Bayesian inference for non linear stochastic SIR epidemic model

Hamid El Maroufy, Tewfik Kernane, Sidali Becheket, Abdelah Ouddadj

Bayesian Inference for non linear stochastic SIR epidemic model

Hamid El Maroufy[1]
Department of Mathematics, Sultan Moulay Slimane University, Morocco.
T. Kernane, S. Becheket, A. Ouddadj
University of Sciences and Technology USTHB, Algiers, Algeria

Inference for epidemic parameters can be challenging, in part due to data that are intrinsically stochastic and tend to be observed by means of discrete-time sampling, which are limited in their completeness. The problem is particularly acute when the likelihood of the data is computationally intractable. Consequently standard statistical techniques can become too complicated to implement effectively. In this work, we develop a powerful method for bayesian paradigm for SIR stochastic epidemic models via data-augmented Markov Chain Monte Carlo. The latter samples missing values as well as the model parameters, where the missing values and parameters are treated as random variables. These routines are based on the approximation of the discrete-time epidemic by diffusion process. We illustrate our techniques using simulated epidemics.

Keywords: Epidemic model; Diffusion process; Conjugate distribution; MCMC simulation.
2010 Mathematics Subject Classification: 92D30, 60J28, 60J20, 60H10, 62F20.

1 Introduction

For readers the epidemics played a fundamental role in human history as one of the causes of misery, poverty and death during different centuries. From economic and human points of view, it is very important to have the means to achieve full understanding of the evolution of these phenomena. Since the late fifties, the development of mathematical and computational tools have enabled a more formal study of epidemics; therefore, several types of epidemic models using various techniques have been proposed such as the SIR (Susceptible-Infected- Removed) model, which is used to understand the evolution of complex infectious diseases in order to predict the impact of public health programs. The results depend on how clear are the estimates of the parameters governing these models. For SIR epidemiological systems, estimation for key parameters of interest (the contact infectious and the removal rates) is a crucial methodological problem. Various methods have been proposed to estimate these parameters from a time data, of which the more sophisticated rely on the calculation of a likelihood function (e.g. Becker (1989), Becker and Britton(1999), Becker et *al*.(2003) and Andersson and Britton (2000)). Likelihood based inference for epidemic models poses many challenges, not least because available epidemic data are often censored or incomplete, but also the likelihood requires the knowledge of the transition probabilities which are unavailable in simple closed form for this type of models (El Maroufy et all. (2012)). At the same time, others authors as Britton (1998), Becker and Hasofer (1997), Becker and Britton (1999) drew attention to this problem of estimation within the context of classical inference and described a martingale technique for non-likelihood approaches. However, although the estimator based on martingale estimating functions

[1]Address correspondence to Hamid El Maroufy, Department of Applied Mathematics, Faculty of Sciences and Technology, C. P : 523, Béni-Mellal, Morocco; E-mail : maroufy@fstbm.ac.ma

wich is consistent and asymptotically normally distributed, it is inefficient in general. Furthermore, computation of the standard errors of the resulting estimates is difficult and requires techniques such as parametric bootstrapping. Another alternative approach is the Bayesian paradigm. This approach is particularly suited to the context of epidemic modelling since the parameters of interest are usually defined in terms of an individual, which naturally leads one to consider the distributions of these parameters over a whole population. Up to now, almost all of the literature concerning Bayesian statistical inference use numerical techniques such as Markov Chain Monte Carlo (MCMC), which is being increasingly employed in epidemic modelling (e.g. O'Neill and Roberts (1999); Demiris and O'Neill (2005a, 2005b); and references therein), but the Bayesian methods used are distinct from that we present in this paper.

First, we approximate the discreet epidemic by diffusion process. Next, supposing that the diffusion model is given in parametric form, so the problem of parameter estimation can be brought back using inference for discretely observed diffusion process which has been widely studied through the like-lihood function by many authors (see for example Sorenønsen (2004), Beskos et *al.* (2006), Gilioli et *al.* (2008) and the reference therein). The maximum likelihood estimator, using discreet data, is consistent and asymptotically normal if the time intervals between consecutive observations are suf-ficiently small and under suitable conditions as the processes is ergodic (see, Prakasa Rao (1999)). However, for SIR epidemic models, the likelihood function is unknown, and in many situations the inter-observation times are large or latent; hence, the likelihood estimation may be unreliable and the ergodicity is very hard to prove. To overcome this difficulty, we will adopt the bayesian method proposed by Eraker (2001) in the context of financial models. The method consist of augmenting the low-frequency observations by insertion of finite number of latent data between two consecutive real observations. This task is performed by application of Markov Chain Monte Carlo (MCMC) tech-nique which alternately update the data and are usually feasible within moderate computing time. In particular we state that under the family of prior independent gamma distributions this leads to gen-eralized inverse gamma distributions as posterior distribution. The hybrid accept-reject Metropolis-Hastings (hereafter AR-MH) give more rapid convergence and requires knowledge of the unnormal-ized target density and a proposal density (see, Eraker (2001), Golithly and Wilkinson (2004)). So we suggest it should be with normal density as good proposal density.

The structure of this paper is as follows. In Section 2, we present a temporal discreet and non linear diffusion epidemic model. In section 3, we present a bayesian approach to estimate the parameter. To deal with a small number of observations, in Section 4, we introduce latent data between pairs of observations to improve the results. In Section 5, we illustrate some simulation examples.

