UMass Lowell
Mathematical Sciences Colloquium

Date & Time:  Tuesday February 21 at 4:00 PM
Location:  212 Olney Hall (North Campus)
Speaker:  Jingjing Yang (Biostatistics Department, University of Michigan)

Title:  A scalable Bayesian Method for Integrating Functional Information in Genome-wide Association Studies

Abstract:  Although genome-wide association studies (GWASs) have identified many complex loci, most of which reside in noncoding regions and have unknown biological functions. Integrative analysis that incorporates known functional information into GWAS can help elucidate the underlying biological mechanisms and prioritize causal-variants. We develop a novel, flexible Bayesian variable selection model with efficient computational techniques for such integrative analysis. Different from previous approaches, our method models the effect-size distribution and probability of causality for variants with different annotations and jointly models genome-wide variants to account for linkage disequilibrium (LD), thus prioritizing associations based on the quantification of the annotations and allowing for multiple causal-variants per locus. Our method dramatically improves both computational speed and posterior sampling convergence by taking advantage of the block-wise LD structures in human genomes. In simulations, our method accurately quantifies the functional enrichment and performs more powerful for identifying true causal-variants than alternative methods, where the power gain is especially apparent when multiple causal-variants in LD reside in the same locus. We applied our method to an in-depth GWAS of age-related macular degeneration with 33,976 individuals and 9,857,286 variants. We find the strongest enrichment for causality among non-synonymous variants (54x more likely to be causal, 1.4x larger effect-sizes) and variants in active promoter (7.8x more likely, 1.4x larger effect-sizes), as well as identify 5 potentially novel loci in addition to the 32 known AMD risk loci. In conclusion, our method is shown to efficiently integrate functional information in GWASs, helping identify causal-variants and underlying biology.