Bayesian Analysis (2008) **3**, Number 4, pp. 737–758

Bayesian estimation of the basic reproduction number in stochastic epidemic models

Damian Clancy* and Philip D. O'Neill†

Abstract. In recent years there has been considerable activity in the development and application of Bayesian inferential methods for infectious disease data using stochastic epidemic models. Most of this activity has employed computationally intensive approaches such as Markov chain Monte Carlo methods. In contrast, here we address fundamental questions for Bayesian inference in the setting of the standard SIR (Susceptible-Infective-Removed) epidemic model via simple methods. Our main focus is on the basic reproduction number, a quantity of central importance in mathematical epidemic theory, whose value essentially dictates whether or not a large epidemic outbreak can occur. We specifically consider two SIR models routinely employed in the literature, namely the model with exponentially distributed infectious periods, and the model with fixed length infectious periods. It is assumed that an epidemic outbreak is observed through time. Given complete observation of the epidemic, we derive explicit expressions for the posterior densities of the model parameters and the basic reproduction number. For partial observation of the epidemic, when the entire infection process is unobserved, we derive conservative bounds for quantities such as the mean of the basic reproduction number and the probability that a major epidemic outbreak will occur. If the time at which the epidemic started is observed, then linear programming methods can be used to derive suitable bounds for the mean of the basic reproduction number and similar quantities. Numerical examples are used to illustrate the practical consequences of our findings. In addition, we also examine the implications of commonly-used prior distributions on the basic model parameters as regards inference for the basic reproduction number.

Keywords: Basic reproduction number; Bayesian inference; Epidemics; Linear programming; Stochastic epidemic models

1 Introduction

In recent years there has been considerable activity in both the methodological development and application of methods for Bayesian data analysis of infectious disease outbreak data using stochastic epidemic models. Almost all of this literature employs Markov chain Monte Carlo (MCMC) methodology, which offers enormous power and flexibility compared to other approaches (see e.g. Gibson and Renshaw, 1998; O'Neill and Roberts, 1999; O'Neill *et al.*, 2000; Streftaris and Gibson, 2004; Neal and Roberts, 2005). The methods have been applied to many different human, animal and plant

*Department of Mathematical Sciences, University of Liverpool, Liverpool, UK, D.Clancy@liv.ac.uk

†School of Mathematical Sciences, University of Nottingham, Nottingham, UK, pdo@maths.nott.ac.uk

 DOI:10.1214/08-BA328

pathogens and diseases, examples of which include pneumococcal carriage (Auranen *et al.*, 2000), measles (Li *et al.*, 2002), swine fever (Höhle *et al.*, 2005), influenza (Cauchemez *et al.*, 2004; Demiris and O'Neill, 2005), norovirus (O'Neill and Marks, 2005) and nosocomial infections (McBryde *et al.*, 2006).

Despite the advances mentioned above, one notable absence in the current literature is any treatment of Bayesian inference for one of the most basic stochastic epidemic models, namely the SIR (Susceptible-Infective-Removed) model, without recourse to MCMC methods. This situation is in stark contrast to the classical case, where estimation methods for the SIR model, at least in the case where the model is Markov, are long-established (see e.g. Becker, 1989, Chapter 7; Andersson and Britton, 2000, Chapters 9 and 10). It is therefore of interest to see what can be achieved without computationally intensive methods such as MCMC.

An additional and important motivation for our study is that Bayesian analysis of the SIR model provides insights that are useful in the analysis of more complex and realistic models. SIR models themselves are frequently used as components of more complex epidemic models, for example those featuring populations divided into households (Demiris and O'Neill, 2005) or epidemics on networks (Andersson and Britton, 2000, Chapter 7). An example of such an insight, described in detail below, is the extent to which apparently natural (and commonly used) prior distributions for model parameters can affect the resulting inference.

In this paper, our attention generally focuses on the basic reproduction number R_0, informally defined as the average number of secondary cases caused by a single infective individual in a large susceptible population. This quantity is of enormous importance within epidemic modelling because, roughly speaking, if $R_0 \leq 1$ then an epidemic is highly unlikely to occur. Estimation of R_0, or equivalent parameters in more complex models, can usually be achieved via MCMC methods. Among other things, we derive a closed-form expression for the posterior density of R_0 given suitably complete data, and bounds on various quantites (e.g. the mean) for other data scenarios.

The paper is organised as follows. Section 2 describes the standard SIR epidemic model and recalls some results required in the sequel. Sections 3 and 4 consider two different special cases of the SIR model, corresponding to exponential and constant infectious periods, respectively. In Section 5 we apply our methods to three example data sets from the literature. Some brief concluding comments are given in Section 6.

2 Preliminaries

2.1 The standard SIR model

We begin by recalling the definition of the standard SIR (Susceptible-Infective-Removed) stochastic epidemic model (see e.g. Andersson and Britton, p.11). Consider a population of N individuals, assumed to mix together homogeneously. At each time $t \geq 0$, every individual in the population is either susceptible, infective or removed,

with the numbers in each category denoted $S(t)$, $I(t)$ and $R(t)$, respectively, so that $S(t) + I(t) + R(t) = N$. At time $t = 0$, the population contains only infectives and susceptibles, so that $S(0) \geq 1$, $I(0) \geq 1$ and $R(0) = 0$. Each infective individual remains so for a period of time (called the infectious period) having an arbitrary but specified distribution T_I, before becoming removed. Removed individuals play no further part in the epidemic. The infectious periods of different individuals are assumed to be mutually independent. During its infectious period, an infective individual has infectious contacts with each susceptible individual at times given by the points of a homogeneous Poisson process of rate β/N, with these processes being mutually independent. Each such contact results in the susceptible immediately becoming infective. Since the number of susceptible-infective pairs at time $t \geq 0$ is $S(t)I(t)$, it follows that the overall rate of infection at time t is $\beta S(t)I(t)/N$. The epidemic ends as soon as there are no more infectives left in the population.

2.2 The basic reproduction number

A quantity of major importance within mathematical epidemic theory is the basic reproduction number R_0, heuristically defined as the average number of new infections caused by a single infective in a large susceptible population (see e.g. Dietz, 1993). This quantity is important because roughly speaking, in a large population, a large epidemic outbreak can occur if and only if $R_0 > 1$. When $R_0 > 1$ the epidemic is said to be above threshold. Knowledge of the value of R_0 makes it possible to calculate the proportion of a population that should be vaccinated in order to prevent an epidemic from occurring. Both the definition and threshold interpretation of R_0 can be made mathematically precise by allowing the population size to tend to infinity, so that R_0 essentially becomes the mean offspring size of a branching process of new infections (see e.g. Andersson and Britton, 1999, p. 22.) For the standard SIR model defined above, $R_0 = \beta E[T_I]$.

2.3 Data and notation

In the following, we focus on an epidemic outbreak that results in a total of n removals, where $1 \leq n \leq N$. Our main interest will be in making inference about the parameters of the epidemic, and in particular R_0, given observation of the removal process alone. However, to begin with it is fruitful to consider inference given complete observation of the epidemic process, i.e. observing both infections and removals. In some settings such a detailed level of observation is not that unrealistic, e.g. animal experiments in which the animals are regularly tested for exposure to the pathogen in question.

We start with some notation. Suppose that the epidemic begins with a single infection at time i_1, so that $(S(i_1), I(i_1), R(i_1)) = (N-1, 1, 0)$. Subsequent infections occur at times $i_2 \leq i_3 \leq \ldots \leq i_n$, where $i_2 \geq i_1$, and removals occur at times $r_1 \leq r_2 \leq \ldots \leq r_n$. We suppose that the period of observation is $[i_1, r_n]$, so that it is assumed that the entire epidemic is observed, and define $\tau := r_n$. We write $\boldsymbol{r} = (r_1 \ r_2 \ \ldots \ r_n)$ and $\boldsymbol{i} = (i_2 \ i_3 \ \ldots \ i_n)$.

It is important to note that the infection and removal times must satisfy the inequalities $i_{k+1} \leq r_k$ for $k = 1, 2, \ldots, n-1$. In particular, this constraint ensures that the number of infectives does not reach zero until the time of the last removal, r_n. For a given $\boldsymbol{r}$ define $E_{\boldsymbol{r}}$ to be the set of all infection times $(i_1, \boldsymbol{i})$ satisfying $i_k \leq i_{k+1} \leq r_k$ for $k = 1, 2, \ldots, n-1$. Thus $E_{\boldsymbol{r}}$ contains all possible configurations of infection times for a given set of ordered removal times $\boldsymbol{r}$.