2 Temporal SIR epidemic model

In the basic SIR epidemic model the population under consideration is classified into susceptible, infectious and recovered. We consider in this type of model a closed population of $n+a$ individuals. At time t, we denote the size of each category by $S(t)$, $I(t)$ and $R(t)$, respectively, so that $S(t)+I(t)+R(t)=n+a$. At time $t=0$ the population only contains susceptible and infected individuals with $S(0)=n$, $I(0)=a$. An infected individual remains infected, before being removed, for a random period with mean $1/\mu$. The lengths of stay in the state "infected" are assumed to be mutually independent. In a small time period dt, the probability of an individual to be infected is $I(t)\frac{\beta}{n}dt$. A susceptible individual after contact with an infected individual immediately becomes infectious. The mean number of new infected individuals is then $I(t)S(t)\frac{\beta}{n}dt$. The epidemic ends when there is no

infectuous in the population. This model is often known as the general stochastic epidemic, and is the most widely studied model. This model is analogous to the deterministic SIR model, defined in terms of ordinary differential equations (see, Bailey (1975)). The stochastic epidemic is completely determined by $\{(X(t),Y(t));t\geq 0\}$, which is a continuous time Markov Chain on the state space $\mathbf{E}=\{(i,j);\ 0\leq i\leq n,\ 0\leq j\leq (n-i)+a\}$. The transition probabilities from time t to $t+h$ are given by

$$\begin{array}{ccccc} & \text{Transition} & & & \text{Probability} \\ (i,l) & \longrightarrow & (i-1,l+1) & & \frac{\beta il}{n}h+o(h), \\ (i,l) & \longrightarrow & (i,l-1) & & \mu lh+o(h), \\ (i,l) & \longrightarrow & (i,l) & & 1-(\frac{\beta\ il}{n}h+\mu lh)+o(h). \end{array} \tag{2.1}$$

For $(i,l)\in\mathbf{E}$, we define $P_{il}(t)=P\{S(t)=i,I(t)=l\}$. It follows directly from (2.1) that these transition probabilities satisfy the set of Kolmogorov forward equations:

$$P'_{(i,l)}(t)=\frac{\beta(i+1)(l-1)}{n}P_{(i+1,l-1)}(t)+\mu(l+1)P_{(i,l+1)}(t)-(\mu l+\frac{\beta il}{n})P_{(i,l)}(t)$$

for $(i,l)\in\mathbf{E}$, with $P_{il}(t)\equiv 0$ if $(i,l)\notin\mathbf{E}$ and $P_{na}(0)=1$.
A Markov process with above described dynamics is termed the general stochastic epidemic.
Instead of using S and I, we normalise the process by the transformations $x(t)=S(t)/n$ and $y(t)=I(t)/n$. By setting $f(x,y)=n\beta xy=\frac{\beta}{N}SI$ and $g(x,y)=n\mu I=\mu y$ the Kolmogorov's equations become

$$p'(x,y,t)=f(x+\varepsilon,y-\varepsilon)p(x+\varepsilon,y-\varepsilon,t)+g(x,y+\varepsilon)p(x,y+\varepsilon,t)-[f(x,y)+g(x,y)]p(x,y,t) \tag{2.2}$$

where $\varepsilon=\frac{1}{n}$ and $p(x,y,t)=p_{(x,y)}(t)$. By substracting and adding terms to equation (2.2) and letting $\varepsilon\to 0$, we establish, by setting $\mathbf{y}=(x,y)$ (see Fuchs (2013) for rigorous proof), that

$$\frac{\partial}{\partial t}p(\mathbf{y},t)=-\frac{\partial}{\partial\mathbf{y}}[U(\mathbf{y},\theta)p(\mathbf{y},t)]+\frac{1}{2}\frac{\partial^2}{\partial\mathbf{y}^2}[\Sigma(\mathbf{y},\theta)P(\mathbf{y},t)] \tag{2.3}$$

with $\theta=(\beta;\mu)$, $U(\mathbf{y},\theta)=\binom{-\beta xy}{\beta xy-\mu y}$ and $\Sigma(\mathbf{y},\theta)=\frac{1}{n}\begin{pmatrix}\beta xy & -\beta xy\\ -\beta xy & \beta xy+\mu y\end{pmatrix}$.
Equation (2.3) is the Fokker-Planck equation associated to the diffusion process $(x(t),y(t))$ which is solution, according to Øksendal (1995) (see also, Tory (2000) and Kloeden and Platen (1999)), of the non-linear bivariate Itô stochastic differential equation :

$$\begin{pmatrix}dx\\ dy\end{pmatrix}=\begin{pmatrix}-\beta xy\\ \beta xy-\mu y\end{pmatrix}dt+\sigma(x,y)\begin{pmatrix}dW_1\\ dW_2\end{pmatrix} \tag{2.4}$$

where $\sigma(x,y)=\frac{1}{\sqrt{n}}\begin{pmatrix}\sqrt{\beta xy} & 0\\ -\sqrt{\beta xy} & \sqrt{\beta y}\end{pmatrix}$. The conditions under which the SDE given by (2.4) can be solved for (x,y) are satisfied (see Anderson and Britton (2000)). The right hand side of the differential equation (2.4) consists of the deterministic (see Bailey (1975)) and the stochastic components W_1 and W_2 which are two standard independent Brownian motions, representing stochastically in disease transmission and recovers.
The parameter θ is unknown and has to be estimated. However, the objective of this paper is to illustrate an efficient method to estimate it because the quantity $R_0=\frac{\beta}{\mu}$, which design the basic

reproduction number (the average number of new infections caused by a single infective in a large susceptible population). This quantity is used to interpret the threshold behaviour of the epidemic; a large outbreak can occur if and only if $R_0 > 0$. A prior distribution will be specified on $\theta = (\beta, \mu)$ and inference will be performed based upon its joint posterior distribution using a Bayesian approach.

