① Consider SNPs in a given gene. run Han's code to get their scores in population p ($p=1,\cdots,P$). define the scores as

$(Z_1^{(p)}, Z_2^{(p)}, \cdots, Z_K^{(p)})$. K is the number of SNPs in the gene.

Note: get these scores directly, no need to scale back. let the covariance matrix be $\sum^{(p)}$, (from Han's code, no need to scale it back, just take from the output, this matrix should be the one for the whole block, which includes several genes)

② From $\sum^{(p)}$, generate N copies of Z. put them in the matrix

$$Z^{(p)} = \begin{pmatrix} Z_{1,1}, & Z_{1,2}, & \cdots, & Z_{1,K} \\ Z_{2,1} & Z_{2,2}, & \cdots, & Z_{2,K} \\ \vdots & & & \\ Z_{N,1}, & Z_{N,2}, & \cdots, & Z_{N,K} \end{pmatrix}$$

③ For SNP k, do the following steps, $k=1,\cdots,K$

ⓐ let $t_{0,K} = \max_p \left\{ \frac{|Z_K^{(p)}|}{\sigma_K^{(p)}}, p=1,\cdots,P \right\}$, where

$\sigma_k^{(p)}$ is $\sqrt{\sum_{(k,k)}^{(p)}}$, square root of diagnal term in $\sum^{(p)}$.

ⓑ For each i^{th} row of $Z^{(p)}$, $p=1,\cdots,P$. let

$t_{i,k} = \max_p \left\{ \frac{|Z_{i,k}^{(p)}|}{\sigma_K^{(p)}}, p=1,\cdots,P \right\}$, with the same $\sigma_K^{(p)}$ as in Step ⓐ.

(3) Form the following matrix

$$G = \begin{pmatrix} t_{0,1} & t_{0,2} & \cdots & t_{0,K} \\ t_{1,1} & t_{1,2} & \cdots & t_{1,K} \\ \vdots & & & \\ t_{N,1} & t_{N,2} & \cdots & t_{N,K} \end{pmatrix}$$

(4) In Step 3, the max is over the populations that have scores Z for the SNP.

(5) Convert each element of Q into a p-value. Let the corresponding matrix as R.

More details on Step (5)

$S^{(p)} = \frac{Z}{\sigma}$ as above.

Suppose $t = \text{Max}\{|S^{(p)}|, p=1,\cdots,L\}$, the maximum of L Z-scores. (L can be less than P as some population do not have that SNP.)

Since $Z^{(p)}$, $p=1,\cdots,L$ are independent standard normal, let $X_1, X_2, \cdots X_L$ be i.i.d. normal random variable.

$$P(\max\{|X_p|, p=1,\cdots,L\} < t)$$

$$= \prod_{p=1}^{L} P(|X_p| < t) = (\Phi(t))^L$$

wrong! should be $(2\Phi(t)-1)^L$

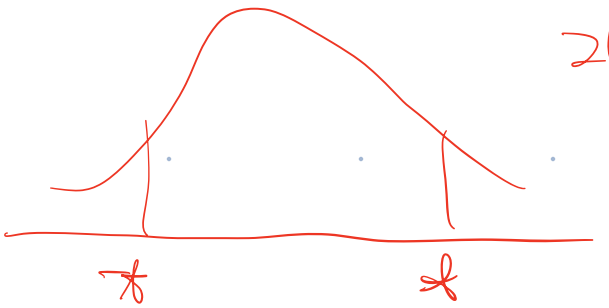
so the p-value for t can be written as

$1-(\Phi(t))^L$ → should be $1-(2\Phi(t)-1)^L$

⑥ Treat R as the matrix for SNP-level p value for ARTP for one population, and proceed as one population pathway analysis.

Note: The algorithm should be applied to a block at a time in each population. So, we should use a common block for all populations.

$t \geq c$



$2(1-\Phi(t))$

$1-2+2\Phi(t)$

$=2\Phi(t)-1$