Hindawi
Journal of Mathematics
Volume 2022, Article ID 1925202, 9 pages
https://doi.org/10.1155/2022/1925202

Research Article

The Basic Reproduction Number for the Markovian SIR-Type Epidemic Models: Comparison and Consistency

Muteb Faraj Alharthi [1,2]

[1]*Department of Mathematics and Statistics, College of Science, Taif University, Taif, Saudi Arabia*
[2]*Visiting Scholar, Syracuse University, Syracuse, NY, USA*

Correspondence should be addressed to Muteb Faraj Alharthi; muteb.faraj@tu.edu.sa

Received 19 May 2022; Revised 23 July 2022; Accepted 8 August 2022; Published 24 September 2022

Academic Editor: Ali Sajid

This paper is concerned with a well-known epidemiological concept to measure the spread of infectious disease, that is, the basic reproduction number. This paper has two major objectives. The first is to examine Bayesian sensitivity and consistency of this measure for the case of Markov epidemic models. The second is to assess the Martingale method by comparing its performance to that of Markov Chain Monte Carlo (MCMC) methods in terms of estimating this parameter and the infection and removal rate parameters given only removal data. We specifically consider the Markovian SIR (Susceptible-Infective-Removed) epidemic model routinely employed in the literature with exponentially distributed infectious periods. For illustration, numerical simulation studies are performed. Abakaliki smallpox data are examined as a real data application.

1. Introduction and Related Work

The basic reproduction number (or ratio) is a fundamental concept in epidemiology, commonly denoted as R_0. For homogeneously mixing stochastic epidemic models, it can be roughly specified as the expected number of infections generated by one infectious individual in a large susceptible population [1, p. 11]. When $R_0 > 1$, each generation of infectious individuals is more likely to infect the next with a larger number until the susceptible population is significantly depleted. Interventions in the infection or removal process, such as vaccination policies, are frequently carried out with the purpose of decreasing R_0 below unity.

One of the approaches to estimate R_0 via fitting stochastic epidemic models to data, that offers a high level of modeling flexibility, is the data augmentation methods [2] such as MCMC method. Fitting stochastic epidemic models to data is a nonstandard problem [3], as data on the infection processes defined in such models is rarely observed directly. As a result, the likelihood of the observed data is intractable in the sense that calculating it is computationally expensive. Although the data-augmented approaches often weaken in large populations due to poor mixing and increasing computation time, they give a solution to this issue. MCMC methods have already been used successfully for model-based analyses of partially temporal data [4–7].

For certain types of epidemic models (models that attain Markov property), the Martingale method [8, 9] can be utilized for inference. The essential concept behind this technique is to build Martingales for counting processes that are included in the epidemic mode from which estimating equations for the parameters of interest may be derived. This technique may be used to temporal and final size data and could be utilized to undertake nonparametric inference in specific scenarios. While this approach is very elegant, it is somewhat specialized and not as widely applicable as most other methods for fitting epidemic models to data. This paper aims to explore the efficiency of this approach in estimating model parameters as well as comparing its estimation performance with the MCMC method outcomes.

When analysing temporal data, the likelihood of the observed data can become very difficult to evaluate, and so is the posterior distribution, as calculating the likelihood

involves integration over all possible unobserved infection times, which is rarely analytically possible. However, simulating realizations from a stochastic epidemic model can be relatively straightforward. Implicit methods such as Approximate Bayesian Computation (ABC) may be suited to make inference for the parameters of epidemic models based on partially observed data, and this has been illustrated when temporal data are available in [10–13]. However, the choice of the tolerance parameter and the summary statistics are nontrivial tasks and may require expensive pilot runs to select the appropriate ones. Moreover, when working with large amounts of data, it often becomes necessary to use simplistic models due to computational restraints, especially when performing time-sensitive analyses. Approximation methods [14, 15] could be a natural solution under which likelihood can be computed. However, the simplifying assumptions used in approximation methods are not always appropriate in epidemic context.

Our main focus in this paper 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. As the emphasis is on the Martingale methods which need models to be Markovian, we specifically consider the SIR model that is routinely employed in the literature, namely, the SIR model with exponentially distributed infectious period. The motivations for this study are that Bayesian analysis of this important quantity (R_0) is of interest in its own right. In addition, the Markovian SIR model offers guidance that is valuable in analysing more complex and realistic models as SIR models themselves are frequently used as components of more complicated epidemic models, for instance those featuring populations divided into households [16] or epidemics on networks [1].

A relative asymptotic efficiency of R_0 was derived, using Martingale method in [17] by comparing the variance of R_0 with that of the maximum likelihood estimator based on complete observation of the epidemic. The asymptotic performance of a Martingale estimator for the infection rate was compared with the maximum likelihood estimator in [18]. However, these comparisons are of limited value since complete observation of epidemic is not possible. Large and finite sample properties for the Martingale estimator of R_0 were studied in [19], where the performance of Martingale estimate with the maximum likelihood estimate based on complete observation and the maximum likelihood estimate based on the eventual size of the epidemic was compared.

Initial values and continuous observation of the removal process are a plausible data set that may be observed during the duration of an epidemic. The time of diagnosis or commencement of symptoms, which is often accessible, might be used to approximate the time of removal for many diseases. Using such type of data, Martingale technique was applied in [9] to obtain estimating equations that enable the model parameters to be estimated separately. Data augmentation technique was implemented to deal with such data by [6] in which the missing data (the infection times) were treated as extra (unknown) parameters to be inferred from the data.

