Final Project Suggestions
Stat 212A, Fall 2015: Topics in Selective Inference
Instructor: Will Fithian

Options for the project are:

- a 25-minute presentation with 5 minutes for questions at the end
- a 10-page report on the topic

Please sign up for a time November 13 or 14 (next Thursday or Friday) to chat with me about what topic you're interested in. A 1–2-page proposal will be due Thursday November 20 (graded for completion). You can present one of the selected papers below, or work on your own research idea, or some combination, e.g. extending an idea in one of the papers.

If you choose to present one of these papers, it is neither expected nor possible in most cases to present the entirety of the paper. Rather, you should understand and synthesize the ideas yourself, and shoot for a maximally educational 25-minute (or 10-page) presentation in which you summarize the motivation and main animating idea of the paper, summarize the main results, possibly give a short proof or proof sketch of an important result, possibly give an example of the method or idea in action, etc.

Some Interesting Papers

An asterisk (*) indicates the paper seems more technically challenging.

Linear regression and the submodel view:

Linear Regression, Submodel View, Random Design, Philosophy of Model Selection:
Andreas Buja, Richard Berk, Lawrence Brown, Edward George, Emil Pitkin, Mikhail Traskin, Linda Zhao, and Kai Zhang. Models as approximations: A conspiracy of random regressors and model deviations against classical inference in regression

Random Forests, Submodel View, Causal Inference:
Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *arXiv preprint arXiv:1510.04342*, 2015 (*)

Asymptotics, high-dimensional inference, and theory:

Theoretical impossibility result about post-selection inference:
Hannes Leeb and Benedikt M Pötscher. Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics*, pages 2554–2591, 2006 (*)

High-Dimensional Regression with Selected Predictors:
Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference for high-dimensional sparse econometric models. *arXiv preprint arXiv:1201.0220*, 2011 (*)

High-Dimensional Regression with Selected Predictors:
Cun-Hui Zhang and Stephanie S Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242, 2014 (*)

Empirical Bayes ideas on selective inference in one-way layout setup:

Consequences of Correlation for BH procedure:

Bradley Efron. Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association*, 102(477), 2007

Empirical null distribution:

Bradley Efron et al. Microarrays, empirical bayes and the two-groups model. *Statistical science*, 23(1): 1–22, 2008

Bias correction via Tweedie's formula:

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011

Replicability:

Replicability in meta-analysis via "partial conjunction nulls":

Yoav Benjamini and Ruth Heller. Screening for partial conjunction hypotheses. *Biometrics*, 64(4): 1215–1222, 2008

Large-Scale Replication of Psych studies:

Open Science Collaboration et al. Estimating the reproducibility of psychological science. *Science*, 349 (6251):aac4716, 2015

Knockoffs:

Extension to FWER control:

Lucas Janson and Weijie Su. Familywise error rate control via knockoffs. *arXiv preprint arXiv:1505.06549*, 2015

Conditional inference:

Improving Selective Inference via Randomized Selection:

Xiaoying Tian and Jonathan E Taylor. Selective inference with a randomized response. *arXiv preprint arXiv:1507.06739*, 2015 (*)

Inference for Square-Root Lasso:

Xiaoying Tian, Joshua R Loftus, and Jonathan E Taylor. Selective inference with unknown variance via the square-root lasso. *arXiv preprint arXiv:1504.08031*, 2015

Work in CS on adaptive / selective inference using differential privacy:

Re-usable holdout set:

Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. Preserving statistical validity in adaptive data analysis. *arXiv preprint arXiv:1411.2664*, 2014

(also see shorter *Science* publication:)

Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. The reusable holdout: Preserving validity in adaptive data analysis. *Science*, 349(6248):636–638, 2015 (*)