# Understanding the Benjamini-Hochberg method

Tan Siew Li, Linda

Department of Statistics and Applied Probability
National University of Singapore

# False discovery rate

- False discovery rate (FDR) is the expected proportion of tests which are incorrectly called significant out of all the tests which are called significant.
- The table below lists all possible outcomes from $M$ hypothesis tests. Thus FDR $= E(V/R)$.

| | Not called significant | Called Significant | Total |
|---|---|---|---|
| $H_0$ true | $U$ | $V$ | $M_0$ |
| $H_0$ false | $T$ | $S$ | $M_1$ |
| Total | $M - R$ | $R$ | $M$ |

- The Benjamini-Hochberg (BH) method is a procedure which controls the false discovery rate so that FDR $\leq \alpha$.

# Benjamini-Hochberg method

- Suppose we have computed the $p$-values for $M$ hypothesis tests:

$$H_{0j} \quad \text{vs} \quad H_{1j}, \quad j = 1, \ldots, M.$$

- The Benjamini-Hochberg method can be performed as follows.

**Benjamini-Hochberg Method**

To control FDR $\leq \alpha$:

1. Let $p_{(1)} \leq \cdots \leq p_{(M)}$ be ordered $p$-values.
2. Define $L = \max\left\{j : p_{(j)} < \alpha j/M\right\}$.
3. Reject all hypotheses $H_{0j}$ for which $p_j \leq p_{(L)}$.

- Why does the Benjamini-Hochberg method work?

# Understanding the Benjamini-Hochberg method

- Here we attempt to provide an intuitive understanding.
- Suppose we wish to perform $M$ hypothesis tests of the form:

  $$H_{0j} : \mu_{1j} = \mu_{2j} \quad \text{vs} \quad H_{1j} : \mu_{1j} \neq \mu_{2j}, \quad j = 1, \ldots, M.$$

  Here $\mu_{1j}$ and $\mu_{2j}$ denote the population means of the two groups in the $j$th test.
- For example, we may have measurements of gene expressions for two groups of patients for $M$ genes, and we want to identify the genes for which expressions of the two groups are different.
- Let $M = 10000$.
- Suppose out of these $M$ tests, there are actually $M_0 = 9000$ tests for which $H_0$ is true and $M_1 = 1000$ tests for which $H_0$ is false.

## Data simulation

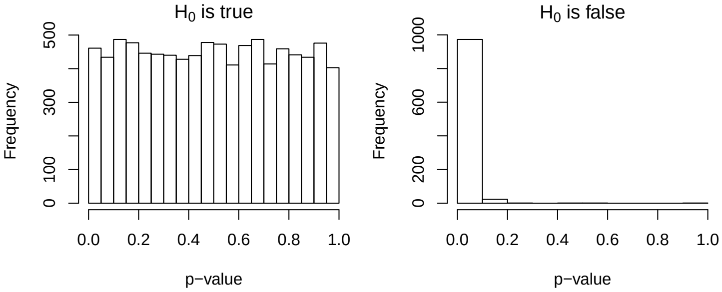
- Let us simulate some data and examine how the $p$-values are distributed.
- Suppose we have two groups, each of size $25$.
- For $j = 1, \ldots, 10000$, we generate a $\mu_{1j}$ from $N(0, 1)$.
  - For $j \leq 9000$, we set $\mu_{2j} = \mu_{1j}$.
  - For $j > 9000$, we set $\mu_{2j} = \mu_{1j} + 1$.
- Finally, we generate the observations such that,

$$x_{ij} \sim \begin{cases} N(\mu_{1j}, 1) & \text{if observation } i \text{ belongs to group 1,} \\ N(\mu_{2j}, 1) & \text{if observation } i \text{ belongs to group 2,} \end{cases}$$

for $j = 1, \ldots, M$ and $i = 1, \ldots, 50$.

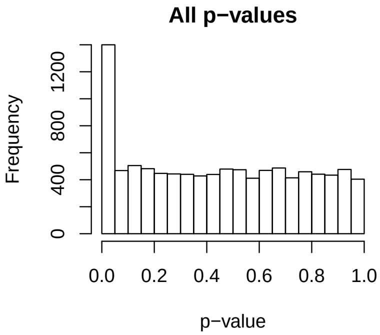
# Distribution of $p$-values

- The $p$-values are then computed using two-sample independent $t$-tests.




