Assignment 8 - False Discovery Rate

Naveen Narayanan Meyyappan

22/11/2019

## False Discovery Rate

The false discovery rate (FDR) is a method of conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons. FDR-controlling procedures are designed to control the expected proportion of “discoveries” (rejected null hypotheses) that are false (incorrect rejections).The FDR value of 0.05 means that 5% of declared positive results are truly negative. If many p values fall into the range where the null hypothesis of no association should be rejected, the FDR is much less conservative than Bonferroni false positive rate.

## Algorithm/Pseudo Code

1.Sort P values

2.Count tests (m)

3.Set Q

4.Plot sorted Pvalues(smallest to largest) vs line Q\*c(1:m)/m

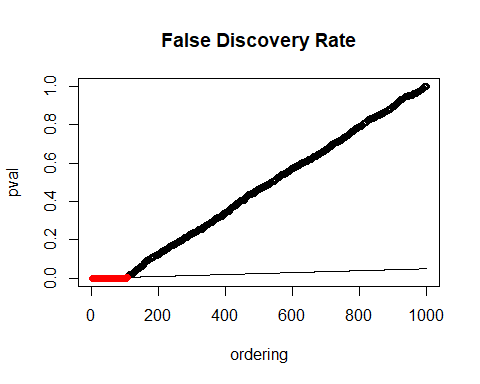
5.(If not independent, Q*c(1:m)/(m*(sum(1/i)i=1,…m)))

6.Find P\* =largest P value <line

7.Every P<=P\* is “interesting”

Let us try to perform the false discovery rate on an input test vector v1<-c(1e-5\*runif(100),runif(900)) and with Q=0.05. We need the list of hypothesis which are interesting by number in the original unsorted list of p values.

v1=c(1e-5\*runif(100),runif(900)) # Using the input vector  
Q=0.05 # Setting the Q value  
o1<-order(v1) # Generating the order for the input test vector  
pvec<-v1[o1] # Sorting the vector and storing the sorted array in pvec  
m<-length(v1) # Obtaining the number of hypothesis by calculating the length of test vector  
qline<-Q\*c(1:m)/m # Generating Q line for given length  
plot(c(c(1:m),c(1:m)),c(qline,pvec),type ="n",xlab ="ordering",ylab="pval",main = "False Discovery Rate") # Plotting the sorted p values with Q Line. Here we are setting up the axes  
lines(c(1:m),qline) # Adding Q line  
points(c(1:m),pvec) # Adding the sorted test vector points  
dv<-pvec-qline   
I1<-(dv<0)  
pstar<-max(pvec[I1]) # Calculating p\*  
I2<-pvec<=pstar # Every p that is less than p\* is interesting  
points(c(1:m)[I2],pvec[I2],col="red") # Marking those points with red



o1[I2] # This gives the pvalues that are intersting fro the ordered list. In other words it gives the index of the v1 vector for which the p values are interesting

## [1] 49 84 85 18 92 78 11 3 70 94 100 16 15 44 40 1 74  
## [18] 54 58 91 12 67 83 30 45 28 35 95 99 4 33 82 97 53  
## [35] 17 62 14 72 20 48 64 80 66 79 8 68 76 69 2 21 41  
## [52] 5 13 29 43 51 10 98 96 25 73 39 47 71 37 93 77 27  
## [69] 60 42 75 61 34 81 9 90 32 22 6 55 19 57 26 65 23  
## [86] 52 36 31 24 87 56 88 50 59 89 7 46 63 86 38 966 446  
## [103] 658 799 578 565

v1[o1[I2]]# These are the p values that are interesting from the original unsorted test vector

## [1] 2.027265e-07 3.030792e-07 5.978187e-07 6.203567e-07 7.140948e-07  
## [6] 7.686845e-07 8.723287e-07 1.011378e-06 1.014613e-06 1.445569e-06  
## [11] 1.602813e-06 1.636923e-06 1.664457e-06 1.869890e-06 1.928601e-06  
## [16] 1.943210e-06 2.176679e-06 2.197262e-06 2.202909e-06 2.358342e-06  
## [21] 2.494675e-06 2.559656e-06 2.652320e-06 2.733864e-06 3.024490e-06  
## [26] 3.144722e-06 3.185512e-06 3.217886e-06 3.277409e-06 3.330892e-06  
## [31] 3.435031e-06 3.590390e-06 3.710060e-06 3.767741e-06 3.784635e-06  
## [36] 3.859025e-06 3.886279e-06 3.998962e-06 4.008061e-06 4.018624e-06  
## [41] 4.041745e-06 4.192411e-06 4.334787e-06 4.505489e-06 4.505607e-06  
## [46] 4.506796e-06 4.549493e-06 4.595868e-06 4.718650e-06 4.775095e-06  
## [51] 4.842561e-06 4.869688e-06 4.889943e-06 4.896661e-06 4.963882e-06  
## [56] 5.023267e-06 5.136517e-06 5.414102e-06 5.476980e-06 5.561055e-06  
## [61] 5.930380e-06 6.066542e-06 6.137945e-06 6.311471e-06 6.507858e-06  
## [66] 6.632625e-06 6.660984e-06 6.664309e-06 6.726914e-06 6.938911e-06  
## [71] 7.137520e-06 7.423887e-06 7.658022e-06 8.155454e-06 8.288332e-06  
## [76] 8.314021e-06 8.314069e-06 8.444827e-06 8.561011e-06 8.620436e-06  
## [81] 8.642936e-06 8.651556e-06 8.688980e-06 8.760735e-06 8.793647e-06  
## [86] 8.817356e-06 8.901278e-06 8.908649e-06 8.993080e-06 9.003920e-06  
## [91] 9.092722e-06 9.115272e-06 9.194829e-06 9.389950e-06 9.464488e-06  
## [96] 9.574775e-06 9.583298e-06 9.619617e-06 9.696563e-06 9.706568e-06  
## [101] 1.851274e-04 4.331155e-04 1.227707e-03 1.871433e-03 2.870093e-03  
## [106] 4.957519e-03

## References

1. <https://en.wikipedia.org/wiki/False_discovery_rate>
2. <https://www.sciencedirect.com/topics/neuroscience/false-discovery-rate>