the number of susceptibles S_t within an observation period can be neglected and so $S_t \approx \bar{S_0}, \forall t \in [0, T]$. This assumption is reasonable if S_t is sufficiently large.

(ii) The infection rate is constant during an observation period and so $\beta(t) \approx \beta_0, \forall t \in [0, T]$, where $\beta(t)$ is infection rate at time t which depends on time.

The estimation of $\bar{S_0}$ is given in appendix C in Cauchemez & Ferguson (2008). Under assumptions (i) and (ii), $\{I_t : 0 \leq t \leq T\}$ can be considered as a birth and death process over time $[0, T]$, with birth rate $\beta_0 \bar{S_0} I_t$ and death rate γI_t. Assuming I_t is continuous, we then have a stochastic differential equation, namely

$$\mathrm{d}I_t = r_0 I_t \mathrm{d}t + \sigma_0 \sqrt{I_t} \mathrm{d}W_t, \tag{4.2}$$

where $r_0 = \beta_0 \bar{S_0} - \gamma$, $\sigma_0^2 = \beta_0 \bar{S_0} + \gamma$ and W_t is Brownian motion. The equation (4.2) represents the Cox-Ingersoll-Ross diffusion process and fortunately, the exact solution of (4.2) is available and has a non-central χ^2 distribution with zero degrees of freedom (Cox et al. 1985, siegel 1979). Given I_0 and S_0, the probability density function of I_t is

$$c_0(u_0/(c_0 I_t))^{0.5} \exp(-(u_0 + c_0 I_t))\mathbf{I_1}(\sqrt{2u_0 c_0 I_t}),$$

where

$$c_0 = 2r_0/[(\exp(r_0 T) - 1)\sigma^2],$$

$$u_0 = c_0 \exp(r_0 T) I_0$$

and $\mathbf{I_1}$ is the modified Bessel function of the first kind. The derivation of the p.d.f. of I_t is described below. As $I_0 > 0$, the first part of (4.1) is

$$P(I_T | I_0, S_0) = 2c_0 f_{2u_0}(2c_0 I_T)$$

and $f(\cdot)$ is defined below.

Siegel (1979) stated that $(2c_0 I_T | I_0, S_0)$ follows a non-central χ^2 distribution with zero degrees of freedom and with non-centrality parameter $2u_0$. The non-central χ^2 distribution with zero degrees of freedom has a mass at 0, which corresponds to the probability of extinction of the outbreak. Suppose $X \sim \chi_0^2(\lambda)$ where λ is the non-centrality parameter, then we have $P(X = 0) = \exp(-\lambda/2)$. The positive part of the distribution has a density f_λ if $X \sim \chi_0^2(\lambda)$ and $0 \leq a < b$, then

$$P(a < X < b) = \int_a^b f_\lambda(x) \mathrm{d}x,$$

$$f_\lambda(x) = 0.5(\lambda/x)^{0.5} \exp(-0.5(\lambda + x)) \mathbf{I_1}(\sqrt{\lambda x}).$$

Note that, $\mathbf{I_1}(\sqrt{\lambda x})$ may become extremely large when λ is large which can cause a numerical problem. In the application discussed later, the value of $\mathbf{I_1}(\sqrt{\lambda x})$ may reach 10^{308} which is beyond the range of double precision number storage in C++. Therefore, a Normal distribution with the same mean, λ, and same variance, 4λ, i.e. $N(\lambda, 4\lambda)$ can be used to approximate the non-central χ^2 distribution when λ is large. Rather than using the method above mentioned in Cauchemez & Ferguson (2008), we adopt a different method which approximates the modified Bessel function, and hence approximates the non-central χ^2 distribution. Abramowitz & Stegun (1970 p.377) developed asymptotic expansions for large arguments for the modified Bessel function and the method is adopted for our case to approximate $\mathbf{I_1}(\sqrt{\lambda x})$. Specifically,

$$\begin{aligned} I_\nu(y) \sim & \frac{e^y}{\sqrt{2\pi y}} \{1 - \frac{4\nu^2 - 1}{8y} + \frac{(4\nu^2 - 1)(4\nu^2 - 9)}{2!(8y)^2} \\ & - \frac{(4\nu^2 - 1)(4\nu^2 - 9)(4\nu^2 - 25)}{3!(8y)^3} + \cdots \} \text{ for } (|\arg y| < 1/2\pi), \end{aligned}$$

where $\nu = 1$ and $y = \sqrt{(\lambda x)}$ for our case.

In order to evaluate the second part of (4.1), the probability of U_0 given I_T, I_0 and S_0, Cauchemez & Ferguson (2008) adopted a negative Binomial distribution to approximate the variable $U_0|I_T, I_0, S_0$. In fact, the number of new infection cases, U_0, occurring in $[0, T]$ is Poisson distributed with mean $E_0 = \beta_0 \bar{S_0} \int_0^T I_t \mathrm{d}t$. If we approximate $E_0|I_T, I_0, S_0$ using a Gamma distribution, then a negative Binomial distribution is constructed. In other words, a negative Binomial can be considered as a Poisson distribution with mean, say λ, where λ is itself a random variable, distributed as a Gamma distribution. Formally, the probability mass function of the negative Binomial distribution, $f(k; r, p)$, can be derived through the method discussed above as follows:

$$\begin{aligned}
f(k; r, p) &= \int_0^\infty f_{\text{Poisson}(\lambda)}(k) \cdot f_{\text{Gamma}\left(r, \frac{p}{1-p}\right)}(\lambda)\ \mathrm{d}\lambda \\
&= \int_0^\infty \frac{\lambda^k}{k!} e^{-\lambda} \cdot \lambda^{r-1} \frac{e^{-\lambda(1-p)/p}}{\left(\frac{p}{1-p}\right)^r \Gamma(r)}\ \mathrm{d}\lambda \\
&= \frac{(1-p)^r p^{-r}}{k!\,\Gamma(r)} \int_0^\infty \lambda^{r+k-1} e^{-\lambda/p}\ \mathrm{d}\lambda \\
&= \frac{(1-p)^r p^{-r}}{k!\,\Gamma(r)}\ p^{r+k}\,\Gamma(r+k) \\
&= \frac{\Gamma(r+k)}{k!\,\Gamma(r)}\ (1-p)^r p^k \text{ for } k = 0, 1, 2, \cdots
\end{aligned}$$

where r represents the number of failures and p represents the probability of success in a sequence of independent Bernoulli trials.

Supposing that the Gamma distribution used to approximate $E_0|I_T, I_0, S_0$ has mean M_0 and variance V_0, the approximation for the distribution $P(U_0|I_T, I_0, S_0)$ is

$$P(U_0|I_T, I_0, S_0) = \int_0^\infty P(U_0|E_0) P(E_0|I_T, I_0, S_0)\ \mathrm{d}E_0,$$

which follows the negative Binomial distribution with mean M_0 and variance $M_0 + V_0$. To determine the mean M_0 and the variance V_0 of $E_0|I_T, I_0, S_0$, a linear model suggested by Cauchemez & Ferguson (2008) is used to derive M_0 and V_0 with the Laplace transform method shown below:

$$E_0 = x_0 + y_0 I_T + \epsilon_0,$$

where ϵ_0 is the error. If we denote by $\tilde{x}_0$ and $\tilde{y}_0$ the values which minimise $E[(E_0 - x_0 - y_0 I_T)^2]$ and $\tilde{v}_0$ the variance of $E_0 - \tilde{x}_0 - \tilde{y}_0 I_T$, the mean M_0 and

the variance V_0 can be approximated by $\tilde{x}_0 + \tilde{y}_0 I_T$ and $\tilde{v}_0$ respectively. Detailed analysis of the Laplace transform and the estimated $\tilde{x}_0$, $\tilde{y}_0$ and $\tilde{v}_0$ are given below (Cauchemez & Ferguson, 2008).

The Laplace transform of $(I_T, \int_0^T I_t \mathrm{d}t)|I_0, S_0$ is

$$\begin{aligned} F(m,n) &= E\left\{\exp\left(-xI_T - y\int_0^T I_t \mathrm{d}t\right)|I_0, S_0\right\} \\ &= \exp(-Y(m,n)I_0), \end{aligned}$$

where $\xi(y) = \sqrt{r_0^2 + 2y\sigma_0^2}$,

$$Y(m,n) = \frac{m(\xi(n) - r_0 + e^{\xi(n)T}(\xi(n) + r_0)) + 2n(e^{\xi(n)T} - 1)}{(x\sigma_0^2 + \xi(n) - r_0)(e^{\xi(n)T} - 1) + 2\xi(n)}.$$

Define $(\tilde{v}, \tilde{w})$ the scalars that minimise the function

$$L(v,w) = E\left[\left(\int_0^T I_t \mathrm{d}t - v - wI_T\right)^2 |I_0, S_0\right].$$

The solution can be obtained analytically by setting $\frac{\partial L}{\partial v}(\tilde{v}, \tilde{w}) = 0$ and $\frac{\partial L}{\partial w}(\tilde{v}, \tilde{w}) = 0$, i.e.

$$\tilde{w} = \frac{\mathrm{cov}(I_T, \int_0^T I_t \mathrm{d}t|I_0, S_0)}{\mathrm{var}(I_T|I_0, S_0)} = \frac{\partial^2 F/\partial m \partial n}{\partial^2 F/\partial^2 m}\bigg|_{(0,0)} = \frac{1}{r_0} - \frac{T}{e^{r_0 T} - 1},$$

$$\begin{aligned} \tilde{v} &= E(\int_0^T I_t \mathrm{d}t|I_0, S_0) - \tilde{w}E(I_T|I_0, S_0) \\ &= -\frac{\partial F}{\partial n}\bigg|_{(0,0)} + \tilde{w}\frac{\partial F}{\partial m}\bigg|_{(0,0)} \\ &= \frac{I_0(e^{r_0 T} - 1)}{r_0} - \tilde{w}I_0 e^{r_0 T}. \end{aligned}$$

Denoting $A_T = \int_0 T I_t \mathrm{d}t - \tilde{v} - \tilde{w}I_T$, the residual, the Laplace transform of $(I_T, A_T)|I_0, S_0$ is

$$\begin{aligned} G(m,n) &= E\{\exp(-mI_T - nA_T)|I_0, S_0\} \\ &= E\left\{\exp\left(-mI_T - n\left(\int_0^T I_t \mathrm{d}t - \tilde{v} - \tilde{w}I_T\right)|I_0, S_0\right)\right\} \\ &= F(m - n\tilde{w}, n)\exp(m\tilde{v}). \end{aligned}$$

The mean and variance of the residual are 0 and $\tilde{z}$, where

$$\tilde{z} = \frac{\partial^2 G}{\partial^2 n} = \frac{I_0\sigma_0^2(1 + e^{2r_0 T} - e^{r_0 T}(2 + r_0^2 T^2))}{r_0^3(e^{r_0 T} - 1)}$$

Then scalars $\{\tilde{x}_0, \tilde{y}_0, \tilde{v}_0\}$ can be obtained from $\{\tilde{v}, \tilde{w}, \tilde{z}\}$ and the definition of E_0

$$\tilde{y}_0 = \beta_0 \bar{S}_0 (1/r_0 - T/(e^{r_0 T} - 1)),$$

$$\tilde{x}_0 = \beta_0 \bar{S}_0 I_0 (e^{r_0 T} - 1)/r_0 - I_0 e^{r_0 T} \tilde{y}_0,$$

$$\tilde{v}_0 = (\beta_0 \bar{S}_0)^2 \frac{I_0 \sigma_0^2 (1 + e^{2r_0 T} - e^{r_0 T}(2 + r_0^2 T^2))}{r^3 (e^{r_0 T} - 1)}.$$

