Bayesian Analysis (2018) **TBA**, Number TBA, pp. 1–30

Bayes Factors for Partially Observed Stochastic Epidemic Models

Muteb Alharthi[§*], Theodore Kypraios[†], and Philip D. O'Neill[‡]

Abstract. We consider the problem of model choice for stochastic epidemic models given partial observation of a disease outbreak through time. Our main focus is on the use of Bayes factors. Although Bayes factors have appeared in the epidemic modelling literature before, they can be hard to compute and little attention has been given to fundamental questions concerning their utility. In this paper we derive analytic expressions for Bayes factors given complete observation through time, which suggest practical guidelines for model choice problems. We adapt the power posterior method for computing Bayes factors so as to account for missing data and apply this approach to partially observed epidemics. For comparison, we also explore the use of a deviance information criterion for missing data scenarios. The methods are illustrated via examples involving both simulated and real data.

Keywords: Bayes factor, power posterior, stochastic epidemic model.

MSC 2010 subject classifications: Primary 62P10; secondary 62F15.

1 Introduction

This paper is concerned with the problem of choosing between a small number of competing infectious disease transmission models, given partial observation of an epidemic outbreak through time. A key reason to consider such a problem is that the models represent different hypotheses about disease spread, such as the infection mechanism, the nature of the contact structure between individuals in the population, or other disease characteristics. A secondary reason to compare models is that they are often used to simulate potential future outbreaks, perhaps with a view to designing control strategies, in which case it is computationally more efficient to employ the simplest possible model which can represent observed data reasonably well. Note that it is typically the case that we only have a few models under consideration, in contrast to variable-selection problems that occur in regression modelling. Also, our focus here is not on the assessment of model fit, itself concerned with whether or not a single specific epidemic model adequately describes the data to hand.

Within the epidemic modelling literature, there is to date no definitively preferred method for model choice. In the Bayesian setting, approaches include the use of Bayes

*Department of Mathematics and Statistics, College of Science, Taif University, P.O. Box 888, Hawiyah, Taif, 26571, Kingdom of Saudi Arabia, muteb.faraj@tu.edu.sa

†School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK, theodore.kypraios@nottingham.ac.uk

‡School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK, philip.oneill@nottingham.ac.uk

§Funding from Taif University.

 https://doi.org/10.1214/18-BA1134

factors, (e.g. Neal and Roberts, 2004; O'Neill and Marks, 2005), criteria such as the Deviance Information Criterion (DIC) (e.g. Worby, 2013; Lau et al., 2014; Deeth et al., 2015), and methods based on the predictive distribution of future outbreaks (Zhang, 2014). Here we focus on the use of Bayes factors. The most common approach to calculating Bayes factors for epidemics has been to use reversible jump Markov chain Monte Carlo (MCMC) methods (Neal and Roberts, 2004; O'Neill and Marks, 2005), although these are often problematic in practice due to the challenge of designing efficient algorithms. Touloupou et al. (2018) used a combination of MCMC methods and importance sampling to estimate marginal likelihoods, from which Bayes factors can be calculated. Knock and O'Neill (2014) used a path-sampling method (Gelman and Meng, 1998) to compare epidemic models given data on the final outcome of the epidemic. This is somewhat related to the methods we develop, albeit for different kinds of data.

In this paper we adapt the power posterior method for calculating Bayes factors (Friel and Pettitt, 2008; Friel et al., 2014) to a missing-data situation that commonly occurs in epidemic modelling. In addition to this computational method, we also derive analytic expressions for Bayes factors in the setting where an epidemic outbreak is completely observed. Although such detailed observation is not that common in practice, our results are of theoretical interest and also provide practical insight into the choice and influence of prior distributions for parameters of the competing models. For comparison, we also consider a form of DIC suitable for missing data.

Throughout the paper we focus on the Susceptible-Infective-Removed (SIR) epidemic model, the most widely-studied stochastic epidemic model. However, the computational methods we develop could be applied to more complex models. For illustration we consider two specific kinds of model choice question: one in which models with different infectious period distributions are compared, and one in which models with different infection mechanisms are compared. Again, other comparisons are possible using our methods.

The paper is arranged as follows. Section 2 recalls preliminary information on epidemic models, Bayes factors, DIC methods and power posterior methods, adapting the latter to a missing-data situation. In Section 3 we derive analytical expressions for Bayes factors given completely observed outbreaks, and in Section 4 we describe computational methods for partially observed outbreaks. Section 5 contains illustrative examples of the methods, and concluding comments are found in Section 6.

2 Preliminaries

2.1 The stochastic SIR epidemic model

The stochastic SIR epidemic model is defined as follows (see e.g. Andersson and Britton, 2000). Consider a closed population of N individuals. At any time, each individual is either susceptible, infective or removed. Susceptible individuals have not contracted the disease but are able to do so. Infective individuals have the disease and can pass it on to others. Removed individuals are no longer infective and play no further part

in disease spread. In applications, the removed state is context-specific and could correspond to immunity, isolation, death, or similar outcomes. The population initially consists of n susceptible individuals and a infectives who have just become infective. Each infective individual remains so for a period of time, called the infectious period, which is drawn from a specified non-negative probability distribution T_I. At the end of its infectious period an individual is immediately removed. All infectious periods are independent.

Every pair of individuals in the population has contacts at times given by the points of a Poisson process of rate βn^{-1}. The Poisson processes for different pairs are mutually independent. If a contact occurs when one member of a pair is infective and the other susceptible, then the susceptible immediately becomes infective. All other contacts (e.g. between two infectives, two susceptibles, etc.) have no effect on the states of the individuals concerned. The epidemic ends as soon as there are no more infectives remaining in the population.

For any time t, denote by $X(t)$ and $Y(t)$ respectively the numbers of susceptible and infective individuals currently in the population. It follows that infections occur in the population at rate $\beta n^{-1} X(t)Y(t)$, meaning that the probability of an infection occurring in $(t, t+\delta t)$ is $\beta n^{-1} X(t)Y(t)\delta t + o(\delta t)$. We will also consider a variant of the SIR model in which the overall infection rate is $\beta n^{-1} X(t)Y^p(t)$ for $p \in (0,1)$. Such models were first introduced by Severo (1969) and relax the assumption that the overall infection rate increases linearly with $Y(t)$ (see also O'Neill and Wen, 2012, and references therein). We will focus on two particular choices of infectious period distribution, namely exponential and gamma, both of which frequently appear in the epidemic modelling literature. Finally, we will assume throughout that there is one initial infective, although this constraint can easily be relaxed.

2.2 Bayes factors

Suppose we have two competing epidemic models, m_1 and m_2, with parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively, and that we have observed data $\boldsymbol{y}$. The Bayes factor for m_1 relative to m_2 is defined by

$$BF_{12} = \frac{\pi(\boldsymbol{y}|m_1)}{\pi(\boldsymbol{y}|m_2)} = \frac{\int \pi(\boldsymbol{y}|\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\int \pi(\boldsymbol{y}|\boldsymbol{\theta}_2)\pi(\boldsymbol{\theta}_2)d\boldsymbol{\theta}_2},$$

where here and throughout the paper, π denotes a probability mass or density function, as appropriate. For partially observed epidemic models, the likelihood term $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ is typically intractable, and a common approach is to introduce auxiliary variables, $\boldsymbol{x}$ say, such that the augmented likelihood $\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})$ is available in closed form and can be computed efficiently. Estimation of the posterior distribution of $\boldsymbol{\theta}$ can then be achieved via Markov chain Monte Carlo methods (Gibson and Renshaw, 1998; O'Neill and Roberts, 1999). In most situations, $\boldsymbol{x}$ will describe the infection process, since this is usually unobserved.

The computation of Bayes factors is, in general, a challenging problem. Many approaches exist, but here we focus specifically on the power posterior method. One attractive aspect of this approach is that it is relatively prescriptive, meaning that the

user does not have that many implementation choices which could seriously affect the performance of the resulting algorithm. Also, as we now explain, the method can be adapted to cover the kind of missing data problem that usually arises in the context of epidemic modelling.

2.3 The power posterior method for models incorporating missing data

We now adapt the power posterior approach (Lartillot et al., 2006; Friel and Pettitt, 2008; Friel et al., 2014) to models incorporating missing data. The method itself provides a way of calculating the marginal density $\pi(\boldsymbol{y}|m_i)$, $i = 1, 2$, from which the Bayes factor can be obtained.

Let $\boldsymbol{y}$ and $\boldsymbol{x}$ denote the observed and the missing data respectively with $\boldsymbol{\theta}$ representing the model parameters. We refer to $(\boldsymbol{y}, \boldsymbol{x})$ as the complete data. Note that in the epidemic settings that we will consider, neither $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ nor $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ are typically tractable but we can compute the augmented likelihood function $\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})$. Specifically, (i) $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ could be obtained by integrating $\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})$ with respect to $\boldsymbol{x}$, but this integral is typically analytically and numerically intractable,; (ii) $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ is proportional to $\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})$, but the normalising constant $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ is the typically-intractable integral of $\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})$ with respect to $\boldsymbol{y}$.

For $t \in [0, 1]$, we define the power posterior for the missing data scenario as

$$\pi_t(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta}),$$

with the normalizing constant

$$z_t(\boldsymbol{y}) = \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}\, d\boldsymbol{x}.$$

Thus, noting that $z_{t=1}(\boldsymbol{y}) = \pi(\boldsymbol{y})$ and $z_{t=0}(\boldsymbol{y}) = 1$,

$$\log(\pi(\boldsymbol{y})) = \log\left[\frac{z_{t=1}(\boldsymbol{y})}{z_{t=0}(\boldsymbol{y})}\right] = \int_0^1 E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] dt, \tag{1}$$

where the second equality in (1) can be derived by adapting the arguments of Lartillot et al. (2006), as follows:

$$\begin{aligned}
\frac{d}{dt}\log(z_t(\boldsymbol{y})) &= \frac{1}{z_t(\boldsymbol{y})}\frac{d}{dt} z_t(\boldsymbol{y}) \\
&= \frac{1}{z_t(\boldsymbol{y})} \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \frac{d}{dt} \pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}\, d\boldsymbol{x} \\
&= \frac{1}{z_t(\boldsymbol{y})} \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})^t \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}\, d\boldsymbol{x} \\
&= \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \frac{\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta})}{z_t(\boldsymbol{y})} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] d\boldsymbol{\theta}\, d\boldsymbol{x}
\end{aligned}$$

$$\begin{aligned} &= \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] \pi_t(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{\theta}\, d\boldsymbol{x} \\ &= E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right]. \end{aligned}$$

Note that the above argument requires a regularity condition, specifically permitting the exchange of order of integration and differentiation.

