Application of High Dimensional Covariance and Precision Matrices in Clinical Research

Tianxi Cai*
Harvard University, USA - tcai@hsph.harvard.edu

Large datasets containing both a wealth of clinical and experimental data now exist as a result of the increasing adoption of electronic medical records (EMR) and availability of clinical research registries linked with specimen bio-repositories. These datasets allow for deriving data driven classification and prediction of sub-phenotypes with high dimensional genomic and phenotypic patient level data. In this talk, I'll discuss the application of various large scale covariance matrix and precision matrix testing/estimation methods that are useful for such purposes and illustrate both the challenges and potential opportunities that arise from analyzing large scale clinical research data. For example, linking the genomic and biological markers to a wide range of disease phenotypes in the EMR enables us to conduct phenome-wide association studies (PheWAS) to rigorously study genome-phenome association networks. This approach could allow the discovery of new subtypes of disease, along with their genetic causes. The high dimensional genomic data also enables us study gene-by-gene interactions and how they impact disease phenotypes. I will discuss some high dimensional precision matrix testing methods that can be used to address such questions.