4.3 Inference

We consider the situation that observations are only the time series $\{U_k^*\}_{k=0}^K$, $\{B_k\}_{k=0}^K$, where $\{U_k^*\}$ is the number of reported cases, $\{B_k\}$ is the number of births and K is the number of observation periods and the removal rate, γ is assumed to be known. Then the data can be augmented with $\{U_k\}_{k=0}^K$, $\{I_{kT}\}_{k=0}^{K+1}$ and $\{S_0\}$ considered as parameters. Note that given S_0 and $\{U_k, B_k\}_{k=0}^{K-1}$, $\{S_{kT}\}_{k=1}^K$ can be calculated deterministically via the relationship

$$S_{kT} = S_0 + \sum_{i=0}^{k-1} (B_i - U_i).$$

Assume $\{U_k^*\}$ follows a Binomial distribution

$$U_k^* | U_k \sim Bin(U_k, \rho),$$

where ρ is the proportion of the reported cases. Note that, the proportion, ρ, is a constant value throughout the whole epidemic, i.e. the proportion of the reported cases in each observation period is the same. Finkenstädt & Grenfell (2000) adopted local regressions, leading to the estimation of a sequence of reporting rates $\{\rho_k\}_{k=0}^K$ in order to tackle a problem caused by changes in the structure of the population (through changes in birth rates). However, we do not explore this topic as the structure of the population is not modelled explicitly in our case. Denote by β_k the infection rate for the $(k+1)^{\text{th}}$ observation period. The joint distribution of the observations, augmented data and parameters is as follows.

$$\begin{aligned}
&\pi(I_{(K+1)T}, \{I_{kT}, U_k\}_{k=0}^K, S_0, \rho, \{\beta_k\}_{k=0}^K | \{U_k^*, B_k\}_{k=0}^K) \\
&= \prod_{k=0}^K \{P(U_k^* | U_k, \rho) P(I_{(K+1)T}, U_k | I_{kT}, S_0, \{U_i, B_i\}_{i=0}^{k-1}, \{\beta_k\})\} \\
&\quad \times P(S_0, I_0 | \rho, \{\beta_k\}_{k=0}^K) \pi(\rho, \{\beta_k\}_{k=0}^K). \qquad (4.3)
\end{aligned}$$

The last term in the formulation (4.3) is the prior density for the parameters ρ and $\{\beta_k\}_{k=0}^K$. The posterior distribution of augmented data and parameters then can be written down easily and explored by MCMC sampling, as demonstrated in the following section.

4.3.1 MCMC sampling

We now describe an MCMC algorithm for sampling the reporting rate, ρ, the infection rates, $\{\beta_k\}_{k=0}^K$, the total number of new infections, $\{U_k\}_{k=0}^K$, the initial number of susceptibles, S_0, and the number of infectives, $\{I_{kT}\}_{k=0}^{K+1}$. We use the same method for sampling all the parameters as described in Cauchemez & Ferguson (2008) except sampling the reporting rate, ρ.

4.3.1.1 Sampling the reporting rate

As the reporting rate, ρ only depends on U_k^* and U_k, considering the joint distribution above, the full conditional density for ρ is

$$\pi(\rho|\{U_k^*, U_k\}_{k=0}^K) \propto \prod_{k=0}^{K} P(U_k^*|U_k, \rho) \times \pi(\rho).$$

Based on the posterior distribution of ρ, we could assume the prior of the reporting rate follows a Beta distribution, e.g. $Beta(1,1)$, in which case the posterior distribution of the reporting rate hence follows a Beta distribution too, i.e. $Beta(1 + \sum_{k=0}^K (U_k^*), 1 + \sum_{k=0}^K (U_k - U_k^*))$. We then use a Gibbs step to update the reporting rate, as opposed to using a Metropolis-Hastings algorithm and updating ρ by a random walk on the real line, as adopted by Cauchemez & Ferguson (2008).

4.3.1.2 Sampling the infection rates

We adopt a reparametrisation technique to reduce correlation between the infection rates and the initial number of susceptibles, S_0, by assuming $\beta_k^* = S_0\beta_k$. Therefore, we update $\{\beta_k^*\}_{k=0}^K$ instead of $\{\beta_k\}_{k=0}^K$. We assign a prior distribution $U[0, 10000]$ on β_k^*. The Jacobian term must be calculated when the posterior distribution is computed as the transformed parameter is updated. In particular, we assume $\beta_k^* = S_0\beta_k$ and $S_0^* = S_0$ and transform from β, S_0 to β^*, S_0^*, then

we have

$$\pi_{\beta^*,S_0^*}(\beta^*, S_0^*) = \pi_{\beta,S_0}(\beta^*/S_0^*, S_0) \left| \frac{\partial(\beta,S_0)}{\partial(\beta^*,S_0^*)} \right|,$$

where the Jacobian term

$$\left| \frac{\partial(\beta,S_0)}{\partial(\beta^*,S_0^*)} \right| = \begin{vmatrix} \frac{\partial\beta}{\partial\beta^*} & \frac{\partial S_0}{\partial\beta^*} \\ \frac{\partial\beta}{\partial S_0^*} & \frac{\partial S_0}{\partial S_0^*} \end{vmatrix}.$$

Therefore, the Jacobian term for this case is $1/S_0$. Considering the joint distribution, the full conditional density for β_k is

$$\begin{aligned} &\pi(\beta_k|\{I_{(k)T}, I_{(k+1)T}, S_0, \{U_i, B_i\}_{i=0}^{k-1}\}) \\ &\quad \propto P(I_{(k+1)T}, U_k|I_{(k)T}, S_0, \{U_i, B_i\}_{i=0}^{k-1}, \beta_k^*) \times \pi(\beta_k^*). \end{aligned}$$

We use a Metropolis-Hastings step to sample from the transformed infection rates, $\{\beta_k^*\}_{k=0}^K$. We iterate over each of the $K+1$ transformed infection rates for each observation period and propose a candidate value by a random walk on the real line.

4.3.1.3 Sampling the total number of new infections

We obtain a full conditional density for U_k by considering the joint distribution in the formulation (4.3), yielding

$$\begin{aligned} &P(U_k|I_{(K+1)T}, S_0, \{I_{iT}, U_i, B_i, \beta_i\}_{i=k+1}^K) \\ &\propto P(U_k^*|\rho, U_k) P(U_k|I_{(k+1)T}, I_{kT}, S_0, \{U_i, B_i\}_{i=0}^{k-1}, \beta_k) \\ &\times \prod_{i=k+1}^{K} P(I_{(i+1)T}, U_i|I_{iT}, S_0, \{U_j, B_j\}_{j=0}^{i-1}, \beta_i). \end{aligned}$$

Following the method used by Cauchemez & Ferguson (2008), we update the total number of new infections, $\{U_k^*\}_{k=0}^K$, using an independence sampler (Brooks 1998). As discussed in Section 4.2.2, $U_k|U_k^*$ is assumed to follow the negative Binomial distribution, so we propose a new candidate $\tilde{U}_k$ as follows: $\tilde{U}_k = U_k^* + X_k$, where X_k is drawn from the negative Binomial distribution $(U_k^* + a, ((\rho + b)/(1 + a)))$ with fixed $a = b = 10^{-5}$.

4.3.1.4 Sampling the initial number of susceptibles

We assign a prior distribution $U[0, 10^7]$ on S_0. We then have a full conditional density for S_0 by considering the joint distribution in the formulation (4.3), giving

$$\begin{aligned} &P(S_0|I_{(K+1)T}, \{I_{kT}, U_k, B_k, \beta_k\}_{k=1}^K) \\ &\propto \prod_{k=1}^K P(I_{(k+1)T}, U_k|I_{kT}, S_0, \{U_i, B_i\}_{i=0}^{k-1}, \beta_k) \times P(S_0). \end{aligned}$$

As we adopt the reparametrisation technique and assume $\beta_k^* = S_0\beta_k$, according to the equation of the Jacobian term given in Section 4.3.1.2, we need to add $\log(S_0/S_0')$ for this case, to the acceptance ratio on the log scale, where S_0' is the proposed initial number of susceptibles. We also wish to sample from the posterior distribution on the initial number of susceptibles, S_0 and we use a Metropolis-Hastings step to perform this sampling and update S_0 with a random walk proposed on the real line.

4.3.1.5 Sampling the number of infectives

We put a prior distribution $U[0, 10^7]$ on I_0 which is the same as for S_0. We then derive the full conditional density for I_{kT} by considering the joint distribution in the formulation (4.3), yielding

$$\begin{aligned} &P(I_{kT}|I_{(k-1)T}, I_{(k+1)T}, S_0, \{U_i, B_i\}_{i=0}^{k-1}, \beta_k, \beta_{k-1}) \\ &\propto P(I_{(k+1)T}, U_k|I_{kT}, S_0, \{U_i, B_i\}_{i=0}^{k-1}, \beta_k) \\ &\quad \times P(I_{kT}, U_{k-1}|I_{(k-1)T}, S_0, \{U_i, B_i\}_{i=0}^{k-2}, \beta_{k-1}) \times P(I_0). \end{aligned}$$

We update the number of infectives, $\{I_{kT}\}_{k=0}^{K+1}$, with a random walk proposed on the log scale, e.g. $I_0' = e^{\log I_0 + Z}$, where $Z \propto N(0, \sigma^2)$. Supposing that the probability density function of Z is φ, then we have

$$\pi_{I_0'}(I_0') = \varphi\left(\log(I_0'/I_0)\right) \left|\frac{\mathrm{d}I_0'}{\mathrm{d}Z}\right|^{-1}.$$

As $I_0'(Z) = I_0 e^Z$, the Jacobian term $\left|\frac{\mathrm{d}I_0'}{\mathrm{d}Z}\right|^{-1} = 1/(I_0 e^Z) = 1/I_0'$. Therefore, $\log(I_{kT}'/I_{kT})$ needs to be added to the acceptance ratio on the log scale.

For the MCMC algorithms discussed above, the standard deviations of the proposals were tuned to obtain an acceptance rate of 20-30%.

4.4 Simulation study using Cauchemez & Ferguson's method

The epidemic process is simulated for 10 years from the true SIR model with a parameter setting below:

- initial number of susceptibles: $S_0 = 160,000$,
- initial number of infectives: $I_0 = 900$,
- birth rate: 2152 births per two weeks which is the average birth rate in London between 1944 and 1964,
- reporting rate: $\rho = 45\%$,
- mean infectious period: $1/\gamma = 14$,
- infection rates: $\{\beta_k\}_{k=0}^{25}$, shown in Table 4.1, are set under level of 10^{-7} with a similar shape in simulation study section in Cauchemez & Ferguson's paper (2008).

Table 4.1: Infection rates for each observation period under level of 10^{-7}.

β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
5.5	4.4	5.5	5.6	5.1	4.9	5.0	4.6	3.9
β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}	β_{17}
4.0	4.4	4.6	4.1	3.9	3.7	3.2	2.6	2.9
β_{18}	β_{19}	β_{20}	β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	
3.7	5.5	5.6	5.1	4.6	5.0	5.1	4.6	

We consider the situation where observed data are collected every two weeks, and where infection rates vary every two weeks with a period of 1 year, i.e. the infection rates are seasonal (1 calendar year per season) and mathematically, $\beta_k = \beta_{k+(T_2/T_1)*p}$, for $p = 0, 1, 2, ..., 9$, where $T_1(= 2)$ represents length of the observation period in weeks, $T_2(= 52)$ shows there are 52 weeks a season and p gives season number, e.g. it is the second season if $p = 1$. Therefore, for this case, we have 26 infection rates to estimate.

By using the existing method proposed by Cauchemez & Ferguson (2008), estimation results are shown in Figures 4.3, 4.4, 4.5 and 4.6. We can see, from Figure 4.6, the MCMC chain converges to the truth rapidly and from Figures 4.3, 4.4, 4.5, the reasonable estimation results indicate that we successfully reproduced Cauchemez & Ferguson's method for the 10 years simulation data generated from the SIR model.

