Bayesian variable selection in generalized linear mixed models with non-local priors for non-Gaussian GWAS data

Marco Ferreira
November 10 | 3.30 PM (ET)

In-person Seminar

Join Online: Zoom Meeting
https://virginiatech.zoom.us/j/86237780926?pwd=cjVDTThKTUpnUnBQSGG4Q1RFVzMxQT09

Meeting ID: 862 3778 0926, Passcode:_STATSROCKS

Abstract: We present a novel Bayesian method to find single nucleotide polymorphisms (SNPs) associated with non-Gaussian phenotypes from genome-wide association studies (GWAS). We call our method Bayesian GLMMs for GWAS (BG2). BG2 solves a variable selection problem in generalized linear mixed models (GLMMs) where the number of regressors \( p \) is two to three orders of magnitude larger than the sample size \( n \). In addition, the observations are correlated and the SNPs regressors are highly correlated. To deal with these challenges, we propose novel nonlocal priors specifically tailored to GLMMs and develop related fast approximate computations for Bayesian model selection. To search through hundreds of thousands of possible SNPs, BG2 uses a two-step procedure: first, BG2 screens for candidate SNPs; second, BG2 performs model search that considers all screened candidate SNPs as possible regressors. A simulation study shows favorable performance of BG2 when compared to other methods widely used in the GWAS literature. We illustrate the use of BG2 with case studies based on real GWAS datasets: a case study with count data on the number of root-like structures in the model plant A. Thaliana, and a case study on binary data related to human health. This is joint work with Shuangshuang Xu and Jake Williams.