Multi-SKAT

Diptavo Dutta

Multi-SKAT R-package

This package produces kernel regression based rare-variant association tests for multiple phenotypes. The functions aggregate variant-phenotype score statistic in a particular region/gene and computes corresponding pvalues efficiently accounting for different models of association between the region and the battery of phenotypes.

Overview

In this vignette we display an elementary workflow to obtain the association test results corresponding to different `Multi-SKAT` tests (omnibus and with prespecified kernel)

Unrelated individuals

Multi-SKAT with pre-specified kernels

An example dataset `MultiSKAT.example.data` has a genotype matrix `Genotypes` of 5000 individuals and 56 SNPs, a phenotype matrix `Phenotypes` of 5 continuous phenotypes on those individuals, and a covariates vector `Cov` denoting intercept.

The first step is to create a null model using `MultiSKAT_NULL` function (for unrelated individuals) with the phenotype matrix and covariate matrix. Subsequently, `MultiSKAT` function is used with appropriate kernel to obtain the association p-value.

```
library(MultiSKAT)
```

```
## Loading required package: SKAT
```

```
## Loading required package: nlme
```

```
## Loading required package: copula
```

```
data(MultiSKAT.example.data)
attach(MultiSKAT.example.data)

obj.null <- MultiSKAT_NULL(Phenotypes,Cov)

### Phenotype-Kernel: PhC; Genotype-Kernel: SKAT
out1 <- MultiSKAT(obj.null,Genotypes,Sigma_p = cov(Phenotypes),verbose = FALSE)
out1$p.value
```

```
## [1] 1.359056e-31
```

```
### Phenotype-Kernel: Het; Genotype-Kernel: SKAT
out2 <- MultiSKAT(obj.null,Genotypes,Sigma_p = diag(5),verbose = FALSE)
out2$p.value
```

```
## [1] 6.326384e-25
```

```
### Phenotype-Kernel: Hom; Genotype-Kernel: SKAT
out3 <- MultiSKAT(obj.null,Genotypes,Sigma_p = matrix(1,ncol = 5,nrow = 5),verbose = FALSE)
out3$p.value
```

```
## [1] 9.227556e-06
```

```
### Phenotype-Kernel: PC-Sel; Genotype-Kernel: SKAT
### Select top 4 principal components

sel <- 4
L <- obj.null$L
V_y <- cov(Phenotypes)
V_p <- cov(Phenotypes%*%L)
L_sel <- cbind(L[,1:sel],0)
pc_sel_kernel <- V_y%*%L_sel%*%solve(V_p)%*%solve(V_p)%*%t(L_sel)%*%V_y
out4 <- MultiSKAT(obj.null,Genotypes,Sigma_p = pc_sel_kernel,verbose = FALSE)
out4$p.value
```

```
## [1] 1.674126e-25
```

```
### Phenotype-Kernel: PhC; Genotype-Kernel: Burden
out5 <- MultiSKAT(obj.null,Genotypes,Sigma_p = cov(Phenotypes),verbose = FALSE,r.corr = 1)
out5$p.value
```

```
## [1] 0.9504494
```

```
### Phenotype-Kernel: Het; Genotype-Kernel: Burden
out6 <- MultiSKAT(obj.null,Genotypes,Sigma_p = diag(5),verbose = FALSE,r.corr = 1)
out6$p.value
```

```
## [1] 0.9039228
```

Assign weights to variants

It is assumed that rarer variants are more likely to be causal variants with large effect sizes. The default version of `MultiSKAT` uses $w_i = Beta(MAF_i, 1, 25)$ as per Wu et al(2011)(via). Other weighting schemes can also be incorporated in the `MultiSKAT` function through the `weights` and `weights.beta` option.

```
### No weights
MultiSKAT(obj.null,Genotypes,Sigma_p = cov(Phenotypes),
          weights.beta = c(1,1),verbose = FALSE)$p.value
```

```
## [1] 3.361046e-69
```