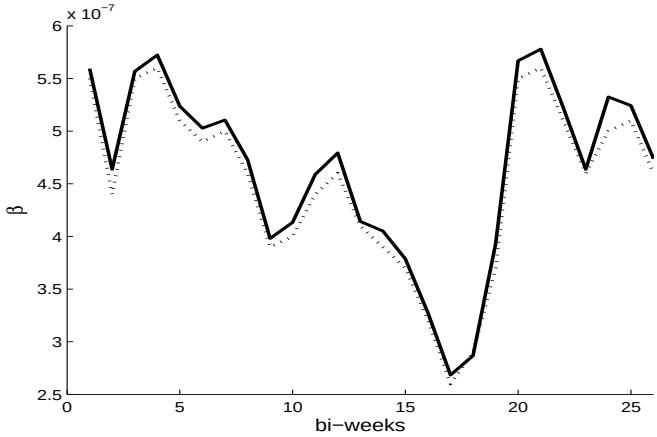


Figure 4.3: 10 years simulation data analysed by Cauchemez & Ferguson's method. Posterior mean (solid line) and true value (dashed line) of the infection rates over 1 year for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days. The infection rates vary every two weeks with a period of 1 year.

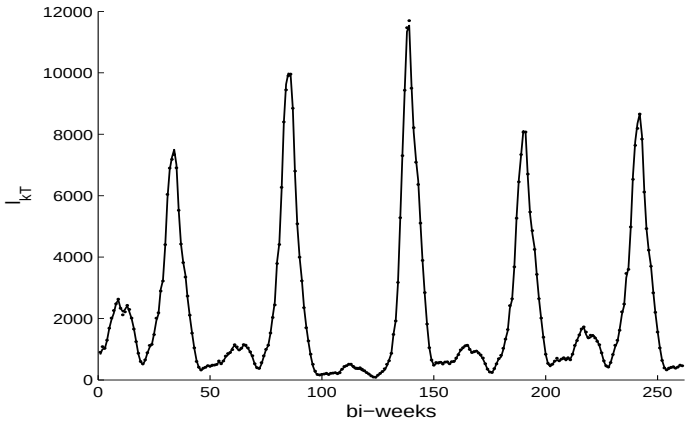


Figure 4.4: 10 years simulation data analysed by Cauchemez & Ferguson's method. Posterior mean (solid line) and true value (dot) of the number of infectives over 10 years for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days.

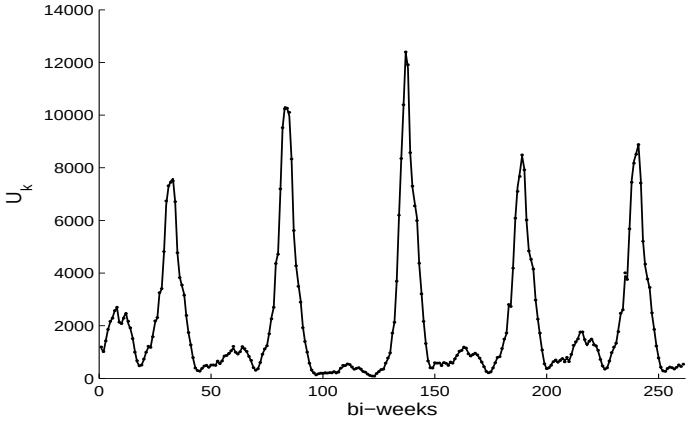


Figure 4.5: 10 years simulation data analysed by Cauchemez & Ferguson's method. Posterior mean (solid line) and true value (dot) of the total number of new infections over 10 years for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days.

4.5 Bayesian nonparametric estimation on infection rates

We now estimate infection rates, $\{\beta_k\}_{k=0}^{K}$, under a Bayesian nonparametric framework. For the convenience, we shall abbreviate Cauchemez & Ferguson's method by the CF method and our Bayesian nonparametric method by the BNP method. The key difference between the CF method and the BNP method is that in CF method, the infection rate for the $(k+1)^{\text{th}}$, β_k, is estimated independently, i.e. the model does not impose any constraints on β_{k-1}, β_k, β_{k+1} despite the fact that, in reality, they should be related. But our method, the BNP method, does enable us to add such constraints. In the previous chapter, we analysed small-scale epidemics under the Bayesian nonparametric framework by considering the epidemic process as an inhomogeneous Poisson process. Function values for the Gaussian process can be drawn from a multivariate Gaussian distribution corresponding to the Gaussian inputs, i.e. the infection times for each individual. However, in the present problem, it is infeasible to place a Gaussian process prior on the infection rates corresponding to the infection times for each individual in the population since if the infection times for each individual are all augmented, the covariance matrix will be extremely large and the computation of the matrix inverse will take too long. Fortunately, with the help of the

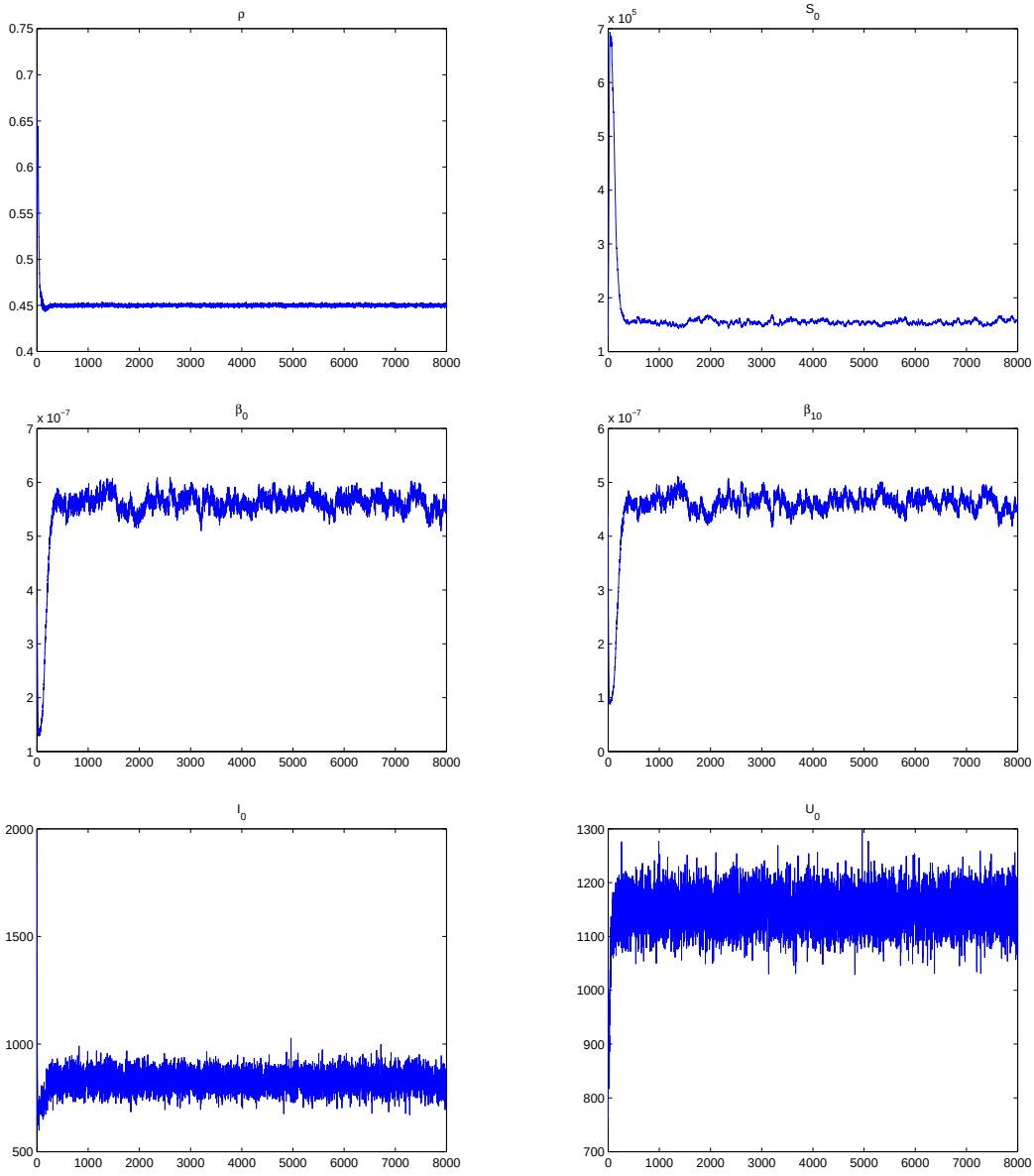


Figure 4.6: Convergence of the MCMC algorithm for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days under Cauchemez & Ferguson's method. The plots show MCMC trace of the reporting rate, ρ, the initial number of susceptibles, S_0, the infection rate for the first observation period, β_0, the infection rate for the 11^{th} observation period, β_{10}, the initial number of infectives, I_0 and the total number of new infections for the first observation period.

method that approximates the SIR model for large-scale epidemics, we are able to place a Gaussian prior on the infection rates corresponding to some times points of each observation period, e.g. the middle time points of each observation period can be chosen as Gaussian process inputs. Then we can perform inference for the infection rates by implementing the MCMC algorithm. Note that the number of the Gaussian process inputs is fixed. For the CF method, the seasonal infection rates are assumed independent from each other and are given independent uniform prior distributions and updated separately. For the BNP method, however, we put a Gaussian process prior on the joint distribution of all infection rates. Hence, the infection rates depend on each other and are updated together simultaneously.

4.5.1 Seasonal assumption on infection rates

We again put a seasonal assumption on the infection rates as for the CF method, i.e. $\beta_k = \beta_{k+26p}$. The reparametrisation technique is used again to reduce correlation between the infection rates and the initial number of susceptibles, S_0 by assuming $\beta_k^* = S_0\beta_k$. We denote the middle time points of each observation period within a season, i.e. the Gaussian process inputs, by $\{s_j\}_{j=1}^J$ and denote the transformed infection rates, by $\{\beta_k^*\}_{k=0}^K$, within a season, the Gaussian process function values, by $\mathbf{g}_J$, where J is 26 and K is 25 in this case. We also denote by respectively α and θ the hyperparameters in the covariance function of the Gaussian process. We now place a Gaussian process prior with a squared exponential covariance function on the transformed infection rates, $\{\beta_k^*\}_{k=0}^K$, within a season, where the squared exponential covariance function is

$$K(x, x') = \alpha^2 \exp\left(-\frac{1}{2\theta^2}(x - x')^2\right).$$

The GP prior can be written as $\pi(\mathbf{g}_J|\{s_j\}_{j=1}^J, \alpha, \theta)$. As we put a prior on all the infection rates throughout the whole season, we have the posterior density of $\{\beta_k\}_{k=0}^K$ as

$$\begin{aligned}
&\pi(\{\beta_k\}_{k=0}^K|I_{(k+1+26\times9)T}\{I_{kT}, U_k, B_k\}_{k=0}^{K+26\times9}, S_0) \\
&\quad\propto \prod_{p=0}^{9}\prod_{k=0}^{K} P(I_{(k+1+26p)T}, U_{k+26p}|I_{(k+26p)T}, S_0, \{U_i, B_i\}_{i=0}^{k-1+26p}, \beta_k^*) \\
&\qquad\times \pi(\mathbf{g}_J|\{s_j\}_{j=1}^J, \alpha, \theta). \qquad (4.4)
\end{aligned}$$

The value of p is chosen from 0 to 9 in the formulation (4.4) since 10 years of time series data are assumed to be observed.

