Statistics and Data Science Seminar

Testing goodness-of-fit and conditional independence with approximate co-sufficient sampling

Rina Foygel Barber (University of Chicago)

Abstract: Goodness-of-fit (GoF) testing is ubiquitous in statistics, with direct ties to model selection, confidence interval construction, conditional independence testing, and multiple testing, just to name a few applications. While testing the GoF of a simple (point) null hypothesis provides an analyst great flexibility in the choice of test statistic while still ensuring validity, most GoF tests for composite null hypotheses are far more constrained, as the test statistic must have a tractable distribution over the entire null model space. A notable exception is co-sufficient sampling (CSS): resampling the data conditional on a sufficient statistic for the null model guarantees valid GoF testing using any test statistic the analyst chooses. But CSS testing requires the null model to have a compact (in an information-theoretic sense) sufficient statistic, which only holds for a very limited class of models; even for a null model as simple as logistic regression, CSS testing is powerless. In this paper, we leverage the concept of approximate sufficiency to generalize CSS testing to essentially any parametric model with an asymptotically-efficient estimator; we call our extension “approximate CSS” (aCSS) testing. We quantify the finite-sample Type I error inflation of aCSS testing and show that it is vanishing under standard maximum likelihood asymptotics, for any choice of test statistic. We apply our proposed procedure both theoretically and in simulation to a number of models of interest to demonstrate its finite-sample Type I error and power. This work is joint with Lucas Janson.