2.4 Ratio of independent Gamma distributions

In the sequel we will make some use of the following facts (see e.g. Bhoj and Schiefermayr, 2001). Denote by $\Gamma(a, b)$ a Gamma random variable with shape and scale parameters a and b, respectively (i.e. with mean and variance a/b and a/b^2; our choice of parameterisation is for later convenience). Let $X \sim \Gamma(a, b)$ and $Y \sim \Gamma(c, d)$ be independent, and define $W = X/Y$. Then W has a scaled F-distribution with probability density function given by

$$f_W(w) = \left(\frac{b}{d}\right)^a \frac{\Gamma(a+c)}{\Gamma(a)\Gamma(c)} \frac{w^{a-1}}{(\frac{bw}{d}+1)^{a+c}}, \quad w \geq 0,$$

and

$$E[W^k] = \left(\frac{d}{b}\right)^k \frac{\Gamma(a+k)\Gamma(c-k)}{\Gamma(a)\Gamma(c)}, \quad k = 1, 2, \ldots [c],$$

where $[c]$ denotes the largest integer less than or equal to c. Furthermore, W has mode $0 \vee d(a-1)/b(c+1)$, where $\vee$ denotes maximum, and distribution function

$$F_W(w) = \left(\frac{b}{d}\right)^a \frac{\Gamma(a+c)}{\Gamma(a)\Gamma(c)} \frac{w^a}{a} \, {}_2F_1(a+c, a; a+1; -bw/d), \quad w \geq 0, \tag{1}$$

where ${}_pF_q(n_1, \ldots, n_p; m_1, \ldots, m_q; x)$ denotes the hypergeometric function defined by

$${}_pF_q(n_1, \ldots, n_p; m_1, \ldots, m_q; x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \frac{(n_1)_k (n_2)_k \ldots (n_p)_k}{(m_1)_k (m_2)_k \ldots (m_q)_k},$$

where $(x)_0 = 1$ and for $k = 1, 2, \ldots$, $(x)_k = (x)(x+1)\ldots(x+k-1)$.

3 Exponential infectious period

In this section we suppose that the infectious period distribution is exponential with mean $E[T_I] = \gamma^{-1}$. This model is often known as the general stochastic epidemic, and is the most widely-studied SIR stochastic epidemic model. This model is also the natural analogue of the deterministic SIR epidemic model, defined in terms of differential equations (see e.g. Bailey, 1975, p. 82), which is itself a component of many deterministic epidemic models studied in the literature.

3.1 Likelihood

The likelihood of the infection and removal times given the model parameters β, γ and i_1 is

$$\begin{aligned}\pi(\boldsymbol{i},\boldsymbol{r}|\beta,\gamma,i_1) &= \left(\prod_{j=2}^{n}\beta N^{-1}S(i_j-)I(i_j-)\right)\left(\prod_{j=1}^{n}\gamma I(r_j-)\right)\\ &\quad\times\exp\left(-\int_{i_1}^{\tau}\beta N^{-1}S(t)I(t)+\gamma I(t)\,dt\right)\mathbb{1}_{\{(i_1,\boldsymbol{i})\in E\boldsymbol{r}\}},\end{aligned}\tag{2}$$

where $S(t-)=\lim_{s\uparrow t}S(s)$, *etc.*, see for example O'Neill and Roberts (1999).

3.2 Parameter prior distributions

Suppose that β and γ are, *a priori*, independent and respectively distributed as $\Gamma(m_\beta,\lambda_\beta)$ and $\Gamma(m_\gamma,\lambda_\gamma)$. As recalled below, this choice of prior distributions is convenient in terms of Bayesian inference due to conjugacy (O'Neill and Roberts, 1999). Furthermore, the flexibility of the Gamma distribution means that it is frequently used in practice as a prior distribution for rate parameters in epidemic models (see e.g. Auranen *et al.*, 2000; Cauchemez *et al.*, 2004; Streftaris and Gibson, 2004). However, the induced prior for R_0 is rarely mentioned in the literature, and we now consider this for the current model.

Since $R_0=\beta/\gamma$, applying the results in Section 2.4 yields that R_0 has prior density

$$f(R_0)=\left(\frac{\lambda_\beta}{\lambda_\gamma}\right)^{m_\beta}\frac{\Gamma(m_\beta+m_\gamma)}{\Gamma(m_\beta)\Gamma(m_\gamma)}\frac{R_0^{m_\beta-1}}{(\frac{\lambda_\beta R_0}{\lambda_\gamma}+1)^{m_\beta+m_\gamma}},\quad R_0\geq 0,$$

and prior mean, mode and variance given, when they exist, by

$$E[R_0]=\frac{m_\beta\lambda_\gamma}{(m_\gamma-1)\lambda_\beta},\quad \text{mode}(R_0)=\frac{\lambda_\gamma(m_\beta-1)}{\lambda_\beta(m_\gamma+1)}\vee 0,$$
$$\text{var}[R_0]=\frac{m_\beta(m_\beta+m_\gamma-1)}{(m_\gamma-1)^2(m_\gamma-2)}\left(\frac{\lambda_\gamma}{\lambda_\beta}\right)^2.$$

In practice, it is often the case that uninformative $\Gamma(m,\lambda)$ prior distributions are assigned, popular choices being $(m,\lambda)=(1,\epsilon)$ or $(m,\lambda)=(\epsilon,\epsilon)$ where ϵ is a small positive number, or zero. In the present case, we note that if $m_\gamma\leq 1$ then R_0 has infinite mean *a priori*, and if $m_\gamma\leq 2$ then R_0 has infinite prior variance. In other words, such vague priors on β and γ yield a vague prior for R_0.

As recalled in Section 2.2, the question of whether or not $R_0>1$ is often of interest. Suppose now that β and γ are assigned the same prior distribution, so that $m=m_\beta=m_\gamma$, $\lambda=\lambda_\beta=\lambda_\gamma$; a typical case in practice would be $m=1$, λ some small positive number. However, it then follows that $E(R_0)>1$. This suggests some need for caution in using the posterior mean of R_0 as the sole means of assessing whether or not an

epidemic is above threshold, a point that we shall return to in the sequel. However, here it is also the case that $P(R_0 > 1) = 0.5$, so that *a priori* the epidemic is equally likely to be above or below threshold. Of course, the posterior mean of R_0 is still of interest in its own right. First, it is a frequently-cited and convenient summary measure of the epidemic which gives a natural indication of how fast the epidemic spreads. Additionally, knowledge of R_0 is important to inform control measures - for example, the minimum vaccination coverage required to prevent epidemics is a function of R_0 - and so knowledge of the posterior mean of R_0 is useful in determining how such control measures should be implemented.

3.3 Parameter posterior distributions

By Bayes' Theorem, the joint posterior density of β and γ given $\boldsymbol{i}$, $\boldsymbol{r}$ and i_1 is defined by $\pi(\beta, \gamma|\boldsymbol{i}, \boldsymbol{r}, i_1) \propto \pi(\boldsymbol{i}, \boldsymbol{r}|\beta, \gamma, i_1)\pi(\beta)\pi(\gamma)$. It follows from (2) that

$$\pi(\beta|\boldsymbol{i}, \boldsymbol{r}, i_1) \sim \Gamma\left(n + m_\beta - 1, \lambda_\beta + N^{-1}\xi_{SI}\right), \tag{3}$$

$$\pi(\gamma|\boldsymbol{i}, \boldsymbol{r}, i_1) \sim \Gamma\left(n + m_\gamma, \lambda_\gamma + \xi_I\right) \tag{4}$$

where

$$\xi_I = \int_{i_1}^{\tau} I(t)\, dt, \quad \xi_{SI} = \int_{i_1}^{\tau} S(t)I(t)\, dt,$$

see O'Neill and Roberts (1999). Moreover, the posterior densities of β and γ are independent, and thus the distribution of R_0 given $\boldsymbol{i}$, $\boldsymbol{r}$ and i_1 is a ratio of two independent Gamma random variables. It follows that

