1 Warfarin Data

1.1 Data description

For this data set, the sample size is 304. Two phenotypes are measured, labeled as ydose and yinr, respectively. Also, some covariant are recorded: sex, age, drug, height and weight. The genotypes for two genes are provided. Gene A contains 7 SNPs and gene B contains a 4-allelic marker.

1.2 Data Analysis

Two Trait-Gene Similarity regression models with 3 tests are performed for this data set. The first model contains the Similarity between 2 genes and the similarity for the interaction, as described below:

$$Z_{ij} = b^A S^A_{ij} + b^B S^B_{ij} + b^{AB} S^{AB}_{ij} + e_{ij}$$

where Z_{ij} is the normalized trait similarity between individual i and j, S^A_{ij} and S^B_{ij} record the gene similarity between individual i and j for gene A and gene B respectively. S^{AB}_{ij} describes the genetic similarity of the interaction between gene A and gene B. b^A, b^B and b^{AB} are the effects of S^A_{ij}, S^B_{ij} and S^{AB}_{ij}, respectively.

Since the data is a little different from the usual 2-allelic marker data. The genetic similarity matrix for a gene is computed in the following way:

Define M is number of the SNPs the gene contains. For ith individual's mth marker, two SNPs are recorded, denoted as G^m_{i1} and G^m_{i2}, then we have the genetic similarity s^m_{ij} between individual i and j for the mth marker in a particular gene:

$$s^m_{ij} = \begin{cases} 1, & G^m_{i1} = G^m_{i2} = G^m_{j1} = G^m_{j2} \\ 0, & G^m_{i1} \neq G^m_{j1} \text{ and } G^m_{i1} \neq G^m_{j2} \text{ and } G^m_{i2} \neq G^m_{j1} \text{ and } G^m_{i2} \neq G^m_{j2} \\ 0.5, & o.w \end{cases}$$

and the similarity of the gene S_{ij} is

$$S_{ij} = \frac{1}{M} \sum_{m=1}^{M} s^m_{ij}$$

and $S^{AB}_{ij} = S^A_{ij} \times S^B_{ij}$.

There are two tests under this model. One the joint test under the hypothesis: $b^A = b^B = b^{AB} = 0$, the other one is the interaction test under the hypothesis $b^{AB} = 0$.

Other model for this data set is shown below:

$$Z_{ij} = b^A S^A_{ij} + b^B S^B_{ij} + e_{ij}$$

In this model, we just consider the main effect of the gene similarity. The goal of this model is to test $b^A = 0$ with the adjustment of Gene B.

Traits	Joint test	Epi test	One Gene Test Gene A	Gene B
ydose	3.15×10^{-21}	1	2.6×10^{-22}	3.38×10^{-8}
yinr	0.41	0.54	0.46	0.59

This table shows that for either trait, the epi test is not significant. However, trait ydose has a very significant in joint test, and One gene test for both gene A and gene B. Can we use it as an application of One gene test?

2   Debug

There is a weird problem I met when I debug the code for Epi test. Although now I directly use the similarity matrix S_A and S_B to estimate τ_A and τ_B through EM algorithm, which should get rid of the haplotype H and the variance matrix R for the haplotype effect, I still use the way you used in your one gene Trait-Similarity regression model to compute a fake H and R to facilitate the computation for V^{-1}. The H_B and R_B from S_B can not fully satisfy the restriction that $S_B = H_B R_B H_B^T$, I think that is caused by the numerical calculation error. The weird part is, if I directly use S_B to estimate tau_B in the loop, τ_B can not converge. However, if I compute $S_B^* = H_B R_B H_B^T$ and replace S_B by S_B^* in the loop, everything is fine. I don't know whether it is big question since S_B^* is quite close to S_B. But it is better to point out.