Despite the extensive literature about R_0, as far as the author knows, Bayesian sensitivity and consistency are not covered in much detail. Furthermore, given only the removal times, the performance of the Martingale and MCMC methods in terms of estimating R_0 and the infection and removal rates is not assessed yet. This gap in the literature needs to be filled.

The main purpose of this paper is to study Bayesian sensitivity and consistency for the posterior distribution of R_0 when Markovian SIR model is considered. Another important aim of this work is to explore the performance of the Martingale method in terms of estimating the basic reproduction number and the infection and removal rates, and compare it with the performance of the MCMC method performance as the most widely used method in epidemic inference.

The rest of this paper is organized as follows: In Section 2, relevant preliminary information will be presented which includes the definition of our stochastic epidemic model. Then, MCMC methods for the Markovian SIR model based on removal data will be introduced in Section 3. In Section 4, Martingale methods for the Markovian SIR model will be discussed. Bayesian consistency and sensitivity for R_0 will be investigated in Section 5. A comparison study of Bayesian and Martingale estimation for R_0 and the infection and removal rates will be conducted using both simulated and real data sets in Section 6. Some concluding remarks will be given in Section 7.

2. Preliminaries

2.1. The Stochastic SIR Epidemic Model. The stochastic SIR epidemic model is defined as follows [1, 6]. Consider a closed population of size N individuals divided to n and k initial susceptible and infected individuals, respectively. At any given time during the epidemic, each individual is either susceptible, infected, or removed. Susceptible individuals have not yet developed the disease, but they are at risk. Infected individuals carry the disease and can spread it to others. Individuals who have been removed are no longer infectious and have no role in disease transmission. The removed state is context-specific in applications, and it might equate to immunity, isolation, death, or other comparable outcomes. Each infective individual remains so for a period of time, called the infectious period, which is drawn from a specified nonnegative probability distribution T_I. At the end of its infectious period, an individual is immediately removed. During its infectious period, a given infectious has contacts with a given susceptible in the population at times given by the points of a Poisson process of rate βn^{-1}, where β is the infection rate parameter. The first such contact, if it occurs, results in the susceptible immediately becoming infected, and later contacts have no effect. All infectious periods and all Poisson processes are mutually independent. The epidemic ends as soon as there are no more infections remaining in the population.

For any time t, let $X(t), Y(t)$, and $Z(t)$ denote, respectively, the numbers of susceptible, infected, and removed individuals in the population. The above model is known as the Markovian SIR model when the infectious periods follow an exponential distribution with rate γ, say. In

this case, the epidemic process $\{(X((t), Y(t)): t \geq 0)\}$ is a bivariate continuous time Markov chain with the following transition rates:

$$\begin{aligned}(i, j) &\longrightarrow (i-1, j+1): \beta n^{-1} X(t) Y(t),\\ (i, j) &\longrightarrow (i, j-1): \gamma Y(t).\end{aligned} \tag{1}$$

The corresponding transition probabilities to an infection and removal are

$$\begin{aligned}&P\left[X(t+\delta t) - X(t) = -1, Y(t+\delta t) - Y(t) = 1 | \mathscr{H}_t\right]\\ &= \beta n^{-1} X(t) Y(t) + o(\delta t), P[X(t+\delta t) - X(t)\\ &= 0, Y(t+\delta t) - Y(t) = -1 | \mathscr{H}_t]\\ &= \gamma Y(t) + o(\delta t),\end{aligned} \tag{2}$$

where all other transitions having probability $o(\delta t)$, $\mathscr{H}_t$ is the sigma-algebra generated by the history of the process up to time t, and γ denotes the removal rate parameter.

2.2. The Basic Reproduction Number R_0. For the Markovian SIR model, the basic reproduction number is defined as $R_0 = \beta E(T_I) = \beta \gamma^{-1}$ (see for example [1]). The importance of this measure comes from its relation to the spread of epidemic. In other words, in an infinitely large population, if $R_0 \leq 1$, then with probability one, only a finite number of individuals will become infected, i.e., the epidemic will die out, whereas when $R_0 > 1$ there is a positive probability that the epidemic will take off. Consequently, and more importantly, knowing the value of R_0 enables us to compute the proportion of a population that should be vaccinated in order to prevent an epidemic from spreading [1]. That is to say, vaccinating a fraction $\nu_c = 1 - 1/R_0$ (critical vaccination coverage with perfect vaccine) of the community chosen uniformly at random is sufficient to prevent a large outbreak.

2.3. Epidemic Outbreak Data. Although the details of outbreak data vary considerably from study to study, broadly speaking, there are three types of epidemic data, namely, complete temporal data, partial (incomplete) temporal data, and final size data [20]. The complete temporal data consist of the time of infection and time of removal of all infected individuals in the population. The final size data do not contain explicit temporal information regarding the disease propagation throughout the population but instead consist of the initial number of susceptible individuals and which of these individuals were ultimately infected during the course of the epidemic. The partial temporal data are less detailed than complete temporal data, but more detailed than final size data. This type of data is more common in practice and typically consists of removal times.

3. MCMC Methods for the Markovian SIR Model Based on Removal Data

In the context of epidemic modeling, MCMC techniques are appealing for two reasons. First, they provide a great deal of modeling flexibility. Second, when used in combination with the Bayesian framework, they allow for the analysis of all model parameters or any function of them.

