Summary of “On set-based association tests: insights from a regression using summary statistics”

Questions to answer:

1. What is “Set-based” meaning in genetics?

An association study with only a set of SNPs in a specific gene or pathway.

2. What does sparse signal means in association study?

Rare variant means that type of genetic variant has small MAF (minor allele frequency), whereas the sparse signal refers to the problem that in the alternative hypothesis, there are only a few variants are associated with the phenotype (very few of the constraints in the null hypothesis are not true).

3. How is the exponential family’s score vector derived?

Refer to the wiki page.

4. How to understand the “minor allele frequency” measure?

The frequency of the less frequent allele in the population.

Summary:

1. The proposed method only requires the information of variant-specific statistics, and the individual level data’s statistics can be shown to a special case of the proposed method.

2. The proposed method can improve the power of detecting sparse alternatives by transforming the variant-specific statistics using its precision matrix.

3. For set-based analysis of multiple rare genetic variants, the statistical tests are linear(burden) test, quadratic(variance-component) test, and hybrid test. But these methods may not perform well in the sparse signal setting (require most of the variants being causal to detect any effect).

4. The proposed method has a mixed effect unifying model that has all the three kinds of tests as special cases.

5. Linear tests are only powerful if all the effects have the same direction, otherwise there might be some cancellation in the computation.