

Department of Applied and Computational Mathematics and Statistics Colloquium



Hongzhe Li
Department of Biostatistics and Statistics
University of Pennsylvania

Hypothesis Testing for Phylogenetic Composition: A Minimum-cost Flow Perspective

Quantitative comparison of microbial composition from different populations is a fundamental task in various microbiome studies. We consider two-sample testing for microbial compositional data by leveraging the phylogenetic tree information. Motivated by existing phylogenetic distances, we take a minimum-cost flow perspective to study such testing problems. We first show that multivariate analysis of variance with permutation using phylogenetic distances, one of the most commonly used methods in practice, is essentially a sum-of-squares type of tests and has better power for dense alternatives. However, empirical evidence from real data sets suggests that the phylogenetic microbial composition difference between two populations is usually sparse. Motivated by this observation, we propose a new maximum type test, Detector of Active Flow on a Tree, and investigate its properties. We show that the proposed method is particularly powerful against sparse phylogenetic composition difference and enjoys certain optimality. The practical merit of the proposed method is demonstrated by simulation studies and an application to a human intestinal biopsy microbiome data set on patients with ulcerative colitis.

Tuesday, March 3, 2020

4:15PM – 5:15 PM

127 Hayes-Healy Center

Colloquium Tea 3:45 PM to 4:15 PM 101A Crowley Commons Room