

Accumulation tests: gene dosage data experiment

Ang Li and Rina Foygel Barber

May 27, 2015

Introduction

This demo reproduces the gene dosage data experiment in the paper:

- Ang Li and Rina Foygel Barber, “Accumulation tests for FDR control in ordered hypothesis testing” (2015). Available from <http://arxiv.org/abs/1505.07352> (<http://arxiv.org/abs/1505.07352>)

The demo was run using R version 3.2.0, using the GEOquery package version 2.34.0 to obtain the data.

Setup

In this section we define and describe some functions that will be used in this demo.

Two-sample t-tests on a data matrix

This function inputs a data matrix X with n columns. The columns indexed by g_1 and g_2 belong to sample 1 and sample 2, respectively (e.g. cases and controls). For each row of X , the function runs a two-sample t-test for that row.

```
ttest_mat=function(X,g1,g2){
  # g1 & g2 give columns for groups 1 & 2
  n1=length(g1);n2=length(g2)
  means1=rowMeans(X[,g1]);means2=rowMeans(X[,g2])
  vars1=rowSums((X[,g1]-means1)%*%t(rep(1,n1)))^2)/(n1-1);vars2=rowSums((X[,g2]-means2)%*%t(rep(1,n2)))^2)/(n2-1)
  sds_diff = sqrt(vars1/n1 + vars2/n2)
  tstats = (means1-means2)/sds_diff
  dfs = (vars1/n1+vars2/n2)^2 / ((vars1/n1)^2/(n1-1) + (vars2/n2)^2/(n2-1))
  pvals=2*(1-pt(abs(tstats),dfs))
  output=list()
  output$tstats=tstats
  output$dfs= dfs
  output$pvals= pvals
  output
}
```

The next function is the same, except that instead of performing two-sided t-tests, we perform a one-sided t-test for each row of X . The signs s_i specify, for the i th row of X , whether we are testing for a positive effect or a negative effect.

```
signed_ttest_mat=function(X,g1,g2,s){
  n1=length(g1);n2=length(g2)
  means1=rowMeans(X[,g1]);means2=rowMeans(X[,g2])
  vars1=rowSums((X[,g1]-means1)%*%t(rep(1,n1)))^2)/(n1-1);vars2=rowSums((X[,g2]-means2)%*%t(rep(1,n2)))^2)/(n2-1)
  sds_diff = sqrt(vars1/n1 + vars2/n2)
  tstats = s*(means1 - means2)/sds_diff
  dfs = (vars1/n1+vars2/n2)^2 / ((vars1/n1)^2/(n1-1) + (vars2/n2)^2/(n2-1))
  pvals=(1-pt(tstats,dfs))
  output=list()
  output$tstats=tstats
  output$dfs= dfs
  output$pvals= pvals
  output
}
```

Multiple comparisons methods

The functions below implement the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) and Storey's modification of the BH procedure (Storey 2002), both of which are methods for false discovery rate control in a multiple comparisons problem. The main inputs are an *unordered* sequence of p-values p_1, \dots, p_n and a desired FDR level α .

```
BH_method = function(pvals,alpha){
  khat=max(c(0,which(sort(pvals)<=alpha*(1:length(pvals))/length(pvals))))
  which(pvals<=alpha*khat/length(pvals))
}

Storey_method = function(pvals,alpha,thr){
  est_num_nulls=sum(pvals>thr)/(1-thr)
  khat=max(c(0,which(sort(pvals)<=alpha*(1:length(pvals))/est_num_nulls)))
  which(pvals<=alpha*khat/length(pvals))
}
```

Accumulation tests for ordered hypothesis testing

The functions in `accumulation_test_functions` run different accumulation tests on an *ordered* sequence of p-values, p_1, \dots, p_n , given a desired FDR level α . The accumulation tests are the SeqStep and SeqStep+ methods (Barber and Candès 2014), the ForwardStop method (G'Sell et al. 2013), and the HingeExp method (see main paper).

```
source('accumulation_test_functions.R')
```

Download gene expression vs dosage data

Data obtained from the GEOquery package (Davis and Meltzer 2007) (version 2.34.0).

```

source("http://bioconductor.org/biocLite.R")
biocLite("GEOquery")
library(GEOquery)
gds2324 = getGEO('GDS2324')
eset = GDS2eSet(gds2324, do.log2=TRUE)
Data = exprs(eset)