Suppose that an epidemic outbreak which results in a total of m removals in a closed population of size N, n initial susceptibles and one initial infectious is observed on $(\mathrm{t}_1^I, \mathrm{t}_m^R]$, so that all removals are assumed to be observed. Let $0 = \mathrm{t}_1^R, \mathrm{t}_2^R, \ldots, \mathrm{t}_m^R$ be the observed ordered removal times and write $\mathbf{t}^{\mathbf{R}} = (\mathrm{t}_1^R, \ldots, \mathrm{t}_m^R)$. Also, let $\mathrm{t}_1^I < 0$ be the first unobserved infection time and $\mathbf{t}^I = (\mathrm{t}_2^I, \ldots, \mathrm{t}_m^I)$ be the rest of unobserved ordered infection times, where $\mathrm{t}_h^I < \mathrm{t}_{h+1}^I < \mathrm{t}_h^R$ for $h = 1, 2, \ldots, m-1$. Then, the joint likelihood of $\mathbf{t}^{\mathbf{I}}$ and $\mathbf{t}^{\mathbf{R}}$ given the model parameters and t_1^I is [6, 21].

$$\begin{aligned}L\left(\mathbf{t}^I, \mathbf{t}^R | \beta, \gamma, \mathrm{t}_1^I\right) &= \prod_{j=2}^{m} \beta n^{-1} X\left(\mathrm{t}_j^I -\right) Y\left(I_j -\right) \times \prod_{j=1}^{m} \gamma Y\left(\mathrm{t}_j^R -\right)\\ &\times \exp\left(-\int_{\mathrm{t}_1^I}^{\mathrm{t}_m^R} \left(\beta n^{-1} X(t) Y(t) + \gamma Y(t)\right) \mathrm{d}t\right),\end{aligned} \tag{3}$$

where t_j^I- denotes the time just prior to t_j^I and $\mathrm{t}_j^R-$is defined similarly.

In practice, the infection times are hard to observe which in turn makes the likelihood of observing only the removal times $\mathbf{t}^R$ given the model parameters intractable. One possible solution to make the likelihood tractable is to use the data augmentation technique [2, 22] by treating the missing data as extra (unknown) parameters to be estimated from the data [6]. By doing so, we can sample from the conditional posterior distribution of the infection times in order to produce samples from the conditional posterior distributions of the model parameters.

Now, we assign an independent gamma prior distribution for each rate parameter, namely, $\mathrm{Gamma}(\lambda_\zeta, \nu_\zeta)$, where $\zeta = \beta, \gamma$, and assume a priori that $-\mathrm{t}_1^I \sim \mathrm{Exp}(\theta)$, where $\mathrm{Gamma}(a, b)$ represents a gamma distribution with mean a/b and variance a/b^2 and $\mathrm{Exp}(a)$ denotes an exponential distribution with mean a^{-1}. Then, by multiplying the likelihood and the priors, we obtain the following full conditional posterior densities for β, γ, and t_1^I, that are

$$\begin{aligned}\beta | \gamma, \mathrm{t}_1^I, \mathbf{t}^I, \mathbf{t}^R &\sim \mathrm{Gamma}\left(\lambda_\beta + m - 1, \nu_\beta + n^{-1} \int_{\mathrm{t}_1^I}^{\mathrm{t}_m^R} X(t) Y(t) \mathrm{d}t\right),\\ \gamma | \beta, \mathrm{t}_1^I, \mathbf{t}^I, \mathbf{t}^R &\sim \mathrm{Gamma}\left(\lambda_\gamma + m, \nu_\gamma + \int_{\mathrm{t}_1^I}^{\mathrm{t}_m^R} Y(t) \mathrm{d}t\right),\end{aligned} \tag{4}$$

and

$$\mathrm{t}_1^I | \beta, \gamma, \mathbf{t}^I, \mathbf{t}^R \sim \mathrm{t}_2^I - \mathrm{Exp}(\beta + \gamma + \theta). \tag{5}$$

Also, we have the following proportional full posterior density for $\mathbf{t}^{\mathbf{I}}$:

$$\begin{aligned}\pi\left(\mathbf{t}^I | \beta, \gamma, \mathrm{t}_1^I, \mathbf{t}^R\right) &\propto \prod_{j=2}^{m} X\left(\mathrm{t}_j^I -\right) Y\left(\mathrm{t}_j^I -\right) \times \prod_{j=1}^{m} Y\left(\mathrm{t}_j^R -\right)\\ &\times \exp\left(-\int_{\mathrm{t}_1^I}^{\mathrm{t}_m^R} \left(\beta n^{-1} X(t) Y(t) + \gamma Y(t) \mathrm{d}t\right)\right).\end{aligned} \tag{6}$$

The model parameters β, γ, and the initial infection time t_1^I, all can be updated using Gibbs sampling steps as they have closed forms of the posterior density. However, the infection times vector, $\mathbf{t}^I$, needs to be updated using Metropolis-Hastings steps as its posterior density is not available explicitly. Any function of these parameters, such as the basic reproduction number $R_0 = \beta/\gamma$, can be obtained easily by sampling from the parameters posterior distributions. Hence, we can produce an MCMC estimate for the basic reproduction number, say $\widehat{R_{01}}$.

4. Martingale Methods for the Markovian SIR Model Based on Removal Data

Maximum likelihood estimation of parameters of the Markovian epidemic model is tedious when only partial information is available, as is often the case in practice. This motivates the development of easier methods of estimation such as the one established in [8] which can lead to simple but nevertheless efficient estimators.