By integrating with respect to $t \in [0, 1]$, we have

$$\begin{aligned} \log(\pi(\boldsymbol{y})) &= \log(z_{t=1}(\boldsymbol{y})) - \log(z_{t=0}(\boldsymbol{y})) \\ &= \log(z_{t=1}(\boldsymbol{y})) \\ &= \int_0^1 E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] dt. \end{aligned}$$

We shall evaluate the final integral numerically, by evaluating it at a finite number of t values, namely $0 = t_0 < t_1 < \ldots < t_r = 1$. To reduce the resulting approximation error, we follow Friel et al. (2014) and make use of the fact that the gradient of the expected log-likelihood curve equals its variance. Specifically,

$$\begin{aligned} \frac{d}{dt} E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] &= \frac{d}{dt} \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] \pi_t(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{\theta}\, d\boldsymbol{x} \\ &= \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] \frac{d}{dt} \pi_t(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{\theta}\, d\boldsymbol{x}, \end{aligned}$$

where

$$\begin{aligned} \frac{d}{dt} \pi_t(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) &= \frac{d}{dt} \left[\frac{\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta})}{z_t(\boldsymbol{y})} \right] \\ &= \frac{z_t(\boldsymbol{y}) \pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta}) \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] - \pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta}) \frac{d}{dt} z_t(\boldsymbol{y})}{z_t^2(\boldsymbol{y})} \\ &= \frac{\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta})}{z_t(\boldsymbol{y})} \left[\log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] - \frac{1}{z_t(\boldsymbol{y})} \frac{d}{dt} z_t(\boldsymbol{y}) \right] \\ &= \pi_t(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) \left[\log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] - \frac{d}{dt} \log(z_t(\boldsymbol{y})) \right]. \end{aligned}$$

Hence,

$$\begin{aligned} \frac{d}{dt} E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] &= \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \left(\log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right]\right)^2 \pi_t(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{\theta}\, d\boldsymbol{x} \\ &- \frac{d}{dt} \log(z_t(\boldsymbol{y})) \int_{\boldsymbol{x}} \int_{\boldsymbol{\theta}} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right] \pi_t(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{\theta}\, d\boldsymbol{x} \\ &= E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t} \left(\log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right]\right)^2 - \left(E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right]\right)^2 \\ &= V_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t} \log\left[\pi(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\theta})\right]. \end{aligned}$$

Using the corrected trapezoidal rule form (Atkinson and Han, 2004), namely

$$\int_a^b f(y) dy \approx \frac{(b-a)}{2} \left[f(a) + f(b)\right] - \frac{(b-a)^2}{12} \left[f'(b) - f'(a)\right],$$

we have the following adapted power posterior estimate of $\log(\pi(\boldsymbol{y}))$:

$$\begin{aligned}\log(\pi(\boldsymbol{y})) &\approx \sum_{j=1}^{r} \frac{1}{2}(t_j - t_{j-1}) \\ &\times \left\{E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t_j} \log\left[\pi(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta})\right] + E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t_{j-1}} \log\left[\pi(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta})\right]\right\} \\ &- \sum_{j=1}^{r} \frac{1}{12}(t_j - t_{j-1})^2 \\ &\times \left\{V_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t_j} \log\left[\pi(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta})\right] - V_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t_{j-1}} \log\left[\pi(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta})\right]\right\}. \qquad (2)\end{aligned}$$

Algorithm 1 can be used to implement the adapted power posterior method for missing data models. We follow the recommendation of Friel and Pettitt (2008) for the choice of t_j values in (2). This choice ensures that many of the t_j values are close to zero, where the expected log-likelihood curve often changes rapidly in practice. The t_j values are often called temperatures, and the collection of values called a temperature ladder, this terminology arising because the power posterior method is a form of so-called thermodynamic integration.

Algorithm 1 MCMC algorithm for estimating the marginal likelihood via the missing-data power posterior approach.

1. Initialise algorithm with $\boldsymbol{x}^0$ and $\boldsymbol{\theta}^0$.
2. For $j = 0, \ldots, r$:
   a. Set $t_j = (j/r)^c$, where $c > 1$ is a constant.
   b. Generate a sample $\{(\boldsymbol{\theta}_j^{(1)}, \boldsymbol{x}_j^{(1)}), \ldots, (\boldsymbol{\theta}_j^{(M)}, \boldsymbol{x}_j^{(M)})\}$ from $\pi_{t_j}(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})$ via an MCMC sampling scheme.
   c. Estimate $E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t_j} \log\left[\pi(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta})\right]$ and $V_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y},t_j} \log\left[\pi(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta})\right]$ using the sample from b.
   d. While $j < r$, initialise the next chain at the previous posterior mean of $\pi_{t_j}(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{x})$.
3. Estimate $\log(\pi(\boldsymbol{y}))$ using (2).

2.4 DIC for models with missing data

Although Bayes factors are our primary focus, for comparison we will also compute a form of DIC suitable for missing data situations. Celeux et al. (2006) propose various options; the one best-suited to our setting, in the sense that it is suitable for situations where the missing data are not our main focus, and we can compute it, is

$$\mathrm{DIC}_6 = -4E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y}}[\log(\pi(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta}))] + 2E_{\boldsymbol{x}|\boldsymbol{y},\hat{\boldsymbol{\theta}}}[\log(\pi(\boldsymbol{y},\boldsymbol{x}|\hat{\boldsymbol{\theta}}))]. \qquad (3)$$

Calculation of this quantity requires two runs of an MCMC algorithm. The first is to derive $E_{\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y}}[\log(\pi(\boldsymbol{y},\boldsymbol{x}|\boldsymbol{\theta}))]$ and $E_{\boldsymbol{\theta}|\boldsymbol{y}}(\boldsymbol{\theta}|\boldsymbol{y}) = \hat{\boldsymbol{\theta}}$, and the second run is to obtain $E_{\boldsymbol{x}|\boldsymbol{y},\hat{\boldsymbol{\theta}}}[\log(\pi(\boldsymbol{y},\boldsymbol{x}|\hat{\boldsymbol{\theta}}))]$ setting $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and allowing $\boldsymbol{x}$ to vary. In our epidemic setting, $\hat{\boldsymbol{\theta}}$

contains posterior point estimates of the model parameters, the identity of the initial infective and the initial infection time. The preferred model from those under consideration is the one with the lowest DIC_6 value.

3 Model selection given complete outbreak data

In this section we show that Bayes factors for the epidemic models of interest can be computed explicitly if complete data are available. This situation is rare in practice, although it can arise when outbreaks are being closely monitored (e.g. in the early stages of a suspected major epidemic, or in experimental settings for animal diseases). Nevertheless, we gain some insight into the value of Bayes factors as a tool for model choice, particularly with respect to the choice of within-model prior distribution.

3.1 The SIR model with different infectious periods

Suppose we observe an epidemic among a population of N individuals of whom initially n are susceptible and one is infective. Denote by n_R the total number of individuals ever infected, including the initial infective, and label these n_R individuals $1, \ldots, n_R$. The remaining individuals are labelled $n_R + 1, \ldots, N$. For $j = 1, \ldots, N$ let I_j and R_j denote, respectively, the infection and removal time of individual j, with $I_j = R_j = \infty$ for $j > n_R$. Let z denote the label of the initial infective, so that $I_z < I_j$, for all $j \neq z$. Finally let $\boldsymbol{I} = (I_1, \ldots, I_{z-1}, I_{z+1}, \ldots, I_{n_R})$ denote the vector of infection times of infected individuals other than the initial infective, and let $\boldsymbol{R} = (R_1, \ldots, R_{n_R})$ denote the vector of removal times of all infected individuals.

We consider two competing SIR models with identical infection mechanisms but different choices of infectious period distribution T_I. Specifically, model m_1 has $T_I \sim \mathrm{Exp}(\gamma)$ and model m_2 has $T_I \sim \mathrm{Gamma}(\alpha, \delta)$ with shape parameter α assumed known. The likelihoods of $(\boldsymbol{I}, \boldsymbol{R})$ under the two models are

$$\pi(\boldsymbol{I}, \boldsymbol{R}|\beta, \gamma, I_z, z, m_1) = \beta^{n_R - 1} \prod_{j=1, j\neq z}^{n_R} n^{-1} Y(I_j-) \times e^{-\beta n^{-1} A} \times \gamma^{n_R}\, e^{-\gamma B}$$

and

$$\begin{aligned} \pi(\boldsymbol{I}, \boldsymbol{R}|\alpha, \beta, \delta, I_z, z, m_2) &= \beta^{n_R - 1} \prod_{j=1, j\neq z}^{n_R} n^{-1} Y(I_j-) \times e^{-\beta n^{-1} A} \\ &\times \Gamma^{-n_R}(\alpha) \times \prod_{j=1}^{n_R} (R_j - I_j)^{\alpha - 1} \times \delta^{\alpha n_R}\, e^{-\delta B}, \end{aligned}$$

where

$$A = \int_{I_z}^{R_{n_R}} X(t) Y(t) dt = \sum_{j=1}^{n_R} \sum_{k=1}^{N} (R_j \wedge I_k - I_k \wedge I_j),$$

$$B = \int_{I_z}^{R_{n_R}} Y(t)dt = \sum_{j=1}^{n_R}(R_j - I_j),$$

$\Gamma^{-n_R}(\alpha)$ denotes $(\Gamma(\alpha))^{-n_R}$ and $Y(t-) = \lim_{s\uparrow t} Y(s)$, see e.g. Kypraios (2007).

