Speaker: Yufan LI

Title: Spectrum-aware adjustment: a new debiasing framework with application to principal component regression

Abstract: We introduce a new debiasing framework for high-dimensional linear regression that bypasses the restrictions on covariate distributions imposed by modern debiasing technology. We study the prevalent setting where the number of features and samples are both large and comparable. In this context, state-of-the-art debiasing technology uses a degrees-of-freedom correction to remove the shrinkage bias of regularized estimators and conduct inference. However, this method requires that the observed samples are i.i.d., the covariates follow a mean zero Gaussian distribution, and reliable covariance matrix estimates for observed features are available. This approach struggles when (i) covariates are non-Gaussian with heavy tails or asymmetric distributions, (ii) rows of the design exhibit heterogeneity or dependencies, and (iii) reliable feature covariance estimates are lacking.

To address these, we develop a new strategy where the debiasing correction is a rescaled gradient descent step (suitably initialized) with step size determined by the spectrum of the sample covariance matrix. Unlike prior work, we assume that eigenvectors of this matrix are uniform draws from the orthogonal group. We show this assumption remains valid in diverse situations where traditional debiasing fails, including designs with complex row-column dependencies, heavy tails, asymmetric properties, and latent low-rank structures. We establish asymptotic normality of our proposed estimator (centered and scaled) under various convergence notions. Moreover, we develop a consistent estimator for its asymptotic variance. Lastly, we introduce a debiased Principal Components Regression (PCR) technique using our Spectrum-Aware approach. In varied simulations and real data experiments, we observe that our method outperforms degrees-of-freedom debiasing by a margin.