Omnibus Tests:

minP

To combine MultiSKAT tests with pre-specified phenotype kernels `minP` function can be used given the genotype kernels remain the same.

```
### Combining PhC, Het and Hom with genotype kernel being SKAT
obj.list = list(out1,out2,out3)
obj.minP = minP(obj.null,obj.list,Genotypes)
```

```
## [1] "The region has 56 variants"
## [1] "The region has 46 rare variants"
```

```
obj.minP$p.value
```

```
## [1] 4.077168e-31
```

```
### Combining PhC and Het with genotype kernel being Burden
obj.list = list(out5,out6)
obj.minP2 = minP(obj.null,obj.list,Genotypes,r.corr = 1)
```

```
## [1] "The region has 56 variants"
## [1] "The region has 46 rare variants"
```

```
obj.minP2$p.value
```

```
## [1] 0.9463782
```

minPcom

To combine MultiSKAT tests with pre-specified phenotype kernels with simultaneously varying genotype kernels`minPcom` function can be used.

```
### Getting minPcom p-value combining PhC, Het and Hom kernels
Sigma_Ps = list(cov(Phenotypes),diag(5),matrix(1,ncol = 5,nrow = 5))
obj.com = minPcom(obj.null,Sigma_Ps,Genotypes,verbose = FALSE)
obj.com$p.value
```

```
## [1] 8.154335e-31
```

Related individuals

`Multi-SKAT` functions can analyse related individuals by incorporating kinship correction. An example dataset with related individuals `MultiSKAT.Kinship.example.data` includes a kinship matrix `Kinship` of 500 individuals and 20 SNPs, a co-heritability matrix `V_g`, residual covariance matrix `V_e` in addition to 5 phenotypes, genotype matrix and covariates. Additionally, if the kinship matrix can be written as

$$Kinship = UDU^T$$

, with D being a diagonal matrix of eigen values, the dataset contains `U` and `D`. The workflow for the related individuals remains the same. First the construction of the null model through `MultiSKAT_NULL_Kins` followed by obtaining the p-value through `MultiSKAT`

```
detach(MultiSKAT.example.data)
data(MultiSKAT.Kinship.example.data)
attach(MultiSKAT.Kinship.example.data)

Kinship[1:6,1:6]
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]  1.0  0.5  0.0  0.0  0.0  0.0
## [2,]  0.5  1.0  0.0  0.0  0.0  0.0
## [3,]  0.0  0.0  1.0  0.5  0.0  0.0
## [4,]  0.0  0.0  0.5  1.0  0.0  0.0
## [5,]  0.0  0.0  0.0  0.0  1.0  0.5
## [6,]  0.0  0.0  0.0  0.0  0.5  1.0
```

```
V_g
```

```
##             [,1]       [,2]      [,3]      [,4]      [,5]
## [1,] 0.14010271 0.01793845 0.1424442 0.1720590 0.2137881
## [2,] 0.01793845 0.33529520 0.1160337 0.2552043 0.1140791
## [3,] 0.14244417 0.11603368 0.2794992 0.3066876 0.1807799
## [4,] 0.17205899 0.25520432 0.3066876 0.4844014 0.2450426
## [5,] 0.21378813 0.11407914 0.1807799 0.2450426 0.4240207
```

```
V_e
```

```
##            [,1]      [,2]       [,3]       [,4]      [,5]
## [1,] 0.8580322 0.3504979 0.30109816 0.19109055 0.1422359
## [2,] 0.3504979 0.6627300 0.29984708 0.17956632 0.3050065
## [3,] 0.3010982 0.2998471 0.71851052 0.09732203 0.1880925
## [4,] 0.1910906 0.1795663 0.09732203 0.51359888 0.1675392
## [5,] 0.1422359 0.3050065 0.18809247 0.16753917 0.5739799
```

```
eig <- eigen(Kinship)
U <- eig$vectors; D <- diag(eig$values);

obj.null <- MultiSKAT_NULL_Kins(Phenotypes,Cov,U,D,V_g,V_e)

### Phenotype-Kernel: PhC; Genotype-Kernel: SKAT
out1 <- MultiSKAT(obj.null,Genotypes,Sigma_p = cov(Phenotypes),verbose = FALSE)
out1$p.value
```

```
## [1] 8.280796e-14
```