

When: Friday 12:40 – 13:30, February 7, 2020

Where: ETB 1020

Speaker: **Abhishek Chakraborty, Ph.D.**
Assistant Professor
Department of Statistics
Texas A&M University



Title: **High Dimensional M -Estimation with Missing Outcomes:
A Semi-Parametric Framework**

Abstract: The abundance of observational data in the big data era equips us with rich amount of information, but also brings with it several statistical challenges, including high dimensionality, incompleteness and selection bias, among others. In this talk, I will discuss high dimensional (HD) M -estimation problems in settings where the response Y is possibly missing at random and the covariates X can be HD compared to the sample size, settings that are of great relevance in various modern studies. The parameter of interest is defined as the risk minimizer for a convex loss, under a non-parametric model, and is itself HD which is a key distinction from most existing work. Standard HD regression and series estimation problems with possibly misspecified models and missing Y are included as special cases, as well as their counterparts in causal inference based on the ‘potential outcomes’ framework. Assuming that the parameter is sparse, we propose to estimate it via an L1-regularized debiased and doubly robust (DDR) estimator based on a HD adaptation of the traditional double robust (DR) estimators’ construction. Under mild assumptions and only some high-level conditions on the nuisance estimators implicitly involved, we establish finite sample performance bounds for the DDR estimator showing its optimal L2 error rate and its DR properties. Further, we propose a desparsified version of the DDR estimator that satisfies an asymptotic linear expansion and facilitates HD inference via Gaussian approximation. All results are validated via simulations.

Bio: Abhishek Chakraborty is an Assistant Professor at the Dept. of Statistics, Texas A&M University, prior to which he was a postdoctoral fellow at the Dept. of Statistics, University of Pennsylvania. He received his Ph.D. in Biostatistics from Harvard University, and his Bachelors and Masters in Statistics from the Indian Statistical Institute. His research broadly focuses on robust and efficient inference with high dimensional and/or incomplete observational data. Some of his methodological research interests include high dimensional semi-parametric inference, missing data and causal inference, semi-supervised or weakly supervised learning, regularized estimation, concentration inequalities, empirical processes etc., with applications in the analysis of data from modern biomedical studies.