4.5.1.1 Sampling the function values

Given the GP inputs $\{s_j\}_{j=1}^J$ and the current hyperparameters α and θ, we wish to sample from the distribution on the function values $\mathbf{g}_J$. We propose new function values $\mathbf{g}'_J$ by a proposal of the form $\mathbf{g}'_J = \delta\mathbf{g}_J + \sqrt{1-\delta^2}\mathbf{h}_J$, where $\mathbf{h}_J$ is drawn from the Gaussian process at $\{s_j\}_{j=1}^J$ and δ is in $[0,1)$. We set δ to 0.99 for our case to obtain a good acceptance ratio. We define

$$L = \prod_{p=0}^{9}\prod_{k=0}^{K} P(I_{(k+1+26p)T}, U_{k+26p} | I_{(k+26p)T}, S_0, \{U_i, B_i\}_{i=0}^{k-1+26p}, \beta_k^*)$$

and

$$L' = \prod_{p=0}^{9}\prod_{k=0}^{K} P(I_{(k+1+26p)T}, U_{k+26p} | I_{(k+26p)T}, S_0, \{U_i, B_i\}_{i=0}^{k-1+26p}, \beta_k'^*),$$

then we derive the Metropolis-Hastings ratio below

$$\begin{aligned}
&\frac{\pi(\mathbf{g}'_J)}{\pi(\mathbf{g}_J)} \times \frac{q(\mathbf{g}'_J \to \mathbf{g}_J)}{q(\mathbf{g}_J \to \mathbf{g}'_J)} \\
&= \frac{L' \times \pi(\mathbf{g}'_J | \{s_j\}_{j=1}^J, \alpha, \theta)}{L \times \pi(\mathbf{g}_J | \{s_j\}_{j=1}^J, \alpha, \theta)} \times \frac{\pi(\mathbf{g}_J | \{s_j\}_{j=1}^J, \alpha, \theta)}{\pi(\mathbf{g}'_J | \{s_j\}_{j=1}^J, \alpha, \theta)} \\
&= \frac{L'}{L}.
\end{aligned}$$

4.5.1.2 Sampling the hyperparameters

We assign prior distributions $U[0, 10^7]$ on the hyperparameters α and θ. We then have respectively the full conditional densities of α and θ, $\pi(\alpha | \mathbf{g}_J, \{s_j\}_{j=1}^J, \theta)$ and $\pi(\theta | \mathbf{g}_J, \{s_j\}_{j=1}^J, \alpha)$. Given the GP function values, we use a Metropolis-Hastings step and update α and θ by random walks on the real line. Denote by $\Sigma_{\alpha,\theta}$ the Gaussian process covariance matrix with current value of hyperparameters α and θ. The Metropolis-Hastings ratios for α and θ are

$$\frac{|\Sigma_{\alpha,\theta'}|^{-\frac{1}{2}} \times \exp(-{\mathbf{g}_J}^T {\Sigma_{\alpha,\theta'}}^{-1} \mathbf{g}_J / 2)}{|\Sigma_{\alpha,\theta}|^{-\frac{1}{2}} \times \exp(-{\mathbf{g}_J}^T {\Sigma_{\alpha,\theta}}^{-1} \mathbf{g}_J / 2)},$$

and

$$\frac{|\Sigma_{\alpha',\theta}|^{-\frac{1}{2}} \times \exp(-\mathbf{g}_J{}^T \Sigma_{\alpha',\theta}{}^{-1} \mathbf{g}_J/2)}{|\Sigma_{\alpha,\theta}|^{-\frac{1}{2}} \times \exp(-\mathbf{g}_J{}^T \Sigma_{\alpha,\theta}{}^{-1} \mathbf{g}_J/2)}.$$

4.5.1.3 Results

To compare with the CF method, we apply the BNP method to the same data simulated in Section 4.4. Estimation results for the BNP method are shown in Figures 4.7, 4.8, 4.9 and 4.10. Compared to the CF method, the BNP method also provides reasonable estimations of infection rates of the epidemic as well as all the other parameters which motivates us to explore the method more.

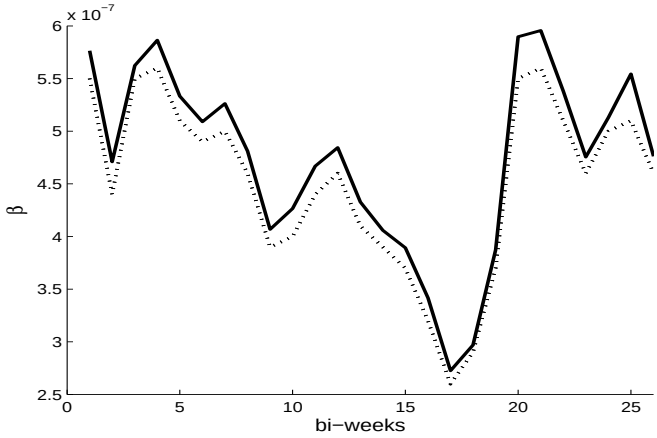


Figure 4.7: Posterior mean (solid line) and true value (dashed line) of the infection rates over 1 year for the SIR epidemic simulated with mean infectious period $1/\gamma = 14$ days under the BNP method. The infection rates vary every two weeks with a period of 1 year.

4.5.2 Seasonal assumption on infection rates relaxed

We now relax the seasonal assumption on infection rates. In other words, we no longer assume that $\beta_k = \beta_{k+26p}$. Thus the number of the infection rates increases to 260 from 26 given 10 years data, i.e. the infection rates are represented as $\{\beta_k\}_{k=0}^{259}$ for this case. Under this assumption, we applied the CF method to the data generated in Section 4.4 but the MCMC chain failed to converge to the true values which implied that the method is no longer working if the seasonal

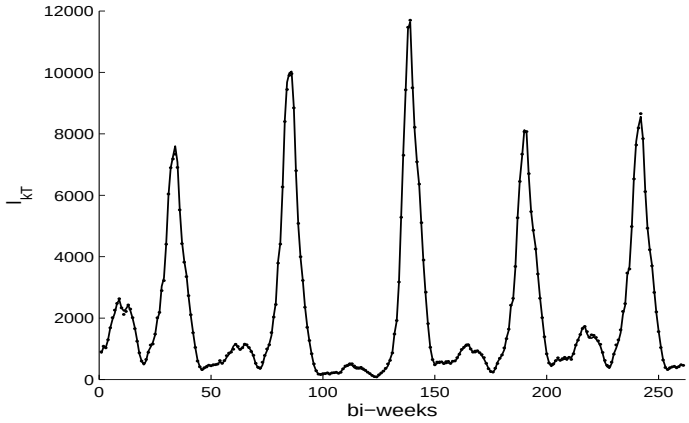


Figure 4.8: Posterior mean (solid line) and true value (dot) of the number of infectives over 10 years for the SIR epidemic simulated with mean infectious period $1/\gamma = 14$ days under the BNP method.

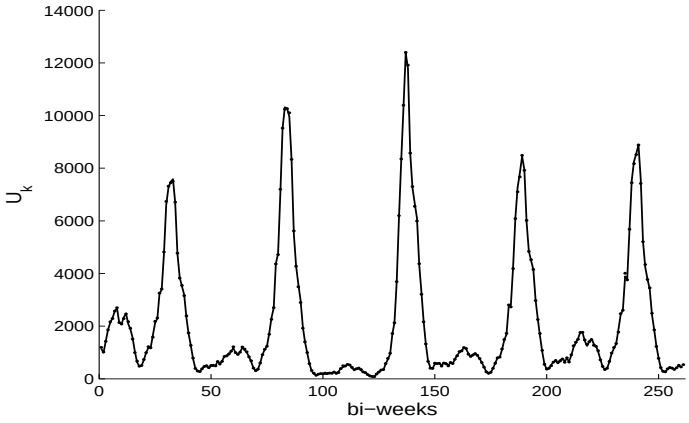


Figure 4.9: Posterior mean (solid line) and true value (dot) of the total number of new infections over 10 years for the SIR epidemic simulated with mean infectious period $1/\gamma = 14$ days under the BNP method.

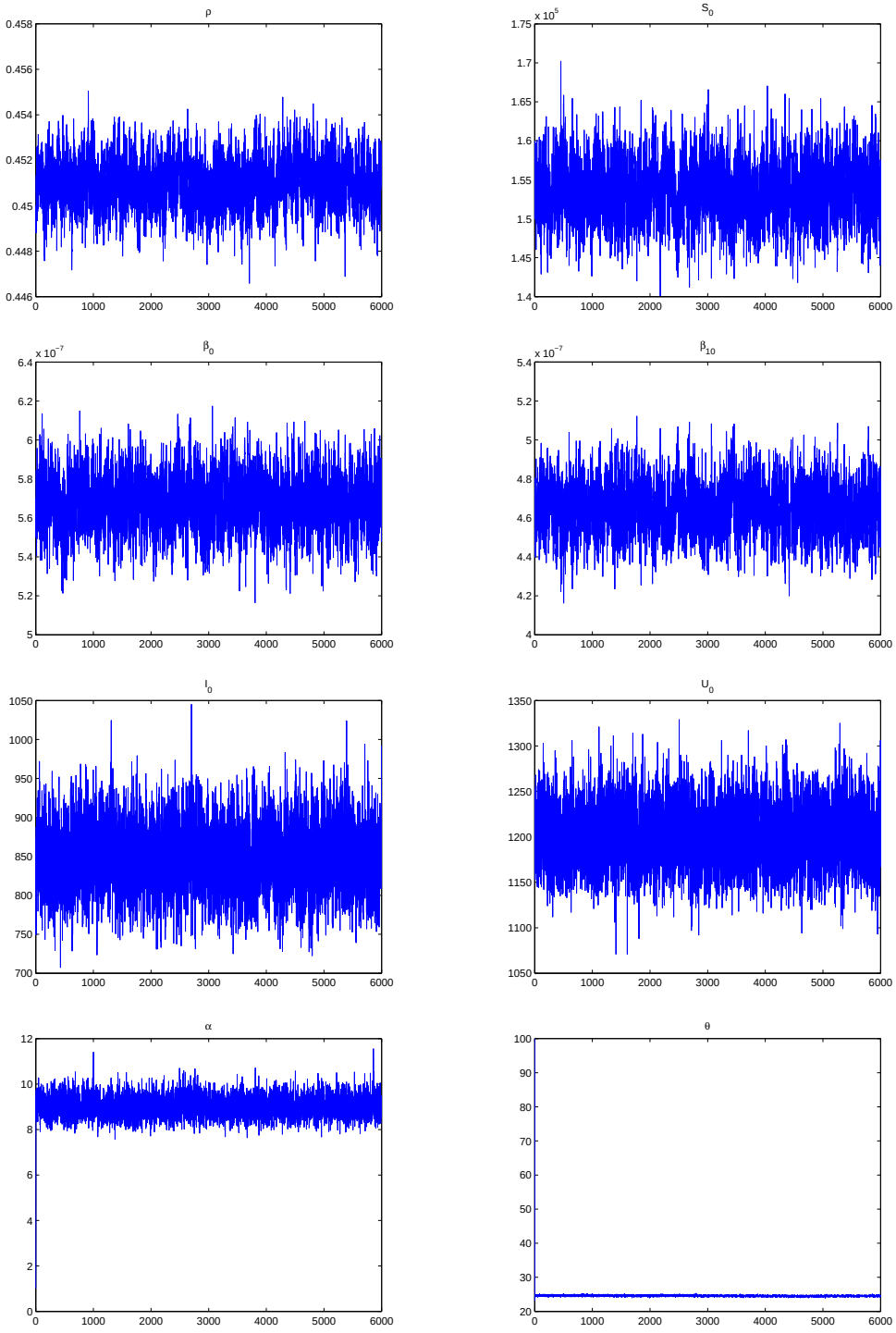


Figure 4.10: Convergence of the MCMC algorithm for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days under the BNP method. The plots show MCMC trace of the reporting rate, ρ, the initial number of susceptibles, S_0, the infection rate for the first observation period, β_0, the infection rate for the 11^{th} observation period, β_{10}, the initial number of infectives, I_0, the total number of new infections for the first observation period and the hyperparameters α and θ.

assumption is relaxed. For instance, the estimated reporting rate converged to 1. We then applied the BNP method used in Section 4.5.1 and placed a Gaussian process prior with the squared exponential covariance function on the infection rate throughout the whole epidemic. The posterior density of $\{\beta_k\}_{k=0}^{259}$ is given by

$$\begin{aligned}
&\pi(\{\beta_k\}_{k=0}^{259}|I_{(k+1)T}\{I_{kT}, U_k, B_k\}_{k=0}^{259}, S_0) \\
&\quad \propto \prod_{k=0}^{259} P(I_{(k+1)T}, U_k|I_{kT}, S_0, \{U_i, B_i\}_{i=0}^{k-1}, \beta_k^*) \\
&\qquad \times \pi(\mathbf{g}_J|\{s_j\}_{j=1}^{J}, \alpha, \theta),
\end{aligned}$$

where J is equal to 260 in this case which is different from what it is in Section 4.5.1. The MCMC chain again failed to converge to the true values. The failure of the convergence indicates that methods which do not capture the seasonal feature on infection rates lead to bad estimations of key quantities of interest.