By assigning an independent gamma prior distribution for each of the model parameters, namely Gamma$(\lambda_\zeta, \nu_\zeta)$, where $\zeta = \beta,\ \gamma,\ \delta$, the Bayes factor can be derived explicitly in this case as follows:

$$\begin{aligned}
BF_{12} &= \frac{\pi(\boldsymbol{I},\boldsymbol{R}|m_1)}{\pi(\boldsymbol{I},\boldsymbol{R}|m_2)} = \frac{\int_\gamma\int_\beta \pi(\boldsymbol{I},\boldsymbol{R}|\beta,\gamma)\pi(\beta)\pi(\gamma)d\beta\ d\gamma}{\int_\delta\int_\beta \pi(\boldsymbol{I},\boldsymbol{R}|\beta,\alpha,\delta)\pi(\beta)\pi(\delta)d\beta\ d\delta} \\
&= \frac{\nu_\beta^{\lambda_\beta}\ \nu_\gamma^{\lambda_\gamma}\ \Gamma(\lambda_\beta)\,\Gamma(\lambda_\delta)}{\nu_\beta^{\lambda_\beta}\ \nu_\delta^{\lambda_\delta}\ \Gamma(\lambda_\beta)\,\Gamma(\lambda_\gamma)} \times \frac{\Gamma^{n_R}(\alpha) \times \prod_{j=1,j\neq z}^{n_R} n^{-1}Y(I_j-)}{\prod_{j=1,j\neq z}^{n_R} n^{-1}Y(I_j-) \times \prod_{j=1}^{n_R}(R_j - I_j)^{\alpha-1}} \\
&\times \frac{\int_\beta \beta^{n_R-1} \times e^{-\beta n^{-1}A} \times \beta^{\lambda_\beta-1}e^{-\nu_\beta\beta}d\beta}{\int_\beta \beta^{n_R-1} \times e^{-\beta n^{-1}A} \times \beta^{\lambda_\beta-1}e^{-\nu_\beta\beta}d\beta} \\
&\times \frac{\int_\gamma \gamma^{n_R}\ e^{-\gamma\sum_{j=1}^{n_R}(R_j-I_j)} \times \gamma^{\lambda_\gamma-1}e^{-\nu_\gamma\gamma}d\gamma}{\int_\delta \delta^{\alpha n_R}\ e^{-\delta\sum_{j=1}^{n_R}(R_j-I_j)} \times \delta^{\lambda_\delta-1}e^{-\nu_\delta\delta}d\delta} \\
&= \frac{\nu_\gamma^{\lambda_\gamma}\ \Gamma(\lambda_\delta) \times \Gamma^{n_R}(\alpha)}{\nu_\delta^{\lambda_\delta}\ \Gamma(\lambda_\gamma) \times \prod_{j=1}^{n_R}(R_j - I_j)^{\alpha-1}} \\
&\times \frac{(\nu_\delta + \sum_{j=1}^{n_R}(R_j - I_j))^{\alpha n_R+\lambda_\delta} \times \Gamma(n_R + \lambda_\gamma)}{(\nu_\gamma + \sum_{j=1}^{n_R}(R_j - I_j))^{n_R+\lambda_\gamma} \times \Gamma(\alpha n_R + \lambda_\delta)}.
\end{aligned} \tag{4}$$

The resulting Bayes factor is independent of the infection rate prior parameters. This is a consequence of the fact that the likelihood expressions can be factorized into parts corresponding to the infection and removal processes, and the former are the same in both models. Note also that the Bayes factor is identical to that obtained by comparing exponential and gamma distributions for a sequence of independent and identically distributed observations $R_1 - I_1, \ldots, R_{n_R} - I_{n_R}$, although in our setting things are slightly different because the observations and n_R are not independent.

In the epidemic modelling literature, it is often the case that prior parameters for positive quantities such as rate parameters are assigned the same vague prior distributions. Here, this assumption gives $\lambda_\gamma = \lambda_\delta = \lambda$ and $\nu_\gamma = \nu_\delta = \nu$, where $\lambda \geq 1$ and ν is a small positive number. The Bayes factor in (4) then becomes

$$BF_{12} = \frac{\Gamma(n_R + \lambda)\ \ \Gamma^{n_R}(\alpha)}{\Gamma(\alpha n_R + \lambda)\ \prod_{j=1}^{n_R}(R_j - I_j)^{\alpha-1}} \times \left(\nu + \sum_{j=1}^{n_R}(R_j - I_j)\right)^{n_R(\alpha-1)}. \tag{5}$$

Reformulating (5) in terms of the mean and variance of the prior distribution yields

$$BF_{12} = \frac{\Gamma\left(n_R + \frac{\mu^2}{\sigma^2}\right)\ \ \Gamma^{n_R}(\alpha)}{\Gamma\left(\alpha n_R + \frac{\mu^2}{\sigma^2}\right)\ \prod_{j=1}^{n_R}(R_j - I_j)^{\alpha-1}} \times \left(\frac{\mu}{\sigma^2} + \sum_{j=1}^{n_R}(R_j - I_j)\right)^{n_R(\alpha-1)},$$

True model	α	β	$E[\log BF_{12}]$ $N=30$	$E[\log BF_{12}]$ $N=50$	$P(BF_{12}>1)$ $N=30$	$P(BF_{12}>1)$ $N=50$
m_1	10	1.5	42.6	69.5	0.92	0.92
m_1	5	1.5	15.8	23.3	0.84	0.87
m_1	2	1.5	1.8	3.2	0.65	0.67
m_1	10	2.0	62.6	107.2	0.92	0.94
m_1	5	2.0	22.5	39.3	0.91	0.91
m_1	2	2.0	2.9	4.9	0.75	0.83
m_2	10	1.5	-9.4	-14.6	0.02	0.02
m_2	5	1.5	-5.7	-8.9	0.04	0.04
m_2	2	1.5	-1.5	-2.3	0.15	0.12
m_2	10	2.0	-15.0	-25.3	0.01	0.01
m_2	5	2.0	-8.4	-14.8	0.02	0.02
m_2	2	2.0	-2.1	-3.7	0.09	0.07

Table 1: Expected log Bayes factors for models with different infection periods, assuming diffuse prior distributions as in equation (6). Model m_1 is an SIR model with $T_I \sim$ Exp(1) while m_2 has $T_I \sim$ Gamma(α, α), so that both models have the same mean infectious period. Each row gives parameter values and results from 1000 simulated epidemics from the true model in which at least one new infection occurred.

where $\mu = \lambda/\nu$ and $\sigma^2 = \lambda/\nu^2$. Thus as $\sigma^2 \to \infty$, the prior becomes increasingly diffuse and BF_{12} converges to its lower limit, that is

$$BF_{12} \to \frac{\Gamma(n_R)\ \Gamma^{n_R}(\alpha)}{\Gamma(\alpha n_R)\ \prod_{j=1}^{n_R}(R_j - I_j)^{\alpha-1}} \times \left(\sum_{j=1}^{n_R}(R_j - I_j)\right)^{n_R(\alpha-1)}. \tag{6}$$

However, as $\sigma^2 \to 0$ the prior gets increasingly concentrated at μ, and the Bayes factor becomes more decisive in supporting m_1, that is $BF_{12} \to \infty$.

These results suggest that diffuse priors are more appropriate. Table 1 shows the results of a simulation exercise in which data sets were simulated under either m_1 or m_2, and in each case BF_{12} was calculated using (6). The resulting mean values, and the proportion of times that m_1 was favoured, are presented. It can be seen that the Bayes factor discriminates effectively between the two models, even in the relatively small population sizes of $N = 30$ and $N = 50$. As expected, increasing β increases the outbreak size which in turn makes the comparison more decisive.

3.2 The SIR model with different infection mechanisms

Consider now the situation where we have two competing SIR models with different infection mechanisms but the same infectious period distribution, the latter being essentially arbitrary. Specifically, m_1 is the standard SIR model in which infections occur at rate $\beta n^{-1} X(t)Y(t)$, and m_2 the model in which infections occur at rate $\beta n^{-1} X(t)Y^p(t)$. We assume $p \in (0, 0.5)$, so that m_1 and m_2 are clearly distinct, and that p is known. We

assume, *a priori*, that (i) $\beta \sim \text{Gamma}(\lambda_\beta, \nu_\beta)$, and (ii) the parameters of the infectious period distribution are independent of β.

Adopting the notation and arguments of the previous section leads to

$$BF_{12} = \prod_{j=1, j\neq z}^{n_R} Y^{1-p}(I_j-) \times \left(\frac{\nu_\beta + n^{-1} A_p}{\nu_\beta + n^{-1} A} \right)^{n_R + \lambda_\beta - 1}, \tag{7}$$

where $A = \int_{I_z}^{R_{n_R}} X(t)Y(t)\, dt$, $A_p = \int_{I_z}^{R_{n_R}} X(t)Y^p(t)\, dt$. As expected, the Bayes factor only involves the infection process part of the likelihoods since the removal processes in the two models are assumed to be the same. Rewriting the prior distribution for β in terms of its mean and variance, σ^2 say, we find that

$$BF_{12} \to (A_p/A)^{n_R - 1} \prod_{j=1, j\neq z}^{n_R} Y^{1-p}(I_j-) \text{ as } \sigma^2 \to \infty \tag{8}$$

and

$$BF_{12} \to \prod_{j=1, j\neq z}^{n_R} Y^{1-p}(I_j-) \text{ as } \sigma^2 \to 0.$$

It is natural to suppose that the diffuse prior setting is a natural candidate for consideration. It is evident from (8) that larger $Y(I_j-)$ values in the product term will improve model discrimination; conversely, if all these values equal 1 then the product term is independent of p. This in turn suggests that we require either larger or faster-growing epidemics to effectively discriminate between m_1 and m_2. This is illustrated by the results in Table 2, which shows that we require larger values of N than the infectious period distribution comparison of the previous section in order to obtain clear evidence in favour of the true model, and that increasing β also improves discrimination. Also, as expected, as p decreases then m_1 and m_2 become less similar, which also makes discrimination easier.

4 Model selection given incomplete outbreak data

We now consider the situation in which we observe removal times but not infection times, which in turn means that the Bayes factors of interest are no longer analytically tractable. In this section we describe how to apply the adapted power posterior methods in section 2.3 to the model comparison scenarios in sections 3.1 and 3.2. In both cases, Algorithm 1 requires an MCMC scheme that provides samples from the power posterior distribution for any given value of t_j. Since the infection times are unobserved, these are included as additional components of the posterior distribution. Thus the required MCMC algorithm may be specified by defining the updates for each of the model parameters, the details of which are given below.

We also briefly explore the performance of the power posterior methods and DIC_6, via simulation studies. It should be noted that such simulations are highly computationally expensive and time-consuming, since we require separate runs of an MCMC

True model	p	β	$E[\log BF_{12}]$ $N = 50$	$N = 200$	$P(BF_{12} > 1)$ $N = 50$	$N = 200$
m_1	0.5	2.0	0.5	6.6	0.48	0.72
m_1	0.3	2.0	1.7	15.7	0.58	0.77
m_1	0.0	2.0	4.4	39.7	0.67	0.77
m_1	0.5	4.0	2.9	18.5	0.77	0.92
m_1	0.3	4.0	6.1	41.5	0.85	0.94
m_1	0.0	4.0	15.0	98.2	0.92	0.94
m_2	0.5	2.0	-0.8	-1.3	0.09	0.06
m_2	0.3	2.0	-1.2	-1.4	0.07	0.05
m_2	0.0	2.0	-1.2	-1.5	0.06	0.05
m_2	0.5	4.0	-1.9	-4.9	0.06	0.02
m_2	0.3	4.0	-2.4	-5.4	0.06	0.02
m_2	0.0	4.0	-3.2	-5.2	0.03	0.02

Table 2: Expected log Bayes factors for models with different infection mechanisms, assuming diffuse prior distributions as in equation (8). Model m_1 is a standard SIR model while m_2 has a modified infection mechanism of the form $\beta n^{-1} X(t) Y^p(t)$. Both models have $T_I \sim \text{Exp}(1)$. Each row gives parameter values and results from 1000 simulated epidemics from the true model in which at least one new infection occurred.

algorithm for every single t_j value in the temperature ladder. Throughout we set $c = 5$ so that $t_j = (j/r)^5$. In all cases, including calculation of DIC_6, the results are based on MCMC runs of 27,000 iterations of which the first 2,000 were discarded as burn-in, and then thinned by taking every 5th value. Convergence and mixing were assessed visually, and found to be satisfactory.