$$\pi(R_0|\boldsymbol{i}, \boldsymbol{r}, i_1) = \left(\frac{\lambda_\beta + N^{-1}\xi_{SI}}{\lambda_\gamma + \xi_I}\right)^{n+m_\beta-1} \times \frac{\Gamma(2n + m_\beta + m_\gamma - 1)}{\Gamma(n + m_\beta - 1)\Gamma(n + m_\gamma)} \frac{R_0^{n+m_\beta-2}}{\left(\left(\frac{\lambda_\beta+N^{-1}\xi_{SI}}{\lambda_\gamma+\xi_I}\right) R_0 + 1\right)^{2n+m_\beta+m_\gamma-1}}, \quad R_0 \geq 0, \tag{5}$$

$$E[R_0|\boldsymbol{i}, \boldsymbol{r}, i_1] = \left(\frac{n + m_\beta - 1}{n + m_\gamma - 1}\right)\left(\frac{\lambda_\gamma + \xi_I}{\lambda_\beta + N^{-1}\xi_{SI}}\right), \tag{6}$$

$$\text{mode}[R_0|\boldsymbol{i}, \boldsymbol{r}, i_1] = \left(\frac{n + m_\beta - 2}{n + m_\gamma + 1}\right)\left(\frac{\lambda_\gamma + \xi_I}{\lambda_\beta + N^{-1}\xi_{SI}}\right) \vee 0, \tag{7}$$

and, for $n + m_\gamma > 2$,

$$\text{var}[R_0|\boldsymbol{i}, \boldsymbol{r}, i_1] = \frac{(2n + m_\beta + m_\gamma - 2)(n + m_\beta - 1)}{(n + m_\gamma - 1)^2(n + m_\gamma - 2)}\left(\frac{\lambda_\gamma + \xi_I}{\lambda_\beta + N^{-1}\xi_{SI}}\right)^2. \tag{8}$$

Note from (5) that the posterior density of R_0 is only dependent on the infection and removal times via the quantity $(\lambda_\gamma + \xi_I)/(\lambda_\beta + N^{-1}\xi_{SI})$. This is in accord with

the estimator of R_0 given by the ratio of the maximum likelihood estimators of β and γ given $i_1, \boldsymbol{i}$ and $\boldsymbol{r}$, namely $\hat{R}_0 = N(n-1)\xi_I/n\xi_{SI}$ (Andersson and Britton, 2000, p. 93).

Now since $S(t) \leq N-1$ for $i_1 < t \leq \tau$, it follows that $\xi_{SI} \leq (N-1)\xi_I < N\xi_I$ (the final inequality requires $i_1 < \tau$, which we shall assume true). It follows from (6) that if the prior distributions of β and γ are identical then $E[R_0|\boldsymbol{i}, \boldsymbol{r}, i_1] > 1$. Such a conclusion seems to be an artefact of the choice of gamma distributions as priors for β and γ, along with the induced prior distribution on R_0 as discussed in Section 3.2, and reinforces the need for caution in interpreting the posterior mean of R_0. Arguing similarly for the posterior mode of R_0 defined in (7) yields that $\text{mode}[R_0|\boldsymbol{i}, \boldsymbol{r}, i_1] > (n+m-2)/(n+m+1)$, where $m = m_\beta = m_\gamma$, while the classical estimator $\hat{R}_0$ mentioned above satisfies $\hat{R}_0 > (n-1)/n$.

3.4 Bounding the posterior mean of R_0

In practice, infection times are rarely observed, and so we now turn our attention to summary measures of R_0 given data on removal times alone. We assume initially that the start time of the epidemic, i_1, is also known, but this assumption is relaxed later. Without loss of generality, we set $i_1 = 0$.

We now show how to compute bounds on the posterior mean of R_0. First note that

$$\min_{\boldsymbol{i}} E[R_0|\boldsymbol{i}, \boldsymbol{r}, i_1] \leq E[R_0|\boldsymbol{r}, i_1] \leq \max_{\boldsymbol{i}} E[R_0|\boldsymbol{i}, \boldsymbol{r}, i_1],$$

and thus we can obtain bounds by minimising or maximising (6) over all possible infection times $\boldsymbol{i}$. This is equivalent to minimising or maximising the function $h(\boldsymbol{i})$ defined (for given $\boldsymbol{r}$, and $i_1 = 0$) by

$$h(\boldsymbol{i}) \quad = \quad \frac{\lambda_\gamma + \xi_I}{\lambda_\beta + N^{-1}\xi_{SI}}. \tag{9}$$

In fact, maximising $h(\boldsymbol{i})$ is relatively straightforward in some cases, as follows. For $i_1 \leq t \leq \tau$ we have $S(t) \geq N - n$, so that $\xi_{SI} \geq (N-n)\xi_I$, and hence

$$h(\boldsymbol{i}) \leq \frac{\lambda_\gamma + \xi_I}{\lambda_\beta + ((N-n)/N)\xi_I}. \tag{10}$$

It is straightforward to show that the right hand side of (10) is non-decreasing in ξ_I if and only if $(\lambda_\beta/\lambda_\gamma) \geq (N-n)/N$. Now ξ_I is maximised when all infections occur at time $i_1 = 0$, in which case $\xi_I = \sum_{k=1}^n r_k$. It follows that, for $(\lambda_\beta/\lambda_\gamma) \geq (N-n)/N$,

$$h(\boldsymbol{i}) \leq \frac{\lambda_\gamma + \sum_{k=1}^n r_k}{\lambda_\beta + ((N-n)/N)\sum_{k=1}^n r_k}, \tag{11}$$

and moreover this bound is attained in the 'soon as possible' scenario $i_2 = i_3 = \cdots = i_n = 0$.

It is tempting to conjecture that the 'late as possible' scenario, in which $i_{k+1} = r_k$ for $k = 1, 2, \ldots, n-1$, will provide the minimal value of $h(\boldsymbol{i})$, at least when $\lambda_\beta = \lambda_\gamma = 0$ corresponding to vague priors. That this is not the case is demonstrated by the following explicit counterexample.

$$N = 11;\ n = 7;\ \boldsymbol{r} = (3, 4, 5, 6, 7, 8, 11);\ \lambda_\beta = \lambda_\gamma = 0.$$

In this case the 'late as possible' process has infection times $\boldsymbol{i} = (3, 4, 5, 6, 7, 8)$, giving $h(\boldsymbol{i}) = 11/7$. If instead we take $\boldsymbol{i} = (1, 2, 5, 6, 7, 8)$ then we find that $h(\boldsymbol{i}) = 165/106 < 11/7$. Note that this example also shows that the minimal infection-time-vector depends on the entire course of the epidemic, since the 'late as possible' process does minimise the ratio of interest for $0 \leq t \leq 3$.

To find the minimal value of $h(\boldsymbol{i})$, it is helpful first to express the integrals ξ_I and ξ_{SI} in terms of the removal times and infection times. Clearly,

$$\xi_I \quad = \quad \sum_{k=1}^{n} (r_k - i_k)\,.$$

Defining $i_{n+1} = i_{n+2} = \cdots = i_N = \infty$, then it can be shown (Neal and Roberts, 2005) that

$$\xi_{SI} \quad = \quad \sum_{k=1}^{n} \sum_{j=1}^{N} (r_k \wedge i_j - i_k \wedge i_j)\,, \tag{12}$$

where we use $\wedge$ to denote minimum.