Let $N(t) = Y(t) + Z(t)$ count the number of infections that occur by time t. Under the assumptions of the Markovian SIR epidemic model, there are two zero-mean Martingale processes [8] defined by

$$\begin{aligned} M_\beta\left(t_m^R\right) - N\left(t_m^R\right) + \beta n^{-1} \int_0^{t_m^R} X(t)Y(t)\mathrm{d}t = 0, \\ M_\gamma\left(t_m^R\right) - Z\left(t_m^R\right) + \gamma \int_0^{t_m^R} Y(t)\mathrm{d}t = 0. \end{aligned} \tag{7}$$

Besides the Martingales considered above, the method of moments and the Martingale process $\left\{X(t)(1+R_0/n)^{Z(t_m^R)-Z(t)}: t \geq 0\right\}$ were used in [9] to derive explicit estimators for the infection rate β and the mean infectious period γ^{-1} based on observing only the removal process. These estimators can be written as

$$\begin{aligned} \widehat{\beta} &= \frac{Z\left(t_m^R\right)\sum_{j=1}^{Z\left(t_m^R\right)-1} n/(n+1-j)Z\left(t_m^R\right)}{Nt_m^R - \int_0^{t_m^R} Z(t)dt - X\left(t_m^R\right)\int_0^{t_m^R}\left(1+\sum_{j=1}^{Z\left(t_m^R\right)-1} n/(n+1-\ j)Z\left(t_m^R\right)/n\right)^{Z\left(t_m^R\right)-Z(t)} dt}, \\ \widehat{\gamma^{-1}} &= \frac{Nt_m^R - \int_0^{t_m^R} Z(t)\mathrm{d}t - X\left(t_m^R\right)\int_0^{t_m^R}\left(1+\sum_{j=1}^{Z\left(t_m^R\right)-1}(n+1-j)Z\left(t_m^R\right)/n\right)^{Z\left(t_m^R\right)-Z(t)}\mathrm{d}t}{Z\left(t_m^R\right)}, \end{aligned} \tag{8}$$

provided at least one individual remains susceptible at time t_m^R. Note that these estimators depend only on the initial number of susceptible individuals n, the final number of removals $Z(t_m^R)$, and the removal times. A Martingale estimate for the basic reproduction number is given by

$$\widehat{R_{02}} = \sum_{j=1}^{Z\left(t_m^R\right)-1} \frac{n}{(n+1-j)Z\left(t_m^R\right)}, \tag{9}$$

with standard deviation given as

$$s.d\left(\widehat{R_{02}}\right) = \frac{1}{Z\left(t_m^R\right)}\sqrt{Z\left(t_m^R\right)\widehat{R_{02}}^2 + n^2 \sum_{j=1}^{Z\left(t_m^R\right)-1}\frac{1}{(n+1-j)^2}}. \tag{10}$$

5. Bayesian Consistency and Sensitivity for R_{01}

An estimator θ_n of a quantity of interest θ is said to be consistent if $\theta_n \longrightarrow \theta$ as $n \longrightarrow \infty$, in other words, if it is guaranteed to converge to the true value. However, the posterior is consistent for θ if the posterior distribution concentrates around a mass point which is the true value of the parameter [23].

Given complete observation of the epidemic, the following explicit posterior density for R_{01} was derived in [24] using prior settings as in Section 3.

$$\pi\left(R_{01}|\mathbf{t}^I, \mathbf{t}^R, t_1^I\right) = \frac{\Gamma\left(2m-1+\lambda_\beta+\lambda_\gamma\right)}{\Gamma\left(m+\lambda_\gamma\right)\Gamma\left(m-1+\lambda_\beta\right)} \times \frac{\left(n^{-1}A+\nu_\beta/B+\nu_\gamma\right)^{m-1+\lambda_\beta} R_{01}^{m-2+\lambda_\beta}}{\left(\left(1+\left(n^{-1}A+\nu_\beta/B+\nu_\gamma\right)\right)R_{01}\right)^{2m-1+\lambda_\beta+\lambda_\gamma}}, \tag{11}$$

given that $R_{01} > 0$, where $A = \int_{t_1^I}^{t_m^R} X(t)Y(t)\mathrm{d}t$ and $B = \int_{t_1^I}^{t_m^R} Y(t)\mathrm{d}t$.

The posterior mean and variance are given as follows:

$$E\left(R_{01}|\mathbf{t}^I, \mathbf{t}^R, \mathrm{t}_1^I\right) = \left(\frac{m-1+\lambda_\beta}{m-1+\lambda_\gamma}\right) \times \left(\frac{B+\nu_\gamma}{n^{-1}A+\nu_\beta}\right),$$
$$\mathrm{Var}\left(R_{01}|\mathbf{t}^I, \mathbf{t}^R, \mathrm{t}_1^I\right) = \frac{\left(2(m-1)+\lambda_\beta+\lambda_\gamma\right)\left(m-1+\lambda_\beta\right)}{\left(m-2+\lambda_\gamma\right)\left(m-1+\lambda_\gamma\right)^2} \times \left(\frac{B+\nu_\gamma}{n^{-1}A+\nu_\beta}\right)^2, \tag{12}$$

given that $m + \lambda_\gamma > 2$.

