Exploiting a gold standard to compare nonparametric estimators of survival in the presence of truncated data

Mauro Gasparini[1] and Martina Gandini[2]

[1] Dipartimento di Matematica, Politecnico di Torino, Italy
[2] Dipartimento di Sistemi di Produzione, Politecnico di Torino, Italy

We present an application of nonparametric estimation of the survival function in the presence of left-truncated and right-censored data, a case where it is well known that simply using the Kaplan-Meier estimator overestimates survival. Appropriate alternatives, starting from the Non-Parametric Maximum Likelihood Estimator (NPMLE), are discussed when estimating survival of thalassemia patients based on a sample of 191 from the Centro Microcitemie in Turin. The peculiarity of our data set is that, due to further retrospective work, it has recently been expanded into a more complete dataset of 249 patients, whose lifetimes are still censored, but not any more truncated. This provides the best approximation to a real-life gold standard and allows for comparison of the approximate sampling distributions of the alternative estimators. Focus is on the NPMLE and on the Iterative Nelson Estimator (INE): contrary to the results of some recent literature, the NPMLE does not underestimate true survival and its sampling distribution is less biased than the one of the alternative INE.