Recalling that $i_k \leq i_{k+1} \leq r_k$ for $k = 1, 2, \ldots, n-1$ and using the fact that $i_k = \infty$ for $k \geq n+1$, we can re-write (12) as

$$\begin{aligned}
\xi_{SI} &= \sum_{k=1}^{n}\sum_{j=1}^{n} r_k \wedge i_j + \sum_{k=1}^{n}(N-n)r_k - \sum_{k=1}^{n}\sum_{j=1}^{k} i_j - \sum_{k=1}^{n}\sum_{j=k+1}^{N} i_k \\
&= \sum_{j=1}^{n} i_j + \sum_{k=1}^{n-1}\sum_{j=1}^{k+1} i_j + \sum_{k=1}^{n-2}\sum_{j=k+2}^{n} r_k \wedge i_j + \sum_{k=1}^{n}(N-n)r_k \\
&\quad - \sum_{j=1}^{n}\sum_{k=j}^{n} i_j - \sum_{k=1}^{n}(N-k)i_k \\
&= \sum_{j=1}^{n} i_j - i_1 + \sum_{j=1}^{n}(n-j+1)i_j + \sum_{k=1}^{n-2}\sum_{j=k+2}^{n} r_k \wedge i_j + \sum_{k=1}^{n}(N-n)r_k \\
&\quad - \sum_{j=1}^{n}(n-j+1)i_j - \sum_{k=1}^{n}(N-k)i_k \\
&= \sum_{k=1}^{n}(N-n)r_k + \sum_{k=1}^{n}(k+1-N)i_k + \sum_{k=1}^{n-2}\sum_{j=k+2}^{n} r_k \wedge i_j,
\end{aligned}$$

since $i_1 = 0$. Thus in the definition (9) of $h(\boldsymbol{i})$, the numerator is an affine function of the infection times $\boldsymbol{i}$, while the denominator is an affine function of $\boldsymbol{i}$ together with the set of variables $\{r_k \wedge i_j : k = 1, 2, \ldots, n-2, j = k+2, k+3, \ldots, n\}$. In order to minimise $h(\boldsymbol{i})$, define

$$h(\boldsymbol{i}, \boldsymbol{a}) = \frac{\lambda_\gamma + \sum_{k=1}^n r_k - \sum_{k=1}^n i_k}{\lambda_\beta + (1-(n/N)) \sum_{k=1}^n r_k + (1/N) \left(\sum_{k=1}^n (k+1-N) i_k + \sum_{k=1}^{n-2} \sum_{j=k+2}^n a_{kj} \right)}, \quad (13)$$

where $\boldsymbol{a} = \{a_{kj} : k = 1, 2, \ldots, n-2, j = k+2, k+3, \ldots, n\}$. Consider the following linear fractional program.

[LFP]: Minimise $h(\boldsymbol{i}, \boldsymbol{a})$ subject to

$$\left.\begin{array}{rcll} i_k & \leq & i_{k+1}, & k = 1, 2, \ldots, n-1, \\ i_{k+1} & \leq & r_k, & k = 1, 2, \ldots, n-1, \\ a_{kj} & \leq & r_k, & k = 1, 2, \ldots, n-2, j = k+2, \ldots, n, \\ a_{kj} & \leq & i_j, & k = 1, 2, \ldots, n-2, j = k+2, \ldots, n. \end{array}\right\} \quad (14)$$

Suppose $\boldsymbol{i}, \boldsymbol{a}$ satisfy the constraints (14) and are such that $a_{kj} < r_k \wedge i_j$ for some k, j. Then from the form of the right hand side of (13) it is clear that we can reduce the value of h without violating any of the constraints (14) by increasing a_{kj} up to $r_k \wedge i_j$ while leaving $\boldsymbol{i}$ unchanged. Hence the minimum in [LFP] must be attained for some $\boldsymbol{i}, \boldsymbol{a}$ satisfying $a_{kj} = r_k \wedge i_j$ for all k, j, and therefore provides a minimum of $h(\boldsymbol{i})$ defined by (9).

Before solving [LFP], we shall show that the minimal value of $h(\boldsymbol{i})$ is attained for some $\boldsymbol{i}$ with $i_k \in \mathcal{A} = \{i_1, r_1, r_2, \ldots, r_{n-1}\}$ for $k = 2, 3, \ldots, n$. First, note that a linear fractional program is known to attain its minimal value at a vertex of the feasible region (Martos, 1965). Now consider a set of infection times $\boldsymbol{i}$ with $i_k \notin \mathcal{A}$ for some $k \in \{2, 3, \ldots, n\}$, and take $a_{kj} = r_k \wedge i_j$ for all k, j. Take $m = \min\{k : i_k \notin \mathcal{A}\}$, and $l = \max\{k : i_k = i_m\}$. Set $q = i_{m-1} \vee \max\{r_k : r_k < i_m\}$ and $p = i_{l+1} \wedge \min\{r_k : i_m < r_k\}$ (with the convention that $i_{n+1} = \infty$). Define sets of infection times $\boldsymbol{i}'$, $\boldsymbol{i}''$ as follows:

$$\begin{array}{lll} i'_k = i_k \text{ for } 2 \leq k \leq m-1; & i'_k = q \text{ for } m \leq k \leq l; & i'_k = i_k \text{ for } l+1 \leq k \leq n, \\ i''_k = i_k \text{ for } 2 \leq k \leq m-1; & i''_k = p \text{ for } m \leq k \leq l; & i''_k = i_k \text{ for } l+1 \leq k \leq n. \end{array}$$

We thus have $\boldsymbol{i}' \neq \boldsymbol{i}''$ and $\boldsymbol{i} = \lambda \boldsymbol{i}' + (1-\lambda)\boldsymbol{i}''$ for some λ such that $0 < \lambda < 1$ (in fact, $\lambda = (p - i_m)/(p - q)$). Further, setting $a'_{kj} = r_k \wedge i'_j$ and $a''_{kj} = r_k \wedge i''_j$ we also have $\boldsymbol{a} = \lambda \boldsymbol{a}' + (1-\lambda)\boldsymbol{a}''$. Hence any set of infection times with $i_k \notin \mathcal{A}$ for some k (together with associated a_{kj} values) cannot lie at a vertex of the feasible region, and so the minimum of [LFP] must be attained with $i_k \in \mathcal{A}$ for all k, as claimed.

As an aside, it is also possible to show that for any values of $\lambda_\beta, \lambda_\gamma$ the maximum value of $h(\boldsymbol{i})$ is attained with $i_k \in \mathcal{A}$ for all k using related methods (as opposed to the direct argument we used previously under the condition that $(\lambda_\beta/\lambda_\gamma) \geq (N -$

$n)/N$). Specifically, $h(\boldsymbol{i})$ can be shown to be quasi-convex, from which it follows that its maximum is attained at a vertex of the feasible region (see Martos, 1965).

Various methods exist for solving linear fractional programs (see Ibaraki, 1981). We follow Charnes and Cooper (1962) in transforming our problem into a linear programming problem, which can then be efficiently solved using the simplex method, software for which is widely available. Thus we consider the problem

[LP]: Minimise $g(\boldsymbol{c}, \boldsymbol{b}, t) = (\lambda_\gamma + \sum_{k=1}^{n} r_k)\, t - \sum_{k=1}^{n} c_k$ subject to

$$\left(\lambda_\beta + (1-(n/N))\sum_{k=1}^{n} r_k\right) t + (1/N)\left(\sum_{k=1}^{n}(k+1-N)c_k + \sum_{k=1}^{n-2}\sum_{j=k+2}^{n} b_{kj}\right) = 1,$$

$$\begin{aligned} c_1 &= 0, \\ c_k &\leq c_{k+1}, \quad k = 1, 2, \ldots, n-1, \\ c_{k+1} &\leq r_k t, \quad k = 1, 2, \ldots, n-1, \\ b_{kj} &\leq r_k t, \quad k = 1, 2, \ldots, n-2, j = k+2, \ldots, n, \\ b_{kj} &\leq c_j, \quad k = 1, 2, \ldots, n-2, j = k+2, \ldots, n, \\ t \geq 0, \quad b_{kj} &\geq 0, \quad k = 1, 2, \ldots, n-2, j = k+2, \ldots, n. \end{aligned}$$

Denoting by $\boldsymbol{c}^{min}, \boldsymbol{b}^{min}, t^{min}$ the (not necessarily unique) values of $\boldsymbol{c}, \boldsymbol{b}, t$ for which the minimum of [LP] is attained, then the minimal value of $h(\boldsymbol{i})$ is attained at $\boldsymbol{i} = \boldsymbol{c}^{min}/t^{min}$ and is given by $g\left(\boldsymbol{c}^{min}, \boldsymbol{b}^{min}, t^{min}\right)/t^{min}$.

3.5 Bounds when the initial infection time is unobserved

So far, we have assumed that the initial infection time i_1 is known, and fixed a time origin by taking $i_1 = 0$. In practice it is likely that the initial infection event will not be observed, as illustrated in Section 5 below. Thus we now allow i_1 to take any value, and seek bounds on $h(i_1, \boldsymbol{i})$ under this relaxation of our assumptions. It should be noted that in most applications, plausible bounds on i_1 will be available, and so the bounds derived below are conservative. Note also that our analysis takes no account of the influence of a prior density on i_1, which from (2) is required to estimate $E[R_0|\boldsymbol{r}]$ itself (see O'Neill and Roberts, 1999). However, as described below, the bounds we derive are still surprisingly good.

