

Optimal stratification of survival data via Bayesian nonparametric mixtures

Bernardo Nipoti (*bernardo.nipoti@unimib.it*)

University of Milano Bicocca, Italy

April 13, 2022

The stratified proportional hazards model represents a simple solution to take into account heterogeneity within the data while keeping the multiplicative effect of the predictors on the hazard function. Strata are typically defined *a priori* by resorting to the values of a categorical covariate. We propose a general framework that allows the stratification of a generic accelerated lifetime model, including, as a special case, the Weibull proportional hazards model. The stratification is determined *a posteriori*, taking into account that strata might be characterized by different baseline survivals, and also by different effects of the predictors. This is achieved by considering a Bayesian nonparametric mixture model and the posterior distribution it induces on the space of data partitions. An optimal stratification is then identified following a decision theoretic approach. In turn, stratum-specific inference is carried out. The performance of this method and its robustness to the presence of right-censored observations are investigated through a simulation study. Further illustration is provided analysing a data set from the University of Massachusetts AIDS Research Unit IMPACT Study.