We now take advantage of the Gaussian process and place a Gaussian process prior with a different covariance function to address what we need, specifically a periodic covariance function. Mathematically, our periodic covariance function is defined as follows:

$$K(x, x') = \alpha^2 \exp\left(-\frac{1-\cos(\frac{2\pi|x-x'|}{\omega})}{\theta}\right),$$

where ω is the length of a single season in the time series and θ is assumed to be positive. Chandola & Vatsavai (2011) pointed out that for periodic time series, the observables at time t are strongly correlated with observables at $t-1$ as well as with the observables at time $t-\omega$. Such periodic dependency cannot be modelled using the squared exponential covariance function for any value of θ. As it is assumed that there are 7 days a week, 52 weeks a year and the infection rates keep a similar pattern for each year, we set 364(days) to ω for our case. Figure 4.11 and 4.12 show the posterior mean of all of the infection rates and the posterior mean of the number of infectives throughout the whole epidemic compared with the truth and Figure 4.13 gives convergence of the MCMC algorithm for the SIR epidemic under BNP method using a periodic covariance function for the Gaussian process. We can see the MCMC chain converges to the truth rapidly from Figure 4.13 and Figure 4.11 and 4.12 give reasonable estimations of the infection rates and the number of infectives which implies that

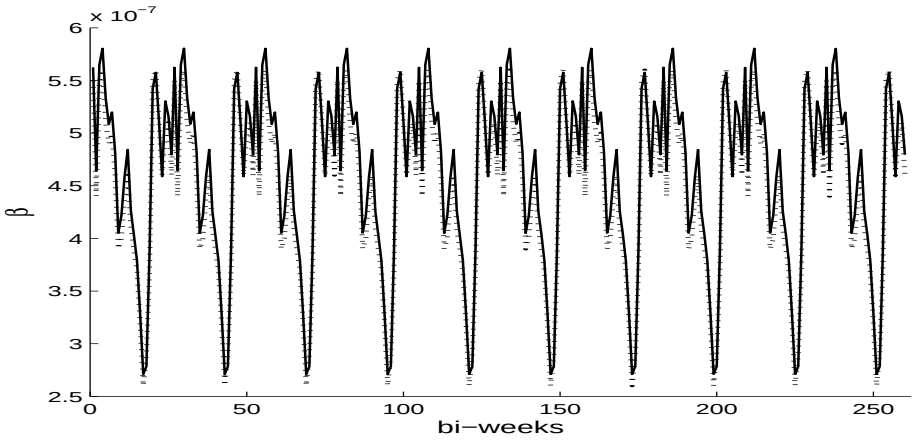


Figure 4.11: Posterior mean (solid line) and true value (dashed line) of the infection rates over 10 years for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days under BNP method using a periodic covariance function for the Gaussian process.

our BNP method using a periodic covariance function for the Gaussian process works under the assumption that the seasonal assumption on infection rates is relaxed.

4.5.3 Different parameter settings for the Gaussian process

4.5.3.1 Strong prior on hyperparameters for Gaussian processes

As our BNP methods provided similar estimation results to the CF method, we are encouraged to see how the Gaussian process behaves with different hyperparameter settings. In order to explore how it affects the estimation of the infection rates within a season if a strong prior is put on the hyperparameter, θ, in the Gaussian process covariance function, we study the same observed data simulated in Section 4.4. We compare the results between the one obtained in Section 4.5.1 and the one where we put strong prior distribution with a bigger mean value on θ. We assume the infection rates are seasonal with a period of 1 year and a squared exponential covariance function is used for the Gaussian process. From the MCMC trace of θ obtained in Section 4.5.1, the posterior mean of θ is around 24. As θ controls the characteristic length-scale of the process and the bigger θ is which implies that the more slowly varying the

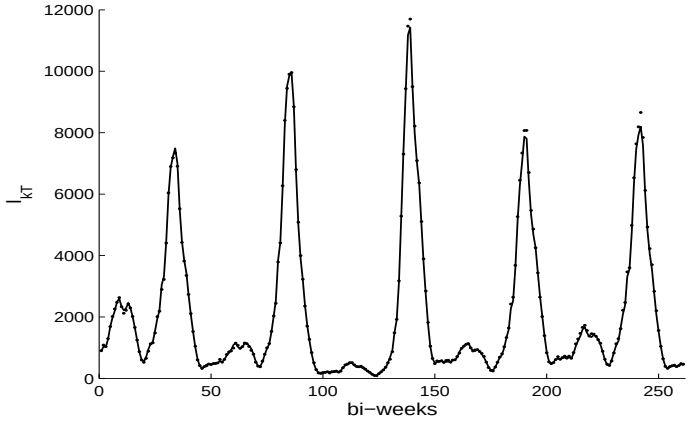


Figure 4.12: Posterior mean (solid line) and true value (dot) of the number of infectives over 10 years for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days under BNP method using a periodic covariance function for the Gaussian process.

function generated will be, we expect a much smoother curve of the estimated mean infection rates in a season. We put two priors, $\Gamma(50, 1)$ and $\Gamma(100, 1)$, on θ respectively. Figure 4.14 and 4.15 give the estimation results and show the effect of strong priors used on the hyperparameter, θ. As expected, the pattern of the infection rates is smoothed and the one is much smoother when the $\Gamma(100, 1)$ prior is put on θ.

4.5.3.2 Different covariance functions for Gaussian processes

For the BNP method, we also use other covariance functions for the Gaussian processes to compare with the squared exponential covariance function under the seasonal assumption on the infection rates, i.e. we place Gaussian process prior with different covariance functions on the transformed infection rates within a season. The covariance functions we use here are as follows:

$$K_(x, x') = \alpha^2 \exp\left(-\left(\frac{|x - x'|}{\theta}\right)^r\right), \text{ for } \theta > 0 \text{ and } 0 < r < 2.$$

Such covariance functions are so-called the r-exponential covariance function (Rasmussen & Williams 2006). Actually, the squared exponential covariance function also belongs to the r-exponential family of the covariance functions when r is set to 2 and it is the smoothest one among the whole family. We set r

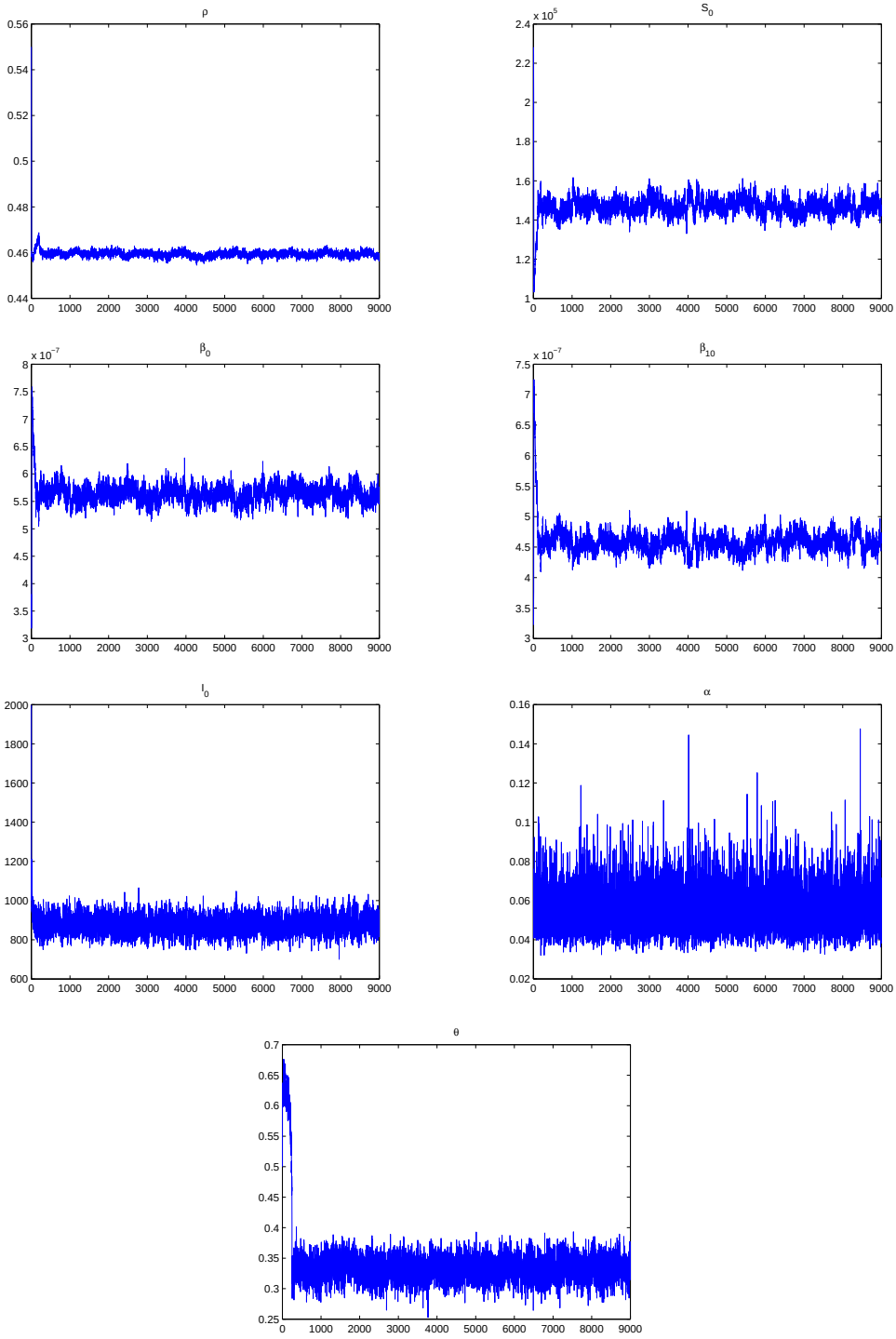


Figure 4.13: Convergence of the MCMC algorithm for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days under BNP method. The plots show MCMC trace of the reporting rate, ρ, the initial number of susceptibles, S_0, the infection rate for the first observation period, β_0, the infection rate for the 11^{th} observation period, β_{10}, the initial number of infectives, I_0 and the hyperparameters in the Gaussian process covariance function α and θ.

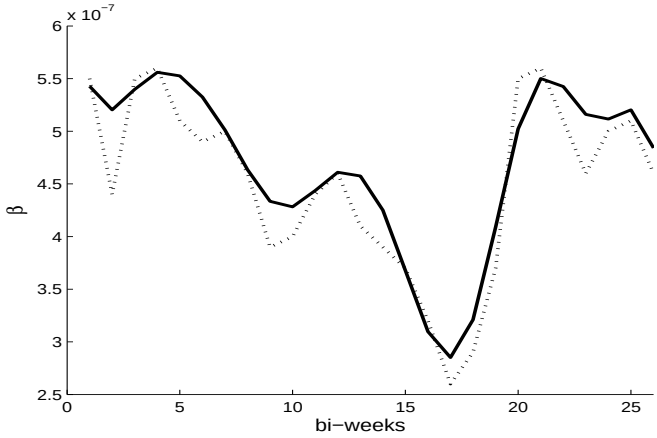


Figure 4.14: Posterior mean (solid line) and true value (dashed line) of the infection rates over 1 year for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days under BNP method using a squared exponential function for the Gaussian process. A strong prior, $\Gamma(50, 1)$ is put on the haperparameter, θ.

