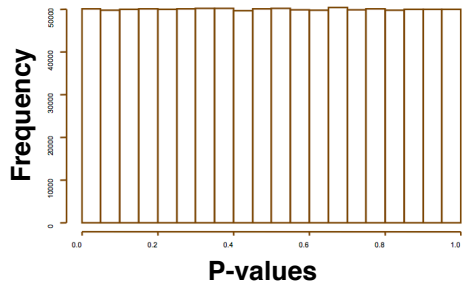


Multiple Testing

Why Multiple Testing Matters ?



P-values are uniformly distributed under the null
If we try 1 Million times a test and use 5% significance,
there will be roughly 50,000 SNPs that will be significant at 5% level

Based on <http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf>

2

Most Common Correction: Bonferroni

Single Step Approach: Bonferroni

- Very simple method for ensuring that the overall Type I error rate of α is maintained when performing m independent hypothesis tests

- Rejects any hypothesis with p-value $\leq \alpha/m$:

$$\tilde{p}_j = \min[mp_j, 1]$$

- For example, if we want to have an experiment wide Type I error rate of 0.05 when we perform 10,000 hypothesis tests, we'd need a p-value of $0.05/10000 = 5 \times 10^{-6}$ to declare significance

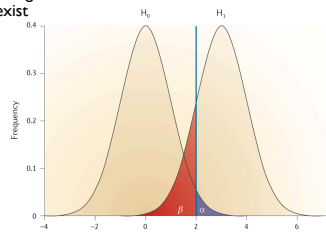
<http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf>

3

Philosophical Objections to Bonferroni Corrections

“Bonferroni adjustments are, at best, unnecessary and, at worst, deleterious to sound statistical inference” Perneger (1998)

- Counter-intuitive: interpretation of finding depends on the number of other tests performed
- The general null hypothesis (that all the null hypotheses are true) is rarely of interest
- High probability of type 2 errors, i.e. of not rejecting the general null hypothesis when important effects exist



<http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf>

False Discovery Rate vs. False Positive Rate

Table 1. Possible outcomes from thresholding m features for significance

	Called significant	Called not significant	Total
Null true	F	$m_0 - F$	m_0
Alternative true	T	$m_1 - T$	m_1
Total	S	$m - S$	m

$$\widehat{\text{FDR}} = \frac{F}{S} \quad \text{vs.} \quad \widehat{\text{FPR}} = \frac{F}{m_0}$$

Storey & Tibshirani, "Statistical significance for genomewide studies," Proceedings of the National Academy of Sciences, Jul. 2003.

5

Benjamini and Hochberg FDR

- To control FDR at level δ :
 1. Order the unadjusted p-values: $p_1 \leq p_2 \leq \dots \leq p_m$
 2. Then find the test with the highest rank, j , for which the p value, p_j , is less than or equal to $(j/m) \times \delta$
 3. Declare the tests of rank 1, 2, ..., j as significant

$$p(j) \leq \delta \frac{j}{m}$$

<http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf>

B&H FDR Example

Controlling the FDR at $\delta = 0.05$

Rank (j)	P-value	(j/m) × δ	Reject H_0 ?
1	0.0008	0.005	1
2	0.009	0.010	1
3	0.165	0.015	0
4	0.205	0.020	0
5	0.396	0.025	0
6	0.450	0.030	0
7	0.641	0.035	0
8	0.781	0.040	0
9	0.900	0.045	0
10	0.993	0.050	0

<http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf>

False Discovery Rate vs. False Positive Rate

Table 1. Possible outcomes from thresholding m features for significance

	Called significant	Called not significant	Total
Null true	F	$m_0 - F$	m_0
Alternative true	T	$m_1 - T$	m_1
Total	S	$m - S$	m

$$\widehat{\text{FDR}} = \frac{F}{S} \quad \text{vs.} \quad \widehat{\text{FPR}} = \frac{F}{m_0}$$

q-value : $P(\text{Null} | \text{significant})$

p-value : $P(\text{significant} | \text{Null})$

Storey & Tibshirani, "Statistical significance for genomewide studies," Proceedings of the National Academy of Sciences, Jul. 2003.

How to Calculate FDR

	Called significant	Called not significant	Total
Null true	F	$m_0 - F$	m_0
Alternative true	T	$m_1 - T$	m_1
Total	S	$m - S$	m

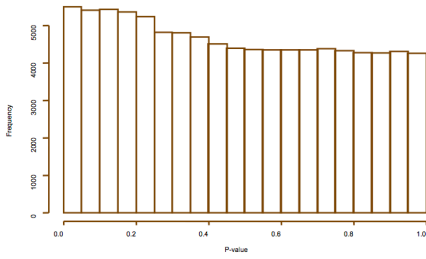
$$\widehat{\text{FDR}}(t) = \frac{F(t)}{S(t)} \quad S(t) = \#\{p_i \leq t; i = 1, \dots, m\}$$

$$F(t) = \#\{\text{null } p_i \leq t; i = 1, \dots, m\}$$

Storey & Tibshirani, "Statistical significance for genomewide studies," Proceedings of the National Academy of Sciences, Jul. 2003.

How to Calculate $S(t)$?

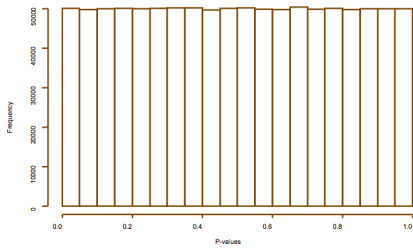
$$S(t) = \#\{p_i \leq t; i = 1, \dots, m\}$$



10

How to Calculate $F(t)$?

$$F(t) = \#\{\text{null } p_i \leq t; i = 1, \dots, m\}$$



Under the null, the probability $p_i \leq t$ is t

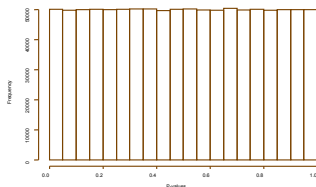
$F(t) = m_0 t$ where m_0 = total number of null tests

11

How Do We Estimate m_0 ?

Estimating The Proportion of Truly Null Tests

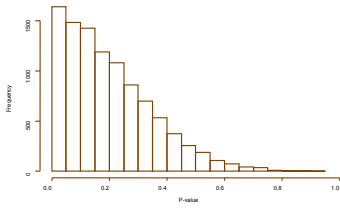
- Under the null hypothesis p-values are expected to be uniformly distributed between 0 and 1



12

Estimating The Proportion of Truly Null Tests

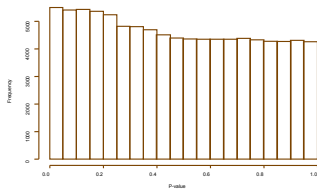
- Under the alternative hypothesis p-values are skewed towards 0



<http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf>

Estimating The Proportion of Truly Null Tests

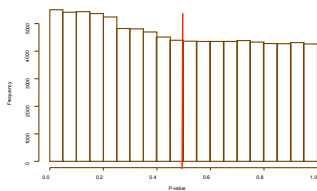
- Combined distribution is a mixture of p-values from the null and alternative hypotheses



<http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf>

Estimating The Proportion of Truly Null Tests

- For p-values greater than say 0.5, we can assume they mostly represent observations from the null hypothesis



<http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf>
