Optimal Sampling for the SIS Epidemic

Daniel E. Pagendam[1] and Philip K. Pollett[1]

[1] Department of Mathematics, The University of Queensland, Australia

Birth-death processes have been used to describe a wide range of phenomena that evolve over time, with applications in epidemiology, ecology and chemistry. One challenge for practitioners when applying these models is to obtain data that will allow estimation of model parameters (such as infection and recovery rates in an epidemic). Typically an experiment is limited to a small number of observations of the process within a given time interval; the number of observations and the time interval are dictated by budgets and time frames. Unfortunately, very little research has focussed upon the design of experiments or sampling methods for these types of processes, because the likelihood is often difficult to evaluate.

We consider a common Markovian model in epidemiology, known as the SIS epidemic, where the population is comprised of individuals that make transitions between “susceptible” and “infective” states. We demonstrate how Gaussian diffusion approximations of birth-death processes can be used to obtain optimal designs for scheduling a series of observations of an epidemic. The method allows us to calculate an approximate likelihood, allowing straightforward calculation of the Fisher Information matrix. Our analytical results are easily coupled with optimisation methods such as the Cross-Entropy method to arrive quickly at optimal sampling schedules for parameter estimation.

Our findings are presented in the context of both natural and controlled experiments. For natural experiments, we consider the optimal scheduling of observations in time to estimate precisely the model parameters. In the case of controlled experiments, we consider a scenario where the practitioner introduces a number of infected individuals into an artificial population in order to observe the spread of the disease. For the latter experiment, our goal is not only to find the optimal sampling times, but also to determine the optimal number of infected organisms to introduce into the population in order to obtain the most precise estimates. We show that there is an optimal starting population density of infected organisms and that use of this initial population density can drastically affect experimental efficiency.

One of the outcomes of our work is that we are able to determine the circumstances under which a simpler design, where observations are taken at regular intervals, results in near optimal precision and, conversely, where such a design is inappropriate. We find that the optimal design greatly outperforms the simpler equidistant design when there are few opportunities to sample and the time available for the sampling is less of a constraint.