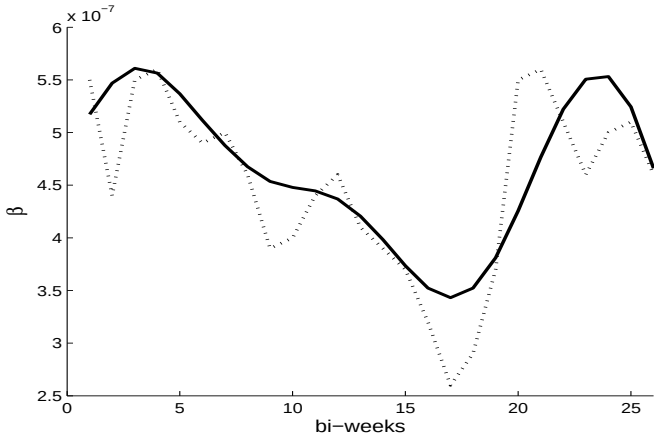


Figure 4.15: Posterior mean (solid line) and true value (dashed line) of the infection rates over 1 year for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days under BNP method using a squared exponential covariance function for the Gaussian process. A strong prior, $\Gamma(100, 1)$ is put on the haperparameter, θ.

to 1 and analysed the same SIR epidemic simulated in Section 4.4 using the BNP method with the different covariance function for the Gaussian process. Figure 4.16 shows posterior mean values of the infection rates within a season using two different covariance functions with $r = 1$ and $r = 2$. We can not see much difference of the estimation results for both covariance functions although the function values generated from the exponential covariance functions are supposed to be less smooth. One explanation could be that there is only 26 observation periods per season, hence we have only 26 Gaussian process inputs as we chose the middle time points of each observation period within a season as the inputs, so functions could not be generated quite differently with different covariance functions unless a strong prior is put on the hyperparameters in the covariance function which significantly changes the model.

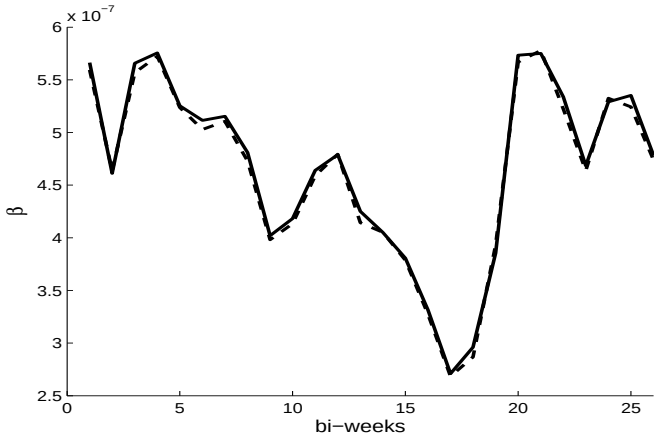


Figure 4.16: 10 years simulation data analysed by the BNP method with different covariance functions. Posterior mean (solid line for $r = 1$ and dashed line $r = 2$) of the infection rates over 1 year for the SIR epidemic simulated with mean infectious period, $1/\gamma = 14$ days. The infection rates vary every two weeks with a period of 1 year.

4.6 Measles epidemics in london (1948-1957)

We also analysed real life time series data, namely measles in London between 1948 and 1957 (`http://www.zoo.cam.ac.uk/zoostaff/grenfell/measles.htm`). Measles is caused by a highly infective single-stranded RNA virus belonging to the morbillivirus group and is mostly specialised on its human hosts. On

infection, the virus passes through a latent period of around 6 to 9 days, followed by a 6 to 7 days infective period. Therefore, the characteristic time scale of the transmission dynamics is around 14 days (Bjørnstad et al. 2002, Grenfell et al. 2002). Based on the information above, we assume the infection period is known, i.e. $1/\gamma = 14$. The measles data were collected bi-weekly and are under-reported as the reporting is not complete during the period and the number of births is, on average, 2152 bi-weekly in London between 1948 and 1957. We do not consider changes in the structure of the population that modify transmission parameters themselves, see Cauchemez & Ferguson (2008).

Given the observed data described above, we now apply the CF method and BNP method to the 10 years measles data and estimate the reporting rate, initial number of susceptibles, total number of the cases for each observation period, number of infectives at the beginning of each observation period and the infection rates for each observation period within a season, namely, a year. For the BNP method, we use a squared exponential covariance function for the Gaussian process and placed a Gaussian process prior on the infection rates within a season, i.e. the infection rates are assumed to be seasonal. We also use the periodic covariance functions for the Gaussian process and placed a Gaussian process prior on all of the infection rates throughout the whole epidemic, i.e. the seasonal assumption is relaxed. Figures 4.17, 4.18, 4.19, 4.20, 4.21 and 4.22 show estimation results of the parameters. The Figure 4.21 shows that the results of the posterior mean of ρ for the CF method and the BNP method assuming the infection rates are seasonal are quite close. The Figure 4.22 also gives similar results for the initial number of susceptibles, S_0 for both methods discussed above. However, for the BNP method without assuming the infection rates are seasonal, the posterior mean of ρ and S_0 is smaller and bigger respectively compared to the results of the methods assuming the infection rates are seasonal. One explanation is that methods assuming the infection rates are seasonal or not may give different estimates for the reporting rate and the initial number of susceptibles.

Table 4.2 gives posterior mean, 95% CI and standard deviation for the parameters ρ, S_0, I_0, β_0 and β_{10}. From the results, we found that under the seasonal assumption on the infection rates, the BNP method provided similar estimation for each parameter compared with the CF method. When such assumption is

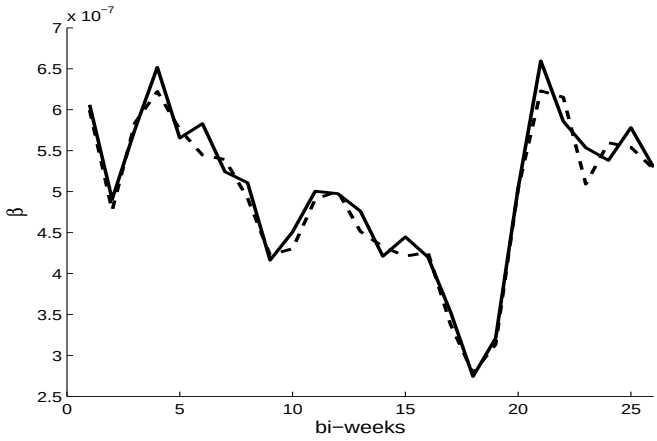


Figure 4.17: Posterior mean of the infection rates over 1 year for the CF method (dashed line) and BNP method (solid line) using the squared exponential covariance function for the Gaussian process for the 10 years measles data.

relaxed, the BNP method using appropriate covariance function for the Gaussian process successfully estimated the infection rates throughout the whole epidemic as well as all of the other parameters.

4.7 Conclusion

We firstly applied the CF method to an SIR epidemic simulated with a similar parameter setting used in Cauchemez and Ferguson's paper (2008) and reproduced reasonable estimation results. Then we applied our BNP method to the same simulated data under the same assumption on the infection rates, i.e. $\beta_k = \beta_{k+26p}$. We also relaxed such seasonal assumption on the infection rates and successfully estimated the infection rates throughout the whole epidemic under the BNP method using a periodic covariance function for the Gaussian process for the same simulated data. Under the seasonal assumption, strong priors on the hyperparameters in the covariance function and different covariance functions were conducted to explore the effect of the estimation of the infection rates as well as other parameters. At the end of the chapter, our methods were illustrated using the real life data, measles in London between 1948 and 1957 and the estimation results were compared with the one obtained using the CF method.

Table 4.2: Posterior mean, 95% CI, standard deviation for the reporting rate, ρ, the initial number of susceptibles S_0 and infectives I_0 and infection rates. CF, BNP1 and BNP2 represent respectively the CF method, the BNP method using the squared exponential covariance function for the Gaussian process assuming the infection rates are seasonal and the BNP method using the periodic covariance function for the Gaussian process without assuming the infection rates are seasonal. Measles in London 1948-1957.

	CF	BNP1	BNP2
ρ (%)	50.87 [50.69, 51.05] (0.11)	50.71 [50.56, 50.87] (0.10)	50.26 [50.09, 50.44] (0.11)
S_0 ($\times 10^3$)	164 [160, 168] (2.6)	161 [158, 164] (2.1)	185 [180, 190] (3.0)
I_0	603 [551, 659] (33.3)	605 [550, 661] (33.7)	602 [547, 658] (33.9)
β_0 ($\times 10^{-7}$)	5.99 [5.81, 6.17] (0.11)	6.08 [5.92, 6.24] (0.10)	5.31 [5.15, 5.47] (0.10)
β_10 ($\times 10^{-7}$)	4.91 [4.74, 5.06] (0.10)	5.02 [4.88, 5.17] (0.09)	4.18 [4.05,4.32] (0.08)

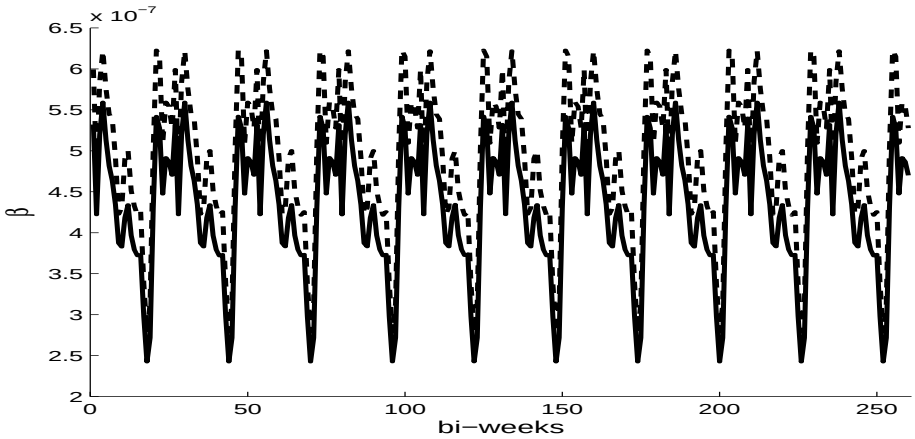


Figure 4.18: Posterior mean of the infection rates over 10 years for the CF method (dashed line) and BNP method (solid line) using the periodic covariance function for the Gaussian process for the 10 years measles data.

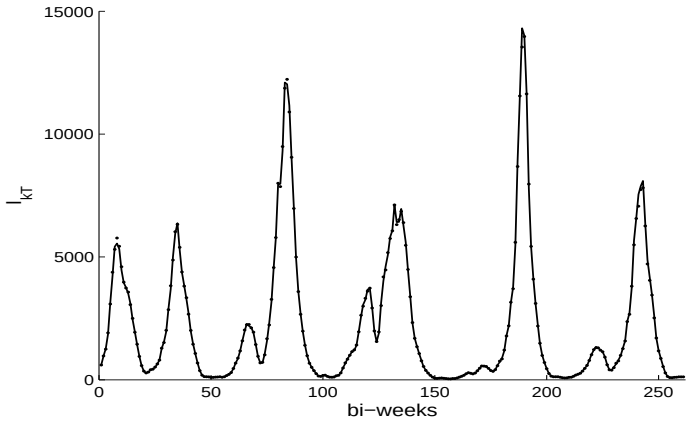


Figure 4.19: Posterior mean of the number of infectives over 10 years for the CF method (dot) and BNP method using the squared exponential covariance function for the Gaussian process (solid line) for the 10 years measles data.

