Statistical Inference for Stochastic Epidemic Models

George Streftaris[1] and Gavin J. Gibson[1]

[1] Department of Actuarial Mathematics & Statistics, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, UK
Email: G.Streftaris@ma.hw.ac.uk

Abstract: We consider continuous-time stochastic compartmental models which can be applied in veterinary epidemiology to model the within-herd dynamics of infectious diseases. We focus on an extension of Markovian epidemic models, allowing the infectious period of an individual to follow a Weibull distribution, resulting in more flexible modelling for many diseases. Following a Bayesian approach we show how approximation methods can be applied to design efficient MCMC algorithms with high acceptance ratios for fitting non-Markovian models to partial observations of epidemic processes. A simulation study is conducted to assess the effects of the frequency and accuracy of diagnostic tests on the information yielded on the epidemic process.

Keywords: Bayesian inference; diagnostic tests; Markov chain Monte Carlo; Metropolis–Hastings acceptance rate; stochastic epidemic modelling.

1 Introduction

The use of stochastic compartmental models in the statistical analysis of infectious diseases is becoming more widespread thanks to the extensive availability of computing power, and the development of sophisticated computational techniques such as Markov chain Monte Carlo (MCMC) (Gelfand and Smith, 1990; Tierney, 1994). These models represent populations of individuals as being partitioned in disjoint subsets - e.g. susceptible, infectious and removed, or recovered, in the case of the general epidemic or SIR model (e.g. Bailey, 1975) - and represent the transitions of individuals between compartments as stochastic processes.

In practical situations, inference for such models is complicated because the processes involved are only partially observed. For example, observations of an epidemic of an infectious disease in a population of humans or animals may record only the time of the appearance of symptoms, with precise infection times of individuals being unobserved. Furthermore, times of those events which are observed may be censored in some way, e.g. due to sporadic or infrequent testing regimes. Several authors have used the

aforementioned advances in stochastic integration methods to tackle similar problems within a Bayesian framework (e.g. Gibson and Renshaw, 1998; O'Neill and Roberts, 1999).
Following the same approach, we first note that for many diseases the time for which an individual remains infectious follows a distribution that is clustered around some central modal value, making the exponential distribution implausible. We therefore consider the Weibull distribution for sojourn times. We remark that an alternative generalisation of the exponential distribution, the gamma distribution, has been considered by O'Neill and Becker (2001). Our distributional assumptions and the incomplete nature of the observed data raise the issue of developing appropriate MCMC methods for estimation. We employ an approximation to full conditional distributions of non-closed form, to obtain high acceptance ratios in a Metropolis–Hastings algorithm.
Finally, when studying epidemics, the timing and frequency of applying diagnostic tests to a population has an important bearing on the information yielded on the epidemic process, as does the sensitivity and specificity of the tests themselves. To gain some initial understanding of the effects of these factors on the resulting information, we conduct a study using simulated epidemics and diagnostic tests of various efficiency and frequency.

2 Non-Markovian SIR model

We consider a population of fixed size N, in which an epidemic is taking place. We assume that one infectious individual initiates the epidemic and thereafter secondary (animal to animal) transmissions of the disease take place according to a stochastic SIR model. That is, at time t each individual is characterised by its current state (susceptible, infectious or removed), and therefore belongs to the S, I or R compartment, with $S(t)$ and $I(t)$ giving the number of individuals in the S and I class respectively (prior to the occurrence of any event at time t). Under the assumption of constant infectiousness over time, transitions from compartment S to I in the infinitesimal time increment $[t, t+dt)$, occur according to a probability given by

$$\Pr\{S(t+dt) = S(t) - 1\} = \beta S(t) I(t) dt$$

with β denoting the rate of infection per possible susceptible-infectious contact. In a departure from the standard Markovian SIR model we assume that individuals remain in the I compartment for a time drawn randomly from a Weibull(ν, λ) distribution with density function given by

$$f(x) = \nu\lambda x^{\nu-1} \exp\left(-\lambda x^{\nu}\right), \qquad x, \nu, \lambda > 0.$$

