Approach

Approach

Overall Research Design

Specific Aim 1: We will use mean field variational methods to provide analytic approximations to the posterior distribution for a Bayesian model with sparsity-inducing priors for polygenic risk scores.

Introduction

Current inference methods for Bayesian polygenic risk score models use sampling-based strategies like Markov chain monte carlo and, thus, pose significant computational burdens for modern human genetics studies with large sample sizes and high-dimensional measurements. The objective of Specific Aim 1 is to use computationally efficient and scalable variational inference methods for posterior inference in polygenic risk score models. Variational inference uses an analytic approximation to the posterior distribution to estimate quantities of interest. To achieve this objective, we will use the polygenic risk score model developed @prive2020ldpred2. However, instead of using the sampling-based inference methods of @prive2020ldpred2, we will apply the variational inference methods of @ray2020spike and @yang2020alpha for approximate inferences from the posterior distribution. The rationale is that our variational inference-based strategy will diminish computing time and memory requirements while maintaining predictive ability of the sampling-based strategy of @prive2020ldpred2. Moreover, with the need to model polygenic risk scores from studies with sample sizes nearing one million subjects, each of which has thousands of phenotype measurements, current sampling-based strategies for posterior inference are inadequate. Variational methods, on the other hand, are computationally efficient and scalable to big data sets. We will compare our method's performance - in terms of predictive ability and computing resource requirements - against that of @prive2020ldpred2. Upon completion of Specific Aim 1, it is our expectation that we will have created a computationally scalable and efficient method for constructing polygenic risk scores. Our open source implementation ensures transparency in our research and provides a valuable analytic tool to human genetics researchers.

Research Design

Study Data