```

Each row of `Data` is a gene (total: $n = 22283$ genes). The 25 columns of `Data` correspond to 5 trials each at 5 different dosages: columns 1-5 are zero dose (control group), followed by columns 6-10 at the lowest dose, etc. The entries of `Data` are log-transformed gene expression levels.

```

n=dim(Data)[1]
highdose=21:25;lowdose=6:10;control=1:5

```

Computing p-values

First, compute p-values for the (unordered) multiple comparisons methods, which search for significant differences between low-dose gene expression level and control-group gene expression level. We will try two-sample t-test p-values, and permutation-test p-values.

```

pvals_lowdose_ttest=ttest_mat(Data,lowdose,control)$pvals

pvals_lowdose_ttest_permuted=matrix(0,choose(10,5),n)
nchoosek=combn(1:10,5)
for(i in 1:choose(10,5)){
  permord=c(nchoosek[,i],setdiff(1:10,nchoosek[,i]))
  pvals_lowdose_ttest_permuted[i,]=ttest_mat(Data[,permord],lowdose,control)$pvals
}
pvals_lowdose_permutation_test=colSums(abs(pvals_lowdose_ttest_permuted)-rep(1,choose(10,5))%*%t(abs(pvals_lowdose_ttest))<=0)/choose(10,5)

```

Next, we will use the highest-dosage data to produce an ordering on the low-dosage vs control-group p-values, resulting in an ordered sequence of p-values which we will use for the accumulation tests. First, for each gene, produce a pval for highest dose compared to mean of lowest dose + control. Record also the sign of the effect (increased or reduced gene expression at the highest dose, compared to pooled low-dose and control-group data).

```

ttest_highdose=ttest_mat(Data,highdose,c(lowdose, control))
pvals_highdose=ttest_highdose$pvals
test_signs=sign(ttest_highdose$tstats)

```

Next, for each gene we will perform a one-sided t-test (using the sign of the estimated high-dose effect), and then use a permutation test to get a final p-value for this gene. These p-values will then be reordered according to the high-dose results above.

```

signed_pvals_lowdose_ttest=signed_ttest_mat(Data,lowdose,control,test_signs)$pvals

signed_pvals_lowdose_ttest_permuted=matrix(0,choose(10,5),n)
nchoosek=combn(1:10,5)
for(i in 1:choose(10,5)){
  permord=c(nchoosek[,i],setdiff(1:10,nchoosek[,i]))
  signed_pvals_lowdose_ttest_permuted[i,]=signed_ttest_mat(Data[,permord],lowdose,control,test_signs)$pvals
}
signed_pvals_permutation_test=colSums(abs(signed_pvals_lowdose_ttest_permuted)-rep(1,choose(10,5))%*%t(abs(signed_pvals_lowdose_ttest))<=0)/choose(10,5)

signed_pvals_reordered=signed_pvals_permutation_test[order(abs(pvals_highdose))]

```

Run all methods

Here we apply multiple comparison methods to one of the unordered sequences of p-values, `pvals_lowdose_ttest` or `pvals_lowdose_permutation_test`. We also apply accumulation test methods to the ordered sequence of p-values, `signed_pvals_reordered`.

```

methods=c(
  'SeqStep (C=2)', 'SeqStep+ (C=2)',
  'HingeExp (C=2)', 'ForwardStop',
  'BH (p-values from t-tests)', 'BH (p-values from permutation tests)',
  'Storey (p-values from t-tests)', 'Storey (p-values from permutation tests)'
)

```

We run methods at a range of desired FDR levels $\alpha = 0.01, 0.02, \dots, 0.90$. For each method, we evaluate its performance by counting the number of discoveries (i.e. number of rejected hypotheses) that it is able to make, at each α .

```

max_alpha=.9; alphalist = seq(0.01,max_alpha,by=0.01); num_alpha=length(alphalist);

NumRej = matrix(0,length(methods),num_alpha)
for(i in 1:num_alpha){
  alpha=alphalist[i]
  NumRej[which(methods=='SeqStep (C=2)'),i]=SeqStep(signed_pvals_reordered,alpha=alpha,C=2)
  NumRej[which(methods=='SeqStep+ (C=2)'),i]=SeqStepPlus(signed_pvals_reordered,alpha=alpha,C=2)
  NumRej[which(methods=='HingeExp (C=2)'),i]=HingeExp(signed_pvals_reordered*(1-1/(1+choose(10,5))),alpha=alpha,C=2)
  NumRej[which(methods=='ForwardStop'),i]=ForwardStop(signed_pvals_reordered*(1-1/(1+choose(10,5))),alpha=alpha)
  NumRej[which(methods=='BH (p-values from t-tests)'),i]=length(BH_method(pvals_lowdose_ttest,alpha))
  NumRej[which(methods=='BH (p-values from permutation tests)'),i]=length(BH_method(pvals_lowdose_permutation_test,alpha))
  NumRej[which(methods=='Storey (p-values from t-tests)'),i]=length(Storey_method(pvals_lowdose_ttest,alpha,max_alpha))
  NumRej[which(methods=='Storey (p-values from permutation tests)'),i]=length(Storey_method(pvals_lowdose_permutation_test,alpha,max_alpha))
}

