

BIostatistics Seminar



Faming Liang, PhD

Professor
Dept. of Statistics
Purdue University

An Imputation-Consistency Algorithm for Biomedical Complex Data Analysis

The dramatic improvement in data collection and acquisition technologies in the past two decades has enabled scientists to collect vast amounts of health-related data in biomedical studies. If analyzed properly, these data can help us to improve contemporary healthcare services from diagnosis to prevention to personalized treatment, and also provide us some insights toward reducing healthcare costs. However, the biomedical data can be rather complex, which are often characterized by some mixture of missing data, high dimensionality, heterogeneity, high variety, high volume, high velocity, etc. How to analyze these data has posed many challenges on existing methods. Toward an efficient use of biomedical complex data, we propose an imputation-consistency (IC) algorithm as a general algorithm for high-dimensional missing data problems. The IC algorithm works by iterating between an imputation step and a consistency step. At the imputation step, the missing data are imputed conditional on the observed data and the current estimate of parameters; and at the consistency step, a consistent estimate is found for the minimizer of a Kullback-Leibler divergence defined on the pseudo-complete data. The consistency of the averaged IC estimate is established under quite general conditions. Then, under the principles of conditioning and consistency, we extend the IC algorithm to address some other challenges encountered in biomedical complex data analysis, such as data heterogeneity in biomarker identifications and eQTL analysis. Some issues on high-dimensional inference will also be addressed in the talk.

Thursday September 20, 2018

3:30 pm - 4:30 pm

Blue Cross and Blue Shield of North Carolina Foundation Auditorium



UNC
GILLINGS SCHOOL OF
GLOBAL PUBLIC HEALTH