4.1 The SIR model with different infectious periods

Recall the models and notation from section 3.1. The parameters β, γ and δ are assigned independent exponential prior distributions $\text{Exp}(\lambda_\zeta)$, where $\zeta = \beta,\ \gamma,\ \delta$. We assume *a priori* that the initial infection time satisfies $I_z = R_{min} - Y$, where $Y \sim \text{Exp}(\psi)$ and $R_{min} = \min\{R_1, \ldots, R_{n_R}\}$, and that the initial infective z is equally likely to be any of the n_R infected individuals.

At temperature t, the full conditional power posterior distributions for β, γ and δ are

$$\beta | t, \gamma, z, I_z, \boldsymbol{I}, \boldsymbol{R} \sim \text{Gamma}\left(1 + t(n_R - 1), \lambda_\beta + t n^{-1} A\right),$$

$$\gamma | t, \beta, z, I_z, \boldsymbol{I}, \boldsymbol{R} \sim \text{Gamma}\left(1 + n_R t, \lambda_\gamma + t \sum_{j=1}^{n_R} (R_j - I_j)\right),$$

$$\delta | t, \beta, z, I_z, \boldsymbol{I}, \boldsymbol{R} \sim \text{Gamma}\left(1 + t\alpha n_R, \lambda_\delta + t \sum_{j=1}^{n_R} (R_j - I_j)\right),$$

and the full conditional power posterior density for $(\boldsymbol{I}, z, I_z)$ under model m_2 is given by

$$\begin{aligned}\pi(\boldsymbol{I}, z, I_z|t, \alpha, \beta, \delta, \boldsymbol{R}) \quad &\propto \quad \left\{ \prod_{j=1, j\neq z}^{n_R} n^{-1} Y(I_j-) \times e^{-\beta n^{-1} A} \right\}^t \\ &\times \quad \left\{ \prod_{j=1}^{n_R} (R_j - I_j)^{\alpha-1} \times e^{-\delta \sum_{j=1}^{n_R} (R_j - I_j)} \right\}^t \times e^{\psi I_z},\end{aligned}$$

while the corresponding expression for model m_1 is obtained by setting $\alpha = 1$ and $\delta = \gamma$. An MCMC algorithm to update the model parameters and infection times then consists of (i) updating β, γ and δ according to their full conditional distributions, and (ii) updating infection times using a suitable Metropolis-Hastings step as in O'Neill and Roberts (1999); full details can be found in Alharthi (2016).

The DIC_6 calculation involves two steps. An initial MCMC run provides point estimates of the model parameters, the identity of the initial infective and the initial infection time. The model parameters and initial infective conditions are fixed for a second MCMC run in which the remaining infection times are allowed to vary. From each run we also obtain an estimate of the posterior mean of the augmented log-likelihood, from which DIC_6 can be computed according to (3).

Simulation study

First, we briefly verify that our proposed algorithm is capable of recovering the correct results when the latter are known, which is possible when the infection times are observed. Following this we revert to the situation where infection times are unobserved, and assess the impact of the temperature ladder (i.e. the set of t_j values in Algorithm 1), the within-model prior distribution, and the size of the observed epidemic on the calculation of Bayes factors. We set $\psi = 1$ in the prior distribution of I_z.

- **Algorithm verification when infection times are observed**

If both infection and removal times are observed then the results of section 3 give explicit formulae for the Bayes factors of interest. This enables us to test the power posterior approach in this scenario. Specifically, samples from the posterior distribution of (β, γ, δ) can be obtained directly from their full conditional distributions given in the first part of section 4.1 without the need for MCMC, replacing step b in Algorithm 1.

We simulated 100 data sets from model m_1 with $N = 50$, $\beta = 2$ and $\gamma = 1$. For model m_2 we set $\alpha = 2$. Model parameters were assigned independent Exp(1) prior distributions. For each data set we calculated BF_{12} using (5) and estimated it using Algorithm 1 with temperature ladder defined by $t_j = (j/20)^5$, $j = 0, \ldots, 20$. The estimates were extremely close to the true values, as illustrated in Figure 1.

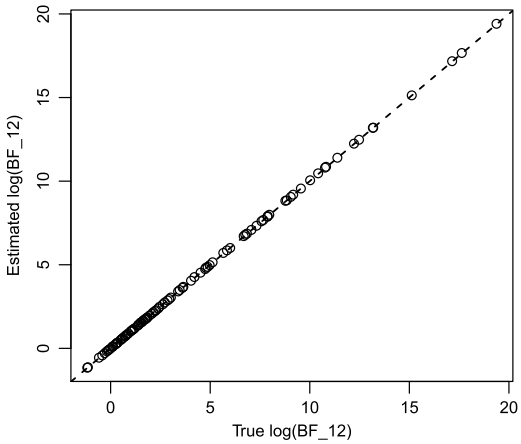

Figure 1: Comparison of 100 pairs of true and estimated log Bayes factors given data generated under the SIR model (m_1) with $T_I \sim \text{Exp}(1)$, $N = 50$ and $\beta = 2$, where model m_2 has $T_I \sim \text{Gamma}(2, \delta)$. Model parameters were assigned independent Exp(1) prior distributions. The co-ordinates of each circle give the two log Bayes factor values. The dashed line shows equality of the co-ordinates.

r	$\log(\pi(\boldsymbol{R}\|m_1))$	$\log(\pi(\boldsymbol{R}\|m_2))$	$\log(BF_{12})$
$\beta, \gamma, \delta \sim \text{Exp}(1)$			
10	-129.86	-152.25	22.39
20	-130.01	-150.87	20.86
40	-130.05	-151.15	21.11
100	-130.04	-150.24	20.19
$\beta, \gamma, \delta \sim \text{Exp}(0.01)$			
10	-131.97	-150.71	18.74
20	-137.07	-151.98	14.91
40	-137.49	-153.04	15.56
100	-137.39	-152.36	14.98

Table 3: Estimates of $\log(\pi(\boldsymbol{R}|m_1))$, $\log(\pi(\boldsymbol{R}|m_2))$ and $\log(BF_{12})$ using data simulated from the standard SIR model with $T_I \sim \text{Exp}(\gamma)$ (m_1), while model m_2 has $T_I \sim \text{Gamma}(\alpha, \delta)$. Parameter values were $N = 30$, $\beta = 1$ and $\gamma = 0.5$, and $n_R = 22$ individuals were infected. The parameters β, γ and δ were assigned Exp(1) (top table) and Exp(0.01) (bottom table) prior distributions.

- **Length of temperature ladder**

Table 3 shows the impact of the number of t_j values, r, on the marginal likelihoods and corresponding Bayes factor BF_{12}. The results are based on a single, but fairly typical, data set simulated from model m_1 with $N = 30$, $\beta = 1$, $\gamma = 0.5$ and in which $n_R = 22$ individuals were infected in total. In model m_2, $\alpha = 10$. Two choices of prior distribution are illustrated. The results show that the estimates are relatively insensitive to the value

of r, although as expected more t_j values are required as the prior distribution becomes more diffuse. Figure 2 shows typical expected log-likelihood curves, and illustrates the sharp change near $t = 0$ which motivates the choice of temperature ladder $t_j = (j/r)^c$. Further details regarding Figure 2 are given below.

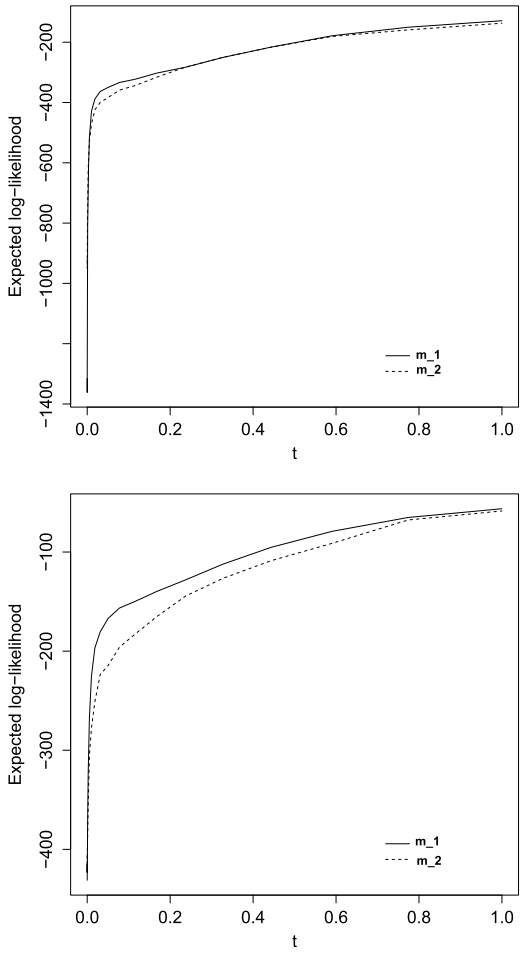

Figure 2: Expected log-likelihood curves given data generated under the SIR model with $T_I \sim \text{Exp}(\gamma)$ (m_1, top plot) and under the SIR model with $T_I \sim \text{Gamma}(\alpha = 10, \delta)$ (m_2, bottom plot). In both cases β, γ and δ were assigned independent Exp(1) prior distributions.

- **Choice of prior distribution**

It is well known that Bayes factors can exhibit strong dependence on the model parameter prior distributions. Here, we explore this issue via two simulated data sets from the

two models under consideration. In both cases the data set itself was fairly typical of epidemics that did not die out quickly. For model m_1 we set $\beta = 2$, $\gamma = 1$ and $N = 50$ and obtained $n_R = 41$ infected individuals. For model m_2 we set $\beta = 2$, $\alpha = 10$, $\delta = 10$ and $N = 30$, and obtained $n_R = 22$. Prior distributions for β, γ and δ were set to be $\text{Exp}(1), \text{Exp}(0.1)$ and $\text{Exp}(0.01)$, for which we used r values of 20, 20 and 40 respectively, inspired by our previous findings regarding the length of temperature ladder.