Common choices of the prior parameters in the literature are $\lambda_\beta = \lambda_\gamma = \lambda$ and $\nu_\beta = \nu_\gamma = \nu$, where $\lambda \geq 1$ and ν is a small positive number. This prior setting yields

$$\pi\left(R_{01}|\mathbf{t}^I, \mathbf{t}^R, \mathrm{t}_1^I\right) = \frac{\Gamma(2m-1+2\lambda)}{\Gamma(m+\lambda)\Gamma(m-1+\lambda)} \times \frac{\left(n^{-1}A+\nu/B+\nu\right)^{m-1+\lambda} R_{01}^{m-2+\lambda}}{\left(\left(1+\left(n^{-1}A+\nu/B+\nu\right)\right)R_{01}\right)^{2m-1+2\lambda}}, \tag{13}$$

with

$$E\left(R_{01}|\mathbf{t}^I, \mathbf{t}^R, \mathrm{t}_1^I\right) = \frac{B+\nu}{n^{-1}A+\nu}, \tag{14}$$

$$\mathrm{Var}\left(R_{01}|\mathbf{t}^I, \mathbf{t}^R, \mathrm{t}_1^I\right) = \frac{2}{(m-2+\lambda)} \times \left(\frac{B+\nu}{n^{-1}A+\nu}\right)^2. \tag{15}$$

It is of our interest to study the behaviour of posterior density, the posterior mean, and the posterior variance in two cases. The first case is when the initial number of susceptibles (n) tends to infinity; and the second case is when the prior distributions of the model parameters (β and γ) get diffuse. We restrict our results on major outbreak situation [18] as it may not be of importance to model and investigate minor outbreak in practical situations.

5.1. Consistency for R_{01}. Suppose that we have a sequence of processes $\{(X_n(t), Y_n(t))\}$ indexed by $X(0) = n$, the initial number of susceptibles. Without the loss of generality, we set $\mathrm{t}_1^I = 0$ and let $\mathrm{t}_m^R = \tau$ be the time when the epidemic ends.

We need to examine the behaviour of the following random variables as $n \longrightarrow \infty$

$$m_n,\ A_n = \int_0^{\tau_n} X_n(t)Y_n(t)\mathrm{d}t, \text{ and } B_n = \int_0^{\tau_n} Y_n(t)\mathrm{d}t. \tag{16}$$

It was shown in [18, 25] that as $n \longrightarrow \infty, m_n/n \longrightarrow^P \rho$, where ρ is a constant satisfies $1 - \rho = e^{-\rho R_0}$.

Also, it was proven that

$$n^{-1}\int_0^{\tau_n} Y_n(t)\mathrm{d}t \xrightarrow{P} \rho/\gamma \text{ and } n^{-2}\int_0^{\tau_n} X_n(t)Y_n(t)\mathrm{d}t \xrightarrow{P} \rho/\beta. \tag{17}$$

To indicate that the posterior mean is specific to n, we rewrite (14) as

$$E_n\left(R_{01}|\mathbf{t}^I, \mathbf{t}^R, \mathrm{t}_1^I\right) = \frac{B_n+\nu}{n^{-1}A_n+\nu} = \frac{B_n/n+\nu/n}{n^{-2}A_n+\nu/n} \longrightarrow \frac{\rho/\gamma}{\rho/\beta} = \frac{\beta}{\gamma} = R_{01},\ as\, n \longrightarrow \infty. \tag{18}$$

Similarly, to indicate that the posterior variance is specific to n, we rewrite (15) as

$$\mathrm{Var}_n\left(R_{01}|\mathbf{t}^{\mathbf{I}}, \mathbf{t}^{\mathbf{R}}, \mathrm{t}_1^I\right) = \frac{2}{(m_n-2+\lambda)} \times \left(\frac{B_n+\nu}{n^{-1}A_n+\nu}\right)^2 = \frac{2/n}{(m_n/n-2/n+\lambda/n)} \times \left(\frac{B_n/n+\nu/n}{n^{-2}A_n+\nu/n}\right)^2 \longrightarrow 0 \text{ as } n \longrightarrow \infty. \tag{19}$$

Therefore, we conclude that as we have more and more data, the posterior mean tends to the true value and the posterior variance tends to zero and so the posterior distribution concentrates around a mass point which is the true value of the parameter. Mathematically, as $n \longrightarrow \infty$ we have

$$\pi\left(R_{01}|\mathbf{t}^I, \mathbf{t}^R, \mathrm{t}_1^I\right) \longrightarrow \delta_{\widetilde{R_{01}}}(R_{01}), \tag{20}$$

where $\delta_{\widetilde{x}}(x)$ is a Dirac delta function and $\widetilde{x}$ is the true value of x.

That is to say, the posterior is consistent at $\widetilde{R_{01}}$, and the Bayes estimate (the posterior mean) is a consistent estimator for R_{01} as $n \longrightarrow \infty$.

5.2. Sensitivity for R_{01}. To see the effect of the prior's diffuseness on both posterior mean and variance of R_{01}, we

reformulate them in terms of the mean and variance of the prior distribution, that are

$$E[\text{Gamma}(\lambda,\nu)] = \lambda/\nu = \mu \text{ and } \text{Var}[\text{Gamma}(\lambda,\nu)] = \lambda/\nu^2 = \sigma^2, \tag{21}$$

which implies $\lambda = \mu^2/\sigma^2$ and $\nu = \mu/\sigma^2$. Substituting these values into (14) and (15) gives

$$E\left(R_{01}|\mathbf{t}^I,\mathbf{t}^R,\mathrm{t}_1^I\right) = \frac{B+\mu/\sigma^2}{n^{-1}A+\mu/\sigma^2} \longrightarrow \frac{nB}{A}, \text{ as } \sigma^2 \longrightarrow \infty. \tag{22}$$

This is in accord with the maximum likelihood estimator of R_0 given by the ratio of the maximum likelihood estimators of β and γ, that is, $\widehat{R}_0 = n(m-1)B/mA$ (see for example [1]).

