Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 12, Number 2 (2016), pp. 1299-1306

http://www.ripublication.com

Parameter Estimation of SIR Epidemic Model Using MCMC Methods

A. S. Talawar and U. R. Aundhakar*

Department of studies in Statistics,
Karnatak University, Dharwad, Karnataka-580003, India.
**Corresponding Author*

Abstract

An understanding of the spread of infection from one individual to another individual is a very important concern in order to prevent major outbreaks. In this work stochastic modelling is used to gain insight into the dynamics of an epidemic. This work deals with a commonly used simple SIR-model; a compartment model with three states: susceptible, infected, and recovered. We have described the changes in the state variable of the model and also estimated model parameters by taking into account recovery and infection times are continuous. A Markov Chain Monte Carlo (MCMC) approach is used to handle the estimation of different parameters.

Keywords: infectious disease, MCMC, SIR model, Monte Carlo simulation, epidemic.

Introduction

Now a day's infectious disease spread has a great concern to public health analytics. Hence it is essential to design strategies for managing disease threat to humans and it is possible through infectious disease modelling. In the present work, we have demonstrated framework for describing infectious diseases and estimate its parameters to fit specific observations on recovery and infection times for continuous time SIR-model. The modelling of disease transmission behavior takes a practical approach to the area of simulation.

Initially, well-known formulas on how to obtain maximum likelihood estimates for β and γ in case of full observations are derived. Hereafter, an approach for handling estimation, in case of missing infection times is presented through Bayesian analysis using Markov Chain Monte Carlo (MCMC) methods. Our goal is to obtain a sufficient understanding for creating an own implementation of the SIR-estimation procedures

with the help of Markov Chain Monte Carlo (MCMC) methods of model parameters. The transmission rate is the most difficult parameter to estimate in model. Some attempts have been made to establish it 'bottom up' from a priori knowledge of host and disease behavior, to predict probable disease dynamics and control in a host in which it had not yet become established [1]. De Leo and Dobson [2] have suggested a method based on 'Allometry' that might provide order of magnitude estimates of the transmission rate in the absence of other data. Two other approaches are more commonly used. One is to deduce the transmission coefficient and the form of the transmission function from results of experiments [3]. The second is to deduce it from observations of disease behaviour in the field, in particular Prevalence and dynamic responses to perturbations such as control of the host (such as culling). Finkenstadt and Grenfell [4] have recently developed a statistically rigorous method for estimating transmission rates from a time series of pathogen prevalence.

Earlier, it is shown that, the behavior of Host–pathogen models is greatly affected by the way in which transmission between infected and susceptible hosts are modelled [5]. Hohle M. and Jorgensen E. [6] did mathematical modelling to gain insight into the dynamics of an epidemic. They have given a rigorous treatment of an existing technique to handle estimation in partially observed epidemics using Markov Chain Monte Carlo (MCMC). The aim of this report is to extend the basic SIR model to handle two common situations: course of the epidemic and population heterogeneity due to the spatial layout of confinement.

Model

The present study consists of stochastic version of a standard susceptible-Infected-Removed (SIR) model through infectious disease which indicates the status of individuals from one state to another state in which susceptible class S to the infective class I to the removed class R.

Here state variables at time t are; $S(t)$ is the number of susceptible individuals, I(t) is the number of infected individuals, $R(t)$ is the number of recovered individuals and $D(t)$ is the number of removed/dead individuals. In a standard SIR model we considered all the Recovered and Dead individuals into a single removed class R with the assumption of population is closed to immigration or emigration so that $S(t) + I(t) + R(t) + D(t) = N$, where N is constant. Now, SIR model [5,7,8] is described as follows

$$\frac{dS}{dt} = -kSln\left(1 + \frac{\beta I}{k}\right) \quad (1)$$

$$\frac{dI}{dt} = kSln\left(1 + \frac{\beta I}{k}\right) - (\gamma + \mu)I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