Results of the simulation study are summarised in Table 4. The expected log-likelihood curves for both SIR models are shown in Figure 2 in the case when the prior distributions are Exp(1).

Model	Prior	$\log(BF_{12})$	DIC_6	
			m_1	m_2
m_1	Exp(1)	22.48	**276.21**	291.01
m_1	Exp(0.1)	10.00	261.25	**248.04**
m_1	Exp(0.01)	9.27	261.30	**247.73**
m_2	Exp(1)	14.53	**123.60**	127.76
m_2	Exp(0.1)	−0.03	115.02	**105.07**
m_2	Exp(0.01)	−0.89	114.63	**102.71**

Table 4: Estimates of $\log(BF_{12})$ and DIC_6 using data simulated from the SIR model with $T_I \sim \text{Exp}(\gamma)$ (m_1; $N = 50$, $\beta = 2$, $\gamma = 1$ and $n_R = 41$ infections) and the SIR model with $T_I \sim \text{Gamma}(\alpha, \delta)$ (m_2; $N = 30$, $\beta = 2$, $\alpha = \delta = 10$, and $n_R = 22$). Bold values for DIC_6 indicate the preferred model. In all cases β, γ and δ were assigned identical independent prior distributions as indicated.

When model m_1 is the true model, the results in Table 4 indicate, in general, that $\log(BF_{12})$ values support the true model for all prior distributions with a noticeable decrease as the prior distribution becomes more diffuse. These findings are in harmony with the behaviour of the Bayes factor for complete data case described in section 3.1. When model m_2 is the true model, the value of $\log(BF_{12})$ varies quite dramatically with different prior distributions. As the latter become more diffuse then m_2 is identified correctly, whereas using an Exp(1) prior gives the opposite conclusion. There is some intuition to explain this conflict, as follows. The true value of the parameter δ used in the simulation is 10, so an Exp(1) prior is a strong prior distribution which is in conflict with the data and in turn results in poor posterior mean estimates for both β and δ, namely $\hat{\beta} = 1.286$ and $\hat{\delta} = 6.629$. This may in turn yield lower values of $\log(\pi(\boldsymbol{R}|m_2))$ and thus m_2 is not identified correctly. However, with Exp(0.1) and Exp(0.01) priors, the estimation was improved giving $\hat{\beta} = 2.053, \hat{\delta} = 11.490$ and $\hat{\beta} = 2.184, \hat{\delta} = 12.133$, respectively, and consequently the correct identification of model m_2 was obtained. These results highlight the sensitivity of Bayes factors to prior distributions, but also suggest that in this situation diffuse priors are likely to be more appropriate.

The values of DIC_6 are also prior-dependent. More seriously, model m_2 is preferred as the prior distributions become more diffuse, which suggests that DIC_6 is not a suitable tool for model discrimination in this setting.

- **Size of outbreak**

It is natural to suppose that, given a small epidemic outbreak, it is hard to effectively distinguish between competing models. Here we show that even small outbreaks can be informative in the setting where we compare infectious period distributions.

We simulated, under model m_1 with $N = 50$, $\beta = 1.15$ and $\gamma = 1$, 20 datasets of size $n_R = 5$ removals each. For each data set we calculated $\log(BF_{12})$ for the complete data (infection and removal times), and for incomplete data (removal times), the latter using $r = 20$, assuming that $\alpha = 10$ in model m_2. Two different prior distributions were used for β, γ and δ, namely Exp(1) and Exp(0.01). The results are summarised Figure 3.

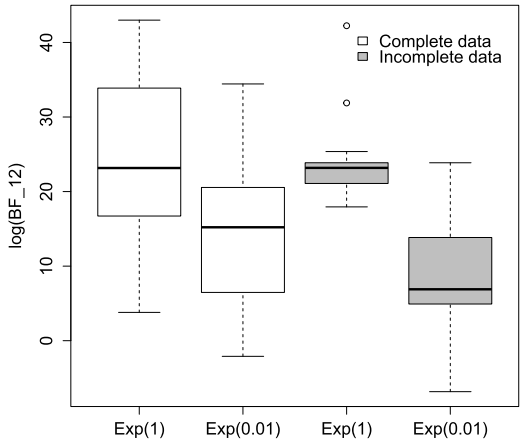

Figure 3: Boxplots of $\log(BF_{12})$ values calculated using 20 simulated data sets of $n_R = 5$ removals each simulated under the SIR model with $T_I \sim \text{Exp}(\gamma)$ (m_1), while m_2 has $T_I \sim \text{Gamma}(\alpha = 10, \delta)$. The $\log(BF_{12})$ values were computed using (5) and the missing-data power posterior method for complete and incomplete data respectively. The parameters β, γ and δ were assigned two choices of prior distribution, namely Exp(1) and Exp(0.01).

Interestingly, the results are not as one might expect with a small outbreak data set, giving decisive support to the true model m_1 under both complete and incomplete data. These findings suggest that, in this particular epidemic setting, a few infectious periods of infected individuals might be enough for the Bayes factor criterion to favour the SIR model with exponential infectious period over the SIR model with gamma infectious period. The impact of the parameter prior distributions is also evident which is in agreement with our previous findings. Note also that there is more variance in the 20 Bayes factors for complete than incomplete data. This is likely to be a consequence of the fact that calculating the marginal likelihoods of the removal times in the incomplete data setting inherently involves averaging over the unobserved infection times, and so the variation one obtains when observing the infection times is removed. This in turn impacts the variability of the resulting Bayes factors.

4.2 The SIR model with different infection mechanisms

We now consider the models in section 3.2, so that m_1 is the standard SIR model and m_2 has infection rate $\beta n^{-1} X(t) Y^p(t)$. Prior distributions are assigned as in section 4.1, and we also set $p \sim U(0, 0.5)$ *a priori*.

For model m_2 at temperature t, we have

$$\beta | t, \gamma, p, I_z, z, \boldsymbol{I}, \boldsymbol{R} \sim \text{Gamma}\left(1 + t(n_R - 1), \lambda_\beta + tn^{-1} \int_{I_z}^{R_{n_R}} X(t) Y^p(t) dt\right),$$

$$\gamma | t, \beta, p, I_z, z, \boldsymbol{I}, \boldsymbol{R} \sim \text{Gamma}\left(1 + n_R t, \lambda_\gamma + t \int_{I_z}^{R_{n_R}} Y(t) dt\right).$$

Conditional densities for p and $\mathbf{I}$ are given by

$$\pi(p | t, \beta, \gamma, I_z, z, \boldsymbol{I}, \boldsymbol{R}) \propto \left\{ \left(\prod_{j=1, j \neq z}^{n_R} Y^p(I_j -) \right) \times \exp\left(-\beta n^{-1} \int_{I_z}^{R_{n_R}} X(t) Y^p(t) dt \right) \right\}^t,$$

$$\begin{aligned} \pi(\boldsymbol{I}, I_z, z | t, \beta, \gamma, p, \boldsymbol{R}) \quad \propto \quad & \left\{ \prod_{j=1, j \neq z}^{n_R} Y^p(I_j -) \right\}^t \\ \times \quad & \left\{ \exp\left(- \int_{I_z}^{R_{n_R}} \left(\beta n^{-1} X(t) Y^p(t) + \gamma Y(t) \right) dt \right) \right\}^t \times e^{\psi I_z}. \end{aligned}$$

The corresponding expressions for model m_1 can be obtained by setting $p = 1$.

Simulation study

We consider the factors described in section 4.1, and again set $\psi = 1$.

- **Algorithm verification when infection times are observed**

If both infection and removal times are observed, then posterior samples for β, γ and p can be obtained using the MCMC algorithm described in section 4.2, omitting the step in which infection times are updated. We simulated 100 data sets from model m_1 with $N = 50$, $\beta = 2$ and $\gamma = 1$. Model parameters β and γ were assigned independent Exp(1) prior distributions and p assigned an independent $U(0, 0.5)$ prior distribution. For each data set we calculated

$$BF_{12} = \frac{\int_\gamma \int_\beta \pi(\boldsymbol{I}, \boldsymbol{R} | \beta, \gamma) \pi(\beta) \pi(\gamma) d\beta \, d\gamma}{\int_p \int_\gamma \int_\beta \pi(\boldsymbol{I}, \boldsymbol{R} | \beta, \gamma, p) \pi(\beta) \pi(\gamma) \pi(p) d\beta \, d\gamma \, dp}$$

by finding the numerator analytically in a similar fashion to the arguments leading to (4), and evaluating the denominator by analytically integrating with respect to β and γ and then numerically integrating with respect to p over the range $p = 0$ to $p = 0.5$,

with $\pi(p) = 2$. We estimated BF_{12} using Algorithm 1 with temperature ladder defined by $t_j = (j/20)^5$, $j = 0, \ldots, 20$. Figure 4 shows excellent agreement between the true and estimated values.

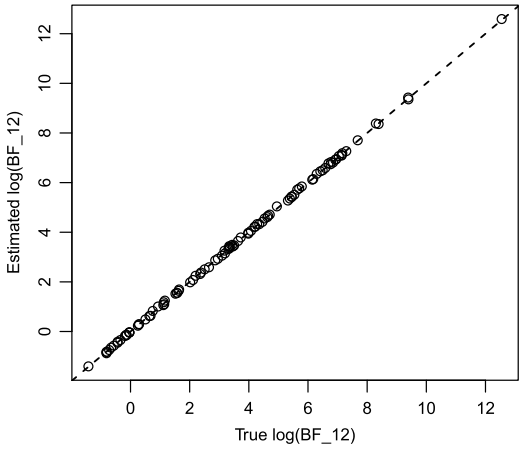

Figure 4: Comparison of 100 pairs of true and estimated log Bayes factors given data generated under the SIR model (m_1) with $T_I \sim \text{Exp}(1)$, $N = 50$ and $\beta = 2$, where model m_2 has infection rate $\beta n^{-1} X(t) Y^p(t)$. Model parameters were assigned independent prior distributions: $\beta, \gamma \sim \text{Exp}(1)$ and $p \sim U(0, 0.5)$. The co-ordinates of each circle give the two log Bayes factor values. The dashed line shows equality of the co-ordinates.

- **Length of the temperature ladder**

Table 5 shows how estimates of marginal likelihoods and Bayes factors vary with r. The results are based on a simulation from model m_2 in which $N = 100$, $\beta = 2$, $\gamma = 0.2$ and $p = 0.3$ which resulted in $n_R = 87$ infections in total. Our findings are the same as those described in section 4.1, namely that as the prior distributions become more diffuse, more temperatures are required to provide accurate estimates.

