Smoothing non equispaced heavy noisy data with wavelets kernels

Anestis Antoniadis, Irène Gijbels[*]*, and Jean-Michel Poggi*

[*]Department of Mathematics and Leuven Statistics Research Center, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3001 Leuven (Heverlee), Belgium
E-mail: irene.gijbels@wis.kuleuven.be

We consider a nonparametric noisy data regression model where the unknown regression function is assumed to belong to a wide range of function classes (including discontinuous functions). The distribution of the noise is assumed to be unknown and satisfying some weak conditions. We allow for error distributions that may have heavy tails, so that, for example, no moments of the noise exist. The error is assumed to have zero median. The design points are assumed to be deterministic points, not necessarily equispaced. Since the functions can be nonsmooth and the noise may have heavy tails, traditional estimation methods cannot be applied directly in this situation. We first use local medians to construct variables that are structured as a Gaussian nonparametric regression, as in Brown *et al.* (2008). The difference here is though that the resulting data are not equispaced. We therefore rely on a wavelet block penalizing procedure (see Amato *et al.* (2006)) adapted to non equidistant designs to construct an estimator of the regression function. Under mild assumptions on the design it is shown that the estimator simultaneously attains the optimal rate of convergence over a wide range of Besov classes, without prior knowledge of the smoothness of the underlying functions or prior knowledge of the error distribution. The performances of the procedure is illustrated via a simulation study covering a broad variety of settings. Applications to real data examples are also given.

REFERENCES

Amato, U., Antoniadis, A. and Pensky, M. (2006). Wavelet kernel penalized estimation for non-equispaced design regression. *Statistics and Computing*, **16**, 37–56.

Brown, L.D., Cai, T. and Zhou, H. (2008). Robust nonparametric estimation via wavelet median regression. *The Annals of Statistics*, to appear.