



Department of Decision Sciences

Statistics Seminars

Tree-based models for high-dimensional compositional data in microbiome studies

Li Ma

Duke University

Thursday, 3rd March 2022
12:00 pm Room AS02 Via Roentgen 1 Milano

Abstract

Human gut microbiome are associated with various diseases and health outcomes. A key characteristic of microbiome compositional data is its large and complex cross-sample heterogeneity. Appropriately accounting for these functional “variance components” is critical for several common inference tasks, including identifying latent structures, carrying out hypothesis testing on cross-group differences, and modeling dynamics, but is complicated by the key features of microbiome compositional data including high-dimensionality, sparsity, and compositionality. These characteristics incur the need for structural constraints on covariance modeling while maintaining the analytical and computational tractability of the resulting models and methods. In this talk, I will present several recently proposed methods that aim to utilize a tree structure---namely the phylogeny of the microbial species---to incorporate flexible covariance components while maintaining computational scalability. In particular, I will present probabilistic models for microbiome compositional data based on the Dirichlet-tree (DT) distribution and the logistic-tree normal (LTN) distribution, and demonstrate their wide applicability in a range of applications including cross-sample comparison, mixed-effects modeling, covariance estimation, clustering analysis, and subcommunity identification. Their performance will be contrasted with previous models such as the Dirichlet-multinomial models and log-ratio based models. A connection to models in the Bayesian nonparametric literature will also be drawn. This talk is based on joint work with my students Jialiang Mao, Zhuoqun Wang, and Patrick LeBlanc.