$$\frac{dD}{dt} = \mu I \quad (4)$$

Where the model parameters are described as β is transmission rate, k is over dispersion

parameter, μ is death rate and γ is rate of recovery to the immune class. The negative binomial distribution can be taken as a compound stochastic process in which encounters between infected and susceptible individuals occur randomly. This SIR model formulation leads to a natural discrete time approximation for the number of infections($\hat{I}$), recoveries($\hat{R}$), and deaths $(\hat{D})$ arising in the unit time interval from t to $(t+1)$. Assuming the total number of infected individuals I, is approximately constant and integrating Eq. (1) over a unit time interval gives

$$S(t+1) = S(t)\left(\frac{k}{k+\beta I(t)}\right)^k, \tag{5}$$

so that the fraction of susceptible individuals surviving throughout a unit time interval is $\left(\frac{k}{k+\beta I(t)}\right)^k$. When viewed as a discrete time stochastic process, where the mean number of remaining susceptible individuals is given by Eq. (5), the mean number of new infections occurring between time t and $(t+1)$ is

$$S(t)\left[1-\left(\frac{k}{k+\beta I(t)}\right)^k\right]. \tag{6}$$

Therefore, if S(t) = s and I(t) = i, we may sensibly take the new infections Î at time (t + 1) to follow

$$\hat{I}|s,i \sim Bin(s, p_i(i,\beta,k)), \text{ where } p_i(i,\beta,k) = 1-\left(\frac{k}{k+\beta i}\right)^k \tag{7}$$

Similarly, by integrating Eqs (2 and 3), the number of recoveries and deaths occurring between time t and (t + 1) can be described by

$$\hat{R}|i \sim Bin(i, p_r), \text{ where } p_r = 1-e^{-\gamma} \tag{8}$$

$$\hat{D}|i,\hat{r} \sim Bin(i-\hat{r}, p_d), \text{ where } p_d = 1-e^{-\mu} \tag{9}$$

Here, (s,i,r) denote the realized value of the associated capital letter random variable. In this discrete time approximation we have assumed a particular ordering of events, namely that recoveries occur first, followed by deaths from among those infected individuals who did not recover, followed by new infections. Simulation studies indicated that these assumptions, as well as other possible orderings, resulted in system dynamics that were equal in expectation to the deterministic solutions to the continuous time SIR model.

Parameter Estimation

Particularly, we have used Markov Chain Monte Carlo (MCMC) [9, 10] to find the posterior distributions of β, k, γ and μ.

Let $\hat{\imath}_t = S(t-1) - S(t)$ be the number of new infected at time T, and similarly for the newly recovered and dead individuals $\hat{r}_t$ and $\hat{d}_t$ so that $\hat{r}_t + \hat{d}_t = I(t-1)$. Then, the likelihood function for the observed data up to time T is given below

$$\prod_{t=1}^{T} Bin\left(\hat{\imath}_t \middle| S(t-1), p_i(I(t-1),\beta,k)\right) \times \sum_{t=1}^{T} Bin(\hat{r}_t|p_r) \times \prod_{t=1}^{T} Bin\left(\hat{d}_t \middle| I(t)-\hat{r}_t, p_d\right) \tag{10}$$

Above equation (10) consists of three mutually independent components when

conditioning on the course of the epidemic. Conditional conjugacy can be exploited for γ and μ via beta priors for p_r and p_d. A $beta(a, b)$ prior for p_r implies that

$$p(v) = (1 - e^{-v})^{\alpha_r - 1} e^{-v\beta_r} \quad (11)$$

Conjugate updating leads to the posterior conditional

$$p_r| \ldots \sim Beta(a + \sum_{t=1}^{T} \hat{r}_t\,, b + \sum_{t=1}^{T} I(t) - \hat{r}_t) \quad (12)$$

