

STATISTICS COLLOQUIUM

TUESDAY, OCTOBER 30, 2012

TALK: 4:00 PM — SCIENCE CENTER LECTURE HALL D

“Stability”

Bin Yu

**Departments of Statistics and EECS
University of California, Berkeley**

ABSTRACT

Reproducibility is imperative for any scientific discovery. Often than not, modern scientific findings rely on statistical analysis of high-dimensional data. At a minimum, reproducibility manifests itself in stability of statistical results relative to “reasonable” perturbations to data and to the model used. Jackknife, bootstrap, and cross-validation are based on perturbations to data, while robust statistics methods deal with perturbations to models.

In this lecture, a case is made for the importance of stability in statistics. Firstly, we motivate the necessity of stability of interpretable encoding models for movie reconstruction from brain fMRI signals. Secondly, we find strong evidence in the literature to demonstrate the central role of stability in statistical inference.

Thirdly, a smoothing parameter selector based on estimation stability (ES), ES-CV, is proposed for Lasso, in order to bring stability to bear on cross-validation (CV). ES-CV is then utilized in the encoding models to reduce the number of predictors by 60% with almost no loss (1.3%) of prediction performance across over 2,000 voxels. Last, a novel “stability” argument is seen to drive new results that shed light on the intriguing interactions between sample to sample variability and heavier tail error distribution (e.g. double-exponential) in high dimensional regression models with p predictors and n independent samples. In particular, when $p/n \rightarrow \kappa \in (0.3, 1)$ and error is double-exponential, OLS is a better estimator than LAD.