



THE UNIVERSITY OF CHICAGO

Department of Statistics

DISSERTATION PRESENTATION AND DEFENSE

YUNFAN TANG

Department of Statistics
The University of Chicago

Models and Inference for Microbiome Data

THURSDAY, April 19, 2018, at 9:00 AM
Jones 304, 5747 S. Ellis Avenue

ABSTRACT

Microbiome refers to the full collection of all microorganisms in a community. Recent advances in sequencing technologies has allowed scientists to quantify the microbiome compositions at an unprecedented resolution. However, high-throughput sequencing data have unique characteristics such as high dimensional, sparse, and compositional. Moreover, there exists a phylogenetic tree that quantifies the evolutionary similarity among all taxa. This thesis presents novel statistical models to analyze microbiome data by leveraging these unique characteristics. The first problem we consider is testing equality of microbial compositions among groups of populations. We apply the Dirichlet-tree multinomial distribution (DTM), a generalization of the traditional Dirichlet-multinomial (DM) distribution, and design a scan statistic that takes advantage of the phylogenetic relations among taxa. We provide an upper bound of p-value using this scan statistic and show that this method has improved power in an empirical dataset as well as in simulation. The second problem continues the investigation of DTM vs DM by introducing a penalization method that selects the best model along the DM-DTM spectrum. The last problem we address is estimating heritability when the input is a matrix of pairwise dissimilarities calculated from beta diversities. Beta diversity is an ecologically meaningful way to measure pairwise compositional differences by taking variations in all taxa into account. We extend the traditional ACE variance component model to the matrix case using Wishart distribution. We also present a new beta diversity measurement, named root-Unifrac, that matches the positive definiteness requirement of the Wishart distribution. This Wishart ACE model allows us to directly measure community heritability, which quantifies the contribution of additive genetics to overall variations in beta diversity.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: <https://lists.uchicago.edu/web/subscribe/statseminars>.