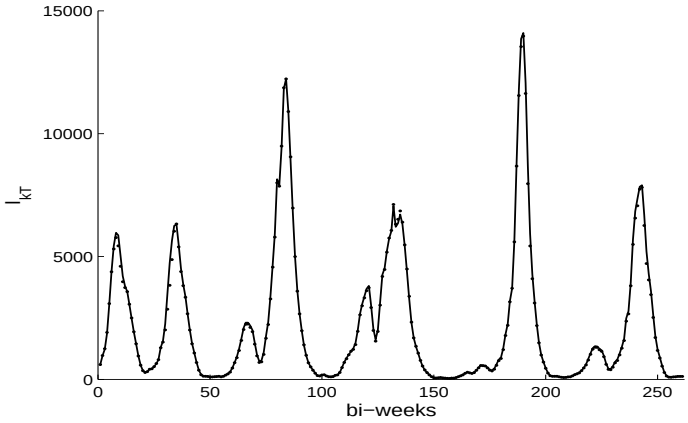


Figure 4.20: Posterior mean of the number of infectives over 10 years for the CF method (dot) and BNP method using the periodic covariance function for the Gaussian process (solid line) for the 10 years measles data.

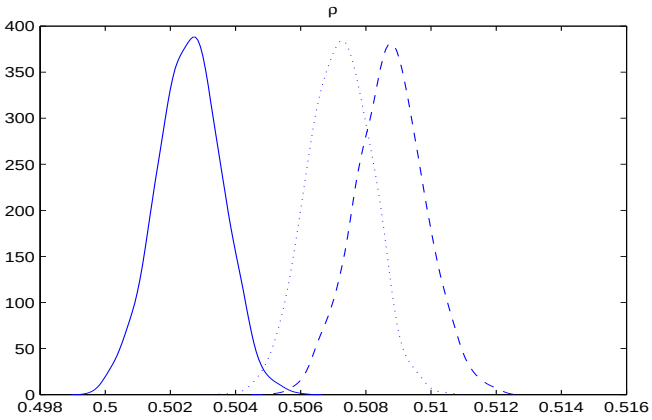


Figure 4.21: The figure shows the density plots of the reporting rate, ρ, for the CF method (dashed line), BNP method using the squared exponential covariance function for the Gaussian process (dotted line) and BNP method using the periodic covariance function for the Gaussian process (solid line) for the 10 years measles data.

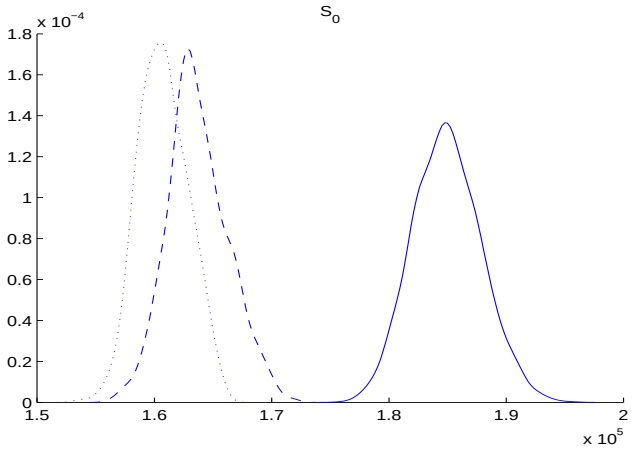


Figure 4.22: The figure shows the density plots of the initial number of susceptibles, S_0, for the CF method (dashed line), BNP method using the squared exponential covariance function for the Gaussian process (dotted line) and BNP method using the periodic covariance function for the Gaussian process (solid line) for the 10 years measles data.

In this chapter, we successfully applied BNP methods to the CF model, and the BNP methods enabled us to (i) put a Gaussian process prior on the infection rate as a function of time, as opposite to assuming independent β_k values; (ii) relax assumption of identical infection rates over the whole epidemic period.

CHAPTER 5

Conclusions

In this thesis we successfully developed nonparametric methods for the estimation of the overall force of infection and the infection rate in small-scale epidemics in a Bayesian framework. We also investigated the large-scale epidemics and developed Bayesian nonparametric methods for the estimation of the infection rate from epidemic times-series data.

The data we used to apply our methodology were simulated data generated from the SIR epidemic model and the real outbreak data.

In Chapter 2, we firstly introduced a Gaussian process-based Bayesian nonparametric inference for the estimation of the inhomogeneous Poisson process intensity which was presented by Adams et al. (2009). We then adapted the work of Adams et al. to the epidemics and developed our Bayesian nonparametric methods for the stochastic SIR models. We explored the overall force of infection of the epidemic process and assumed that it has the form, $h(t)$, which is a function of time without considering any information of the number of susceptibles and infectives. We inferred $h(t)$ by placing Gaussian process priors on it. Standard MCMC methods were used to sample the overall force of infection as well as other parameters of interest from the desired posterior distributions. We applied our methods to several simulated datasets, i.e. SE-Data 1, SE-Data 2 and SE-Data 3. Our methods work fairly well although some dramatic changes within a small time period in the data were not captured due to the fact that Gaussian process priors generally produce smooth estimates. We also analysed the Smallpox data obtained from Bailey (1975, p.125) and compared results with ones obtained from O'Neill & Roberts (1999). It turned out that the posterior

mean of the infection rate is larger than O'Neill & Roberts' estimate at each estimated infection time although the estimated infection rate from O'Neill & Roberts is fully covered in our 95% credible intervals. In terms of the estimate of the removal rate, we have similar results with ones from O'Neill & Roberts. At the end of the chapter, we carried out analysis of the effect of different Gaussian process priors by using different covariance functions and we found that different Gaussian process priors produce different processes from level of smoothness point of view, otherwise they do not make much differences. In this chapter, we assumed that the hyperparameter, α, of any covariance functions is fixed to 2 in order to obtain efficient mixing of the MCMC algorithms and we did not explore the topic of the choice of α.

In Chapter 3, we further investigated our Bayesian nonparametric methods for the SIR model in small-scale epidemics. We assumed that the overall force of infection is of the form, $\tilde{\beta}(t)X_tY_t$. In other words, the infection rate is assumed to be a function of time, and the information of the number of susceptibles and infectives is considered for the overall force of infection. We placed a Gaussian process prior on $\tilde{\beta}(t)$ and treated it as an inhomogeneous Poisson process intensity. We applied our methods to simulated datasets SE-Data 2, SE-Data 3 where the data were generated with a constant infection rate. We also simulated data which were generated with infection rates varying during the epidemics and analysed the dataset SE-Data 4 with our methods. We found that our methods successfully recovered the truth. The estimates of the infection rate in the case where we only observe incomplete data have larger uncertainty than the case where complete data are observed. It is possibly because the estimation of the removal rate results the extra uncertainty. We also illustrated our methods with the Smallpox data. In this chapter, we introduced an extended multi-group SIR model discussed in Hayakawa et al. (2003). In the multi-group SIR model, the population is no longer homogeneous, but partitioned into groups according to level of susceptibility. Hayakawa et al. (2003) developed Bayesian methods for SIR epidemic models with several kinds of susceptible. We applied our Bayesian nonparametric methods to multi-group epidemic data (simulated data and real life data) and then compare the results with ones obtained from Hayakawa et al. It turned out that our methods also worked well for the multi-group SIR model. Finally, we compared the approach with the one developed

in Chapter 2. Specifically, we compared the estimate of the infection rate in two different Bayesian nonparametric methods, i.e. $\hat{\beta}(t)$ and $\tilde{\beta}(t)$. We found that the approach discussed in Chapter 3, which considering the information of the number of susceptibles and infectives, is a better way to estimate the infection rate.

In Chapter 4, we adapted an approximation method due to Cauchemez & Ferguson (2008) to produce a Bayesian nonparametric method of inference for epidemics in large populations. We firstly applied a method developed by Cauchemez and Ferguson (2008) to an SIR epidemic and reproduced reasonable estimation results. Then we applied our Bayesian nonparametric methods to the same simulated dataset under the same assumption on the infection rates. We also investigated the case where we relaxed seasonal assumptions on the infection rates and successfully estimated the infection rates throughout the whole epidemic using our methods with a periodic covariance function for the Gaussian process prior for the same simulated data. At the end of the chapter, our methods were illustrated using the real life data, measles in London between 1948 and 1957 and we compared our results with the ones obtained using the Cauchemez and Ferguson's method. In general, we developed Bayesian nonparameteric methods enabled us to (i) put a Gaussian process prior on the infection rate as a function of time, as opposite to assuming independent values; (ii) relax assumption of identical infection rates over the whole epidemic period.

In the thesis, we performed inference for Gaussian process using standard MCMC algorithms. One might want to use Hamiltonian Monte Carlo (Duane et al., 1987) or slice sampling (Murray et al., 2010) to improve mixing and convergence speed. However, all the methods mentioned above may have computational difficulties for large data sets. It is notable that given n input points, Gaussian processes are very computationally demanding, i.e. $O(n^3)$ time complexity. Fortunately, enormous attention has been given to the development of efficient approaches to Gaussian process models. Snelson (2007) developed sparse approximation techniques to reduce the complexity to $O(nm^2)$, where m is a user chosen number much smaller than n. Csató & Opper (2001) developed an approach for a sparse representation for Gaussian process models in order to gain a computational advantage when using large data sets. For more dis-

cussion of topics related to sparse approximations, see Csató (2002), Keerthi & Chu (2008) or Seeger (2003).

Overall, this thesis provides new methodology which enables nonparametic estimation of parameters governing transmissions under a Bayesian framework. We have successfully developed Bayesian nonparametric methods for the estimation of the stochastic SIR models in small-scale and large-scale epidemics respectively. To our knowledge, there are not any other studies in the literature that considered nonparametric methods under a Bayesian framework. In general, taking a nonparamatric approach not only offers greater potential flexibility than parameterised models but also provides one a tool to assess the goodness of fit of parametric epidemic models, and hence to quantify the extent to which the underlying model assumptions are in line with observed data.

Epidemics are assumed to be complete throughtout the whole thesis. However, our methods are also appropriate to ongoing disease outbreaks analysis. In addition, condition on the current function values, one can obtain new time series that are drawn from the predictive distribution using our methods. Therefore, in practice, epidemiologists can use our methods to answer questions about an ongoing disease outbreak. Using available information and observed data, they can make predictions about how fast an epidemic will move, how it is likely to spread, and how many people will be involved. This information can be used in an attempt to contain the epidemic and to stay one step ahead of it. The Bayesian nonparameteric models can be constantly adjusted as new data rolls in.

The Bayesian nonparameteric methods have potential limitations and disadvantages. Siegel & Castellan (1988) argued that nonparametric methods may lack power as compared with more parametric approaches, especially if the sample size is small or if the assumptions for the corresponding parametric method hold. Hoff (2013) pointed out that for most Bayesian nonparametric approaches in practice, the prior does not represent actual prior beliefs, making it hard to interpret the corresponding posterior distributions as posterior beliefs. Standard Bayesian nonparametric priors include hyperparameters that directly control things that we are unlikely to have prior information about and only indirectly control things we might have information about (like means, variances and correlations). Therefore, it is not easy to imagine that such a prior repre-

sents actual prior information. He also commented that a posterior distribution does not provide an honest assessment of uncertainty. Another limitation of the nonparametric approaches is that appropriate computer software packages for the approaches can be limited.

References

Abramowitz, M. and Stegun, I. A. (1965). Handbook of mathematical functions. Dover, New York.

Abramowitz, M. and Stegun, I. A. (1970). Handbook of mathematical functions with formulas, graphs, and mathematical tables. National bureau of standards - U.S. government printing office.

Adams, R. P. (2009). Kernel methods for nonparametric Bayesian inference of probability densities and point processes. PhD thesis, University of Cambridge.

Adams, R. P., Murray, I., and MacKay, D. J. C. (2009). Tractable nonparametric Bayesian inference in poisson processes with Gaussian process intensities. New York: ACM Press.

