Adaptive Minimax Density Estimation for Huber's Contamination Model under L_p losses

Zhao Ren

University of Pittsburgh
E-mail: zren@pitt.edu

Abstract: Today's data pose unprecedented challenges as it may be incomplete, corrupted or exposed to some unknown source of contamination. In this talk, we address the problem of density function f estimation under L_p losses ($1leq p <infty$) for Huber's contamination model in which one observes i.i.d. observations from $(1-epsilon)f+epsilon g$ and g represents the unknown contamination distribution. We investigate the effects of contamination proportion $epsilon$ among other key quantities on the corresponding minimax rates of convergence for both structured and unstructured contamination classes: for structured contamination, $epsilon$ always appears linearly in the optimal rates while for unstructured contamination, the leading term of the optimal rate involving $epsilon$ also relies on the smoothness of target density class and the specific loss function.

We further carefully study the corresponding adaptation theory in contamination models. Two different Goldenshluger-Lepski-type methods are proposed to select bandwidth and achieve L_p risk oracle inequalities for structured and unstructured contaminations respectively. It is shown that the proposed procedures lead to minimax rate-adaptivity over a scale of the anisotropic Nikol'skii classes for most scenarios except that adaptation to both contamination proportion $epsilon$ and smoothness of density class for unstructured contamination is shown to be impossible. Our technical analysis in adaptive procedures relies on some uniform bounds under the L_p norm of empirical processes developed by Goldenshluger and Lepski.