3 Inference for non-linear diffusion model

For simplicity, we write the system (2.4) in vectorial form:

$$dY_t = U(Y_t, \theta)dt + \sigma(Y_t, \theta)dW(t) \tag{3.1}$$

where $U : \mathbb{R}^2 \times \Theta \to \mathbb{R}^2$ and $\sigma : \mathbb{R}^2 \times \Theta \to \mathbb{R}^{2\times 2}$, for some compact parameter set $\Theta \subseteq \mathbb{R}^2$. We assume that the process $Y(t)$ will be observed at a finite integer of times and the objective is to conduct inference for the (unknown) parameter vector θ on the basis of these partial and discrete observations on $Y(t)$. In practice it is necessary to work with the discretized version of (3.1), given by the Euler approximation

$$\Delta Y_t = U(Y_t, \theta)\Delta t + \sigma(Y_t, \theta)\Delta W(t), \tag{3.2}$$

where $\triangle W_t$ is two dimensional iid $N(0_2, I_2 \triangle t)$ random vector.
Diffusion processes provide an analytical tractability, but the parameters that govern their dynamics are often difficult to estimate from the data obtained after discretization of time. In short, the estimation problem is that the model is formulated in continuous time, while the sampling data are naturally available only at discrete frequencies and perhaps with a low frequency. The Eraker's approach is based on the introduction between each pair of observations of $m-1$ latent data points. The method is based on the MCMC (Monte-Carlo-Markov-Chain) technique applied to a wide class of models, including models with incomplete data. It is assumed that the observations are collected in discrete integer times but by do not putting $\triangle t = 1$ this situation does not approximate well the transition density. Hence, to ensure that the discretization bias is arbitrarily small we put $\triangle t = 1/m$, for conveniently chosen positive integer m. The time interval $[0,T]$ is subdivided into $m \times T$ equidistant points $0 = t_0 < t_1 < .... < t_{n-1} < t_N = T$. Let

$$\hat{Y} = \begin{pmatrix} x_{t_0} & \widehat{x}_{t_1} & . & . & . & \widehat{x}_{t_{m-1}} & x_{t_m} & \widehat{x}_{t_{m+1}} & . & . & . & x_{t_N} \\ y_{t_0} & \widehat{y}_{t_1} & . & . & . & \widehat{y}_{t_{m-1}} & y_{t_m} & \widehat{y}_{t_{m+1}} & . & . & . & y_{t_N} \end{pmatrix}$$

be the matrix of all (real and latent) data. We shall denote by $Y_{t_i} = (x_{t_i}, y_{t_i})$ the real datum at time t_i when i is a multiple integer of m and $\widehat{Y}_{t_i} = (\widehat{x}_{t_i}, \widehat{y}_{t_i})$ the latent datum and, generically, by $\hat{Y}_{t_i}$ a real or latent datum depending on the context. Altogether, $2 \times T \times (m-1)$ data points are missing from the system.
Conditionally to the first observation the joint posterior density is given by

$$\pi(\hat{Y}, \theta) \propto \prod_{i=1}^{N} p(\hat{Y}_i|\theta)p(\theta) \tag{3.3}$$

where

$$p(\hat{Y}_i|\theta) = |\Sigma_{i-1}^{-1}|^{12} \mathbf{e}^{-\frac{1}{2}(\triangle \hat{Y}_i - U_{i-1}\triangle t)'(\Sigma_{i-1}\triangle t)^{-1}(\triangle \hat{Y}_i - U_{i-1}\triangle t)}, \tag{3.4}$$

with $\triangle \hat{Y}_i = \hat{Y}_i - \hat{Y}_{i-1}$, $U_{i-1} = U(\hat{Y}_{i-1}; \theta)$, $\Sigma_{i-1} = \sigma(\hat{Y}_{i-1}; \theta)\sigma^T(\hat{Y}_{i-1}; \theta)$, $|\Sigma_i^{-1}|$ is the determinant of the inverse matrix of Σ_i and $p(\theta)$ is the prior density of the parameter. From now on, we adopt the notation where $\pi(\theta)$ denotes all proper densities, p denotes π in an unnormalized form. Note that all posterior densities of the parameters conditionally to the observations are proportional to (3.3).

4 Gibbs sampler

We have formulated in (3.3) the joint posterior for the model parameters as well as observed and unobserved data but our real interest is in the distribution of $(\theta, \hat{Y} \setminus \hat{Y}_{obs} | \hat{Y}_{obs})$ where Y_{obs} denotes the observed data up to time t_n. As discussed in Golihtly and Wilkinson (2004) (see also the related references therein) the inference may proceed by alternating between the simulated data and parameters conditional on augmented data and the current state of the parameters. As in the epidemiological context the number of missing data is relatively large, a Gibbs sampler is suitable for sampling independently the quantities. At first step, we consider a guess value for θ drawn from the prior distribution $\pi(\theta)$ and generate latent data as shown in the next section. At this point, we generate a new value of θ, then we update latent observations. By repeating, this process generates a Markov chain which has the desired posterior, $\pi(\theta, \hat{Y} \setminus \hat{Y}_{obs} | \hat{Y}_{obs})$, as its stationary distribution. For an overview of the use of Markov chains for exploring the posterior distributions we refer you back to the works of Tierney (1994).