- The histogram on the left shows only the $p$-values for which $H_0$ is true. We see that the $p$-values are uniformly distributed between 0 and 1.
- The histogram on the right shows only the $p$-values for which $H_0$ is false. Most of the $p$-values are very small and lie between 0 and 0.1.
- What happens if we combine all the $p$-values?

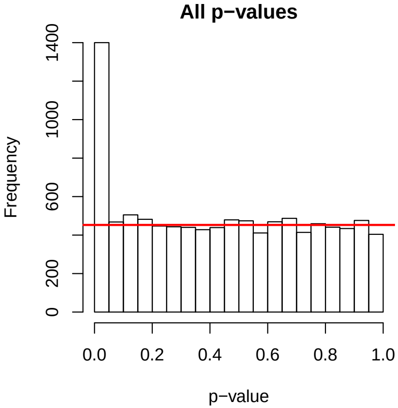
# Distribution of $p$-values




- The histogram shows that most of the $p$-values are uniformly distributed between 0 and 1 but there is a spike to the left close to zero. This spike is due to those $p$-values for which $H_0$ is false.
- Using this histogram, we can obtain an estimate of the number of hypotheses which are false and should be rejected.

# Distribution of $p$-values

- Height of the first block is 1400.
- Average height of the 2nd to 20th block is 453 and a red line is drawn at this height.
- Hence an estimate of the number of hypotheses which are false and should be rejected is $1400 - 453 = 947$.



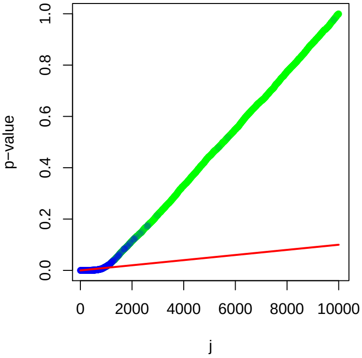

- This estimate is quite close to the true value of 1000.
- However the first block contains 1400 $p$-values. Which of these should we reject? A good rule of thumb is to reject the smallest 947 $p$-values.

# Applying Benjamini-Hochberg method

- If we sort the $p$-values in order from small to large and reject the smallest 947 values, 116 of these are actually $p$-values for which $H_0$ is true. Thus, there are 116 false discoveries and the proportion of false discoveries is $116/947 = 0.12$.

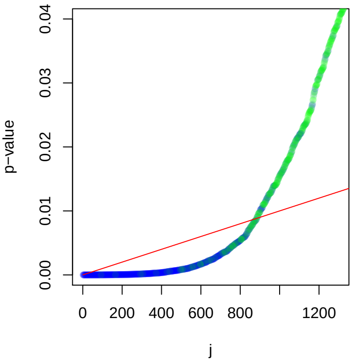
- To keep FDR $\leq \alpha = 0.10$. We can apply the Benjamini-Hochberg procedure.

- Graphically, we plot the sorted $p$-values, draw a line with gradient $\alpha/M$ (red), find the largest $p$-value, $p_{(L)}$ that falls below the line and reject all $p$-values less than or equal to $p_{(L)}$.



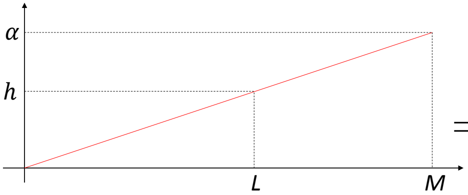
# Applying Benjamini-Hochberg method

- Here $L = 883$.
- The number of false discoveries among these 883 smallest $p$-values is 83.
- Hence the actual proportion is $83/883 = 0.094$, which is indeed smaller than 0.10.




- Why does the Benjamini-Hochberg method ensure FDR $\leq \alpha$?
- Recall that there are $M_0 = 9000$ $p$-values for which $H_0$ is true and they are uniformly distributed between 0 and 1. Thus the expected no. of $p$-values lying in any interval $[0, h]$ is $M_0 h$.

# Why Benjamini-Hochberg method works?




$$\frac{h}{L} = \frac{\alpha}{M}$$

$$\Rightarrow h = \frac{\alpha L}{M}$$

- Consider the red line with gradient $\alpha/M$. Let $L$ be rank of the largest $p$-value that falls below this line. So we reject the smallest $L$ $p$-values.
- The expected no. of $p$-values among these for which $H_0$ is true (false discoveries) is less than or equal to $M_0 h = M_0 \frac{\alpha L}{M}$. Thus

$$\text{FDR} \leq \frac{M_0 \frac{\alpha L}{M}}{L} = \alpha M_0/M \leq \alpha.$$