

Use of a hidden semi-Markov model for recurrent events

Trevor Bailey¹, Theodoros Economou¹

¹ School of Engineering, Computing and Mathematics, University of Exeter, UK

There is a large and rich literature on modelling recurrent events in the context of survival or reliability studies. The typical underlying assumption for many modelling approaches is that the recurrent events arise from a counting process, such as the non-homogeneous Poisson process (NHPP), with a time dependent intensity function (occurrence rate). Recent work in this area includes frailty models, joint modelling of the counting process and the censoring times, semiparametric models and mixture models. In some cases it may also be valuable to use a hidden Markov process formulation to allow for subjects in the study undergoing changes which may possibly depend upon the events themselves. In such cases it is customary to use a non-homogeneous Markov process in which transition probabilities may vary with time. We argue that it is also preferable to use a hidden semi-Markov instead of a Markov process, since the latter implies that the transition times between each state are geometrically distributed which may be an unrealistic assumption in some applications. Use of a semi-Markov process avoids this and allows the specification of alternative inter-event time distributions thus increasing flexibility. In this paper we propose a NHPP whose occurrence rate differs according to states of such a hidden semi-Markov process.

Our particular interest concerns analysis of data such as that commonly arising in studies of tumour recurrences following treatment in humans, or in experimental trials on animals. Correct and proper interpretation of such cancer prevention or cancer blocking trials is clearly important and requires statistical models which are sufficiently versatile to capture the potential complexity of the disease process. In our model, the states in the hidden semi-Markov process relate to the improvement/deterioration of the individual and the transition probabilities are characterised not only by covariates, but also by previous occurrences of tumours. The latter aspect allows the model to adequately capture the possibility that subsequent tumours will depend on previous ones even if those have been removed (because carcinogenic cells could still be present). The model permits individual heterogeneity to be handled as usual by inclusion of frailties into the occurrence rate to handle unmeasured factors which may affect susceptibility to the disease. We illustrate use of our model in reanalysis of a data set taken from an often cited paper of Wei *et al* [1] relating to a bladder cancer study conducted in the late 1980's.

References

- [1] Wei LJ, Lin DY, Weissfeld L (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association* 84(408):1065-1073