This has previously been applied in simulation studies of veterinary epidemic processes for bovine tuberculosis. It implies that our model no longer

retains the Markovian (lack-of-memory) property. We cannot therefore characterise the system at time t merely by the numbers of individuals in each compartment. This poses problems for carrying out inference with such models. In the case that we are able to observe the precise times of all infections and removals in the population during the observation period in a Markovian model (e.g. Becker, 1983), it is sufficient in the statistical sense to know the times and nature of all events occurring in the course of an epidemic. Knowledge of the individuals to which each event applies would not change the parameter likelihood.

However, in a non-Markovian model the explicit history of each individual must be represented in the model, since the absence of the lack-of-memory property implies that removal times of specific individuals now depend on the time of their infection, with the infectious periods modelled as random variables from a Weibull(ν, λ) distribution. Removed individuals play no further role in the spread of the epidemic. We set $t_1 = 0$ as the time of the first infection of a susceptible individual and we observe the population for a time period of length T. We let n_I and n_R denote the number of infected and removed individuals respectively, and define $\mathbf{t} = (t_1, \ldots, t_{n_I+n_R})$ as the ordered set of all events. Then, by associating each individual j with an infection time $s_j \in \mathbf{s}$ and, if appropriate, a removal time $r_j \in \mathbf{r}$, and defining $\mathcal{I}, \mathcal{R}$ as the sets of individuals respectively infected or removed at the end of the observation period T, the likelihood for the model parameters can be written as

$$L(\beta, \nu, \lambda; \mathbf{s}, \mathbf{r}) = \prod_{j \in \mathcal{I}^*} \{\beta I(s_j)\} \exp\left\{ - \sum_{i=2}^{n_I+n_R} \{\beta S(t_i) I(t_i)(t_i - t_{i-1})\right\}$$
$$\times \prod_{j \in \mathcal{R}} \left[\nu\lambda(r_j - s_j)^{\nu-1} \exp\left\{-\lambda(r_j - s_j)^{\nu}\right\}\right] \prod_{j \in \mathcal{I} \cap \bar{\mathcal{R}}} \exp\left\{-\lambda(T - s_j)^{\nu}\right\}, \tag{1}$$

where $\mathcal{I}^*$ denotes set $\mathcal{I}$ with the initially infectious individual omitted, and $\bar{\mathcal{R}}$ is the complement of set $\mathcal{R}$. The last term in (1) is associated with infected individuals whose removal time has been censored, and therefore vanishes in the case of a complete epidemic.

Perfect knowledge of the infection and removal times would allow direct use of the likelihood function (1) to obtain estimates for the parameters of interest, e.g. by the method of maximum likelihood. However, epidemics are only partially observed, with the precise infection times $s_j \in \mathbf{s}$ of individuals being unknown. In such cases the data consist of the removal times $r_j \in \mathbf{r}$ and possibly diagnostic test results. The latter allow the unobserved infection times to be restricted within intervals between successive testing times. Nevertheless, inferences using (1) require the exact values $s_j \in \mathbf{s}$ to be known or estimated.

3 Bayesian formulation and MCMC method

In the Bayesian approach, the hidden aspects of the epidemic process are treated as additional unknown parameters. The joint posterior density of these parameters and the model parameters (given the observations) is then investigated. Inferences on the parameters of interest are made by considering the corresponding marginal densities. To formulate the Bayesian model we need to assume prior distributions for the contact parameter β and the shape and scale parameters, ν and λ respectively, of the Weibull distribution of the infectious periods. We choose the independent gamma priors $\pi(\beta) \sim \text{Ga}(a, b)$; $\pi(\nu) \sim \text{Ga}(c, d)$; and $\pi(\lambda) \sim \text{Ga}(m, \phi)$. Combining the likelihood function in (1) with these prior distributions, we obtain the posterior density

$$p(\beta, \nu, \lambda|\mathbf{s}, \mathbf{r}) \propto \; L(\beta, \nu, \lambda; \mathbf{s}, \mathbf{r}) \; \beta^{a-1} \; \nu^{c-1} \; \lambda^{m-1} \; \exp\{-b\beta - d\nu - \phi\lambda\}. \quad (2)$$

