

MICHIGAN STATE UNIVERSITY
Department of Statistics and Probability

COLLOQUIUM

Snigdha Panigrahi
University of Michigan

Carve Your Data for Adaptive Inference

Tuesday, April 9, 2019
10:20 AM - 11:10 AM
Refreshments 10:00 AM
C405 Wells Hall

Abstract

Any meaningful statistical problem consists of a model and a set of matching parameters for which decisions are required. In the classical, “textbook” paradigm, the model and this set of target parameters are assumed to be chosen independently of the data subsequently used for statistical inference (e.g., estimating the parameters). In practice, however, more often than not, data analysts examine some aspect of the data before deciding on a model and/or the target parameters. For example, in fitting multiple linear regression, one might decide to discard variables with large p-values and re-fit the smaller model before reporting any findings; and in multiple hypothesis testing, the analyst might be tempted to report confidence intervals only for rejected nulls. Of course, ignoring such form of adaptivity in choosing the model and/or the target parameters, may result in the loss of inferential guarantees and lead to flawed conclusions.

My talk today will be based on addressing problems of overcoming selection-bias in a modern, iterative framework of science- where, the researcher begins with an initial data set which she might use for selection. But, usually further observations are made available at a future point in time either because the researcher decided to collect more data after seeing the outcome of the initial analysis, or simply because another data set comes in later on. At this point, the researcher is confronted with the question of how to combine the two data sets to provide inference for parameters selected based only on the first data set. A compellingly simple way is to consider inference based on a split-likelihood, using only on the second data set which has not been used for selection.

I will take a more optimal and less wasteful approach to this problem, dubbed as “carving” and cast the described two-staged scientific procedures into a conditional framework. Introducing a carved-likelihood, I will talk about a mathematical framework to conduct model-free inference based upon the same and discuss an implementation toolbox for practitioners.

To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.