The conditional posterior distribution for γ is similar to Eq. (11) and can be simulated by first drawing p_r using Eq. (12) and then applying the inverse transformation $\gamma = -\log(1 - p_d)$. Sampling for μ proceeds similarly with

$$p_d| \ldots \sim Beta\big(c + \sum_{t=1}^{T} \hat{d}_t\,, d + \sum_{t=1}^{T} I(t) - \hat{r}_t - \hat{d}_t\big) \quad (13)$$

Then it is possible to consider Gibbs samples for γ and μ so long as appropriate hyperparameters a, b, c, d can be found to represent our prior beliefs. We have considered informative prior as uniform for p_r and p_d.

Obtaining samples for β and k requires the Metropolis-Hastings algorithm. So, the conditional on a previous sample (β, k) the next sample (β', k') can be obtained by Metropolis-within-Gibbs steps using:

$$p(\beta'|k, \ldots) \propto \lceil(\beta'|\alpha_b, \beta_b) \prod_{t=1}^{T} Bin(\hat{\imath}_t|S(t-1), p_i(I(t-1), \beta', k)), \quad (14)$$

And

$$p(k'|\beta', \ldots) \propto \lceil(k'|\alpha_k, \beta_k) \prod_{t=1}^{T} Bin(\hat{\imath}_t|S(t-1), p_i(I(t-1), \beta', k')) \quad (15)$$

The prior parameterization sets are $(\alpha_b, \beta_b) = (\alpha_k, \beta_k) = (1, 3)$ which turns out to be uninformative on the scale of the support of the posterior. We use random walk uniform proposals on the positive real line, i.e. $\beta' \sim U[3b/4, 4b/3]$, which gives reasonably good mixing in the Markov chain.

Application

To illustrate the above model we have considered influenza disease which posed a new challenge to the public health systems and community all over the world. Influenza is an emerging infectious disease and influenza-like illness (ILI) is a clinical illness caused by the influenza virus, which occurred throughout history. There have been four major outbreaks since 1918, each with different characteristics, such as 1918-19 Spanish flu, 1957-58 Asian flu, 1968 Hong Kong flu, 2004-05 Bird flu and most recently 2009-10 Swine flu pandemic. Influenza, commonly known as "the flu", is an infectious disease caused by an influenza virus [11].

The model samples, desired realizations of model parameters in a stochastic SIR model for influenza. We have considered $N = 1000$ individuals from time 0 to T (40 Days). To initialize this process for evaluation of epidemic growth over time, initial values of transition rates are considered as $\beta = 0.00218, \gamma = 0.4, k = 10$ and $\mu = 0$ [12]. These Monte-Carlo algorithms were developed and analyzed with R software.

We have performed 10000 iterations for each run of the MCMC algorithm following 2000 burnin. In order to avoid autocorrelation within successive samples, we have allowed every 10th observation to participate in making inference (i.e. thinning) [10].

Initialized the program by choosing model parameters as $\beta = 0.2, \gamma = 0.3,\ k = 10$ and $\mu = 0.3$. We have also verified that estimates were robust to a change in the initial values. We have considered the prior distribution as beta (a, b) with mean $a/(a+b)$ and variance $ab/((a+b)^2\,(a+b+1))$. The seed used in the simulations was given by the computer clock. The joint posterior distribution of parameters was explored by MCMC sampling, and characterized by means and equal tailed 90% credible intervals (CrI).

Results and Conclusions

By taking recourse of Monte Carlo simulation, we have simulated the population trajectories for the above model, which gives us the stochastic SIR model with mean trajectories for each compartment in the experimental population, as well as the 5^{th} and 95^{th} quantiles, are recorded. Figure 1 shows the distribution of the susceptible, infected, and recovered individuals in the population as the epidemic progresses.