The upper bound (11) remains valid, except that each removal time r_k must now be replaced by $r_k - i_1$. If $(N-n)\lambda_\gamma \leq N\lambda_\beta$ then (11) implies that $h(i_1, \boldsymbol{i}) \leq N/(N-n)$, and this bound is attained in the limit as $i_1 \to -\infty$ with $i_2 = i_3 = \cdots = i_n = i_1$.

For the lower bound, note that $S(t) \leq N-1$ on the interval (i_1, τ), so that $\xi_{SI} \leq (N-1)\xi_I$ and hence

$$\begin{aligned} h\,(i_1, \boldsymbol{i}) &\geq \frac{\lambda_\gamma + \xi_I}{\lambda_\beta + ((N-1)/N)\xi_I} \\ &= \frac{N}{N-1} + \frac{\lambda_\gamma - (N/(N-1))\lambda_\beta}{\lambda_\beta + ((N-1)/N)\xi_I}. \end{aligned}$$

For $(N-1)\lambda_\gamma \geq N\lambda_\beta$ this implies that $h(i_1, \boldsymbol{i}) \geq N/(N-1)$, and this lower bound is attained in the limit as $i_1 \to -\infty$ with $i_2, i_3, \ldots, i_n$ held fixed.

Note that the bounds of this section do not depend upon the observed removal times. The upper bound depends only upon the final size n of the epidemic, while the lower bound does not depend on the progress of the epidemic at all but only upon the total population size N.

In the case of non-informative priors $\lambda_\beta = \lambda_\gamma = 0$, $m_\beta = m_\gamma = 1$, we have

$$\frac{N}{N-1} \quad \leq \quad E\left[R_0|\boldsymbol{r}\right] \quad \leq \quad \frac{N}{N-n}.$$

Note in particular that we always have $E[R_0|\boldsymbol{r}] > 1$, and that if only one removal is observed, so $n = 1$, then $E[R_0|\boldsymbol{r}] = N/(N-1)$.

3.6 Distributional bounds

From equation (8), it is clear that bounding the integral ratio $h(\boldsymbol{i})$ allows us to bound not only the posterior mean of R_0, but also the posterior variance. Furthermore, we now show that we can bound the whole posterior distribution of R_0 in the sense of likelihood ratio ordering of distributions.

We have observed previously that the posterior density of R_0 is dependent on the infection and removal times only via the quantity h defined by equation (9). Thus we can write the posterior density (5) in the form

$$\pi(R_0|h) = \frac{\Gamma(2n + m_\beta + m_\gamma - 1)}{\Gamma(n + m_\beta - 1)\Gamma(n + m_\gamma)} \frac{h^{1-n-m_\beta} R_0^{n+m_\beta-2}}{(h^{-1}R_0 + 1)^{2n+m_\beta+m_\gamma-1}}, \quad R_0 \geq 0.$$

For $R_{01}, R_{02} > 0$, with h fixed, we have the likelihood ratio

$$\frac{\pi\left(R_{01}|h\right)}{\pi\left(R_{02}|h\right)} \quad = \quad \frac{\left(R_{01}/R_{02}\right)^{n+m_\beta-2}}{\left[\left(R_{01}+h\right)/\left(R_{02}+h\right)\right]^{2n+m_\beta+m_\gamma-1}},$$

so that for $h_1, h_2 > 0$,

$$\frac{\pi\left(R_{01}|h_1\right)}{\pi\left(R_{02}|h_1\right)} \Big/ \frac{\pi\left(R_{01}|h_2\right)}{\pi\left(R_{02}|h_2\right)} \quad = \quad \left(\frac{\left(R_{01}+h_2\right)\left(R_{02}+h_1\right)}{\left(R_{01}+h_1\right)\left(R_{02}+h_2\right)}\right)^{2n+m_\beta+m_\gamma-1}.$$

If $R_{01} > R_{02}$ and $h_1 > h_2$, then $\left(R_{01}+h_1\right)\left(R_{02}+h_2\right) < \left(R_{01}+h_2\right)\left(R_{02}+h_1\right)$, and so

$$\frac{\pi\left(R_{01}|h_1\right)}{\pi\left(R_{02}|h_1\right)} \quad > \quad \frac{\pi\left(R_{01}|h_2\right)}{\pi\left(R_{02}|h_2\right)}.$$

That is, $R_0|h_2 \leq_{\mathrm{LR}} R_0|h_1$, where $\leq_{\mathrm{LR}}$ denotes likelihood ratio ordering of distributions.

Likelihood ratio ordering implies the standard stochastic ordering between distributions (see Kijima and Seneta, 1991), so that for any $r_0 > 0$ and $h_1 > h_2$ we have

$$P\left(R_0 \leq r_0 | h_1\right) \quad \leq \quad P\left(R_0 \leq r_0 | h_2\right),$$

and for any non-decreasing real-valued function $\theta\left(R_0\right)$,

$$E\left[\theta\left(R_0\right)|h_1\right] \quad \geq \quad E\left[\theta\left(R_0\right)|h_2\right].$$

In particular, if h^{min}, h^{max} are the minimal and maximal possible values, respectively, of h given the observed data, then the probability that the epidemic is below threshold may be bounded by

$$P\left(R_0 \leq 1 | h^{max}\right) \quad \leq \quad P\left(R_0 \leq 1 | \boldsymbol{r}\right) \quad \leq \quad P\left(R_0 \leq 1 | h^{min}\right).$$

From (1), taking $\lambda_\beta = \lambda_\gamma = 0$ and $m_\beta = m_\gamma = 1$ to give non-informative priors, we have

$$P\left(R_0 \leq 1 | h\right) \quad = \quad \frac{\Gamma(2n+1)}{\Gamma(n)\Gamma(n+1)} \frac{{}_2F_1\left(2n+1, n; n+1; -h^{-1}\right)}{nh^n}.$$

When the removal times $\boldsymbol{r}$ are observed, but the initial infection time i_1 is not, then with these non-informative priors we have seen that $h^{min} = N/(N-1)$ and $h^{max} = N/(N-n)$.

4 Constant infectious period

In this section we briefly consider the standard SIR model in which the infectious period is simply a constant, so that $T_I = c$ almost surely. Such a choice of infectious period is often more realistic than the exponential infectious period considered in the previous section. In this case, the basic reproduction number is $R_0 = \beta c$.

For this model, and a given set of removal times $r_1 \leq r_2 \leq \ldots \leq r_n$, we have $i_k = r_k - c$, $k = 1, \ldots, n$, which automatically implies the required ordering $i_k \leq i_{k+1}$, $k = 1, \ldots, n-1$. However, a necessary and sufficient condition for $(i_1, \boldsymbol{i}) \in E_{\boldsymbol{r}}$ (in other words, for the epidemic to contain at least one infective during (i_1, τ)) is that $c \geq \bar{r} := \max_{1 \leq k \leq n-1}(r_{k+1} - r_k)$. To see this, note that the condition implies that $i_{k+1} = r_{k+1} - c \leq r_k$ for $k = 1, \ldots, n-1$, so that $(i_1, \boldsymbol{i}) \in E_{\boldsymbol{r}}$. Conversely if $c < r_{k+1} - r_k$ for some $1 \leq k \leq n-1$ then we have $i_{k+1} > r_k$.

Thus the likelihood is given by

$$\begin{aligned} \pi(\boldsymbol{i}, \boldsymbol{r} | \beta, c, i_1) \quad &= \quad \left(\prod_{j=2}^{n} \beta N^{-1} S(i_j-) I(i_j-) \right) \exp\left(- \int_{i_1}^{\tau} \beta N^{-1} S(t) I(t)\, dt \right) \\ &\times \quad \mathbb{1}_{\{\bar{r} \leq c\}} \mathbb{1}_{\{i_k = r_k - c,\, k = 1, \ldots, n\}}. \end{aligned}$$

It follows that, if $\boldsymbol{r}$, $\boldsymbol{i}$ and i_1 are all observed, then inference for c is trivial. Specifically, c is either a point mass at the value dictated by $\boldsymbol{r}$, $\boldsymbol{i}$ and i_1, or else the observations have zero probability density and the model is inappropriate. If just the removals are observed, then $\beta \sim \Gamma(m_\beta, \lambda_\beta)$ *a priori* yields that, for $c \geq \bar{r}$,

$$\pi(\beta|\boldsymbol{r}, c) \sim \Gamma\left(n + m_\beta - 1, \lambda_\beta + N^{-1}\xi_{SI}\right),$$

where here ξ_{SI} is dependent on the value of c. It follows that for $c \geq \bar{r}$,

$$\pi(R_0|\boldsymbol{r}, c) \sim \Gamma\left(n + m_\beta - 1, (\lambda_\beta + N^{-1}\xi_{SI})c^{-1}\right). \tag{15}$$

The posterior density $\pi(R_0|\boldsymbol{r})$ can in principle be obtained from (15) by integrating c out of the product $\pi(R_0|\boldsymbol{r}, c)\pi(c)$, where $\pi(c)$ is the prior density for c. In general this must be done numerically: analytical expressions are hard to obtain because of the way that ξ_{SI} depends on c.

