

# "Empirical processes in survey sampling"

## Abstract:

This talk is devoted to the study of the limit behavior of extensions of the empirical process, when the data available have been collected through an explicit survey sampling scheme. Indeed, in many situations, statisticians have at their disposal not only data but also weights arising from some survey sampling plans. These weights correspond either to true inclusion probabilities, as is often the case for institutional data, or to some calibrated or post-stratification weights (minimizing some discrepancy with the inclusion probabilities subject to some margin constraints, for instance).

Our main goal is here to investigate how to incorporate the survey scheme into the inference procedure dedicated to the estimation of a distribution function  $P(dx)$  (viewed as a linear operator acting on a certain class of functions  $F$ ), in order to guarantee its asymptotic normality. This problem has been addressed by Breslow and Wellner (2006) in the particular case of a stratified survey sampling, where the individuals are selected at random (with replacement) in each stratum, by means of bootstrap limit results. Our approach is different and follows that of Hajek (1964), extended next by Berger (1998) and is applicable to more general sampling surveys, namely those with unequal first order inclusion probabilities which are close to the Poisson scheme (sequential/rejective, Rao-Sandford or Pareto sampling plans). The main result of this work is a Functional Central Limit Theorem describing the limit behavior of an adequate Horvitz-Thompson type of the empirical process in a superpopulation statistical framework. We show an application to the Kolmogorov-Smirnov statistics in this framework.