

Statistics and Data Science Seminar

Bayesian regression of genome-wide association summary statistics

Xiang Zhu (Pennsylvania State University)

Abstract: Large-scale genome-wide association studies (GWAS) have markedly improved our understanding of how common variation in the human genome affects complex traits and diseases. Regression models have been widely used to analyze GWAS, but existing methods often require input data at the individual level, which are hard to obtain due to many administrative issues. Here we provide a Bayesian framework for multiple regression without the need of individual-level data. Specifically, we derive a "Regression with Summary Statistics" (RSS) likelihood function of the multiple regression coefficients based on the univariate regression summary statistics, which are easily available in GWAS. We combine the RSS likelihood with prior distributions that are specifically designed for a wide range of genetic applications, such as heritability estimation, phenotype prediction, pathway enrichment and gene prioritization. To estimate posterior distributions, we develop efficient Markov chain Monte Carlo and variational inference algorithms that scales well with millions of genetic variants. Applying RSS to a host of real-world GWAS summary statistics, we demonstrate that RSS not only achieves similar performance in settings where existing methods work, but also enables many novel analyses and discoveries that existing methods cannot deliver.

Wednesday, April 26 at 4:00 PM in 636 SEO