However, it is possible to show that R_0 is stochastically non-decreasing in c, as follows. Recall that for $j = 1, \ldots, n$ we have $i_j = r_j - c$, while if $j \geq n+1$ then $i_j = r_j = \infty$. Substitution into (12) and a few lines of algebra yields that, for $c \geq \bar{r}$,

$$\xi_{SI}(c) = Nnc + \sum_{j=1}^{n}\sum_{k=1}^{n}\left\{[r_k \wedge (r_j - c)] - (r_k \wedge r_j)\right\}, \tag{16}$$

whence

$$c\,\xi'_{SI}(c) = Nnc - \sum_{j=1}^{n}\sum_{k=1}^{n} c\mathbb{1}_{\{r_k > r_j - c\}} \tag{17}$$

(except at the finite set of values $c = r_j - r_k$ for $1 \leq j, k \leq n$, where the derivative is undefined).

Next, for $c \geq \bar{r}$ define $\psi(c) = c/[\lambda_\beta + N^{-1}\xi_{SI}(c)]$, and observe that if $\psi(c)$ is non-decreasing in c, then $R_0|\boldsymbol{r}, c$ will be stochastically non-decreasing in c. Now for any $\lambda_\beta \geq 0$, $\psi'(c) \geq 0$ if $\xi_{SI}(c) \geq c\,\xi'_{SI}(c)$. By (16) and (17), this last inequality holds if and only if

$$\sum_{j=1}^{n}\sum_{k=1}^{n} c\mathbb{1}_{\{r_k > r_j - c\}} \quad \geq \quad \sum_{j=1}^{n}\sum_{k=1}^{n}\left\{(r_k \wedge r_j) - [r_k \wedge (r_j - c)]\right\}.$$

Now for $1 \leq j, k \leq n$, $r_k \leq r_j - c$ implies that $(r_k \wedge r_j) - [r_k \wedge (r_j - c)] = r_k - r_k = 0$, and thus $\xi_{SI}(c) \geq c\,\xi'_{SI}(c)$ if and only if

$$\sum_{j=1}^{n}\sum_{k=1}^{n} c\mathbb{1}_{\{r_k > r_j - c\}} \geq \sum_{j=1}^{n}\sum_{k=1}^{n}\left\{(r_k \wedge r_j) - [r_k \wedge (r_j - c)]\right\}\mathbb{1}_{\{r_k > r_j - c\}}.$$

However, for $1 \leq j, k \leq n$, $r_k > r_j - c$ implies that $0 \leq (r_k \wedge r_j) - [r_k \wedge (r_j - c)] \leq c$, and the result follows.

Now if $c > r_n - r_1$, we have that $r_k > r_j - c$ for all $1 \le j, k \le n$. It follows from (16) that

$$\xi_{SI}(c) = (N-n)nc + \sum_{j=1}^{n}\sum_{k=1}^{n}[r_j - (r_k \wedge r_j)],$$

and in particular $\xi_{SI}(c)/c \to n(N-n)$ as $c \to \infty$. Therefore, $\psi(c) \to N/[n(N-n)]$ as $c \to \infty$, which along with the fact that $\psi(c) \ge \psi(\bar{r})$ yields distributional bounds on $R_0|\boldsymbol{r}, c$. For example, whatever the prior density for c, we have

$$\frac{\bar{r}(n+m_\beta-1)}{\lambda_\beta + N^{-1}\xi_{SI}(\bar{r})} \le E[R_0|\boldsymbol{r}] \le \frac{N(n+m_\beta-1)}{n(N-n)}, \tag{18}$$

$$P(R_0 \le 1|\boldsymbol{r}, c=\infty) \le P(R_0 \le 1|\boldsymbol{r}) \le P(R_0 \le 1|\boldsymbol{r}, c=\bar{r}). \tag{19}$$

Note that the upper and lower bounds in (19) are straightforward to evaluate numerically via equation (15).

5 Numerical examples

In this section we illustrate our methods via three data sets, all of which have been analysed before using MCMC methods. We compare our results with those previously obtained for both identical models to those we consider and also more complex models. We also demonstrate how our methods can be adapted for different settings.

5.1 Abakaliki Smallpox data

We begin with a widely-studied data set obtained from a smallpox outbreak in a closed community of $N = 120$ individuals in Abakaliki, Nigeria (see Bailey, 1975, p125). The use of the general stochastic epidemic to model these data is not entirely appropriate, since smallpox has an appreciable latent period; however, we follow O'Neill and Roberts (1999) in using these data to illustrate our methods. O'Neill and Roberts (1999) used a Markov chain Monte Carlo approach to evaluate the joint posterior distribution of (β, γ), treating the unknown infection times $\boldsymbol{i}$ and i_1 as additional unknown parameters.

The data consist of the following 29 inter-removal times, measured in days:

$$13, 7, 2, 3, 0, 0, 1, 4, 5, 3, 2, 0, 2, 0, 5, 3, 1, 4, 0, 1, 1, 1, 2, 0, 1, 5, 0, 5, 5.$$

Note that for these data, the initial infection time i_1 is not observed. Since O'Neill and Roberts (1999) found the posterior mean of γ with non-informative priors to be 0.098, so that the mean infectious period is $1/0.098 \approx 10.2$ days, we shall assume for purposes of illustration that the first removal occurs 10 days after the first infection. With this assumption, the set of removal times become

$$\begin{aligned}\boldsymbol{r} \;=\; & (10, 23, 30, 32, 35, 35, 35, 36, 40, 45, 48, 50, 50, 52, 52, 57, 60, 61, 65, 65, 66, 67, 68,\\ & 70, 70, 71, 76, 76, 81, 86).\end{aligned}$$

For the case of exponential infectious periods we take non-informative priors, with $\lambda_\beta = \lambda_\gamma = 0$ and $m_\beta = m_\gamma = 1$. Using the Matlab `linprog` command to solve [LP] for the minimum, and reading off the maximum from (11), we find that $1.0804 \leq E[R_0|\boldsymbol{r}, i_1] \leq 1.333$, the maximal value being achieved with $i_2 = i_3 = \cdots = i_{30} = 0$, the minimal value with

$$\boldsymbol{i} = (0,0,0,10,30,35,35,36,40,45,48,50,50,52,52,57,60,61,65,65,66,67,68, 70,70,71,76,76,81).$$

If the initial infection time is assumed to be unknown, then we obtain the bounds $1.008 \leq E[R_0|\,\boldsymbol{r}] \leq 1.333$. For comparison, Boys and Giles (2007) use MCMC methods to obtain $E[R_0|\,\boldsymbol{r}] = 1.24$ for the present model. Regarding the probability that R_0 exceeds unity, we find that $0.1616 \leq P(R_0 \leq 1|\,\boldsymbol{r}) \leq 0.5384$, while if the initial infection time i_1 is observed and such that $r_1 - i_1 = 10$ days, then we obtain the slightly tighter bounds $0.1616 \leq P(R_0 \leq 1|\,\boldsymbol{r}, r_1 - i_1 = 10) \leq 0.4319$.