4.1 Simulating the latent data

For the univariate diffusion processes, the MCMC method have been extensively examined, for example in financial models Roberts and Stramer (2001), Elerian et *al.* (2001) and Durham and Gallant (2002) employ block updating schemes to simulate the latent data. For our bivariate partially observed model, due to the high dimensionality (large number missing data and parameters) it is convenient to use Gibbs sampler. The first step in the Gibbs sampler involves the simulation of missing data. We use Eraker's method (see Eraker (2001)) to generate one latent observation $\hat{Y}_i$ given $\hat{Y}_{i-1}, \hat{Y}_{i+1}$. So we know

$$\pi(\hat{Y}_i | \hat{Y}_{\setminus i}, \theta) \propto p(\hat{Y}_i | \hat{Y}_{i-1}, \hat{Y}_{i+1}; \theta) \tag{4.1}$$

where

$$\begin{aligned} p(\hat{Y}_i | \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta) \; &= |\Sigma_{i-1}^{-1}|^{12} \times |\Sigma_i^{-1}|^{12} \times \\ &\mathbf{e}^{-\frac{1}{2}\left[(\triangle\hat{Y}_i - U_{i-1}\triangle t)'(\Sigma_{i-1}\triangle t)^{-1}(\triangle\hat{Y}_i - U_{i-1}\triangle t) + (\triangle\hat{Y}_{i+1} - U_i\triangle t)'(\Sigma_i\triangle t)^{-1}(\triangle\hat{Y}_{i+1} - U_i\triangle t)\right]} \end{aligned} \tag{4.2}$$

and $\hat{Y}_{\setminus i}$ designates the matrix $\hat{Y}$ without the ith column.
The density in (4.2) is not a density of standard form, so we suggest to use the AR-MH algorithm, because this algorithm is fast and only requires knowledge of the unormalized proposal density q that can be sampled from. As motivated in Eraker (2001), when i is not a multiple of m, $\hat{Y}_i$ is updated using the normal distribution $N(\frac{1}{2}(\hat{Y}_{i-1} + \hat{Y}_{i+1}), \frac{1}{2}\Sigma_{i-1}\triangle t)$ as a good proposal density.

4.2 Sampling from the full conditional for the parameter

The remaining step in either the Gibbs sampler is to sample $\theta^{(h)}$ the value of θ, conditional to its current state and the augmented data, at h-th iteration. We consider the step $h+1$, once we have simulated the matrix, the parameter $\theta^{(h+1)}$ must be generated. If the posterior density of the parameter does not have a known usual distribution, we might still impose a Metropolis-Hastings step, but it is not the case her. Due to the form of likelihood function derived in (3.3), a family of independent gamma distribution is seen a natural set of conjugate priors chiefly in the context of epidemic model where the parameters are positive (see, Streftaris and Gibson (2004), Demeris and O'Neill (2005a, 2005b)

and the reference therein). Under this family of priors the application of Bayes' theorem produces as posteriors generalized inverse gaussian distributions. More formally, we have the following

Proposition 4.1. *If the* β *and* μ *follow independent gamma prior distributions:* $\pi(\beta) \propto \Gamma(m_\beta, \lambda_\beta, \beta)$ *and* $\pi(\mu) \propto \Gamma(m_\mu, \lambda_\mu, \mu)$. *Then*

$$\pi(\beta|\hat{Y}) \propto \beta^{m_\beta - \frac{N}{2} - 1} \mathbf{e}^{-\frac{1}{2}\left[(B_2 + 2\lambda_\beta)\beta + \frac{B_1}{\beta}\right]}, \tag{4.3}$$

and

$$\pi(\mu|\hat{Y}) \propto \mu^{m_\mu - \frac{N}{2} - 1} \mathbf{e}^{-\frac{1}{2}\left[(C_2 + 2\lambda_\mu)\mu + \frac{C_1}{\mu}\right]}, \tag{4.4}$$

where B_2, C_2, B_1 *and* C_1 *are constants calculated from data and* $\Gamma(m, \lambda, \alpha) = \alpha^{m-1}\mathbf{e}^{-\lambda\alpha}$ *for* $\alpha \in \mathbb{R}_+^*$ *and* m, λ *are positive constants.*

Proof: The proof of the previous statement is presented in the Appendix **1**.

