The use of Chapman-Kolmogorov Equations and Markov Chain Monte Carlo Techniques to Analyse Infections using a Susceptible-Infectious-Susceptible Framework.

Innocent G.T.[1], O'Reilly K.M.[2], Marion G.[1], Matthews L.[2], McKendrick, I.J.[1]

[1]Biomathematics and Statistics Scotland (BioSS), Edinburgh, UK
[2]Boyd Orr Centre for Population and Ecosystem Health, University of Glasgow, UK

ABSTRACT

In this paper we demonstrate a method for analysing longitudinal Susceptible-Infectious-Susceptible (SIS) data. We suppose that data are available recording the number of animals that test positive at a number of time points. Under the assumption of thorough mixing it is possible to define a set of ordinary differential equations (ODEs) that define the rate of change of the probabilities that the system is in each of a number of states. These ODEs are the Chapman-Kolmogorov equations. If we define the states as representing all possible test results from the group of animals then we can compare these probabilities with the true test result at each time point to derive a likelihood. By applying a Metropolis-Hastings algorithm we can determine the distribution of transmission parameters from the likelihood, or Bayesian posterior distributions, if we wish to provide prior distributions for these parameters. The use of this approach is demonstrated by estimating parameters used in simulation over a range of realistic values. Finally, the method is applied to some data representing bacterial isolations in a group of calves housed together.

KEYWORDS

Chapman-Kolmogorov Equations, MCMC, VTEC, SIS model

INTRODUCTION

Several diseases of interest do not elicit a protective response in affected animals. Such diseases can be modelled using a Susceptible-Infectious-Susceptible (SIS) framework. Typically infection is only partially observed, for example, animals may be tested at regular intervals for infection. The exact timing of transition events, for example from susceptible to infectious, are not observed. The number of such transitions between observation times is therefore unknown. This uncertainty affects estimation of model parameters and therefore predictions of the efficacy of control strategies.

Here we estimate the transmission parameters of a model using a set of ordinary differential equations (ODEs) to describe the change over time of the probability distribution of the number of infectious animals. Inference on the values taken by the model parameters is made within a Markov chain Monte-Carlo (MCMC) framework. The advantage of the MCMC approach is that it makes no assumptions about the asymptotic nature of the parameter distributions, but samples directly from the likelihood itself, with or without prior information.

MATERIALS AND METHODS

The approach that we have taken is to use a set of ODEs to describe the evolution in time of the probability distribution of states in which the system must lie. In a population of size n, we consider n+1 states, since the number of infectious individuals must come from the set {0, 1, 2, ..., n}. The rate of change of the probability of state k is defined in terms of the probability that the system is in state k-1 and the instantaneous rate at which animals become infectious; the probability that it is in state k and nothing changes; and the probability that it is in state k+1 and the instantaneous rate at which individuals recover. This leads to a system of n+1 equations used to describe the rates of change of the probabilities that the state lies in each of the n+1 states. These are the Chapman-Kolmogorov equations for the process. The general case is given below:

$$\begin{aligned}\frac{dP(I=m)}{dt} = &(\alpha(n-m+1)+\beta(m-1)(n-m+1))P(I=m-1)\\ &-\gamma m P(I=m)\\ &-(\alpha(n-m)+\beta(m)(n-m))P(I=m)\\ &+\gamma(m+1)P(I=m+1)\end{aligned}$$

Note that P(I=m) is zero for all m not between 0 and n. The model uses three transmission parameters, the rate of infection from the environment, α, the rate of transmission, β, and the rate of recovery, γ. We wish to estimate parameters for the above model using data describing a time-series of test results on individual animals. One approach is to solve this set of ODEs over time to define a set of probabilities that the group exists in each of the n+1 states at a general time T+t, given their state at time T. If T+t coincides with a subsequent test, then the probability associated with the test result can contribute to a likelihood for the parameters of the ODEs. Resetting the probabilities to reflect the actual test result at each time that the test is performed, the likelihood of the parameters is the product of the probabilities of each test result. This is demonstrated in Figure 1.

