Polygenic score construction with mean field variational inference to model SNP-SNP interactions

2023-10-12

Polygenic score

- Uses genome-wide SNP risk allele counts to produce a number that summarizes genetic risk for a disease of interest, like coronary artery disease
- Polygenic scores typically rely on SNP effect estimates from genome-wide association studies
- Current uses include risk stratification for preventive interventions

SNP-SNP interactions

- Existing polygenic score construction methods consider only SNP main effects and neglect SNP-SNP interactions
- We hypothesize that our variational inference methods, and the associated reduction in computing requirements, will enable construction of polygenic risk scores that model SNP-SNP interactions
- We expect polygenic risk scores that model SNP-SNP interactions (and include SNP main effects) to outperform current methods in predictive accuracy

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

Rationale for Aim 1

Our variational inference-based strategy will diminish computing time and memory requirements while maintaining predictive ability of the sampling-based strategy of Privé, Arbel, and Vilhjálmsson (2020)

Outcomes for Aim 1

- A computationally scalable and efficient method for constructing polygenic risk scores
- Reproducible, open source implementation ensures transparency in our research and provides a valuable analytic tool to human genetics researchers

Specific Aim 2

We will develop a Bayesian statistical model for polygenic risk scores based on SNP effect estimates and estimates for SNP-SNP interaction effects

Rationale for Aim 2

- SNP-SNP interactions, in some diseases, explain a non-negligible proportion of the trait variance
- Therefore, polygenic risk scores that explicitly model SNP-SNP interactions for these diseases is expected to be more predictive of disease than scores from current models that ignore interactions

Outcomes for Aim 2

- A new class of Bayesian statistical models for polygenic risk scores using SNP main effects and SNP-SNP interactions
- Well designed, thoroughly tested, reproducible, open source software package that implements our methods