As mentioned above, the question of whether or note $R_0 > 1$ is often of interest, we will consider this concept in practice. We will use as the prior of R_0. In order to do this, the following facts are used in the sequel. Since $R_0 = \beta/\mu$ then its prior density is given by Clancy and O'Neill (2008)

$$\pi(R_0) = \left(\frac{\lambda_\beta}{\lambda_\mu}\right)^{m_\beta} \times \frac{\Gamma(m_\beta + m_\mu)}{\Gamma(m_\beta) + \Gamma(m_\mu)} \times \frac{R_0^{m_\beta - 1}}{\left(\frac{\lambda_\beta}{\lambda_\mu} R_0 + 1\right)^{m_\beta + m_\mu}} \text{ with } R_0 > 0,$$

with prior mean and variance given by

$$\mathbb{E}(R_0) = \frac{m_\beta \lambda_\mu}{(m_\mu - 1)\lambda_\beta}$$

and

$$\mathbb{V}ar(R_0) = \frac{m_\beta(m_\beta + m_\mu - 1)}{(m_\beta - 1)^2(m_\mu - 2)} \left(\frac{\lambda_\beta}{\lambda_\mu}\right)^2.$$

We see that if $m_\mu \leq 1$, R_0 has negative prior mean and if $m_\mu \leq 2$ R_0 has negative prior variance. In other words, such vague priors on β and μ yield inaccurate prior for R_0, this means that it is possible to consider the posterior mean of R_0 as a suitable summary measure. Now, assume that $m_\beta = m_\mu = m$ and $\lambda_\beta = \lambda_\mu = \lambda$; a typical case in practise will be $m > 1$ and λ a small positive number, this gives $\mathbb{E}(R_0) > 1$ and means that the epidemic is above its threshold or maybe not. This suggests some need for wariness in using the mean as the sole means of assessing whether or not an epidemic is above threshold. However, the alternative is to choose m and λ such that $\mathbb{P}(0 \leq R_0 \leq 1) = 1/2$ in this case, the epidemic shall have the same probability to be below or above the threshold.

4.3 Algorithm

The following algorithm summarizes our strategy of simulation:

(*i*)- Initialize all missing data of the matrix $\hat{Y}$ using linear interpolation between two observed values and initialising parameters $\beta^{(0)}$ and $\mu^{(0)}$ according to the principe in the last paragraph.

(*ii*)- For $i = 1,...,n$ at iteration h we use the AR-MH for drawing an observation using a normal distribution $\mathcal{N}(\frac{Y_{i-1}+Y_{i+1}}{2}, \frac{\Sigma_{i-1}\times\Delta t}{2})$.

(*iii*)- Drawing β using (4.3).

(*iv*)- Drawing μ using (4.4).

(*v*)- Increase the value of h and go back to step (*ii*).

As mentioned in Proposition **4.1**, if the prior distribution of parameters is a gamma distribution, then the posterior is generalized inverse Gaussian. It is not easy to sample directly from this distribution (see Hörmann and Leydold (2014) and the reference given there). In this case a Metropolis-Hasting algorithm can be used in the steps 3 and 4. We consider, as proposal density for the M-H algorithm, a gamma distribution with special parameters. We proceed as follows : At each iteration, we calculate the quantity b given by $b = \max_{\beta\in\mathbb{R}_+} p(\beta^{(h)}\widehat{Y}^{(h)})$. Given that the mean and variance of a gamma distribution $\Gamma(\gamma,\eta)$ are given by $\frac{\gamma}{\eta}$ and $\frac{\gamma}{\eta^2}$ respectively, we take $\gamma = k\times b$ and $\eta = k$ where k is a positive integer which we can be calculated by viewing the shape of the posterior distribution after initialising the data.

5 Simulation study

To illustrate the methodology presented in this paper, the MCMC scheme is applied to the general SIR epidemic model with known parameters. The observable part of the data comes from the simulation of original double Markov chain using the exact Gillespie algorithm given in Gibson and Bruck (2000) with true values of the parameters as given in the Table 1. For each data set, due to computational demands, the MCMC sampler is run for 20000 iterations with $m = 5$, $m = 10$ and $m = 15$. Note that the algorithm is coded in R and executed on Laptop i3. The application of the method outlined her improves the estimate of β and μ. Looking at Table **1** and Figure **1** which summarise the posterior distribution; Table **1** give posterior mean and variance. We see that the estimates are close to the true value as the number of latent data increase. There is a considerable improvement up to a certain value of m, while for greater value the improvement is less pronounced. The histograms in Figure **1** reveal the convergence of the algorithm towards a limit distribution. In the case of a true value of parameters is not available, It should be better to choose m as large as possible but this leads to huge computational cost and the balance between the computational cost and goodness of fit steers us to consider a small value of m.

6 Conclusion

In this paper, we have provided a fully Bayesian approach to estimate the SIR epidemic parameters, when the populations of individuals are large, by adopting a diffusion approximation. We are essentially concerned with the Bayesian analysis of nonlinear, discretely observed stochastic differential equations. We have shown that although the SDE approximation is often adequate for simulation in the context of Bayesian inference, this approach is undertaken by adopting a Gibbs sampler.

