On Parameter Estimation for Markov Processes

Joshua V. Ross, Thomas Taimre and Philip K. Pollett
Department of Mathematics and MASCOS, University of Queensland, Australia
http://www.maths.uq.edu.au/~jvr

Australian Research Council
Centre of Excellence for Mathematics
and Statistics of Complex Systems

THE PROBLEM

- Often the most appropriate model for a stochastic system is that of a discrete-state Markov process. However, generally we can only observe the state of the process at successive, and not necessarily equally-spaced, time points. How can we estimate the parameters of our model from such observational data?

EXAMPLE

- An example of where this problem arises is in epidemiological modelling. A model appropriate for the spread of infections that do not confer any long lasting immunity, and where individuals become susceptible again after infection, is the stochastic SIS (*susceptible-infective-susceptible*) logistic model.
- It is a continuous-time Markov chain $(m(t),\, t \geq 0)$ taking values in $S = \{0, 1, \ldots, N\}$ with non-zero transition rates

$$q(m, m+1) = \lambda \frac{m}{N}(N-m) \qquad (m = 1, 2, \ldots, N-1)$$
$$q(m, m-1) = \mu m \qquad (m = 1, 2, \ldots, N),$$

  where λ, the per-interaction rate of infection, and μ, the per-individual rate of recovery, are both positive.
- Commonly we only know the number of people who are infected at successive, and not necessarily equally-spaced, time points.
- How do we estimate λ and μ, and the basic reproduction ratio $R_0 = \lambda/\mu$ from such observational data?

GENERAL APPROACH

- One approach is to calculate the exact likelihood of observing the given data, and then use a numerical search algorithm to compute the maximum likelihood estimators.
- However, the exact likelihood cannot usually be evaluated explicitly, so it must be computed numerically.
- This combination provides a useful tool for fitting continuous-time Markov chains to real systems, but is computationally infeasible if the parameter space or the maximum population size is large, due to the computational intensity (and storage) required to evaluate the exact likelihood.

APPROXIMATION APPROACH

- One way to achieve an approximate likelihood is to use diffusion approximations. If our model is *density-dependent*, that is the transition rates take the form

$$q_N(m, m+l) = N f\left(\frac{m}{N}, l\right), \quad l \neq 0,$$

  for a suitable function f, where $X_N(t) = m_N(t)/N$, $t \geq 0$, then we may derive a deterministic approximation, and a Gaussian diffusion approximation.
- When the deterministic approximation has an asymptotically stable fixed point, we can accurately model the fluctuations about this fixed point by an Ornstein-Uhlenbeck (OU) process. The likelihood of an OU process is simply a Gaussian distribution. Thus, we may approximate the exact likelihood by a Gaussian distribution, resulting in a substantial decrease in computational complexity.

RESULTS

- This table shows a comparison of the General Approach to the Approximation Approach for the stochastic SIS logistic model with $\boldsymbol{N = 50}$ (number of individuals), $\boldsymbol{\lambda = 0.8}$ (per-interaction rate of infection) and $\boldsymbol{\mu = 0.4}$ (per-individual rate of recovery) and using one set of $\boldsymbol{n = 40}$ observations:

$n = 40$	$\widehat{\lambda}$	$\widehat{\mu}$
General Approach	1.14739	0.566294
Approximation Approach	1.10432	0.552027

- Note that the maximum population size is small in the above comparison. When $\boldsymbol{N}$ is increased to $\mathbf{2000}$, where the General Approach is infeasible, the estimates produced by the Approximation Approach improve:

$n = 40$	$\widehat{\lambda}$	$\widehat{\mu}$
Approximation Approach	0.905564	0.447709

- Our new method provides reasonably accurate estimates of parameters. In particular when the maximum population size is large. This is precisely the situation in which the General Approach becomes infeasible, and thus the methods presented complement each other.

References

[1] Ross, J.V., Taimre, T. and Pollett, P.K. On parameter estimation in population models. Under revision for Theoretical Population Biology.