Parameter estimation for SIR epidemics in households

Owen Lyne (Research supported by UK EPSRC, under grant number GR/L56282.)
University of Nottingham, School of Mathematical Sciences
University Park
Nottingham, NG7 2RD, United Kingdom
odl@maths.nott.ac.uk

Frank Ball
University of Nottingham, School of Mathematical Sciences
University Park
Nottingham, NG7 2RD, United Kingdom
fgb@maths.nott.ac.uk

1. Model

Consider the following model (investigated in Ball, Mollison and Scalia-Tomba (1997)) for the spread of an epidemic among a closed finite population partitioned into households. For $n = 1, 2, \ldots$, let m_n be the number of households of size n and $N = \sum_{n=1}^{\infty} nm_n$ be the total number of individuals. The epidemic is initiated by a number of individuals becoming infected at time $t = 0$. The infectious periods of different infectives are assumed to be independent and identically distributed according to a random variable T_I, having an arbitrary but specified distribution. Throughout its infectious period a given infective makes infectious contacts with each susceptible in the population at the points of a homogeneous Poisson process having rate λ^G/N, and, additionally, with each susceptible in its own household at the points of a homogeneous Poisson process having rate λ^L. All the Poisson processes describing infectious contacts (whether or not either or both of the individuals involved are the same), as well as the random variables describing infectious periods, are assumed to be mutually independent. A susceptible becomes infective as soon as it is contacted by an infective and is removed (and plays no further part in the epidemic) at the end of its infectious period. The epidemic ceases as soon as there are no infectives present in the population.

2. Estimation

Addy, Longini and Haber (1991) investigate a related household-based model where the global infection is modelled using the idea of an escape probability π. Specifically, each individual has probability π of avoiding infection from outside their own household during the course of the epidemic, independently of other susceptibles in the population. Addy et al. gives an algorithm for obtaining maximum likelihood estimates for λ^L and π from final size data, that is, the total number of individuals in each household that are ultimately infected by the epidemic.

The Addy et al. model assumes that the progress of the epidemic in one household is independent of the progress in any other household. In our model different households are not mutually independent but nonetheless it is useful to utilise the concept of an escape probability π, which under our model can be calculated from λ^L and λ^G, as in done in Ball et al. For estimation this calculation can be inverted: estimate (λ^L, π) from the data using the Addy et al. method and then calculate an estimate of λ^G, thus obtaining a maximum pseudo likelihood estimate of (λ^L, λ^G).

It is important to study the standard errors associated with these estimates. Using an embedding representation of the epidemic process it is possible to obtain asymptotic normality results for these estimators and show that the variance of these consists of two components — the independent households component and an additional component which takes account of the dependence between households. In typical examples this additional variance component dominates and standard errors are much larger than if the households were independent.

An important practical case where estimation and calculation of associated standard errors is also possible is where the observed data is a random sample from the true epidemic, that is, we only possess data on a sample of the households. In this case the observed households are, loosely speaking, more nearly independent and there is a simple relationship between the standard errors for the independent, fully dependent and partially dependent cases.

Simulation studies confirm the theoretical calculations, as good agreement is obtained between the mean estimates and the true parameters simulated from, as well as between the standard deviation of the simulated estimates and the calculated standard errors of the estimators. The simulations also show that the asymptotics used to calculate the standard errors are accurate for realistic population sizes.

3. Extensions

Consider two (or more) independent epidemics and the problem of estimating λ^L and λ^G from the final size data. It is straightforward to extend the maximum pseudo likelihood technique for a single dataset to this situation and obtain joint parameter estimates, that is an estimate for each of λ^L and λ^G which makes use of both sets of data. It is again possible to both calculate and simulate (with very good agreement) associated standard errors for these estimates. It is important to note that merely aggregating the data does not lead to the correct estimates and standard errors obtained from aggregated data could be quite misleading.

It is possible to extend the above methods to the multitype case, that is, where there are several classes of individual with possibly different local and global infection rates between classes of individual. This extension will be briefly discussed and results presented on estimation of parameters.

REFERENCES

Addy, C. L., Longini, I. M. and Haber, M. (1991) A generalized stochastic model for the analysis of infectious disease final size data. Biometrics 47, 961–974.

Ball, F. G., Mollison, D. and Scalia-Tomba, G. (1997) Epidemics with two levels of mixing. Ann. Appl. Prob. 7, 46–89.

RESUME

Cet article considère le problème d'estimation de paramètre pour les èpidémies stocastiques parmi une population partagée en foyers. On étudie des modèles avec foyers indépendants et dèpendants et on démontre l'influence de la dépendance des erreurs standards associées. Des études de simulations montrent l'utilité de la méthode pour des tailles de populations réalistes.