- **Choice of prior distribution**

We simulated one data set from each model, in both cases fairly typical. For model m_1 we set $N = 100$, $\beta = 0.5$, $\gamma = 0.2$ and obtained $n_R = 83$ infections. For model m_2 we set $N = 100$, $\beta = 2.5$, $\gamma = 0.2$ and $p = 0.3$ and obtained $n_R = 88$. Table 6 shows estimates of $\log(BF_{12})$ and DIC_6 under three choices of prior distribution, where we used r values of 20, 40 and 40 for the $\text{Exp}(1), \text{Exp}(0.1)$ and $\text{Exp}(0.01)$ prior distributions, respectively.

Results of simulations are displayed in Table 6. Again there is evidence of sensitivity of BF_{12} to the choice of prior, although the results themselves show that the correct model is identified in all cases. Furthermore DIC_6 also performs well in this setting.

Figure 5 shows plots of the expected log-likelihoods against the temperature t for the two simulated data sets. In contrast to Figure 2, here the curves for m_1 and m_2

r	$\log(\pi(\boldsymbol{R}\|m_1))$	$\log(\pi(\boldsymbol{R}\|m_2))$	$\log(BF_{12})$
$\beta, \gamma \sim \text{Exp}(1)$			
10	-113.35	-104.29	-9.06
20	-114.05	-106.07	-7.99
40	-114.09	-105.76	-8.33
100	-113.80	-105.87	-7.93
$\beta, \gamma \sim \text{Exp}(0.01)$			
10	-81.087	-85.72	4.63
20	-120.70	-111.03	-9.67
40	-122.22	-111.59	-10.63
100	-122.28	-112.11	-10.17

Table 5: Estimates of $\log(\pi(\boldsymbol{R}|m_1))$, $\log(\pi(\boldsymbol{R}|m_2))$ and $\log(BF_{12})$ using data simulated from the SIR model with modified infection rate (m_2; $N = 100$, $\beta = 2$, $\gamma = 0.2$, $p = 0.3$ and $n_R = 87$ infections), while m_1 is the standard SIR model. The parameters β and γ were assigned Exp(1) (top table) and Exp(0.01) (bottom table) prior distributions.

Model	Prior	$\log(BF_{12})$	DIC_6	
			m_1	m_2
m_1	Exp(1)	3.49	**51.49**	68.89
m_1	Exp(0.1)	1.94	**52.08**	68.09
m_1	Exp(0.01)	2.10	**51.74**	68.66
m_2	Exp(1)	-8.03	76.57	**68.57**
m_2	Exp(0.1)	-9.79	76.98	**69.94**
m_2	Exp(0.01)	-10.70	77.37	**69.60**

Table 6: Estimates of $\log(BF_{12})$ and DIC_6 using data simulated from the standard SIR model (m_1;$N = 100$, $\beta = 0.5$, $\gamma = 0.2$ and $n_R = 83$) and the SIR model with modified infection rate (m_2; $N = 100$, $\beta = 2.5$, $\gamma = 0.2$, $p = 0.3$ and $n_R = 88$). Bold values for DIC_6 indicate the preferred model. In all cases β, γ and δ were assigned identical independent prior distributions as indicated.

are much closer together, indicating that it is harder to effectively discriminate between models with different infection mechanisms than with different infectious period, at least for the settings we have considered.

- **Size of outbreak**

We consider two scenarios for model m_1, corresponding to small and large epidemics. We simulated 20 epidemics for each, with the same number of removals n_R. In the first scenario $N = 50$, $\beta = 1.15$, $\gamma = 1$ and $n_R = 7$. In the second $N = 50$, $\beta = 2$, $\gamma = 1$ and $n_R = 42$. We set $p = 0.3$ in model m_2. Bayes factor calculations were carried out using $r = 20$, and under two different prior distribution assumptions for β and γ.

Figures 6 and 7 illustrate the results. In contrast to the comparison of infectious period distributions, here we see that small outbreaks may not be sufficient to differentiate between models, which again agrees with our earlier findings that this is an

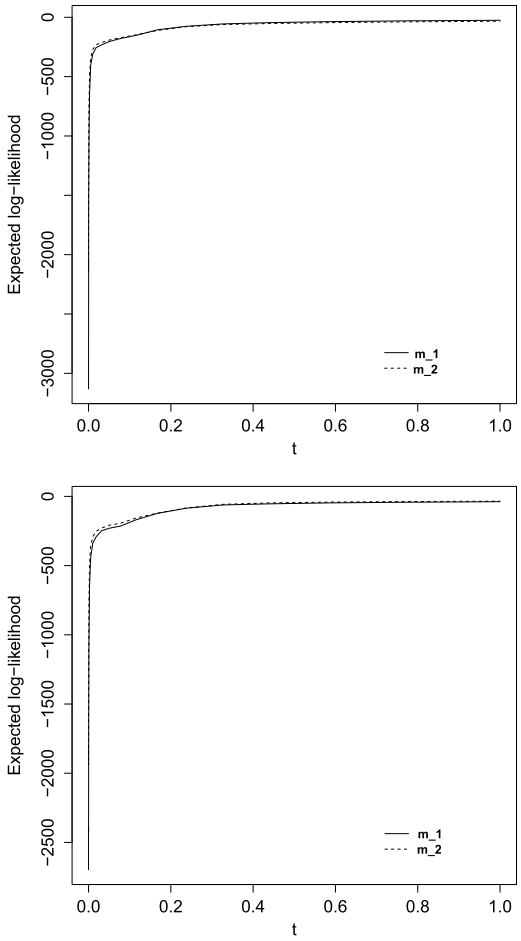

Figure 5: Expected log-likelihood curves given data generated under the standard SIR model (m_1, top plot) and under the SIR model with modified infection rate (m_2, bottom plot). In both cases β and γ are assigned independent Exp(1) prior distributions.

inherently harder problem requiring more data. However, even larger outbreaks appear problematic. The most likely reason for this is that the key to differentiating between m_1 and m_2 is the number of infected individuals present in the population at the time of each infection, as discussed in section 3.2, as opposed to the outbreak size itself. This suggests that data from epidemics that grow quickly would be required to distinguish the competing models, at least for moderate population sizes.

To explore this further, we simulated 20 outbreaks of size $n_R = 47$ from model m_1 with $N = 50$, $\beta = 5$ and $\gamma = 1$. Thus $\beta/\gamma = 5$, in contrast to the value of 2 shown in Figure 7. Figure 8 shows the resulting histogram of $\log(BF_{12})$ values from which we see that the evidence in favour of the true model m_1 is much clearer.

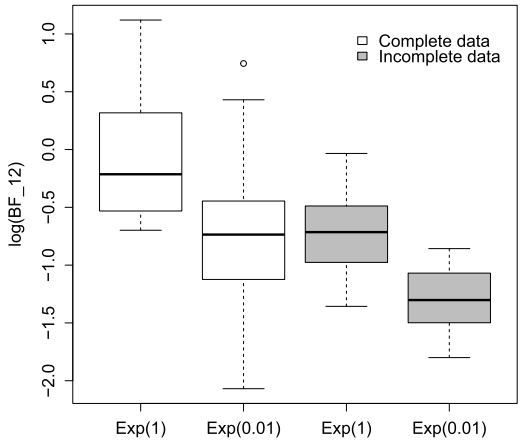

Figure 6: Boxplots of $\log(BF_{12})$ values calculated using 20 simulated data sets of $n_R = 7$ removals each simulated under the standard SIR model (m_1), while m_2 has modified infection rate $\beta n^{-1} X(t) Y^p(t)$, $p = 0.3$. The $\log(BF_{12})$ values were computed using (7) and the missing-data power posterior method for complete and incomplete data respectively. The parameters β, γ and δ were assigned two choices of prior distribution, namely Exp(1) and Exp(0.01).

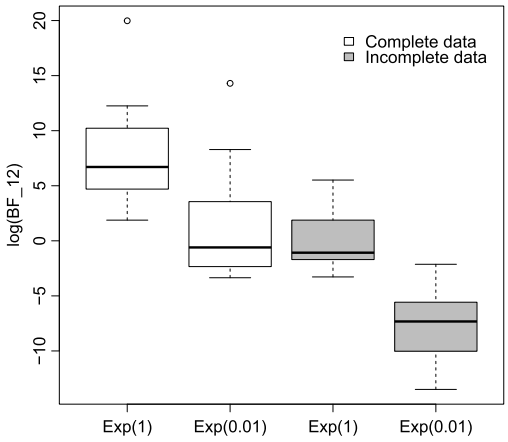

Figure 7: Boxplots of $\log(BF_{12})$ values calculated using 20 simulated data sets of $n_R = 42$ removals each simulated under the standard SIR model (m_1), while m_2 has modified infection rate $\beta n^{-1} X(t) Y^p(t)$, $p = 0.3$. The $\log(BF_{12})$ values were computed using (7) and the missing-data power posterior method for complete and incomplete data respectively. The parameters β, γ and δ were assigned two choices of prior distribution, namely Exp(1) and Exp(0.01).

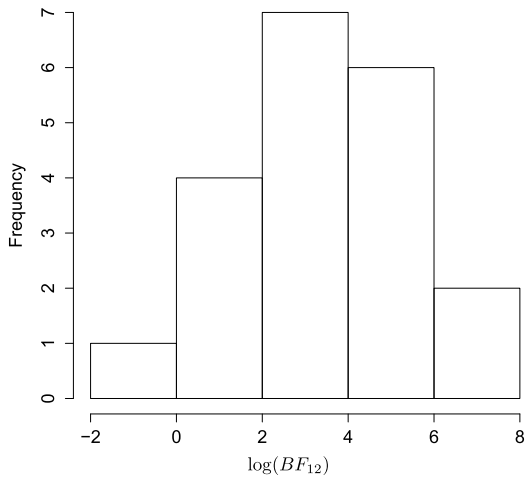

Figure 8: Histogram of the estimated $\log(BF_{12})$ values obtained via the missing-data power posterior approach using 20 simulated data sets of $n_R = 47$ removal times each simulated under the standard SIR model (m_1), while m_2 has modified infection rate $\beta n^{-1} X(t) Y^p(t)$, $p = 0.3$. The parameters β and γ were assigned independent Exp(1) prior distributions.