Also, we have

$$\text{Var}\left(R_{01}|\mathbf{t}^I,\mathbf{t}^R,\mathrm{t}_1^I\right) = \frac{2}{\left(m-2+\mu^2/\sigma^2\right)} \times \left(\frac{B+\mu/\sigma^2}{n^{-1}A+\mu/\sigma^2}\right)^2 \longrightarrow \frac{2n^2B^2}{(m-2)A^2}, \text{ as } \sigma^2 \longrightarrow \infty. \tag{23}$$

6. Comparison of MCMC and Martingale Estimation for R_0, β, and γ

In this section, we explored, via simulation study, the accuracy of Martingale method for estimating R_{02}, β, and γ given only a set of removal times. Moreover, a comparison between MCMC and Martingale estimation methods was performed. Finally, a widely studied removal data set obtained from a smallpox outbreak [26] was considered.

We simulated epidemics under a Markovian SIR model. All outbreaks were in a closed population consisting of different initially susceptible individuals (n) and a single initially infectious case $(k = 1)$. In all simulations, the value of γ was taken (without loss of generality) to be equal to 1, corresponding to a mean infectious period of 1 day, say.

To make our comparisons more useful, we wish to eliminate realizations that resulted in early extinction, thereby presenting results based on major epidemics only. A working algorithm that excluded all simulations for which $R(\tau) < 0.1n + k$ was implemented in [27], where τ is the extinction time. Here, we followed [4, 21] and excluded simulated epidemics in which the final size was not consistent with the theoretical probability of a major epidemic. In other words, we simulated the model until we have the desired number of realizations that have final sizes fit well within the major outbreak part of the simulation-based final size distribution.

6.1. Simulation Study I. There have been several applications of Martingale results to issues of epidemiological parameter inference, but the investigation of their efficacy has been unaddressed so far. To see how efficient the Martingale method is in estimating R_{02}, β, and γ separately given a set of removal times, we conducted this simulation study using different population and final epidemic sizes. We set $\beta = 2.5$ and $\gamma = 1$, so that $R_{02} = 2.5$. 1000 major outbreak realizations were simulated from the model. In Table 1, the column labelled "Avge $(Z(\mathrm{t}_m^R))$" gives the average number of removals over the realizations. The other columns represent the average of the Martingale method outcomes for estimating R_{02} and its standard deviation, β, and γ. It can be seen that the method is very efficient in recovering true values and best results are achieved as both population and final epidemic size get large.

6.2. Simulation Study II. To make a reasonable and meaningful comparison between MCMC and Martingale methods for estimating the basic reproduction number and infection and removal rates given only removal data, we simulated 10 typical (in the sense that the final size fits within the high density of the major outbreak component of the simulation-based final size distribution) epidemic realizations from the model for different population sizes as seen in Table 2 using various values of the threshold parameter R_0 (1.5, 2, and 3). Each outbreak simulation ended when all the infections were removed. For each outbreak, we computed both MCMC and Martingale estimates and then average over the 10 realizations. In Table 2, the column labelled "Avge $(Z(\mathrm{t}_m^R))$" gives the average number of removals over the 10 realizations. Martingale results were obtained using 8, 9, and 10, whereas MCMC results were produced using standard algorithms (see for example [6, 21]).

The findings of this simulation study can be summarized as follows: For both MCMC and Martingale outcomes, the estimation improved as both the population and epidemic final size increase. When the population size gets large, the Martingale method seems to be useful in terms of computational times and precision. When looking to standard deviation as a measure of discrepancy, MCMC results were as good as the Martingale ones. However, as pointed out in [28], in large-scale settings, MCMC methods struggle due to the increased computational burden as a result of high dimensional parameter space.

6.3. Abakaliki Smallpox Data. We now consider a widely studied temporal data set obtained from a smallpox outbreak in a closed population of $N = 120$ individual $(n = 119, k = 1)$ in Abakaliki, Nigeria [26], p. 125. The data consist of times between removals measured in days, which are

Table 1: Results of simulation study I with $R_{02} = 2.5$, $\beta = 2.5$, and $\gamma = 1$. Each row is obtained based on the average of 1000 major outbreak realizations.

N	Avge $(Z(t_m^R))$	Avge (Martingale outcomes)			
		$\widehat{R_{02}}$	$\widehat{\text{s.d}}(\widehat{R_{02}})$	$\widehat{\beta}$	$\widehat{\gamma^{-1}}$
50	42	2.1838	0.55	2.3348	1.0087
100	89	2.5414	0.40	2.2914	1.1608
300	268	2.5401	0.21	2.3685	1.0927
700	625	2.5080	0.16	2.4350	1.0382
1000	892	2.5082	0.13	2.4433	1.0322
1500	1339	2.5075	0.10	2.4646	1.0214

Table 2: Results of simulation study II for different values of R_0, where γ value is set to 1. Each row is obtained based on the average of 10 major outbreak realizations.