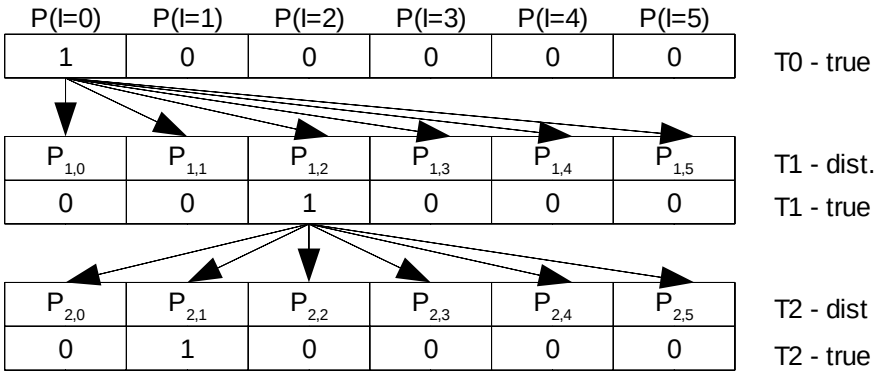

Figure 1. The use of Chapman-Kolmogorov functions to determine the likelihood of parameters. In this case there are 5 animals in the group, all initially susceptible. At time T1 two animals test positive, at time T2 one animal tests positive. The probability at time i of j positive tests is calculated as $p_{i,j}$. In this case the likelihood of the test results is $p_{1,2} * p_{2,1}$.

RESULTS

Initially the model was tested against simulated data. It was able to reproduce the parameters used to simulate the data in a wide variety of situations. In particular, if disease was not constantly present, but occurred and recurred over the period of observation then the model performed well. However, if the parameter set led to an endemic situation, the model was unable to correctly identify the parameters. However, in this case the model could correctly identify the ratio of transmission to recovery. An example of a set of parameters which is well estimated is given in Figure 2.

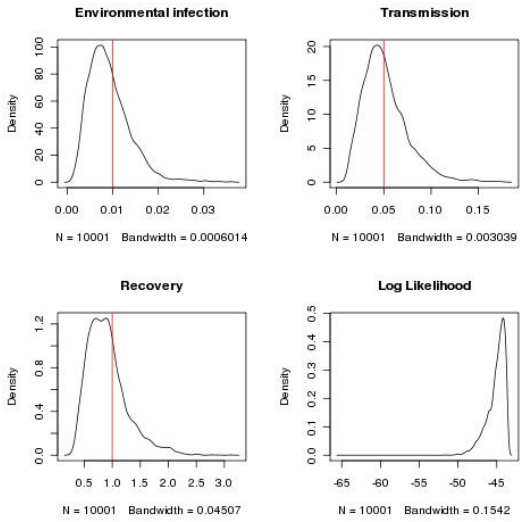

Figure 2. Plot of MCMC distributions for the three parameters of interest. Model tested against simulated data. Vertical red lines indicate the parameters used in simulating the data.

Thereafter, the model was used to analyse data from repeated faecal sampling of a group of calves, tested for the presence of verotoxigenic *Escherichia coli* (VTEC) serotype O26. These data are described more fully by Liu and others (2005 and 2007). Briefly, the data represent the test result of up to 49 animals tested over a period of 17 weeks. Since animals were recruited to the group at birth, not all animals were tested at

all time periods. However, all animals present at any time period were tested. The results determined by the method presented here correspond well with those of Liu and others: Table 1.

Table 1. Table of results of fitting model to VTEC O26 data. Comparison of the Chapman-Kolmogorov MCMC method with those of Liu and others.

Source of results	Background	Transmission	Recovery
Chapman-Kolmogorov MCMC	0.007 (0.001 – 0.022)	0.003 (0.0001 – 0.010)	0.16 (0.07 – 0.49)
Liu 2005	0.005 (0.002 – 0.013)	0.004 (0.002 – 0.012)	0.3 (0.2 – 0.7)
Liu 2007	0.076 (0.055 - 0.11)	0.002 (0 - 0.006)	0.4 (0.25 - 0.6)

CONCLUSION

Where infection is present at all or most observed time points the analysis fails to identify a unique best solution, but can estimate a meaningful function of the parameters. However, when observations at several successive time points show no positive animals, whilst periods of high and low levels of infection are also observed, then the approach is able to produce estimates for each parameter of the model. This reflects the fact that for the model to e estimable, observations must be made on an appropriate time-scale determined by the rates of infection and recovery in the system of interest.

REFERENCES

Liu, WC, Jenkins, C, Shaw, DJ, Matthews, L, Pearce, MC, Low, JC, Gunn, GJ, Smith, HR, Frankel, G and Woolhouse, MEJ (2005), Modelling the epidemiology of Verocytotoxin-producing *Escherichia coli* serogroups in young calves. *EPIDEMIOLOGY AND INFECTION*, **133(3)**, 449-458.

Liu, W.C., Shaw, D. J., Matthews, L., Hoyle, D. V., Pearce, M. C., Yates, C. M., Low, J. C., Amyes, S. G. B., Gunn, G. J. and Woolhouse, M. E. J. (2007) Modelling the epidemiology and transmission of Verocytotoxin-producing *Escherichia coli* serogroups O26 and O103 in two different calf cohorts. *EPIDEMIOLOGY AND INFECTION*, **135(8)**, 1316-1323.