SNP based pathway algorithm (Allowing for different sets in different populations)

(1) Summary data $\{\beta_i^{(m)}, \sigma_i^{(m)}\}$, i is snp id, $m=1,\dots,M$ population id

(2) Do snp filtering separately for each population.

Let $\Lambda^{(m)}$ be the set of SNPs left in population m for a given block. (I assume the block/group is defined the same for each population)

(3) Then for each block,

find the set of SNPs that are union of the sets of SNPs from Step (2). Let this super set as $\Omega = \bigcup_{m=1}^{M} \Lambda^{(m)}$. adjusted by N.

(4). For each population m, find the covariance matrix $\Sigma^{(m)}$ based on the reference data for SNPs in Ω. (for some SNP, $\Sigma_{ii}^{(m)}=0$).

may be adjusted later

(5) For SNP i in Ω, do the following.

(a) If $i \notin \Lambda^{(m)}$, let $\cancel{u_i^{(m)}=0}$ $w_i^{(m)}=0$

If $i \in \Lambda^{(m)}$, let $\cancel{u_i^{(m)} = \frac{1}{\sigma_i^{(m)}}}$ $w_i^{(m)}=1$

(b) Then, ~~define~~ $\cancel{t=\sqrt{\sum_{m=1}^{M}(u_i^{(m)})^2}}$

$\cancel{w_i^{(m)} = \frac{u_i^{(m)}}{t}, \quad m=1,\dots,M}$

(c) Let $z_i = \sum_{m=1}^{M} w_i^{(m)} z_i^{(m)}$. ($z_i^{(m)}$ is score() output from each population so adjusted by $\sqrt{N}$.)

(Note, $z_i^{(m)}$ might not be defined if $w_i^{(m)}=0$. maybe you can assign $z_i^{(m)}=0$ whenever it is not in $\Lambda^{(m)}$)

(6) Based on $w_i^{(m)}$ for $i\in\Omega$, obtain the covariance matrix for z_i, $i\in\Omega$ as $\Sigma = (\Sigma_{ij})_{i\in\Omega, j\in\Omega}$

$$\Sigma_{ij} = \sum_{m=1} w_i^{(m)} w_j^{(m)} \Sigma_{ij}^{(m)}.$$

or just $\Sigma_{ij} = \sum_{m=1}^{M} \Sigma_{ij}^{(m)}$, if you have change $\Sigma^{(m)}$ to the same dimension by adding 0.

⑦ Give Σ and Z, we want to rescale it.

let $N = \max\{N^{(m)}, m=1, \dots M\}$. $N^{(m)}$ is the total sample size (# cases + # controls) for the population m. Then let

$$Z = Z/\sqrt{N}$$

$$\Sigma = \Sigma/N.$$ (to make the variance smaller).