As this is given in a non-closed form, estimation involves intractable integrations and therefore we will employ MCMC techniques.
We suggest a single-component Metropolis–Hastings algorithm. Each model parameter is updated separately in a single step by first generating a candidate value from a proposal distribution. The new value is then accepted with a probability involving the ratio of the full conditional and the proposal distribution. The full conditional distributions of the parameters β and λ are given from (2) as

$$\beta|a, b, \mathbf{r}, \mathbf{s} \sim \text{Ga}\left(n_I + a - 1, \; b + \sum_{i=2}^{n_I+n_R} S(t_i)I(t_i)(t_i - t_{i-1})\right)$$

$$\lambda|\nu, m, \phi, \mathbf{r}, \mathbf{s} \; \sim \text{Ga}\left(n_R + m, \; \sum_{j\in\mathcal{R}}(r_j - s_j)^\nu + \sum_{j\in\mathcal{I}\cap\bar{\mathcal{R}}}(T - s_j)^\nu + \phi\right).$$

These full conditionals can serve as the proposal distributions, leading to Gibbs sampling steps with acceptance probability $\alpha = 1$. However, the density of the full conditional of the parameter ν takes the non-closed form

$$p(\nu|\lambda, c, d, \mathbf{r}, \mathbf{s}) \; \propto \; \nu^{n_R+c-1} \; \prod_{j\in\mathcal{R}}(r_j - s_j)^{\nu-1}$$
$$\times \exp\left[-\lambda\left\{\sum_{j\in\mathcal{R}}(r_j - s_j)^\nu + \sum_{j\in\mathcal{I}\cap\bar{\mathcal{R}}}(T - s_j)^\nu\right\} - d\nu\right]. \quad (3)$$

We update ν employing a proposal distribution $q(\cdot)$ in a Metropolis–Hastings step with acceptance probability given by

$$\alpha = \min\left\{1, \frac{p(\nu^{(new)}|\lambda, c, d, \mathbf{r}, \mathbf{s})/q(\nu^{(new)}|\lambda, c, d, \mathbf{r}, \mathbf{s})}{p(\nu^{(old)}|\lambda, c, d, \mathbf{r}, \mathbf{s})/q(\nu^{(old)}|\lambda, c, d, \mathbf{r}, \mathbf{s})}\right\}. \quad (4)$$

If the new candidate value is not accepted, the value from the previous iteration is retained. The proposal distribution in (4) is not conditional on the value of ν, thus forming an independence-type Metropolis–Hastings algorithm (Tierney, 1994). This implies that we can choose the proposal distribution to be as close to the full conditional as possible, in order to obtain an efficient algorithm in terms of acceptance rate.

3.1 Approximation to full conditional distribution of ν

We use a gamma proposal distribution to approximate $p(\nu|\lambda, c, d, \mathbf{r}, \mathbf{s})$. Its parameters are determined in a way such that the first two moments of the proposal are approximately equal to the corresponding moments of the full conditional distribution. As the latter is fairly symmetrical, we employ its mode and information function as appropriate values for the mean and variance of the gamma approximation. This involves maximising (3) once, and is easily embedded in the MCMC iterative scheme without bearing a prohibitive computational cost. The resulting approximation is illustrated in Figure 1, where the true full conditional was computed from (3) using numerical integration. The suggested method produced accurate approximations, and therefore provided remarkably high acceptance rates ($96\% - 98\%$) for the Metropolis–Hastings algorithm.

