FaST Linear Mixed Models for Genome-Wide Association Studies

Christoph Lippert

Abstract: Factored Spectrally Transformed Linear Mixed Models (FaST-LMM) scales linearly in the number of individuals in both runtime and memory use. The problem of confounding includes population structure, family structure and cryptic relatedness. Statistical methods for correcting these confounders include linear mixed models (LMM), genomic control, family-based association tests, structured association, and Eigenstrat. LMM can capture all of these confounders simultaneously, without knowledge of which are present, and without the need to tease them apart. However, it is computationally expensive relative to the simpler models. The runtime and memory footprint required by these models scales as the cube and square of the number of individuals in the dataset, respectively.

Roughly speaking, LMM tackle confounders by using measures of genetic similarity to capture the probabilities that pairs of individuals share causative alleles. Such measures include those based on IBD and the realized relationship matrix (RRM).