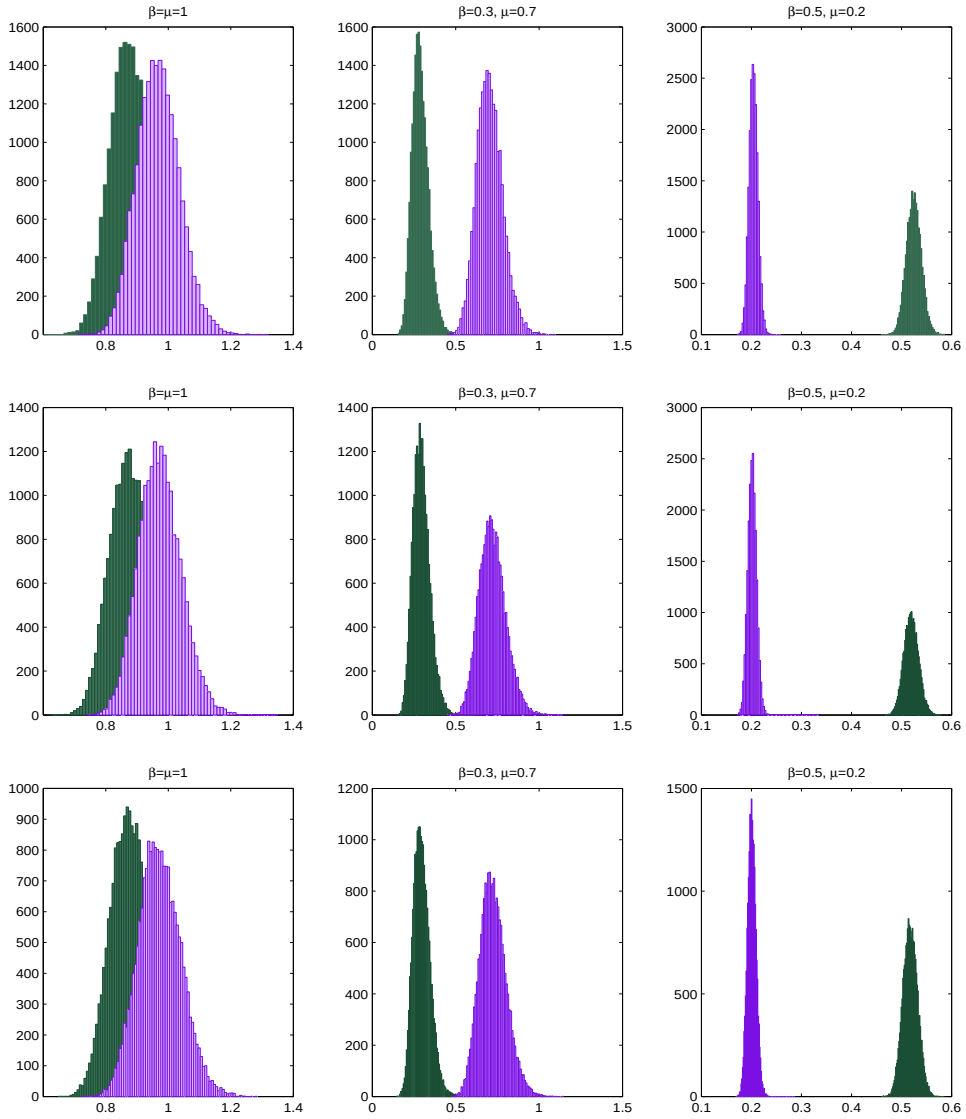

Figure 1: Frequency histograms based on 10000 for posterior densities estimate of β $(green)$, and μ $(violet)$, for $m = 5$, $m = 10$ and $m = 15$ with born-in of 20000 iterations in all cases. (For interpretation of the references to colour in the legend of this figure the reader is referred to the electronic version of this article).

	True values of β	Posterior of β	True values of μ	Posterior of μ	$\hat{R}_0$
		m=5			
mean	1	0.8791245	1	0.9685208	0.9076982
sd		0.06521707		0.06752782	
mean	0.3	0.2886688	0.7	0.704504	0.4097476
sd		0.04927155		0.07741329	
mean	0.5	0.5245903	0.2	0.2040725	2.570608
sd		0.0149452		0.008621898	
		m=10			
mean	1	0.9788925	1	0.9734421	1.0055991
sd		0.0660991		0.06778218	
mean	0.3	0.2949151	0.7	0.7203489	0.409406
sd		0.0507018		0.07960507	
mean	0.5	0.5194062	0.2	0.201533	2.577276
sd		0.01474604		0.00857833	
		m=15			
mean	1	0.9791271	1	0.9739063	1.0053606
sd		0.06481448		0.06833605	
mean	0.3	0.2959734	0.7	0.7238514	0.408887
sd		0.04982702		0.07998687	
mean	0.5	0.5174422	0.2	0.2006646	2.578642
sd		0.01470055		0.00846525	

Table 1: Posterior means and standard deviations for β, μ and R_0 for $\Delta t = 1/m$ in all cases.

Instead a classical normal distribution as prior, which is not realistic because it allows negative value of the parameters, we have proposed Gamma (on the positive real set) distributions. We found that these distributions lead to the best estimation.
The MCMC fit obtained can be considered satisfactory taking into account the computational difficulties. Further research will be devoted to further improving the fit by considering more efficient algorithms based on block updating of missing data (Durham and Gallant (2002)) in addition to consider the hyperbolic diffusion based on the discretized density via the Milstein scheme. Also the method established in this paper can be generalized to other SIR epidemic models.

