

Surveys from large datasets of functional data and estimation of the mean and median curve with full and missing data

Camelia Goga, Université de Franche Comté

Mots-clés : functional data, missing data, survey sampling techniques

In the near future, millions of electricity load curves of French households measured at a very fine scale will be available. All these collected load curves represent a huge amount of information difficult to store due to technical and budgetary constraints. In these situations, survey sampling techniques are attractive alternatives to signal compression techniques since they can offer an interesting trade-off between size of the data and accuracy of estimators of simple indicators such as the mean or the median curves of the electricity consumption. I will present a panorama of different strategies considered to estimate the mean or the median with application on a population test of French electricity consumption curves. Unfortunately, data collection may undergo technical problems resulting in missing values. This problem reduces the accuracy of the estimators and may generate bias. Different approaches can be adapted to deal with missing data in this functional framework: nearest neighbor imputation, kernel smoothing of the discretized trajectories or linear interpolation to the differences around the mean. I will give a comparison of these methods on the estimation of the mean curve of French electricity consumption curves with different scenario of missing data.

Références

- [1] Cardot, H. and De Moliner, A. and Goga, C. Estimation of total electricity consumption curves by sampling in a finite population when some trajectories are partially unobserved. (*à paraître dans Canadian Journal of Statistics, numéro spécial pour le CANSSI project*)
- [2] Cardot, H. and De Moliner, A. and Goga, C. Estimating with kernel smoothers the mean of functional data in a finite population setting. A note on variance estimation in presence of partially observed trajectories. *Statistics & Probability Letters*, 99: 156–166, 2015.
- [3] Cardot, H. and Goga, C. and Lardin, P. Variance estimation and asymptotic confidence bands for the mean estimator of sampled functional data with high entropy unequal probability sampling designs. *Scandinavian Journal of Statistics*, 41: 516-534, 2014.
- [4] Cardot, H. and Goga, C. and Lardin, P. Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electronic Journal of Statistics*, 7: 562-596, 2013.

Camelia Goga, Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, 16 route de Gray 25 000 Besançon

,

,