3.2 Updating the infection times

To update the unobserved infection times the algorithm also needs to take account of the unknown number of infections n_I, proposing feasible moves between subspaces of different dimension and adjusting the acceptance probabilities accordingly. At each MCMC cycle we choose an integer j from $(1, 2, \ldots, N)$. If j corresponds to an infected but not removed individual, we either remove its infection time s_j from the set of infection times $\mathbf{s}$ with probability $p = \frac{1}{2}$, or we move s_j randomly in $(0, T)$ equally likely. In the first case, the new vector $\mathbf{s}$ is accepted with probability

$$\alpha = \min\left\{1, \frac{1}{2T}\frac{L(\beta, \nu, \lambda; \mathbf{r}, \mathbf{s}^{(new)})}{L(\beta, \nu, \lambda; \mathbf{r}, \mathbf{s}^{(old)})}\right\}, \tag{5}$$

while in the latter case it is given by (5) with $\frac{1}{2T}$ omitted. If the jth individual has been removed from the population, we move its infection time randomly in $(0, r_j)$. The acceptance probability is again the same as in the second case of the previous step. Finally, if j indicates a non-infected individual, we randomly add an infection time $s_j \in (0, T)$ to the set of infection times $\mathbf{s}$. The acceptance probability is now given by (5) with $\frac{1}{2T}$ replaced by $\frac{T}{2}$.

The incomplete nature of the data will be reflected in the parameter estimation process both through the precision of the estimates and the performance of the MCMC algorithm. In epidemics where diagnostic testing

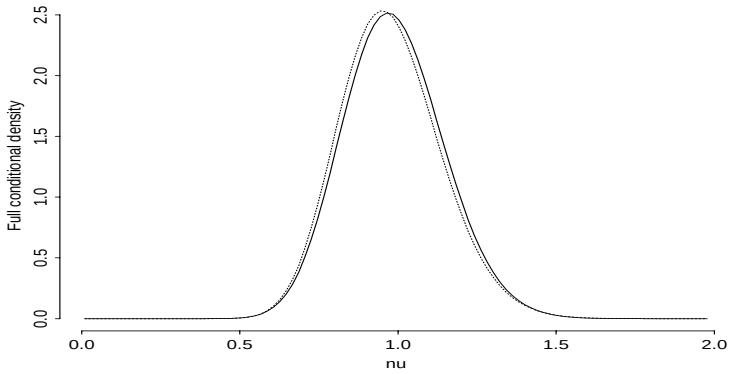

FIGURE 1. Gamma approximation (dotted line) to full conditional distribution (solid line) of ν.

is carried out, the additional information results in better estimation in a degree that depends on the frequency and accuracy (sensitivity and specificity) of the diagnostic tests. In the MCMC algorithm the updating scheme is easily modified to incorporate the additional information, by restricting the individual infection times in appropriate intervals, while also taking account of the test sensitivity and specificity.

4 Application

We use a simulated epidemic to illustrate the described method. A closed population of $N = 100$ individuals is considered and the rate of infection per possible contact is set to $\beta = 0.003$. The infectious periods are drawn from a Weibull distribution with $\nu = 1.1$ and $\lambda = 0.08$. The simulation process produced 91 removals out of 92 infected cases over an observation period of 60 days. Several sets of diagnostic test results were also generated, with varying frequency and accuracy, to investigate their effect on the estimation process.

In the Bayesian model the parameters of the gamma prior distributions considered in Section 3 were assumed to be $a = 10^{-4}, b = 0.1, c = 1, d = 10, m = 1, \phi = 10$. The use of highly vague priors resulted in MCMC convergence difficulties in cases of very infrequent (or no) diagnostic testing, due to the lack of information in the model. The convergence of the algorithm was improved by more frequent updating of the vector of infection times $\mathbf{s}$ (to facilitate the movement of the sampler through the likelihood support), and also by thinning the post-convergence sample to reduce the

autocorrelation in the simulated sequence.
Table 1 presents the posterior estimates of the model parameters, when only removal data are considered and when results of diagnostic tests (of assumed sensitivity and specificity of 85%) are included in the analysis. The posterior density of the transmission parameter β is plotted in Figure 2. Clearly, the posterior distribution is more concentrated around its mean as more information becomes available through diagnostic testing. This is also illustrated in Figure 3, which shows the decrease of the posterior standard deviation of β when the number of conducted tests increases. The improvement is slower for less accurate tests.