Adams, R. P., Murray, I., and MacKay, D. J. C. (2009). The Gaussian process density sampler. *Advances in neural information processing systems*, 21, 9-16.

Andersson, H. and Britton, T. (2000). Stochastic epidemic models and their statistical analysis, volume 151 of Lecture Notes in Statistics. New York: Springer-Verlag.

Andrieu, C., Freitas, N. D., Doucet, A. and Jordan, M. I. (2003). An Introduction to MCMC for Machine Learning. *Machine Learning*, 50, 5-43.

Bailey, N. T. J. (1975). The mathematical theory of infectious diseases and its applications, 2nd ed. London: Griffin.

Balatif, O., Elhia, M., Bouyaghroumni, J., Rachik, M. (2014). Optimal control strategy for a discrete SIR epidemic model. *International Journal of Applied Mathematics and Modeling*, 2(2), 1-8.

Becker, N. G. (1989). Analysis of infectious disease data. London, UK: Chapman and Hall.

Becker, N. G. and Yip, P. S. F. (1989). Analysis of variation in an infection rate. *Australian Journal of Statistics*, 31, 42-52.

Becker, N. G. and Hopper, J. L. (1983). Assessing the heterogeneity of disease spread through a community. *American Journal of Epidemiology*, 117, 362-374.

Bernardo, J. M. and Smith, A. F. M. (1994). Bayesian theory. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley and Sons Ltd., Chichester.

Bjørnstad, O. N., Finkenstädt, B. and Grenfell, B. T. (2002). Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs*, 72(2), 169-184.

Brooks, S. P. (1998). Markov chain monte carlo method and its application. *Journal of the Royal Statistical Society Series D*, 47(1), 69-100.

Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *American Statistician*, 46(3), 167-174.

Cauchemez, S. and Ferguson, N. M. (2008). Likelihood-based estimation of continuous-time epidemic models from time-serires data: application to measles transmission in London. *Journal of the Royal Society Interface 2008*, 5(25), 885-897.

Chandola, V. and Vatsavai, R. R. (2011). A Gaussian process based online change detection algorithm for monitoring periodic time series, proceedings of SIAM international conference on data mining, SDM 2011.

Chen, F., Huggins, R. M., Yip, P.S. and Lam, K. F. (2008). Nonparametric estimation of multiplicative counting process intensity functions with an application to the Beijing SARS epidemic. *Communications in Statistics: Theory and Methods*, 37, 294-306.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society Series B*, 17, 129-164.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53, 385-408.

Csató L. (2002). Gaussian Processes - Iterative Sparse Approximations. PhD thesis, Neural Computing Research Group, Aston University.

Csató L. and Opper, M. (2001).Sparse representation for Gaussian process mod-

els. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems 13*, Cambridge, MA. The MIT Press.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195, 216-222.

Dureau, J., Kalogeropoulos, K. and Baguelin, M. (2013). Capturing the time-varying drivers of an epidemic via stochastic dynamical systems. *Biostatistics*, 14(3), 541-555.

Ebden, M. (2008). Gaussian processes for regression: a quick introduction. [online] Available: http://www.robots.ox.ac.uk/∼mebden/reports/GPtutorial.pdf

Eichner, M. and Dietz, K. (2003). Transmission potential of smallpox: estimates based on detailed data from an outbreak. *American Journal of Epidemiology*, 158, 110-117.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430), 577-588.

Fang, C.T., Hsu, H. M., Twu, S. J., Chen, M. Y., Hwang, J. S., Wang, J. D., Chuang, C. Y., and the Division of AIDS and STD, Center fo Disease Control, Department of Health, Executive Yuan. (2004). Decreased HIV transmission after a policy of providing free access to highly active antiretroviral therapy in Taiwan. *Journal of Infectious Diseases*, 190(5), 879-885.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2), 209-230.

Finkenstädt, B., and Grenfell, B. T. (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Applied Statistics*, 49, 187-205.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410), 398-409.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457-472.

Gelman, A., Carlin, J. B., and Stern, H. S. (2004). Bayesian data analysis. CRC press.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and

the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 6, 721-741.

Gershman, S. J., and Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1), 1-12.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J. M., Berger, J., Dawid, A. P. and Smith, A. F. M. editors, *Bayesian Statistics*, 4, 169-193. Oxford University Press, Oxford, UK.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). Bayesian Nonparametrics. Springer-Verlag, Berlin.

Gibson, G. and Renshaw, E. (1998). Estimating parameters in stochastic compartmental models using Markov chain methods. *IMA Journal of Mathematics Applied in Medicine and Biology*, 15, 19-40.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). Markov chain Monte Carlo in practice. Interdisciplinary Statistics. Chapman and Hall, London.

Grenfell, B. T., Bjørnstad, O. N. and Finkenstädt, B. (2002). Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model. *Ecological Monographs*, 72(2), 185-202.

Hayakawa, Y., O'Neill, P. D., Upton, D. and Yip, P. S. F. (2003). Bayesian inference for a stochastic epidemic model with uncertain numbers of susceptibles of several types. *Australian and New Zealand Journal of Statistics*, 45, 491-502.

Hoff, P. D. (2013). Comment on "Bayesian Nonparametric Inference - Why and How" by Mueller and Mitra. Arxiv preprint arXiv:1304.3676.

Jewell, C. P., Kypraios, T., Neal, P., Roberts, G. O. (2009). Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4(2), 191-222.

Keerthi, S. S. and Chu, W. (2006). A matching pursuit approach to sparse Gaussian process regression. In *Advances in Neural Information Processing Systems 18*.

Kottas, A. (2006). Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *Journal of Statistical Planning and Inference*, 136(3), 578-596.

Kottas, A. and Sansó, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of*

Statistical Planning and Inference, 137, 3151-3163.

Kronberger, G. and Kommenda, M. (2013). Evolution of covariance functions for gaussian process regression using genetic programming. Arxiv preprint arXiv:1305.3794.

Lau, E. H. Y. and Yip, P. S. F. (2008). Estimating the basic reproductive number in the general epidemic model with an unknown initial number of susceptible individuals. *Scandinavian Journal of Statistics*, 35, 650-663.

Lekone, E.P. and Finkenstadt, B. F. (2006). Statistical Inference in a Stochastic Epidemic SEIR Model with Control Intervention: Ebola as a Case Study. *Biometrics*, 62(4), 1170-1177.

Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of a nonhomogeneous Poisson process by thinning. *Naval Research Logistics Quarterly*, 26, 403-413.

Lloyd, A. L. (2001). Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics.*Theoretical Population Biology*, 60, 59-71.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. density estimates. *The Annals of Statistics*, 12(1), 351-357.

Lorio, M. D., Johnson, W. O., Müller, P. and Rosner, G. L. (2009). Bayesian nonparametric non-proportional hazards survival modelling. *Biometrics*, 65(3), 762-771.

MacKay, D. J. C. (1998). Introduction to Gaussian Processes. In Bishop, C. M., editor, *Neural Networks and Machine Learning*. Springer-Verlag.

McKinley, T., Cook, A. R., Deardon, R. (2009). Inference in epidemic models without likelihoods. *The International Journal of Biostatistics*, 5(1), Article 24.

Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25, 451-482.

Moyé, L. A. (2008). Elementary Bayesian Biostatistics. Chapman and Hall, New York, USA.

Murray, I. (2007). Advances in Markov Chain Monte Carlo Methods. PhD thesis, Gatsby Computational Neuroscience Unit, University College London, London.

Murray, I., Adams, R. P. and MacKay, D.J.C. (2010). Elliptical slice sampling. Appeared in *The Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), JMLR W&CP*, 9, 541-548.

Neal, P. and Roberts, G.O. (2005). A case study in non-centering for data augmentation: stochastic epidemics. *Statistics and Computing*, 15, 315-327.

Neal, R. M. (2001). Defining priors for distributions using Dirichlet diffusion trees. Technical Report 0104, Department of Statistics, University of Toronto.

Neal, R. M. (2003). Density modeling and clustering using Dirichlet diffusion trees. *Bayesian Statistics*, 7, 619-629.

Nieto-Barajas, L. E. and Walker, S. G. (2004). Bayesian nonparametric survival analysis via Lévy driven Markov processes. *Statistica Sinica*, 14, 1127-1146.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society Series B*, 40(1), 1-42.

O'Neill, P. D., Balding, D. J., Becker, N. G., Eerola, M., and Mollison, D. (2000). Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series C*, 49(4), 517-542.

O'Neill, P. D. and Becker, N. G. (2001). Inference for an epidemic where susceptibility varies. *Biostatistics*, 2(1), 99-108.

Orbanz, P. and Teh, Y. W. (2010). Bayesian nonparametric models. [online] Available: http://stat.columbia.edu/∼porbanz/reports/OrbanzTeh2010.pdf.

O'Neill, P. D. and Roberts, G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society Series A*, 162, 121-129.

Palacios, J. A. and Minin, V. N. (2013). Gaussian process-based Bayesian nonparametric inference of population size trajectories from gene genealogies. *Biometrics*, 69(1), 8-18.

Pasupathy, R. (2011). Generating homogeneous Poisson processes. *Wiley Encyclopedia of Operations Research and Management Science*, 1-11, John Wiley and Sons, Inc.

Persing, A. F. (2014). Some contributions to particle Markov chain Monte Carlo

algorithms. PhD thesis, Imperial College London.

Pollicott, M. Wang, H. and Weiss, H. (2009). Recovering the time-dependent transmission rate from infection data. Arxiv preprint arXiv:0907.3529 and updates.

Pollicott, M. Wang, H. and Weiss, H. (2012). Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem, *Journal of Biological Dynamics*, 6,509-523.

Rasmussen, C. E. (2000). The infinite Gaussian mixture model. In Solla, S.A., Leen, T.K. and Müller, K.R. editors, *Advances in Neural Information Processing Systems*, 12, 554-560.

Rasmussen, C. E. and Williams, C. K. I. (2006). Gaussian processes for machine learning. Cambridge, MA: MIT Press.

Rathbun, S. L. and Cressie, N. (1994) Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. *Advances in Applied Probability*, 26, 122-154.

Robert, C. P. and Casella, G. (1999). Monte Carlo statistical methods. Springer Texts in Statistics. Springer-Verlag, New York.

Sawyer, S. (2010). The Matropolitan-Hastings algorithm and extensions. Washington University.

Seeger, M. (2003). Bayesian Gaussian Process Models: PAC-Bayesian Generalisation Error Bounds and Sparse Approximations. PhD thesis, Institute of Adaptive and Neural Computation, University of Edinburgh.

Siegel, A. F. (1979). The noncentral chi-squared distribution with zero degrees of freedom and testing for uniformity. *Biometrika*, 66, 381-386.

Siegel, S. and Castellan, N. J. (1988). Non-parametric Statistics for the Behavioural Sciences, second edition. New York: McGraw-Hill.

Smirnova, A. and Tuncer, N. (2014). Estimating time-dependent transmission rate of avian influenza via stable numerical algorithm. *Journal of Inverse and Ill-Posed Problems*, 22(1), 31-62.

Snelson, E. (2007). Flexible and efficient Gaussian process models for machine learning. PhD thesis, Gatsby Computational Neuroscience Unit, University

College London.

Stein, M. L. (1999). Interpolation of Spatial Data. Springer-Verlag, New York.

Tanner, M. A. (1996). Tools for statistical inference. Methods for the exploration of posterior distributions and likelihood functions. Springer Series in Statistics. Springer-Verlag, New York, third edition.

Tildesley, M. J., Ryan, S. J. (2012). Disease Prevention versus Data Privacy: Using Landcover Maps to Inform Spatial Epidemic Models, *PLoS Computational Biology*, 8(11).

Walker, S. G., Damien, P., Laud, P. W. and Smith, A. F. M. (1999). Bayesian nonparametric inference for random distributions and related functions. *Journal of the Royal Statistical Society Series B*, 61(3), 485-527.