

Multiple Testing and Resampling Methods

Applied to GWAS data (one genotype) from lab

Data

```
genos = as.matrix(read.table("./genos.txt"))
phenos = as.matrix(read.table("./phenos.txt"))
#make a binary group label from genotype 1500
table(genos[,1500])
group = ifelse(genos[,1500]==0,0,1)
table(group)
```

Bioconductor multtest package

To implement bootstrap and permutation null distributions

```
#source("http://bioconductor.org/biocLite.R")
#biocLite("multtest")
library(multtest)
#start with just one test
M1=t(as.matrix(phenos))
bootmtp1=MTP(M1,Y=group,B=1000,null="boot.cs")
permmtp1=MTP(M1,Y=group,B=1000,null="perm")
bootmtp1@rawp
permmtp1@rawp
t.test(phenos~group)$p.value
```

Now, let's try multiple tests

```
# make an outcome matrix with 2 rows by adding N(0,1) random noise phenotype
phenos2 = rnorm(length(phenos))
p2 = c(t.test(phenos~group,var.equal=TRUE)$p.value,t.test(phenos2~group,var.equal=TRUE)$p.value)
p2
#marginal methods
methods=mt.rawp2adjp(p2,c("Bonferroni","Holm","BH","BY"))$adjp
methods
```

NOTE: these are sorted smallest to largest and not in the original order!

```
#To get them in the original order
methods2=mt.rawp2adjp(p2,c("Bonferroni","Holm","BH","BY"))
methods2$adjp[order(methods2$index),]
```

Bootstrap - joint null

```
M2=rbind(M1,phenos2)
bootmtp2=MTP(M2,Y=group,B=1000,null="boot.cs",method="sd.maxT")
bootmtp2@rawp
bootmtp1@rawp
bootmtp2@adjp
```

Permutations - joint null

```
permmtp2=MTP(M2,Y=group,B=1000,null="perm",method="sd.maxT")
permmtp2@rawp
permmtp1@rawp
permmtp2@adjp
```

From lecture slides

Simulate 2 groups of 10 individuals with 50 $N(0,1)$ variables each

```
data=matrix(rnorm(20*50),nr=50,nc=20)
label=c(rep(0,10),rep(1,10))
```

Two-sided parametric t-test with Student's T null distribution

```
ttestp=apply(data,1,function(x){t.test(x~label)$p.value})
hist(ttestp) #looks pretty uniform - good
summary(ttestp)
```

Permutation null distribution B=100 iterations

```
B=100
observed=apply(data,1,function(x){t.test(x~label)$statistic})
permt=matrix(0,nr=nrow(data),nc=B)
for(b in 1:B){
  labelb=sample(label)
  permt[,b]=apply(data,1,function(x){t.test(x~labelb)$statistic})
}
permp=NULL
```

```

for(i in 1:50){
  permp[i]=mean(abs(observed[i])<abs(permt[i,]))
}
plot(ttestp,permp) #correlated - good
summary(permp)

```

Bootstrap null distribution B=100 iterations

```

boott=matrix(0,nr=nrow(data),nc=B)
for(b in 1:B){
  labelb=sample(label,replace=TRUE)
  boott[,b]=apply(data,1,function(x){t.test(x~labelb,replace=TRUE)$statistic})
}
bootp=NULL
for(i in 1:50){
  bootp[i]=mean(abs(observed[i])<abs(boott[i,]-mean(boott[i,])))
  #subtracting the mean(boott[i,]) makes it a null distribution
}
plot(ttestp,bootp)
plot(permp,bootp)
summary(bootp)

```