RUTGERS
School of Arts and Sciences

Department of Statistics

DEPARTMENT OF STATISTICS

Zhao Ren
Professor of Statistics
University of Pittsburgh

Robust Estimation and Inference under Huber's Contamination Model

Wednesday, November 11th, 2020
11:45 AM EST

Zoom Meeting: Meeting ID: 992 7807 9741
Password: 515453

Virtual Coffee session before the seminar at 11:30AM EST

Abstract: In the first part of the talk, we study the robust estimation and inference problem for linear models in the increasing dimension regime. Given random design, we consider the conditional distributions of error terms are contaminated by some arbitrary distribution (possibly depending on the covariates) with proportion ϵ but otherwise can also be heavy-tailed and asymmetric. We show that simple robust M-estimators such as Huber and smoothed Huber, with an additional intercept added in the model, can achieve the minimax rates of convergence under the ℓ_2 loss. In addition, two types of confidence intervals with root-n consistency are provided by a multiplier bootstrap technique when the necessary condition on contamination proportion $\epsilon=o(1/\sqrt{n})$ holds. For a larger ϵ, we further propose a debiasing procedure to reduce the potential bias caused by contamination, and prove the validity of the debiased confidence interval. Our method can be extended to the communication-efficient distributed estimation and inference setting in a straightforward way. In the second part of the talk, we address the problem of density function estimation in $\mathbb{R}^d$ under L_p losses ($1\leq p <\infty$) with contaminated data. We investigate the effects of contamination proportion ϵ among other key quantities on the corresponding minimax rates of convergence for both structured and unstructured contamination over a scale of the anisotropic Nikol'skii classes: for structured contamination, ϵ always appears linearly in the optimal rates while for unstructured contamination, the leading term of the optimal rate involving ϵ also relies on the smoothness of target density class and the specific loss function. The corresponding adaption theory is also investigated by establishing L_p risk oracle inequalities via novel Goldenshluger-Lepski-type methods. An interesting feature is that in certain situation adaptive estimation can become a much harder task with the presence of contamination.

Bio: Zhao Ren is an Associate Professor in the Department of Statistics at the University of Pittsburgh. Prior to joining Pitt, Dr. Ren obtained his Ph.D. in Statistics at Yale University in 2014 under the supervision of Harrison Zhou. He is broadly interested in high-dimensional statistical inference, nonparametric function estimation, graphical models, robust estimation, statistical machine learning and applications in statistical genomics.