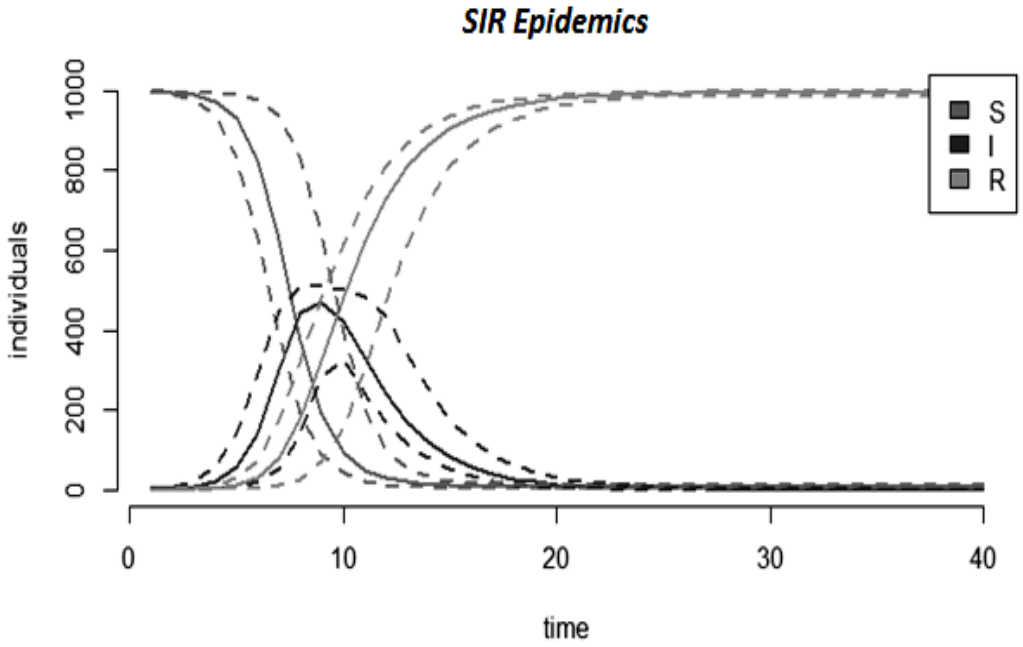

Figure 1: SIR epidemic plot for number of susceptible, infected and recovered individuals with 5th and 95th quantiles are shown.

The output was recorded to constitute samples from the posterior distribution and the convergence was visually assessed through trace plots. Trace plots provide a useful method for detecting problems with MCMC convergence and mixing [10], see Figures 2.

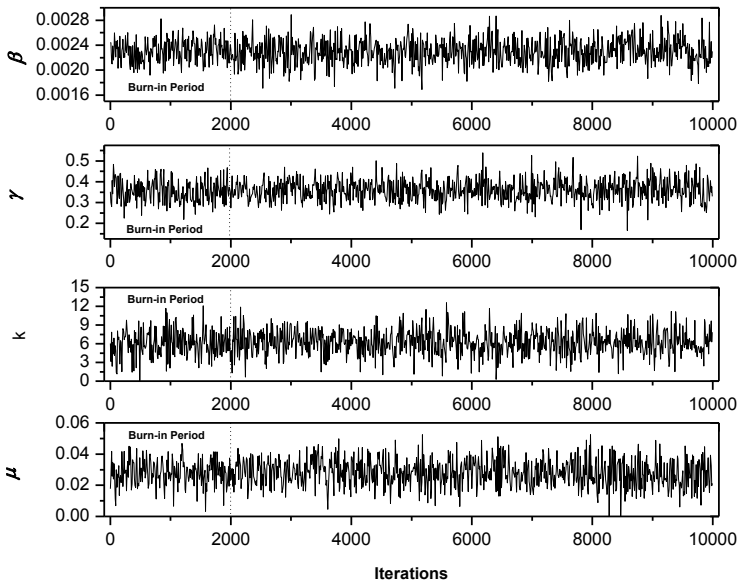

Figure 2: Summary of posterior distribution for a) Realizations of Model parameter β, b) Realizations of over dispersion model parameter γ, c) Realizations of model parameter k and d) Realizations of Model parameter μ.