N	True R_0	Avge $(Z(t_m^R))$	Avge (MCMC outcomes)			Avge (Martingale outcomes)		
			$\widehat{R_{01}}\,(\widehat{s.d})$	$\widehat{\beta}$	$\widehat{\gamma^{-1}}$	$\widehat{R_{02}}\,(\widehat{\text{s.d}})$	$\widehat{\beta}$	$\widehat{\gamma^{-1}}$
50	1.5	33	1.7048 (0.43)	1.8546	1.1042	1.5695 (0.40)	2.0970	0.9207
	2	41	2.2245 (0.52)	2.8048	0.9658	2.0556 (0.49)	2.3263	1.0876
	3	46	3.0847 (0.74)	3.3996	0.9505	2.7089 (0.66)	2.4842	1.1067
100	1.5	67	1.6877 (0.30)	1.7777	1.0219	1.6297 (0.29)	1.6972	0.9858
	2	81	2.0881 (0.34)	2.3713	0.9673	2.0321 (0.34)	1.9176	1.1331
	3	91	2.8392 (0.48)	3.6982	0.8026	2.7349 (0.47)	2.8729	0.9635
300	1.5	182	1.5567 (0.17)	1.6460	1.0022	1.5313 (0.16)	1.7052	0.9057
	2	238	2.0055 (0.19)	2.1683	0.9803	1.9856 (0.19)	2.0097	1.0259
	3	280	2.9007 (0.28)	3.2939	0.8949	2.9288 (0.29)	2.8886	1.0316

$$13, 7, 2, 3, 0, 0, 1, 4, 5, 3, 2, 0, 2, 0, 5, 3, 1, 4, 0, 1, 1, 1, 2, 0, 1, 5, 0, 5, 5, \tag{24}$$

where zero indicates that the case occurred at the same day as the previous one.

Therefore, in our notation, the observed removal times are given by

$$\mathbf{t}^R = \left(\begin{matrix} 0, 13, 20, 22, 25, 25, 25, 26, 30, 35, 38, 40, 40, 42, 42, 47, 50, 51, 55, 55, 56, 57, 58, \\ 60, 60, 61, 66, 66, 71, 76 \end{matrix} \right). \tag{25}$$

The use of the SIR stochastic epidemic to model these data is not entirely appropriate, since smallpox has an appreciable latent period. However, it can be used for illustrative purpose [6]. In the literature, the basic reproductive number for smallpox is about 3 − 6 [29]. However, using the Martingale method, the authors of [8, 9] gave an estimate of 1.1 for R_{02}, while MCMC method produces an estimate of $R_{01} = 1.13$ in [6]. Using the ABC method, an estimate of 1.1 for the basic reproduction number parameter was obtained in [10].

Surprisingly, no study in the literature has provided Martingale estimates for the infection rate β and the mean infectious period γ^{-1} based on these removal times. Using (8), we obtained the following Martingale estimates, $\widehat{\beta} = 0.335$ and $\widehat{\gamma^{-1}} = 3.292$. It is of concern why these estimates are not somehow similar to those obtained using different methods. However, it was pointed out in [14] that assuming that all the members are susceptible and are mixed homogeneously may cause a significant underestimation of the basic reproduction number and hence related parameters.

Table 3 shows estimates of $\widehat{\beta}$, $\widehat{\gamma^{-1}}$ and $\widehat{R}_0$ for Abakaliki smallpox data, using different approaches.

6.4. Comparison Discussion. The simulation studies above revealed the accuracy and the efficiency of the Martingale method when estimating the basic reproduction number and the infection and removal rates of the Markovian SIR model. Although the Martingale method seems promising in estimating all the parameters (see Table 1) and preforms well with large population sizes, it is only applicable when epidemic models are Markov. Unfortunately, most sophisticated epidemic models need to be non-Markov to mimic the mechanism of infectious diseases.

The MCMC method, on the other hand, estimated all the parameters with very good precision. It is known that this method struggles when the population size gets large [28] and for this reason, we considered only $N = 50, 100$, and 300, to make a fair comparison.

In combination with the Bayesian approach, MCMC method has no restriction that models be Markov. Therefore,

Table 3: Estimates of $\widehat{\beta}$, $\widehat{\gamma^{-1}}$ and $\widehat{R}_0$ for Abakaliki smallpox data, using different approaches.

Method	$\widehat{\beta}$	$\widehat{\gamma^{-1}}$	$\widehat{R}_0$
Martingale	0.335	3.292	1.1
MCMC	0.099	11.416	1.13
ABC	0.11	10	1.1

it is an attractive choice in the epidemic context as it permits a huge amount of modeling flexibility, and it enables analysis of all of the model parameters, or any function of them.

7. Concluding Remarks

In this paper, we have considered a quantity of central importance in mathematical epidemic theory, that is, the basic reproduction number. Bayesian sensitivity and consistency of its posterior distribution were discussed, particularly in the case where the infectious period is exponentially distributed as this case is commonly used by modellers, partly for mathematical convenience. The performance of the Martingale method in terms of estimating the basic reproduction number and the infection and removal rates was explored and compared to the performance of the most widely used method, that is, the MCMC method.

The Martingale method's key advantages were that it reduced the stochastic epidemic model's unreasonable assumptions and it simplified the computations. However, this approach seems to be undesirable as it requires models to be Markov. The MCMC method, on the other hand, allows for a great deal of modeling adaptability. In most cases, no such modeling constraint on infectious periods is required. It also allows analysis of all model parameters, or any function of them, when used in conjunction with the Bayesian technique.

This work is significant in the sense that it explores Bayesian sensitivity and consistency for the posterior distribution of the basic reproduction number in the case of the Markovian SIR model. In addition, it showed the interesting performance of the Martingale method in terms of estimating the basic reproduction number and the infection and removal rates.