As well as R_0, our methods are easily adapted to provide bounds for other quantities. For example, from (3) and (4) it is clear that the posterior distributions of β and γ can be stochastically bounded by appropriate minimisation and maximisation of the integrals ξ_{SI} and ξ_I, respectively. The latter integral is trivial to deal with, since it is maximised when all the infection times equal i_1, and minimised when infections occur as late as possible, so that $i_{k+1} = r_k$ for $k = 1, \ldots, n-1$. Thus we have the bounds

$$r_n - i_1 \leq \xi_I \leq \sum_{k=1}^{n} (r_k - i_1).$$

To minimise ξ_{SI}, observe that since removal times are observed then the value of $S + I$ is known at all times. For any given value of the sum $S + I$, we can minimise the product SI subject to $I \geq 1$ by taking $I = 1$. Hence ξ_{SI} is minimised with $i_{k+1} = r_k$ for $k = 1, \ldots, n-1$, giving

$$\xi_{SI} \geq (N - n - 1)(r_n - i_1) + \sum_{k=1}^{n} (r_k - i_1).$$

To maximise ξ_{SI}, observe that for $k = 1, 2, \ldots, n-1$, during the time interval $[r_k, r_{k+1})$ the values of S, I must satisfy the constraints $S + I = N - k$ and $1 \leq I \leq n - k$. Subject to these constraints, the product SI is maximised by taking $I = \lceil (N-k)/2 \rceil \wedge (n-k)$, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. Similarly, within the interval $[i_1, r_1)$ the product SI is maximised with $I = \lceil N/2 \rceil \wedge n$. We can achieve these maximal values of SI throughout the interval $[i_1, r_n)$ by taking infection times $i_k = i_1$ for $2 \leq k \leq k_0$ and $i_k = r_{2(k-k_0)}$ for $k_0 + 1 \leq k \leq n$, where $k_0 = \lceil N/2 \rceil \wedge n$. In the special case when $n \leq (N+1)/2$ we have $i_2 = i_3 = \cdots = i_n = i_1$, giving the bound

$$\xi_{SI} \leq \sum_{k=1}^{n} (N - n)(r_k - i_1).$$

For the Smallpox data, assuming the first removal occurs 10 days after the first infection, we obtain the bounds $86 \leq \xi_I \leq 1,012$ and (observing that $n \leq (N+1)/2$ for these data) $9,266 \leq \xi_{SI} \leq 145,080$. With prior parameters $\lambda_\beta = \lambda_\gamma = 0$ and $m_\beta = m_\gamma = 1$ then we find from (3) and (4) that $0.0248 \leq E[\beta \mid \boldsymbol{i}, \boldsymbol{r}, i_1] \leq 0.3885$ and $0.0306 \leq E[\gamma \mid \boldsymbol{i}, \boldsymbol{r}, i_1] \leq 0.3605$. In both cases the bounds are rather wide, but nevertheless they provide order-of-magnitude information. O'Neill and Roberts (1999) found the posterior means of β and γ to be 0.108 and 0.098, respectively, both of which lie within the bounds we have derived.

If the initial infection time is assumed unknown, then the integrals ξ_I, ξ_{SI} can each be made arbitrarily large by appropriate choice of i_1. For the lower bounds on these integrals, we know that $i_1 \leq r_1$, and so $\xi_I \geq r_n - r_1$ and $\xi_{SI} \geq (N-n-1)(r_n - r_1) + \sum_{k=1}^{n}(r_k - r_1)$, which for the smallpox data gives $\xi_i \geq 76$, $\xi_{SI} \geq 8,076$, and hence we can bound the posterior means as $0 \leq E[\beta \mid \boldsymbol{i}, \boldsymbol{r}] \leq 0.4458$, $0 \leq E[\gamma \mid \boldsymbol{i}, \boldsymbol{r}] \leq 0.4079$.

Suppose now that the infectious period is fixed at length c. For these data we have $c \geq \bar{r} = 13$. Note that this value itself is considerably larger than typical estimates of the mean infectious period in the exponential infectious period model. Assuming $m_\beta = 1$ and $\lambda_\beta = 0$, the values of $E[R_0|\boldsymbol{r}, c]$ for $c = 13, 14, 15$ are, respectively, 1.1774, 1.1796 and 1.1817. From (18) we have the bounds $1.1774 \leq E[R_0|\ \boldsymbol{r}] \leq 1.333$. Note that this range of values lies within the corresponding bounds obtained above for the exponential infectious period case, suggesting that inference for the mean of R_0 is fairly robust across the two different infectious period models. However, such a conclusion does not apply to the probability that the epidemic is below threshold. Specifically, for the constant infectious period model we obtain $0.07474 \leq P(R_0 \leq 1|\boldsymbol{r}) \leq 0.2092$, which contrasts sharply with the range obtained previously for the exponential infectious period. Such a marked difference can be explained in two ways. First, the posterior distribution of R_0 has a larger variance for the exponential infectious period model, which itself is unsurprising since the model contains more inherent variability. Second, recall that the extinction probability of an SIR epidemic model is defined, for $R_0 > 1$, as the smallest root of the equation $f(s) = s$ in $[0, 1]$, where $f(s) = E[s^R]$ is the probability generating function of R, the number of new infections that an infective gives rise to among infinitely many susceptibles (see e.g. Andersson and Britton, Theorem 3.1). In the present case, if the two models have the same $R_0 > 1$ value, then it is straightforward to show that the extinction probability of the exponential infectious period model exceeds that of the constant infectious period model. For example, with $R_0 = 4/3$, the two extinction probabilities are 0.75 and 0.5456, respectively. Such findings illustrate the need for caution when using R_0 alone as a summary measure of an epidemic.

5.2 Gastroenteritis data

Data from an outbreak of gastroenteritis on a hospital ward in South Carolina during 1996 have previously been considered by Britton and O'Neill (2002) and Neal and Roberts (2005). In both cases, the data were used to parameterise a Markov SIR model incorporating a random network representing social structure, using MCMC methods.

It is therefore of interest to see how inference for R_0 obtained by such methods compares with those presented here. The data comprise numbers of cases detected over eight consecutive days, namely 1,0,4,2,3,3,10,5, so that $n = 28$, while $N = 89$. With no knowledge of the initial infection time we find $1.011 \leq E[R_0|\ \boldsymbol{r}] \leq 1.459$, while conditioning on values of i_1 varies the lower bound from 1.082 (with $i_1 = -1$) to 1.019 (with $i_1 = -100$). Such results compare favourably to those obtained by Britton and O'Neill (2002), who found $E[R_0|\ \boldsymbol{r}] = 1.17$ for the random network model.

5.3 Ebola data

Our final example illustrates how our methods can be adapted to more complex models. Data from a well-documented outbreak of Ebola hemorrhagic fever in the Democratic Republic of the Congo in 1995 have been used to parameterise an SEIR model (i.e. an SIR model with an additional stage consisting of an 'exposed' period during which individuals are infected but not yet infectious) by both Chowel *et al.* (2004) and Lekone and Finkenstädt (2006), the latter using MCMC methods. In both cases the authors assume that the infection rate β is time-dependent, being first constant and then decaying exponentially following the introduction of control measures. The data consist of symptom appearance times of 291 individuals and death times for 236 individuals. It is also known that there were a total of 316 cases, so that there are unobserved events of both kinds. The symptom appearance time and time of death for an individual are assumed to correspond to the start of that individual's infectious period and their removal time in the SEIR model, respectively. In what follows we refer to the former time as an infection time, purely for convenience and consistency with the SIR terminology.

Here we consider estimation of R_0, which in this setting is a summary characteristic of the epidemic in the absence of intervention. We therefore focus primarily on the period of time until intervention measures were introduced. First, note that it is straightforward to write down a likelihood similar to (2) for the SEIR model given complete observation of all exposure, infection and removal times, and furthermore it is possible to do so given complete observation up until any fixed time during the epidemic outbreak (cf. O'Neill and Roberts, 1999). It follows that, for example, if both latent and infectious periods are exponentially distributed, then the corresponding rate parameters all have gamma distributed full conditional distributions assuming gamma prior distributions, similar to (3) and (4). Moreover, R_0, which equals β/γ for this SEIR model, is once again distributed as the ratio of independent gamma distributions, given complete observation up until any fixed time. Given complete observation of, for instance, infection and removal times, then as before it is possible to obtain bounds on summary measures of the posterior distribution of R_0 by optimisation over the unobserved exposure times.

