Density estimation in the Current Status model

Geurt Jongbloed
Delft University of Technology
(*joint work with Piet Groeneboom and Birgit Witte*)

Many interesting statistical models involve data that are hidden in some sense. Data can be censored, corrupted by noise or only random fractions of them could be observed. In such situations, there is a hidden- and observation space. Estimating (aspects of) the distribution of random quantities in the hidden space based on data from the observation space is a statistical inverse problem. In this presentation we discus the problem of estimating a density of a hidden variable in the current status (censoring) model.

Consider a positive random variable X with (unknown) distribution function F (think of it as time of onset of a disease) and independent of that another random variable T (think of it as inspection time). In the current status model, one only observes T and an indicator variable $\{X \leq T\}$. In words: one observes the inspection time T and the 'current status' at T of the patient (did the disease already set on or not). Item of interest is the distribution of the random variable X. The Nonparametric Maximum Likelihood estimator (NPMLE) of the distribution function F of X in this model, based on n i.i.d. copies of the observables, is well studied and understood. This estimator is a step function and as such non-smooth.

Under the assumption that F has a density f w.r.t Lebesgue measure, the NPMLE is therefore not a natural estimator. We consider two alternative estimators, a maximum smoothed likelihood estimator and a smoothed maximum likelihood estimator, that take smoothness of F into account. For those we present characterizations and asymptotic properties.