```

Note that for ForwardStop and HingeExp, we must shift the p-values slightly away from 1 to ensure that we do not get values of ∞ .

Plot results

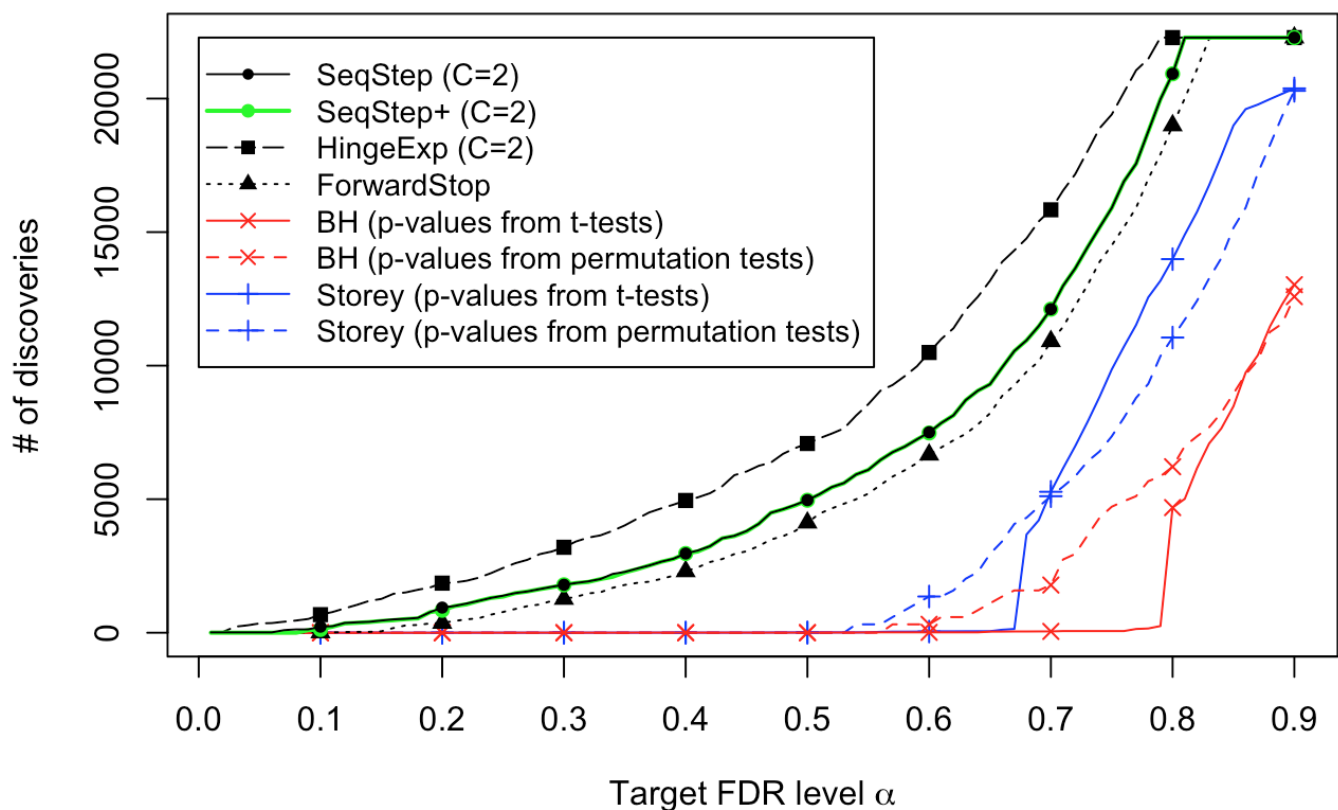
Here we plot the results for each method.

```

cols=c('black','green',rep('black',2),rep('red',2),rep('blue',2))
ltys=c(1,1,5,3,rep(1:2,2))
lwds=c(1,2,rep(1,7))
pch=c(20,16,15,17,4,4,3,3)

# pdf('gene_dosage_results.pdf',width=9,height=5.4)
plot(0:1,0:1,type='n',xlim=range(alphalist),ylim=c(0,n),xlab=expression(paste('Target FDR level ',alpha)),ylab='# of discoveries',axes=FALSE)
axis(side=1,at=0:10/10)
axis(side=2)
alpha_pt=(1:(10*max_alpha))*10
for(i in length(methods):1){
  points(alphalist,NumRej[i,],type='l',col=cols[i],lty=ltys[i],lwd=lwds[i])
  points(alphalist[alpha_pt],NumRej[i,alpha_pt],col=cols[i],pch=pchs[i],lwd=lwds[i])
}
legend(0,n,methods,col=cols,lty=ltys,lwd=lwds,pch=pchs,seg.len=3,cex=0.9)
box(); # dev.off()