It is natural to study Bayesian sensitivity and consistency for other models with different choices of infectious period distribution. However, the posterior distribution of the basic reproduction number given a complete observation may not have a closed form, which in turn makes the analysis harder.

Data Availability

We use simulated data sets for illustration. Abakaliki Smallpox data set is published in the literature.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis*, Vol. 4, Springer, New York, NY, USA, 2000.

[2] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.

[3] P. D. O'Neill, "A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods," *Mathematical Biosciences*, vol. 180, no. 1-2, pp. 103–114, 2002.

[4] M. Alharthi, "Model discrimination for epidemiological SEIR-type models with different transmission mechanisms," *JP Journal of Biostatistics*, vol. 20, no. 27–50, pp. 27–50, 2022.

[5] M. Alharthi, T. Kypraios, and P. D. O'Neill, "Bayes factors for partially observed stochastic epidemic models," *Bayesian Analysis*, vol. 14, no. 3, pp. 907–936, 2019.

[6] P. D. O'Neill and G. O. Roberts, "Bayesian inference for partially observed stochastic epidemics," *Journal of the Royal Statistical Society: Series A*, vol. 162, no. 1, pp. 121–129, 1999.

[7] F. Xiang and P. Neal, "Efficient MCMC for temporal epidemics via parameter reduction," *Computational Statistics & Data Analysis*, vol. 80, pp. 240–250, 2014.

[8] N. G. Becker, *Analysis of infectious disease data*, Vol. 33, CRC Press, Boca Raton, FL, USA, 1989.

[9] N. G. Becker and A. M. Hasofer, "Estimation in epidemics with incomplete observations," *Journal of the Royal Statistical Society: Series B*, vol. 59, no. 2, pp. 415–429, 1997.

[10] T. Kypraios, P. Neal, and D. Prangle, "A tutorial introduction to bayesian inference for stochastic epidemic models using approximate bayesian computation," *Mathematical Biosciences*, vol. 287, pp. 42–53, 2017.

[11] T. McKinley, A. R. Cook, and R. Deardon, "Inference in epidemic models without likelihoods," *International Journal of Biostatistics*, vol. 5, no. 1, 2009.

[12] A. Minter and R. Retkute, "Approximate bayesian computation for infectious disease modelling," *Epidemics*, vol. 29, Article ID 100368, 2019.

[13] J. Peter, "Approximate bayesian computation methods for epidemic models," in *Handbook of Infectious Disease Data Analysis*, pp. 179–197, Chapman and Hall/CRC, Boca Raton, FL, USA, 2019.

[14] M. Eichner and K. Dietz, "Transmission potential of smallpox: estimates based on detailed data from an outbreak," *American Journal of Epidemiology*, vol. 158, no. 2, pp. 110–117, 2003.

[15] J. E. Stockdale, T. Kypraios, and P. D. O'Neill, "Pair-based likelihood approximations for stochastic epidemic models," *Biostatistics*, vol. 22, no. 3, pp. 575–597, 2021.

[16] P. D. O'Neill and P. J. Marks, "Bayesian model choice and infection route modelling in an outbreak of Norovirus," *Statistics in Medicine*, vol. 24, no. 13, pp. 2011–2024, 2005.

[17] P. Yip, "Estimating the initial relative infection rate for a stochastic epidemic model," *Theoretical Population Biology*, vol. 36, no. 2, pp. 202–213, 1989.

[18] W. N. Rida, "Asymptotic properties of some estimators for the infection rate in the general stochastic epidemic model," *Journal of the Royal Statistical Society: Series B*, vol. 53, no. 1, pp. 269–283, 1991.

[19] Q. X. Shao, "Some properties of an estimator for the basic reproduction number of the general epidemic model," *Mathematical Biosciences*, vol. 159, no. 1, pp. 79–96, 1999.

[20] T. Britton, T. Kypraios, and P. D. O'Neill, "Inference for epidemics with three levels of mixing: methodology and application to a measles outbreak," *Scandinavian Journal of Statistics*, vol. 38, no. 3, pp. 578–599, 2011.
[21] M. Alharthi, "Bayesian model assessment for stochastic epidemic models," Ph. D. Thesis, University of Nottingham, Nottingham, UK, 2016.
[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B*, vol. 39, no. 1, pp. 1–22, 1977.
[23] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, V. Aki, and D. B. Rubin, *Bayesian Data Analysis*, CRC Press, Boca Raton, FL, USA, 2013.
[24] D. Clancy and P. D. O'Neill, "Bayesian estimation of the basic reproduction number in stochastic epidemic models," *Bayesian Analysis*, vol. 3, no. 4, pp. 737–757, 2008.
[25] M. Alharthi, "Bayesian analysis of the predictive distribution of the SIR models," *Advances and Applications in Statistics*, vol. 75, no. 1, pp. 1–21, 2022.
[26] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, Charles Griffin & Company Ltd, New York, NY, USA, 1975.
[27] F. Ball, D. Mollison, and G. Scalia-Tomba, "Epidemics with two levels of mixing," *Annals of Applied Probability*, vol. 7, no. 1, p. 46, 1997.
[28] P. D. O'Neill, "Introduction and snapshot review: relating infectious disease transmission models to data," *Statistics in Medicine*, vol. 29, no. 20, pp. 2069–2077, 2010.
[29] R. H. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, New York, NY, USA, 1991.