samsi
NSF·Duke·NCSU·UNC

Sixth Bayesian, Fiducial, and Frequentist (BFF6) Conference on Model Uncertainty
April 28, 2019 – May 1, 2019

SPEAKER TITLES/ABSTRACT

Chiara Sabatti
Stanford University

"Selecting Important Features in Presence of Correlation—a story from Genetics"

A research goal in genetics is to identify which DNA variants influence traits of medical interest. In Genome Wide Association Studies (GWAS), hundreds of thousand polymorphisms are genotyped and statistical approaches are used to identify places in the genome where DNA variation is associated to phenotypic variation, as well as to identify those variants that are more likely to have causal effects. While the ultimate scientific goal is well defined, its translation in the language of statistical inference has been somewhat challenging. On the one hand, in order to rely on available statistical methodology, researchers have often tested hypotheses that are only "approximation" of those of interest. On the other hand, the presence of strong dependence among groups of the studied polymorphisms poses specific challenges in describing interpretable hypotheses that can be investigated with substantial power.

In work with Matteo Sesia, Eugene Katsevich and Emmanuel Candes we tackle this problem aiming to make interpretable and precise findings, using the Knockoffs framework to control the rate of false discoveries. The lessons we learn have applications outside GWAS.