Making sense of complex disorders with semiparametric canonical correlation analysis

This talk discusses the study of complex disorders where certain aspects of the phenotype or outcome are uncharacterized before the start of the study. The majority of the talk will focus on the problem of identifying sets of patients who are at high risk for multiple autoimmune diseases based on their genetic characteristics. We address two primary complicating factors in this type of study. First, a diverse array of phenotypic data, including continuous, discrete, and censored elements, may be necessary to properly capture a multidimensional disorder such as autoimmunity. Second, it is unknown how best to combine phenotypic information to relate it to genetic characteristics. To address these issues, we develop semiparametric canonical correlation analysis (sCCA) which allows us to produce risk scores for optimal combinations of the diverse phenotypic information. A nonparametric method is then used to calibrate the risk scores and identify patients at high risk of multiple (etiologically related) autoimmune diseases. We demonstrate the asymptotic behavior of sCCA estimators. The rest of the talk will touch on other issues in the study of complex disorders, including understanding heterogeneous disorders such as autism spectrum disorder.