TABLE 1. Posterior mean and standard deviation of the model parameters when only removal data are considered and when 2 or 60 (daily) tests are conducted. The sensitivity and specificity of the tests were assumed to be equal to 85%.

	Removal data		*2 tests*		*60 tests*	
	mean	*sd*	*mean*	*sd*	*mean*	*sd*
β	4×10^{-3}	6×10^{-4}	3×10^{-3}	5×10^{-4}	3×10^{-3}	3×10^{-4}
ν	1.140	0.123	1.140	0.107	1.110	0.080
λ	0.098	0.030	0.083	0.025	0.082	0.018

5 Discussion

We considered a non-Markovian compartmental model with Weibull infectious periods for flexibility in biologically realistic situations in veterinary epidemiology. An efficient independence-type Metropolis–Hastings algorithm was developed to tackle parameter estimation within a Bayesian framework, when the precise times of infection are not observed. Where needed, the MCMC technique makes use of appropriate gamma approximations to the full conditional distributions of the parameters, resulting in a particularly high acceptance rate of at least 96% in our applications. Other approximations may also considered (based e.g. on the Laplacian technique), but their computational cost should be taken into account when incorporated in a MCMC iterative scheme.
The methods were adapted to enable diagnostic test data to be used in the analysis. Availability of such data reduces the uncertainty in the unobserved infection times and therefore improves the precision of the estimates. The potential gain depends on the number of tests conducted, as well as on their effectiveness as expressed through their sensitivity and specificity.
Future work includes application to field data, with possible adaptation of the model to the characteristics of a specific disease, and development of suitable methods for model assessment and selection.

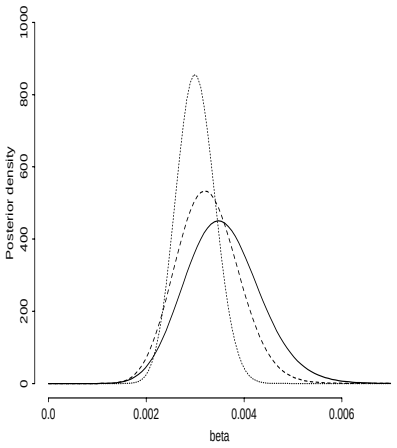

FIGURE 2. Posterior density of β without diagnostic test data (solid line), and when 2 tests (dashed line), or 60 tests (dotted line) are conducted. The sensitivity and specificity of the tests are both assumed to be equal to 85%.

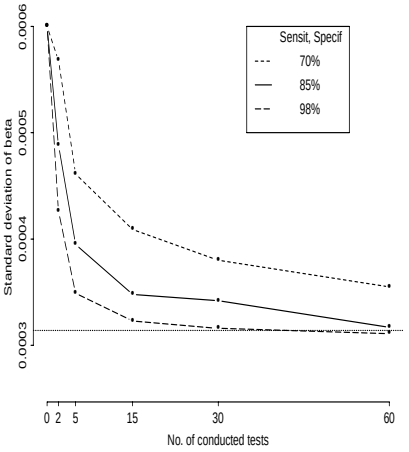

FIGURE 3. Posterior standard deviation of β against the number of tests. Each curve corresponds to a given sensitivity and specificity level. The horizontal dotted line represents complete knowledge of the infection times.

References

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

Becker, N. G. (1983). Analysis of data from a single epidemic. *Austral. J. Statist.* **25**, 191–197.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409.

Gibson, G.J. and Renshaw, E. (1998). Estimating parameters in stochastic compartmental models using Markov chain Monte Carlo methods. *IMA J. Math. Appl. Med. & Biol.* **15**, 19–40.

O'Neill, P.D. and Roberts, G.O. (1999). Bayesian inference for partially observed stochastic epidemics. *J. Roy. Statist. Soc. Ser. A* **162**, 121–129.

O'Neill, P.D. and Becker, N.G. (2001). Inference for an epidemic when susceptibility varies. *Biostatistics* **2**, 99–108.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**, 1701–1762.