```



We see that the accumulation tests have far higher power (greater number of discoveries) than the unordered multiple comparison methods. Among the accumulation tests, HingeExp is most powerful, followed by SeqStep and SeqStep+, followed by ForwardStop. Note that SeqStep and SeqStep+ are nearly indistinguishable in the plot, as their results are nearly identical.

Other methods

While the Benjamini-Hochberg and Storey methods are not guaranteed to control FDR if the (unordered) p-values are not independent, there are other more conservative methods that do give this guarantee. In this setting, however, such methods have extremely low power. We test three such methods here and find that all of them yield 2 or fewer discoveries at the very liberal FDR level $\alpha = 0.9$ (compare to the accumulation test methods, which yield thousands of discoveries). The methods we test are Benjamini & Yekutieli's log-factor correction for the BH procedure (Benjamini and Yekutieli 2001), the Holm-Bonferroni method (Holm 1979), and the Bonferroni correction.

```

BY_method = function(pvals,alpha){
  khat=max(c(0,which(sort(pvals)<=alpha/sum(1/(1:length(pvals)))*(1:length(pvals))/length(pvals))))
  which(pvals<=alpha*khat/length(pvals))
}

HolmBonferroni_method = function(pvals,alpha){
  khat=min(which(sort(pvals)>alpha/(length(pvals)+1-(1:length(pvals)))-1
  which(pvals<=sort(pvals)[khat]))
}

Bonferroni_method = function(pvals,alpha){
  which(pvals<=alpha/length(pvals))
}

alpha=0.9
NumRej_conservative_methods = matrix(
c(length(BY_method(pvals_lowdose_ttest,alpha)),length(BY_method(pvals_lowdose_permutation_test,alpha)),
length(HolmBonferroni_method(pvals_lowdose_ttest,alpha)),length(HolmBonferroni_method(pvals_lowdose_permutation_test,alpha)),
length(Bonferroni_method(pvals_lowdose_ttest,alpha)),length(Bonferroni_method(pvals_lowdose_permutation_test,alpha))),2,3)
rownames(NumRej_conservative_methods)=c('t-test', 'permutation test')
colnames(NumRej_conservative_methods)=c('Benjamini-Yekutieli','Holm-Bonferroni','Bonferroni correction')
NumRej_conservative_methods

```

##	Benjamini-Yekutieli	Holm-Bonferroni	Bonferroni correction
## t-test	0	2	2
## permutation test	0	0	0

References

- Barber, Rina Foygel, and Emmanuel Candès. 2014. “Controlling the False Discovery Rate via Knockoffs.” ArXiv Preprint ArXiv:1404.5609.
- Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B (Methodological)*. JSTOR, 289–300.
- Benjamini, Yoav, and Daniel Yekutieli. 2001. “The Control of the False Discovery Rate in Multiple Testing Under Dependency.” *Annals of Statistics*. JSTOR, 1165–88.
- Davis, Sean, and Paul Meltzer. 2007. “GEOquery: A Bridge Between the Gene Expression Omnibus (GEO) and BioConductor.” *Bioinformatics* 14: 1846–47.
- G’Sell, Max Grazier, Stefan Wager, Alexandra Chouldechova, and Robert Tibshirani. 2013. “False Discovery Rate Control for Sequential Selection Procedures, with Application to the Lasso.” ArXiv Preprint ArXiv:1309.5352.

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics*. JSTOR, 65–70.

Storey, John D. 2002. "A Direct Approach to False Discovery Rates." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (3). Wiley Online Library: 479–98.