4.3 Abakaliki smallpox data

We now briefly consider a widely-studied temporal data set obtained from a smallpox outbreak that took place in the Nigerian town of Abakaliki in 1967. The outbreak resulted in 32 cases, 30 of whom were members of a religious organisation whose 120 members refused vaccination. Numerous authors have considered these data by focussing solely on the 30 cases among the population of 120, and the time series of symptom-appearance times which in our notation is given by

$$\begin{aligned} \boldsymbol{R}^{obs} &= (0, 13, 20, 22, 25, 25, 25, 26, 30, 35, 38, 40, 40, 42, 42, 47, 50, 51, 55, 55, 56, \\ &\quad 57, 58, 60, 60, 61, 66, 66, 71, 76), \end{aligned}$$

where to set a time scale we set $R_1 = 0$. In fact, the original data set includes far more information, particularly on the locations of the homes of the cases and the other members of their households. Analyses of this full data set can be found in Eichner and Dietz (2003) and Stockdale et al. (2017).

Here our purpose is to illustrate our model comparison methods, and so we only consider the partial data set, specifically assuming that the 30 symptom-appearance times correspond to removals in an SIR model. Several authors have previously considered departures from the standard SIR model for these data, and in particular both Becker and Yip (1989) and Xu (2015) considered models in which the infection rate was allowed to vary through time. Here we address this issue by comparing the standard

SIR model, m_1, with a model (m_2) in which the infection rate parameter β is replaced by $\beta(t) = \beta n^{-1} e^{-bt}$, with $b > 0$ a model parameter to be estimated from the data. It is straightforward to modify our methods to this situation.

Results are presented in Table 7 for different prior distributions for β and γ for both SIR models, where b and $R_1 - I_1$ are assigned Exp(1) prior distributions throughout. Figure 9 shows plots of the expected log-likelihoods against the temperature t when the prior distribution is Exp(1), using a temperature ladder with $t_j = (j/r)^5$, $j = 0, \ldots, 20$ and $r = 20$.

Prior	$\log(BF_{12})$	DIC_6	
		m_1	m_2
Exp(1)	-0.51	-105.8	-105.9
Exp(0.01)	-0.86	-105.8	-105.8

Table 7: Results of $\log(BF_{12})$ and DIC_6 for smallpox data based on the standard SIR model (m_1) and the modified SIR model (m_2). For each case, the parameters β and γ were assigned prior distributions as indicated.

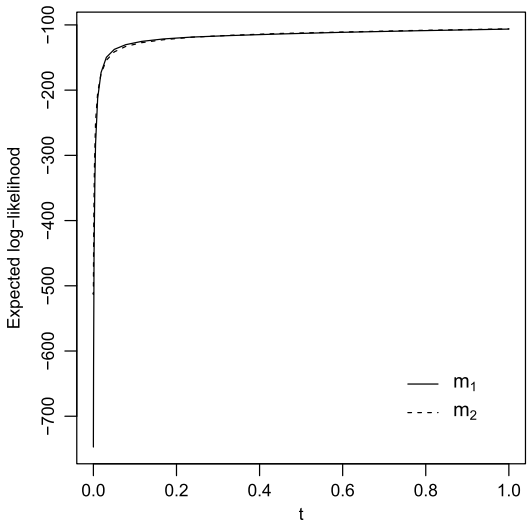

Figure 9: Smallpox data: expected log-likelihood curves calculated using the adapted power posterior approach when $\beta, \gamma \sim \text{Exp}(1)$ *a priori*.

From Table 7, there is clearly little to choose between the two models. This comparison is borne out from parameter estimation from model m_2, specifically Figure 10 which shows that b is close to zero. These findings are in keeping with those in Xu et al. (2016), in which a Bayesian nonparametric time-varying estimate of the infection rate parameter was found to be fairly close to constant over time.

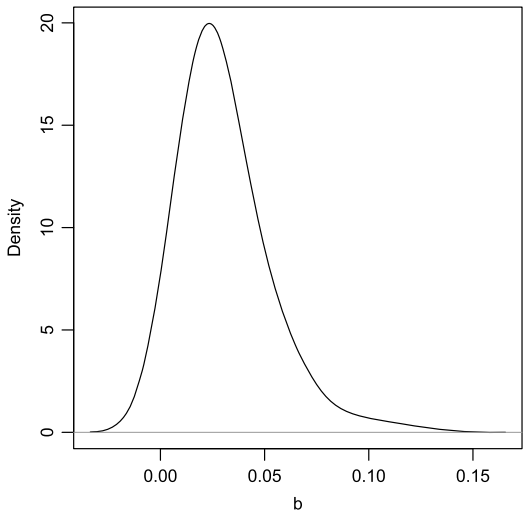

Figure 10: Smallpox data: posterior density of b, when $\beta, \gamma \sim \text{Exp}(1)$ *a priori*.

4.4 Hagelloch measles data

Our second data set describes an historical outbreak of measles in 1861 in the German village of Hagelloch, as described in Pfeilsticker (1861). The outbreak was very severe, as every one of 188 individuals deemed to be susceptible became infected, these individuals all being children. These data have been considered by a number of authors (see e.g Britton et al., 2011, and references therein), and were specifically analysed in a model choice context in Neal and Roberts (2004), where the authors used reversible-jump MCMC methods to evaluate Bayes factors for a number of competing models. In view of our earlier discussion on the possible sensitivity of Bayes factors to within-model prior distributions, it is natural to explore this further for the Hagelloch data.

The data themselves are unusually detailed, consisting for each case of the name, age, sex, date of symptom onset, date of rash onset, class of child in the village school, date of death if this occurred, location of the child's home and several other covariates. We shall adopt the transmission models described in Neal and Roberts (2004), defined as follows.

Each individual belongs to a household and the community at large, and either attends school or is of pre-school age. If an individual becomes infected, they first undergo a symptom-free infectious period T_S which is assumed to be one of two models: either (i) a fixed length of one day, so $T_S = 1$, or (ii) $T_S \sim Gamma(30, \delta)$. At the end of this period, the individual displays symptoms and subsequently develops a rash. The dates of both symptom and rash appearance are given by the data and so we do not model the time between these events. Following the rash appearance, the individual is assumed to be removed three days later, unless they die first (as indicated in the data).

The infectious period is assumed to start upon initial infection, and finish at either removal or death.

While infectious, individual i has infectious contacts with susceptible individual j at rate α_{ij} where α_{ij} depends on the relationship of i and j. Neal and Roberts (2004) first considered a model in which

$$\alpha_{ij} = \beta_H \, 1_{\{\rho(i,j)=0\}} + \beta_C^1 \, 1_{\{L_i=L_j=1\}} + \beta_C^2 \, 1_{\{L_i=L_j=2\}} + \beta_G \exp\{-\theta\rho(i,j)\},$$

where (i) 1_A denotes the indicator function of the event A, (ii) $\rho(i,j)$ denotes the Euclidean distance between the households of individuals i and j, (iii) L_i denotes the school classroom (either 1 or 2) which individual i belongs to, and (iv) $L_i = 0$ if individual i is of pre-school age. The non-negative parameters β_H, β_C^1 and β_C^2 denote the within-household, within-classroom 1 and within-classroom 2 infection rates respectively. Finally, β_G denotes a global infection rate while θ governs the extent to which distance between individuals reduces this infection rate. Neal and Roberts (2004) called the model described above the *full model*, denoted by M.

In order to investigate the relative importance of different transmission routes, Neal and Roberts (2004) considered four other models, each of which is a simplified version of the full model. Specficially, they set each of the parameters $\theta, \beta_H, \beta_C^1$ and β_C^2 to zero in turn in M, yielding models $M[\theta]$, $M[\beta_H]$, $M[\beta_C^1]$ and $M[\beta_C^2]$, say, respectively. Thus for example, $M[\beta_H]$ assumes that

$$\alpha_{ij} = \beta_C^1 \, 1_{\{L_i=L_j=1\}} + \beta_C^2 \, 1_{\{L_i=L_j=2\}} + \beta_G \exp\{-\theta\rho(i,j)\}.$$

Finally, for each parameter ψ define $BF[\psi] = BF_{M,M[\psi]}$, so for example $BF[\theta] = BF_{M,M[\theta]}$ is the Bayes factor for the full model relative to the model in which $\theta = 0$.

We employed an MCMC algorithm which targets the joint power posterior distribution of the parameters $(\theta, \beta_H, \beta_C^1, \beta_C^2$ and β_G when $T_S = 1$, and additionally δ and the unknown infection times when $T_S \sim \text{Gamma}(30, \delta))$ given the observed data. None of the full conditional power posterior distributions for the model parameters are standard, and so parameters were updated using Metropolis-Hastings steps. The variances were tuned manually to achieve an acceptance rate approximately between 25% and 40%. To estimate the marginal likelihood via the power posterior method we used a temperature ladder such that $t_j = (j/r)^5$, where $j = 0, \ldots, 100$ and $r = 100$. Convergence and mixing was assessed visually, and found to be satisfactory.

Results are given in Table 8 for different prior distributions for the model parameters. In Neal and Roberts (2004), an Exp(0.1) prior distribution is assigned to the parameter θ and an Exp(10) to the parameters $\beta_H, \beta_C^1, \beta_C^2$ and β_G. We also consider two alternative sets of prior distributions, specifically assigning an Exp(1) or an Exp(0.1) distribution to *all* the parameters in each model.

The results illustrate a certain amount of sensitivity to both the T_S model and the chosen prior distributions. Regarding the former, the posterior mean of δ was found to be around 4, so that when infection times are not assumed to be known the length of the symptom-free period was estimated to be around 7 days. This is in stark contrast

to the assumption that T_S is one day, and goes some way to explaining the differences between the Bayes factors for the two model scenarios. Note also that our estimate of δ is almost certainly connected to the fact that measles actually has a latent period of around 7–10 days, a feature missing from the SIR model we consider; roughly speaking, increasing the infectious period length is one way to account for the time taken for the whole outbreak. We also briefly investigated what happens if T_S was set equal to integer values from 2 to 7 days, and found that the average posterior log-likelihood increased accordingly, which also supports the view that the $T_S = 1$ model is less appropriate for these data than $T_S \sim \text{Gamma}(30, \delta)$.

Regarding the Bayes factors in Table 8, the only robust conclusion appears to be that there is strong evidence that $\beta_C^1 > 0$, itself a key finding from Neal and Roberts (2004). For the $T_S \sim \text{Gamma}(30, \delta)$ model there is also strong support for the hypotheses that $\beta_C^2 > 0$ and that $\theta = 0$. These findings make sense, since as illustrated in Britton et al. (2011) it is clear from the raw data that the epidemic moves through the two school classes in turn, suggesting that classroom transmission is important, while there is no apparent evidence of purely spatial transmission.

