CHOICE BETWEEN SEMI-PARAMETRIC MARKOV AND NON-MARKOV MULTI-STATE MODELS FROM COARSENED OBSERVATIONS; APPLICATION TO DEMENTIA

D. Commenges[†1], P. Joly[1], A. Gégout-Petit[2], B. Liquet[3]

[1] *INSERM, Bordeaux, France;* [2] *University of Bordeaux Victor Segalen, Bordeaux, France;* [3] *University of Grenoble, Grenoble, France*

[†] E-mail: *daniel.commenges@isped.u-bordeaux2.fr*

We consider models based on multivariate counting processes, including multi-state models. These models are specified semi-parametrically by a set of functions and real parameters. We consider inference for these models based on generally coarsened observations, focusing on families of smooth estimators such as produced by penalized likelihood. An important issue is the choice of model structure, for instance the choice between a Markov and some non-Markov models. We define in a general context the expected Kullback-Leibler criterion and we show that the likelihood based cross-validation (LCV) is a nearly unbiased estimator of it. Our approach unifies the problem of choice of smoothing coefficient and of model structure. We give a general form of an approximate of the leave-one-out LCV. The approach is studied in simulation which shows that it is efficient for model choice between a Markov and a semi-Markov model; the simulation study gives also interesting insight in the variability of LCV. It is illustrated by estimating a Markov and two semi-Markov illness-death models with application to dementia using data of a large cohort study: the non-homogeneous Markov model (with age as basic time-scale) appears to be the best for this application.