Appendix 1

To prove the Proposition 4.1, we first consider the following process

$$\begin{aligned} h(t) &= g(z(t)) = \left(x(t), 1 + \frac{a}{n} - x(t) - y(t)\right) \\ &= (x(t), z(t)), \end{aligned}$$

by Itô formula, we found

$$dh(t) = U(h(t),\theta)\triangle t + \sigma(h(t),\theta)\triangle W_t$$

where

$$\begin{aligned} U(h(t),\theta) &= \begin{pmatrix} -\beta x(t)[1+\frac{a}{n}-x(t)-z(t)] \\ \mu(1+\frac{a}{n}-x(t)-z(t)) \end{pmatrix} \\ \text{and } \sigma(h(t),\theta) &= \frac{1}{\sqrt{n}} \begin{pmatrix} -\sqrt{\beta x(t)[1+\frac{a}{n}-x(t)-z(t)]} & 0 \\ 0 & \sqrt{\mu(1+\frac{a}{n}-x(t)-z(t))} \end{pmatrix}. \end{aligned}$$

Then, by (3.4) it is obvious that

$$p(\hat{Y}_i|\theta) = |\Sigma_{i-1}^{-1}|^{12}\mathbf{e} - \frac{1}{2}(\triangle\hat{Y}_i - U_{i-1}\triangle t)'(\Sigma_{i-1}\triangle t)^{-1}(\triangle\hat{Y}_i - U_{i-1}\triangle t) \tag{6.1}$$

where

$$\Sigma_{i-1} = \begin{pmatrix} \frac{\beta x_{i-1}(t)[1+\frac{a}{n}-x_{i-1}-z_{i-1}]}{n} & 0 \\ 0 & \frac{\mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})}{n} \end{pmatrix}.$$

Since Σ_{i-1} is diagonal, then

$$\begin{aligned} &(\triangle\hat{Y}_i - U_{i-1}\triangle t)^T(\Sigma_{i-1}\triangle t)^{-1}(\triangle\hat{Y}_i - U_{i-1}\triangle t) = \\ &\left(\frac{n}{\triangle t}\right)\left[\frac{\left[\triangle x_i + \beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]\triangle t\right]^2}{\beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]} + \frac{\left[\triangle z_i - \mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})\triangle t\right]^2}{\mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})}\right] \end{aligned}$$

it follows immediately from (6.1), that

$$p(\hat{Y}_i|\theta) = p(\hat{Y}_i|\beta).p(\hat{Y}_i|\mu) \tag{6.2}$$

where

$$p(\hat{Y}_i|\beta) = \sqrt{\frac{n}{\beta x_{i-1}(1+\frac{a}{n}-x_{i-1}-z_{i-1})}}\mathbf{e}^{\frac{-n}{2\triangle t}\times\frac{\left[\triangle x_i+\beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]\triangle t\right]^2}{\beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]}}$$

and

$$p(\hat{Y}_i|\mu) = \sqrt{\frac{n}{\mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})}}\mathbf{e}^{\frac{-n}{2\triangle t}\times\frac{\left[\triangle z_i-\mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})\triangle t\right]^2}{\mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})}}.$$

By independence of β and μ (3.3) becomes

$$\begin{aligned} \pi(\hat{Y},\theta) \quad \propto \quad & \left[\prod_{i=1}^{n}\beta^{-\frac{1}{2}}\mathbf{e}^{\frac{-n}{2\triangle t}\times\frac{\left[\triangle x_i+\beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]\triangle t\right]^2}{\beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]}}p(\beta)\right] \\ & \times\left[\prod_{i=1}^{n}\mu^{-\frac{1}{2}}\mathbf{e}^{\frac{-n}{2\triangle t}\times\frac{\left[\triangle z_i-\mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})\triangle t\right]^2}{\mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})}}p(\mu)\right], \end{aligned} \tag{6.3}$$

So, if replacing $p(\beta)$ by $\beta^{m-1}\exp(-\lambda\beta)$, $\beta\in\mathbb{R}_+$ and $p(\mu)$ respectively by $\mu^{m-1}\exp(-\lambda\mu)$, $\mu\in\mathbb{R}_+$, we obtain from (6.3) the following:

$$\pi(\beta|\hat{Y}) = \beta^{-\frac{N}{2}+(m-1)}\mathbf{e}^{\frac{-n}{2\triangle t}\times\sum_{i=1}^{N}\frac{\left[\triangle x_i+\beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]\triangle t\right]^2}{\beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]}-\lambda\beta}; \beta>0 \tag{6.4}$$

and

$$\pi(\mu|\hat{Y}) = \mu^{-\frac{N}{2}+(m-1)}\mathbf{e}^{\frac{-n}{2\triangle t}\times\sum_{i=1}^{N}\frac{\left[\triangle z_i-\mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})\triangle t\right]^2}{\mu(1+\frac{a}{n}-x_{i-1}-z_{i-1})}-\lambda\mu}; \mu>0. \tag{6.5}$$

We now proceed by regarding the quantities in the exponentials, we have

$$\begin{aligned}
&\sum_{i=1}^{N} \frac{\left[\triangle x_i + \beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]\triangle t\right]^2}{\beta x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]} \\
= \; & \underbrace{\left[\sum_{i=1}^{N} \frac{(\triangle x_i)^2}{x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}]}\right]}_{A_1} \frac{1}{\beta} + \beta \underbrace{\left[\sum_{i=1}^{N} x_{i-1}[1+\frac{a}{n}-x_{i-1}-z_{i-1}](\triangle t)^2\right]}_{A_2} + \underbrace{\sum_{i=1}^{N}(2\triangle x_i \triangle t)}_{A_3}.
\end{aligned}$$

