Department of Biomedical Data Science AT STANFORD MEDICINE

Workshop in Biostatistics

MSOB X303

DATE:	October 17, 2019
TIME:	1:30-2:50pm
TITLE:	Inference on Treatment Effects after Model Selection with application to subgroup analysis
SPEAKER:	**Jingshen Wang, Ph.D.** Assistant Professor in Biostatistics, UC Berkeley

Abstract:
Inferring cause-effect relationships between variables is of primary importance in many sciences. In this talk, we start by discussing two approaches for making valid inference on treatment effects when a large number of covariates are present. The first approach is to perform model selection and then to deliver inference based on the selected model. If the inference is made ignoring the randomness of the model selection process, then there could be severe biases in estimating the parameters of interest. While the estimation bias in an under-fitted model is well understood, a lesser-known bias that arises from an over-fitted model will be addressed. The over-fitting bias can be eliminated through data splitting at the cost of statistical efficiency, and we propose a repeated data splitting approach to mitigate the efficiency loss. The second approach concerns the existing methods for debiased inference. We show that the debiasing approach is an extension of OLS to high dimensions. A comparison between these two approaches provides insights into their intrinsic bias-variance trade-off, and the debiasing approach may lose efficiency in observational studies. For the second part of the talk, we discuss a generalization on how to estimate treatment effects for groups of individuals that share similar features in observational studies. More importantly, even if the estimated treatment effects suggests a promising subgroup (i.e. the group with the maximal treatment effect), we address the question of how good the subgroup really is.

Suggested Reading:

- "Debiased Inference on Treatment Effect in a High-Dimensional Model" https://amstat.tandfonline.com/doi/full/10.1080/01621459.2018.1558062