$T_S = 1$	Priors	$\log(BF[\theta])$	$\log(BF[\beta_H])$	$\log(BF[\beta_C^1])$	$\log(BF[\beta_C^2])$
	NR	-0.03	5.77	33.46	6.72
	Exp(1)	0.78	9.50	30.87	3.82
	Exp(0.1)	-0.64	0.99	28.62	1.89
$T_S \sim \text{Gamma}(30, \delta)$	Priors	$\log(BF[\theta])$	$\log(BF[\beta_H])$	$\log(BF[\beta_C^1])$	$\log(BF[\beta_C^2])$
	NR	-2.72	1.73	70.97	12.71
	Exp(1)	-10.56	18.05	89.77	12.79
	Exp(0.1)	-28.34	-8.23	56.77	20.93

Table 8: Estimates of log Bayes Factors for the Hagelloch data for different models for the symptom-free period T_S and different within-model prior assumptions. NR refers to the priors used in Neal and Roberts (2004).

5 Conclusions

5.1 Advantages and limitations of the methods

We have described a method for computing Bayes factors for epidemic models, specifically by utilising an adaption of the power posterior method to accommodate missing data of a certain kind. The methods appear to work reasonably well in practice. Although we have focussed on single population SIR models, the methods can be applied to more complex epidemic models, as illustrated by our examples featuring the Abakaliki smallpox data and the Hagelloch measles data. In principle at least, the methods could be applied to a wide range of epidemic models and scenarios, including network epidemic models, multi-type models, and models with different levels of mixing such as household models. The procedures are relatively easy to implement, and lend themselves to parallel computation.

In common with the original power posterior method itself, the main limitation with our methods is one of computational efficiency, specifically that it can be quite time-consuming to perform the numerous MCMC runs required for the method. In addition to this, it is likely that the methods would struggle to analyse data on large outbreaks where thousands of individuals are infected. This is because even the basic problem of inferring the model parameters is hard in this setting when using MCMC methods, due to the strong correlations between the model parameters and the missing data (Kypraios, 2007), and matters will be made worse by the need to carry out many MCMC runs. An additional setting in which our methods may struggle is when very diffuse prior distributions are used for the underlying model parameters. This is because such settings typically require a greater number of temperatures in the temperature ladder, with corresponding reductions in the overall computation speed.

5.2 Practical considerations

For typical choices of the prior distribution of the model parameters, such as exponential distributions with low rates, the power posterior becomes flatter as the temperature t decreases towards zero. This in turn means that the MCMC chains for temperatures near zero may exhibit worse mixing properties than those with higher temperatures, simply because the target distribution is more spread out. It is therefore of particular benefit to monitor the mixing of low temperature chains, for example by looking at trace plots for the log-likelihood. It can also be useful to tune any Metropolis-Hastings proposal distributions in the algorithms according to the temperature.

The particular choice of epidemic model, and the nature of the data, can also impact MCMC mixing and the length of chain that is required. For example, reducing the variability of the infectious period in the epidemic model will typically lead to better mixing because there is less inherent uncertainty in the distribution of the unknown infection times. Similarly, larger data settings can also be problematic, as mentioned above.

We have restricted attention to temperature ladders of the form $t_j = (j/r)^c$, but other possibilities exist. For example, Friel et al. (2014) describe a method for sequentially choosing t_{j+2} given t_j and t_{j+1} and the associated estimates of the log-likelihood curve and its variance.

5.3 Analytical results

We have also demonstrated that Bayes factors can be obtained analytically for situations where infection times are known. This in turn enables us to explore the impact of different prior assumptions analytically. Although the infection process is not usually observed in real-life settings, if the infectious period does not exhibit that much variation then we can approximate it by a fixed value. Under this assumption the infection times would be known, which could enable explicit evaluation of Bayes factors for comparing different possible infection mechanisms or routes of infection, depending on the problem in question.

5.4 Future directions

Finally, the power posterior methods that we have adapted are not the only approach to estimating marginal likelihoods. Another related approach that may prove fruitful is the use of stepping-stone algorithms (Xie et al., 2011; Fan et al., 2011), as may other variants of path-sampling algorithms.

References

Alharthi, M. (2016). "Bayesian model assessment for stochastic epidemic models." Ph.D. thesis, University of Nottingham. URL http://eprints.nottingham.ac.uk/id/eprint/33182 12

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and their Statistical Analysis*. New York: Springer. MR1784822. doi: https://doi.org/10.1007/978-1-4612-1158-7. 2

Atkinson, K. and Han, W. (2004). *Elementary Numerical Analysis*. New York: Wiley, 3rd edition. MR2153422. doi: https://doi.org/10.1007/978-0-387-28769-0. 5

Becker, N. and Yip, P. (1989). "Analysis of variations in an infection rate." *Australian Journal of Statistics*, 31(1): 42–52. MR1014889. 22

Britton, T., Kypraios, T., and O'Neill, P. D. (2011). "Inference for epidemics with three levels of mixing: methodology and application to a measles outbreak." *Scandinavian Journal of Statistics*, 38(3): 578–599. MR2833848. doi: https://doi.org/10.1111/j.1467-9469.2010.00726.x. 24, 26

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). "Deviance information criteria for missing data models." *Bayesian Analysis*, 1(4): 651–673. MR2282197. doi: https://doi.org/10.1214/06-BA122. 6

Deeth, L., Deardon, R., and Gillis, D. J. (2015). "Model choice using the Deviance Information Criterion for latent conditional individual-level models of infectious disease spread." *Epidemiologic Methods*, 4(1): 47–68. MR3100975. doi: https://doi.org/10.1515/ijb-2013-0026. 2

Eichner, M. and Dietz, K. (2003). "Transmission potential of smallpox: estimates based on detailed data from an outbreak." *American Journal of Epidemiology*, 158(2): 110–117. 22

Fan, Y., Wu, R., Chen, M.-H., Kuo, L., and Lewis, P. O. (2011). "Choosing among partition podels in Bayesian Phylogenetics." *Molecular Biology and Evolution*, 28(1): 523–532. 28

Friel, N., Hurn, M., and Wyse, J. (2014). "Improving power posterior estimation of statistical evidence." *Statistics and Computing*, 24(5): 709–723. MR3229692. doi: https://doi.org/10.1007/s11222-013-9397-1. 2, 4, 5, 27

Friel, N. and Pettitt, A. N. (2008). "Marginal likelihood estimation via power posteriors." *Journal of the Royal Statistical Society: Series B (Statisti-*

cal Methodology), 70(3): 589–607. MR2420416. doi: https://doi.org/10.1111/j.1467-9868.2007.00650.x. 2, 4, 6

Gelman, A. and Meng, X.-L. (1998). "Simulating normalizing constants: From importance sampling to bridge sampling to path sampling." *Statistical Science*, 13(2): 163–185. MR1647507. doi: https://doi.org/10.1214/ss/1028905934. 2

Gibson, G. J. and Renshaw, E. (1998). "Estimating parameters in stochastic compartmental models using Markov chain methods." *Mathematical Medicine and Biology*, 15(1): 19–40. 3

Knock, E. S. and O'Neill, P. D. (2014). "Bayesian model choice for epidemic models with two levels of mixing." *Biostatistics*, 15(1): 46–59. 2

Kypraios, T. (2007). "Efficient Bayesian inference for partially observed stochastic epidemics and a new class of semi-parametric time series models." Ph.D. thesis, Lancaster University. URL http://eprints.lancs.ac.uk/26392/ 8, 27

Lartillot, N., Philippe, H., and Lewis, P. (2006). "Computing Bayes factors using thermodynamic integration." *Systematic Biology*, 55(2): 195. 4

Lau, M. S., Marion, G., Streftaris, G., and Gibson, G. J. (2014). "New model diagnostics for spatio-temporal systems in epidemiology and ecology." *Journal of The Royal Society Interface*, 11(93): 20131093. 2

Neal, P. J. and Roberts, G. O. (2004). "Statistical inference and model selection for the 1861 Hagelloch measles epidemic." *Biostatistics*, 5(2): 249–261. 2, 24, 25, 26

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(13): 2011–2024. MR2142722. doi: https://doi.org/10.1002/sim.2090. 2

O'Neill, P. D. and Roberts, G. O. (1999). "Bayesian inference for partially observed stochastic epidemics." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1): 121–129. 3, 12

O'Neill, P. D. and Wen, C. (2012). "Modelling and inference for epidemic models featuring non-linear infection pressure." *Mathematical Biosciences*, 238(1): 38–48. MR2947082. doi: https://doi.org/10.1016/j.mbs.2012.03.007. 3

Pfeilsticker, A. (1861). "Beiträge zur Pathologie der Masern mit besonderer Berücksichtigung der Statistischen Verhältnisse." Ph.D. thesis, Eberhard-Karls-Universität, Tübingen. 24

Severo, N. C. (1969). "Generalizations of some stochastic epidemic models." *Mathematical Biosciences*, 4(3): 395–402. MR0245166. doi: https://doi.org/10.1016/0025-5564(69)90019-4. 3

Stockdale, J. E., O'Neill, P. D., and Kypraios, T. (2017). "Modelling and Bayesian analysis of the Abakaliki smallpox data." *Epidemics*, 19: 13–23. 22

Touloupou, P., Alzahrani, N., Neal, P., Spencer, S. E. F., and McKinley, T. J. (2018). "Efficient model comparison techniques for models requiring large scale data aug-

mentation." *Bayesian Analysis*, 13(2): 437–459. MR3780430. doi: https://doi.org/10.1214/17-BA1057. 2

Worby, C. J. (2013). "Statistical inference and modelling for nosocomial infections and the incorporation of whole genome sequence data." Ph.D. thesis, University of Nottingham. URL http://eprints.nottingham.ac.uk/id/eprint/13154 2

Xie, W., Lewis, P. O., Fan, Y., Kuo, L., and Chen, M.-H. (2011). "Improving marginal likelihood estimation for Bayesian Phylogenetic model selection." *Systematic Biology*, 60(2): 150–160. 28

Xu, X. (2015). "Bayesian nonparametric inference for stochastic epidemic models." Ph.D. thesis, University of Nottingham. URL http://eprints.nottingham.ac.uk/id/eprint/29170 22

Xu, X., Kypraios, T., and O'Neill, P. D. (2016). "Bayesian non-parametric inference for stochastic epidemic models using Gaussian Processes." *Biostatistics*, 17(4): 619–633. MR3604269. doi: https://doi.org/10.1093/biostatistics/kxw011. 23

Zhang, L. (2014). "Time-varying Individual-level Infectious Disease Models." Ph.D. thesis, University of Guelph. URL http://hdl.handle.net/10214/7778 2

Acknowledgments

It is a pleasure to thank an Editor, Associate Editor and two reviewers for helpful comments that have improved the content and layout of the paper. We are grateful for access to the University of Nottingham High Performance Computing Facility used in this work.