By setting $B_1 = \frac{n}{\triangle t}A_1$, $B_2 = \frac{n}{\triangle t}A_2$ and $B_3 = \frac{n}{\triangle t}A_3$, then from (6.4) we can easily conclude that, for $\beta \in \mathbb{R}_+$,

$$\pi(\beta|\hat{Y}) \propto \beta^{(m-\frac{N}{2}-1)} \mathbf{e}^{-\frac{1}{2}\left[(B_2+2\lambda)\beta+\frac{B_1}{\beta}\right]}, \tag{6.6}$$

where $m-\frac{N}{2} \in \mathbb{R}$; $B_2+2\lambda \in \mathbb{R}_+^*$; $B_1 \in \mathbb{R}_+^*$. This is the first assertion of the Proposition (4.1).
In the same manner we deduce from (6.5) that

$$\pi(\mu|\hat{Y}) \propto \mu^{m-\frac{N}{2}-1)} \mathbf{e}^{-\frac{1}{2}\left[(C_2+2\lambda)\mu+\frac{C_1}{\mu}\right]}, \tag{6.7}$$

where

$$\begin{aligned}
C_1 &= \frac{n}{\triangle t}\left[\sum_{i=1}^{N} \frac{(\triangle z_i)^2}{[1+\frac{a}{n}-x_{i-1}-z_{i-1}]}\right] \\
C_2 &= \frac{1}{\triangle t}\left[\sum_{i=1}^{N}[1+\frac{a}{n}-x_{i-1}-z_{i-1}](\triangle t)^2\right] = \sum_{i=1}^{N}[1+\frac{a}{n}-x_{i-1}-z_{i-1}](\triangle t).
\end{aligned}$$

■

References

[1] Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis*. Springer Lecture Notes in Statistics, New York.

[2] Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and Its Application*. 2nd edn. Griffin, London.

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

[4] Becker, N. G. and Britton, T. (1999). Statistical studies of infectious disease incidence. Journal of the Royal Statist Soc B 61: 287-308.

[5] Becker, N. G., Britton, T. and O'Neill, P. D. (2003). Estimating vaccine effects on transmission of infection from household outbreak data. Biometrics 59: 467-475.

[6] Becker, N. G. and Hasofer, A. M. (1997) Estimation in epidemics with incomplete observations. J. R. Statist. Soc. B, 59: 415-429.

[7] Beskos, A., Papaspiliopoulos, O., Roberts, G.O and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). Journal of the Royal Statist Soc B 68, 333-382.

[8] Britton, T. (1998). Estimation in multitype epidemics. J. R. Statist Soc. B, 60: 663-679.

[9] Clancy, D; O'Neill Philip D. (2008). Bayesian estimation of the basic reproduction number in stochastic epidemic models. Bayesian Anal. Vol. 3, No 4: 737-757.

[10] Demiris, N. and O'Neill, P. D. (2005a). Bayesian inference for epidemics with two levels of mixing. Scand. J. Stat 32: 265-280.

[11] Demiris, N. and O'Neill, P. D. (2005b). Bayesian inference for stochastic multitype epidemics in structured populations via random graphs. J. Roy. Statist. Soc. B 67: 731-745.

[12] Durham, G. B. and Gallant, R. A. (2002). Numerical techniques for maximum likelihood estimation of continuoustime diffusion processes. Journal of Business and Economic Statistics, 20: 279-316.

[13] Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69: 959-993.

[14] El Maroufy, H., Omari, L. and Taib, Z. (2012). Transition probabilities for generalized SIR epidemic model, *Stoch. Models.* 28(1) : 15-28.

[15] Eraker, B. (2001). MCMC Analysis of diffusion models with application to finance. Journal of Business & Economic Statistics. Vol. 19, No 2: 177-191.

[16] Fuchs, C. (2013). *Inference for diffusion processes with application in life sciences*. Springer.

[17] Gibson, G, J. and Bruck, J. (2000). Efficient exact stochastic simulation of chimical sysytems with many species and many channels. Journal Physical Chemestry Serie A 104: 1876-1889.

[18] Gilioli, G., Pascali, S and Ruggeri, F. (2008). Bayesian inference for function response in stochastic predator-prey system. 70. 358-381.

[19] Golightly, A. and Wilkinson, D. J. (2004). On Bayesian inference for stochastic kinetic models using diffusion approximations. Statistics Preprint STA04, 4, University of Newcastle, U.K.

[20] W. Hörmann, J. Leydold .(2014). Generating generalized inverse Gaussian random variates. Statistics and Computing. 24, 4: 547-557.

[21] Kloeden,P. E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer.

[22] Øksendal, B. (1995). *Stochastic Differencial Equations*. Springer.

[23] O'Neill, P. D. and Roberts, G. O. (1999). Bayesian inference for partially observed stochastic epidemics. J. R. S. S., Series A 162: 121-129.

[24] Prakasa, R. (1999). *S*tatistical Inference for Diffusion Type Processes. Arnold, London.

[25] Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. Biometrika 88: 603-621.

[26] Sørensen, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: a survey. International Statistical Review, vol 72, no. 3, 337-354.

[27] Streftaris, G. and Gibson, G.J. (2004). Bayesian inference for stochastic epidemics in closed populations. Statistical Modelling, 4: 63-75.

[28] Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). The Annals of Statistics 22: 1701-1762.

[29] Tory, E. (2000). Stochastic sedimentation and hydrodynamic diffusion. Chimical Engineering Journal 80: 81-89.