Simulation Protocol

Data generation

1. We will use real genotypes obtained after QC from Blacks and Whites in MESA to mimic real allele frequency distributions and LD patterns.
2. We will sample SNPs will be sampled at random from the overall pool of SNPs, which will be our causal variants or QTL.
3. Phenotypes will be simulated for Whites and Blacks under an additive genetic model

$$y_i = \sum_{j=1}^{5,000} Z_{ij}\beta_j + \varepsilon_i; i = 1, \ldots, n$$

   at different
   - numbers of QTL (1000, 10,000, 50,000)
   - trait heritability (30%, 70%)
   - marker effect correlation (30%, 70%)

4. Model errors $\varepsilon_i \sim N(0, 1)$, marker effects β_j follow bivariate normal distributions with zero mean, and Z_{ij} are the genotype readings at the QTL.
5. All QTL will be drawn from a bivariate normal distribution to mimic a complex trait

$$\begin{pmatrix} \sigma^2_{\beta_1} & \rho\sigma_{\beta_1}\sigma_{\beta_2} \\ \rho\sigma_{\beta_1}\sigma_{\beta_2} & \sigma^2_{\beta_2} \end{pmatrix}$$

Model fitting

- The next step is to fit the interaction model by assuming a Gaussian prior on marker effects (called the interaction G-BLUP).
- Two scenarios will be considered:
  - with only causal loci (i.e. when the likelihood function is correctly specified)
  - with all SNPs (i.e. when the likelihood function is mis-specified)

- In the first scenario, we expect to see low bias and MSE in our estimates of cluster-

specific genomic heritabilities and between-cluster correlations.

- The second scenario will shed light on the influence of excluding causal variants on bias and MSE of the estimates.
- All analyses will be conducted in R using the package BGLR (DE LOS CAMPOS and RODRIGUEZ 2014).

> Written with StackEdit.