Multiple Testing and Adaptive Estimation via the Sorted L-One Norm

In many real-world statistical problems, we observe a large number of potentially explanatory variables of which a majority may be irrelevant. For this type of problem, controlling the false discovery rate (FDR) guarantees that most of the discoveries are truly explanatory and thus replicable. In this talk, we propose a new method named SLOPE to control the FDR in sparse high-dimensional linear regression. This computationally efficient procedure works by regularizing the fitted coefficients according to their ranks: the higher the rank, the larger the penalty. This is analogous to the Benjamini-Hochberg procedure, which compares more significant p-values with more stringent thresholds. Whenever the columns of the design matrix are not strongly correlated, we show empirically that SLOPE obtains FDR control at a reasonable level while offering substantial power.

Although SLOPE is developed from a multiple testing viewpoint, we show the surprising result that it achieves optimal squared errors under Gaussian random designs over a wide range of sparsity classes. An appealing feature is that SLOPE does not require any knowledge of the degree of sparsity. This adaptivity to unknown sparsity has to do with the FDR control, which strikes the right balance between bias and variance. The proof of this result presents several elements not found in the high-dimensional statistics literature.

Refreshments will be served after the seminar in 1181 Comstock Hall.