Returning to the Ebola data, if T denotes the time at which control measures were introduced then we find that $E[R_0|\boldsymbol{i}_T, \boldsymbol{r}_T, \boldsymbol{c}_T, i_1] = h_T(m_\beta + m - 1)/(m_\gamma + n)$, where (i) $\boldsymbol{i}_T, \boldsymbol{r}_T$ and $\boldsymbol{c}_T$ denote infective period start, removal and exposure times during $(i_1, T]$, respectively; (ii) m and n denote the numbers of exposure events and removals during $(i_1, T]$ and (iii) $h_T = (\lambda_\gamma + \int_{i_1}^{T} I(t)dt)/(\lambda_\beta + N^{-1}\int_{i_1}^{T} S(t)I(t)dt)$. Both Chowel *et al.*

(2004) and Lekone and Finkenstädt (2006) take $N = 5,363,500$, and thus if $\lambda_\beta = \lambda_\gamma$ we find that $h_T \approx 1$. Thus if $m_\beta = m_\gamma = 1$ we find that $E[R_0|\boldsymbol{i}_T, \boldsymbol{r}_T, \boldsymbol{c}_T, i_1] \approx m/(n+1)$, and more generally inference about R_0 comes down to knowing the number of exposure and removal events during $(i_1, T]$. Regarding the latter, although we observe deaths in the Ebola data set, it is also known that 80 removal events (whether actual deaths or not) are unrecorded, so that only a proportion $236/316 = 0.75$ of removal events are observed. From the data set, 97 deaths are observed during $(i_1, T]$ and so it is not unreasonable to assume that the $n \approx 97/0.75 \approx 129$. For m we need more assumptions. The observed number of infections before intervention is 169, providing a lower bound for m, but since the exposure period for Ebola is known to be around 6 days we obtain an upper bound by assuming that all observed infections for 6 days after T arose due to exposure events prior to T. There are 48 such infections in the data, so a reasonable upper bound for m is $169+48 = 217$. We thus obtain the bounds $1.3 \leq E[R_0|\boldsymbol{i}_T, \boldsymbol{r}_T, \boldsymbol{c}_T, i_1] \leq 1.7$. Lekone and Finkenstädt (2006) report a posterior mean of R_0 as 1.38, which lies within the bounds we have obtained. Their model is not strictly comparable with our approach here, partly because their SEIR model is discretised in time, but more importantly because the initial infection rate parameter β is also a component of the post-intervention infection rate. Nevertheless, it is interesting to see that our approach rapidly provides bounds that are compatible with estimates derived from MCMC methods.

6 Conclusions

In this paper we have considered Bayesian inference for the standard SIR model, focussing particularly on the basic reproduction number in the case where the infectious period is either exponentially distributed or non-random. These two choices of infectious period distribution are of some practical interest. The exponential case is commonly used by modellers, partly for mathematical convenience, and partly because it provides a natural analogue to deterministic differential equation models, in which a constant rate of removal corresponds to an exponentially distributed infectious period. The constant infectious period case is of interest because, for many diseases, it gives a good approximation to reality. For both models, we have provided either exact expressions or bounds for the posterior distribution and summary statistics of R_0, depending on the assumed data.

It is natural to consider other choices of infectious period distribution, although in general the approaches described in this paper will be harder to adopt. There are two reasons for this. The main difficulty is that the posterior distribution for R_0 given complete observation will not, in general, have a closed form. For example, assuming a $\Gamma(\alpha, \delta)$ infectious period distribution gives $R_0 = \beta\alpha/\delta$, which even in the case of complete observation does not yield a tractable posterior distribution. Of course, this limitation does not necessarily apply when considering inference for other quantities, such as infection rates. A second drawback is that in general it is more natural to work not with a set of ordered infection times $i_1 \leq i_2 \leq \ldots \leq i_n$, but instead define infection time i_k to be that of the individual removed at time r_k for $k = 1, 2, \ldots, n$. In this way, the part of the likelihood that corresponds to the removal process is straightforward to

write down as a product of terms such as $f(r_k - i_k)$, where f is the probability density function of the infectious period. However, in this setting the constraints on the set of possible infection times that ensure there is always at least one infective present become more complicated, since they now rely on the order statistics of the infection times. Nevertheless the methods can, in principle, be applied.

The linear programming approach we have taken to bound certain posterior summary statistics can be applied to related problems. We have illustrated this by considering bounds for related model parameters (such as infection rates) and also by considering different models (such as the SEIR epidemic). The methods are thus likely to be applicable to a rather wider range of models and data sets than we have considered here. An additional practical use of the methods is that they can provide an informal MCMC diagnostic tool, in the sense that the more accurate estimates of model parameters obtained by MCMC must lie within the bounds obtained by our methods. Finally, we remark that our relatively simple approach can provide some indication of how the assumptions of more complex models affect inference for basic model parameters. This can be achieved by, for example, comparing MCMC estimates from such models to the bounds obtained using our simpler models and methods.

References

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and their Statistical Analysis*. Lecture Notes in Statistics 151. New York: Springer-Verlag.

Auranen, K., Arjas, E., Leino, T. and Takala, A. K. (2000). "Transmission of Pneumococcal carriage in families: a latent Markov process model for binary longitudinal data." *Journal of the American Statistical Association*, 95: 1044–1053.

Bailey, N.T.J. (1975). *The Mathematical Theory of Infectious Diseases and its Applications* (2nd ed). London: Griffin.

Becker, N. G. (1989). *Analysis of Infectious Disease Data*. London: Chapman and Hall.

Bhoj, D. S. and Schiefermayr, K. (2001). "Approximations to the distribution of weighted combination of independent probabilities." *Journal of Statistical Computation and Simulation*, 68: 153–159.

Boys, R. J. and Giles, P. R. (2007). "Bayesian inference for stochastic epidemic models with time-inhomogeneous removal rates." *Mathematical Biology*, 55: 223–247.

Britton, T. and O'Neill, P. D. (2002). "Bayesian inference for stochastic epidemics in populations with random social structure." *Scandinavian Journal of Statistics*, 29: 375–390.

Cauchemez, S., Carrat, F., Viboud, C., Valleron, A. J. and Böelle, P. Y. (2004). "A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data." *Statistics in Medicine*, 23: 3469–3487.

Charnes, A. and Cooper, W.W. (1962). "Programming with linear fractional functionals." *Naval Research Logistics Quarterly*, 9: 181–186.

Chowel, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W. and Hyman, J. M. (2004). "The basic reproduction number of Ebola and the effects of public health measures: the cases of Congo and Uganda." *Journal of Theoretical Biology*, 229: 119–126.

Demiris, N. and O'Neill, P. D. (2005). "Bayesian inference for stochastic multitype epidemics in structured populations via random graphs." *Journal of the Royal Statistical Society, Series B*, 67: 731–746.

Dietz, K. (1993). "The estimation of the basic reproduction number for infectious diseases." *Statistical methods in medical research*, 2: 23–41.

Gibson, G. J. 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.

Höhle, M., Jørgensen, E. and O'Neill, P. D. (2005). "Inference in disease transmission experiments by using stochastic epidemic models." *Applied Statistics*, 54: 349–366.

Ibaraki, T. (1981). "Solving mathematical programming problems with fractional objective functions." In *Generalized Concavity in Optimization and Economics*, eds. S. Schaible and W.T. Ziemba, 441–472. New York: Academic Press.

Kijima, M. and Seneta, E. (1991). "Some results for quasistationary distributions of birth-death processes." *Journal of Applied Probability*, 28: 503–511.

Lekone, P. E. and Finkenstädt, B. F. (2006). "Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study." *Biometrics* 62: 1170–1177.

Li, N., Qian, G., and Huggins, R. (2002). "Analysis of between-household heterogeneity in disease transmission from data on outbreak sizes." *Australia and New Zealand Journal of Statistics* 44: 401–411.

Martos, B. (1965). "The direct power of adjacent vertex programming methods." *Management Science*, 12: 241–252.

McBryde, E.S., Gibson, G., Pettitt, A.N., Zhang, Y., Zhao, B. and McElwain, D.L.S. (2006). "Bayesian modelling of the severe acute respiratory syndrome epidemic." *Bulletin of Mathematical Biology*, 68: 889–917.

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

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." *Applied Statistics*, 49: 517–542.

O'Neill, P. D. and Marks, P. J. (2005). "Bayesian model choice and infection route modelling in an outbreak of Norovirus." *Statistics in Medicine*, 24: 2011–2024.

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.

Streftaris, G. and Gibson, G.J. (2004). "Bayesian inference for stochastic epidemics in closed populations." *Statistical Modelling*, 4: 63–75.

Acknowledgments

The authors wish to thank the Editor and referees for helpful comments that have improved the content and layout of the paper.