We have estimated posterior means for transmission rate (β), rate of recovery (γ), over dispersion parameter (k) and death rate (μ). It is characterised by average along with 90% CrI. From the estimates of parameter, we can say that, if the individual is infectious, he/she infects others at rate β (when all individuals are susceptible) per unit time, and the mean duration of the infectious period equals $1/\gamma$. The Table 1 summarizes results from the statistical model See Tables 1.

Table 1: Posterior summary of simple SIR model parameter

Parameters	**Mean**	**5th Quantile**	**95th Quantile**
$\boldsymbol{\beta}$	0.00229	0.00196	0.00262
$\boldsymbol{\gamma}$	0.35948	0.27778	0.44701
$\boldsymbol{k}$	6.22202	2.73678	9.71836
μ	0.02529	0.02229	0.02802

The results from Table 1 shows that, the posterior mean of transmission rate in a community is 0.00229 (0.00196, 0.00262) per day and the posterior mean rate of clearing infection in a community is 0.35948 (0.27778, 0.44701) per day. Also, posterior mean of over dispersion parameter is 6.22202 (2.73678, 9.71836). The posterior mean of death rate in community is 0.02529 (0.02229, 0.02802) per day. Our results are comparable with Murray [12].

Discussion

This model adjusts fairly well to the pattern of the disease spread in the population and efficiently captures effect of the infectious disease transmission. Our model did not take into account all factors that are known to affect the disease spread; For example, this analysis did not include disease sequel, transmission of infection within household members, the infection solidity may depend on serotype of the infection, etc. This work is also compatible with other diseases which are practically fit in SIR model. Since, we have generated desired data through simulation one can perform analysis by collecting real infection data. Despite the inherent limitations of this model, it gives comprehensible idea of disease spread.

Acknowledgement

We are thankful to the Research Fellowship in science for Meritorious Students (RFSMS) UGC, New Delhi, India for research Funding support.

Conflict of interest: No conflict of interest.

References

[1] Pech, R. P., and Hone, J., 1988, "A model of the dynamics and control of an outbreak of foot and mouth disease in feral pigs in Australia," J. Appl. Ecol., 25, pp. 63–77.

[2] De Leo, G. A., and Dobson, A. P., 1996, "Allometry and simple epidemic models for microparasites," Nature, 379, pp. 720–722.

[3] Hone, J., 1994, "Analysis of Vertebrate Pest Control," Cambridge University Press.

[4] Finkenstädt, B. F., and Grenfell, B., 2000, "Time series modelling of childhood diseases: a dynamical systems approach," Appl. Stat., 49, pp. 187–205.

[5] McCallum, H., Barlow, N., and Hone, J., 2001, "How Should Pathogen Transmission be Modelled," Trends in Ecology and Evolution, 16, pp. 295–300.

[6] Hohle, M., and Jørgensen, E., 2002, "Estimating Parameters for Stochastic Epidemics," Dina Research Report No. 102, Nov 2002, Available at: http://staff.math.su.se/hoehle/pubs/dina102.pdf.

[7] Godfray, H., and Hassell, M., 1989, "Discrete and Continuous Insect Populations in Tropical Environments," Journal of Animal Ecology, 58, pp. 153–174.

[8] Hetchote, H. W., 2000, "The Mathematics of Infectious Diseases," SIAM Review, 42, pp. 599– 653.

[9] Gamerman, D., and Lopes, H., 2006, "Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference," Chapman & Hall / CRC.

[10] Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., 1995, "Markov Chain Monte Carlo in Practice," Chapman & Hall, London.

[11] Pandemic Flu History, http://www.flu.gov/pandemic/history
[12] Murray, J. D., 2002, "Mathematical Biology I: An Introduction," Springer Verlag.