Multi-state Markov modelling with R

Christopher Jackson
Department of Epidemiology and Public Health
Imperial College, London

April 15, 2004

A *multi-state model* expresses the movement of an individual between a finite set of states. It is most commonly used in the form of a *Markov model*, where the movement follows a Markov process. It can be seen as an extension of survival or time-to-event models, where there are several events.

A major application of multi-state models is investigating the progression of chronic diseases. It is of interest to estimate average transition rates between disease states, and also to investigate explanatory variables for the rates of transition. Usually the data are only observed as a series of snapshots of the process in continuous time, such as irregular doctor or hospital visits. Then we do not see when each transition occured, only a set of intervals in which certain transitions must have occurred.

The *hidden Markov model* is a powerful extension to the basic multi-state model. This means that the underlying progression of the individual through the states is not observed, while the observed data are generated conditionally on the underlying states. This can be useful in disease screening applications, where the state of disease is only observed through some error-prone biological marker, and the screening test is subject to error.

The R package `msm` [1] can be used to fit continuous-time multi-state Markov models to irregularly observed longitudinal data, using likelihood maximisation. Any form of transition intensity matrix can be estimated, as long as there are data available to identify the model. Transition rates can be modelled in terms of explanatory variables through a proportional-intensities model. `msm` also handles a special case of the hidden Markov model, where the observed states are misclassifications of the true states. Interested users are encouraged to download `msm` from CRAN, try it out and offer suggestions for improvement.

References

[1] C. H. Jackson, L. D. Sharples, S. G. Thompson, S. W. Duffy, and E. Couto. Multistate Markov models for disease progression with classification error. *Journal of the Royal Statistical Society, Series D: The Statistician*, 52(2):1–17, 2003.