Indiana University-Purdue University Indianapolis

Department of Mathematical Sciences

STATISTICS SEMINAR

12:15pm—1:15pm, Tuesday, April 03, 2018
SL 137

Speaker: **Dr. Sha Cao** (Assistant Professor)
Department of Biostatistics, School of Medicine, Indiana University

Title: Computational decomposition of tumor tissue heterogeneity using sparse dictionary learning

Abstract:

Tumor tissue expression data provide substantial information to study cancer biology and evolution in its actual microenvironment. However, it is very challenging for information discovery from tissue data because of their compositional complexity each represents a mixture of gene-expression data from multiple cell types. Hence, meaningful tissue-data analyses require to first sort out the detailed contributions by different cell types. However, the computational challenge in solving the tissue data deconvolution problem stems from the reality that the number of free variables to be estimated is huge it is a multiple of total number of genes (20,000) and cell types (10). Existing algorithms for tissue data deconvolution have been published with very limited applications. We have recently developed a deconvolution algorithm and demonstrated its effectiveness on simulated data, i.e., in silico mixtures of gene-expression data from multiple cell types with varying proportions of each cell type. The unique ideas of our approach are: for each cell type, a sparse representation of the basic elements were derived on a large scale training expression data measured on the pure cells, based on dictionary learning scheme and prior information that over 20,000 gene sets, whose elements were curated to share certain co-regulation or co-expression relationships. This not only captured the complex relations among genes preserved in a cell type, but also greatly reduced the number of free variables to be estimated. Then, a mixture tissue expression was represented as a (non-negatively) weighted sum of the cell type specific basic elements derived as aforementioned.