Multiple Hypothesis Testing

Tiandong Wang

Department of Statistics
Texas A&M University

- The **multiple hypothesis testing** problem is the situation when we wish to consider many hypotheses simultaneously.
- Example: Suppose we have n genes and data about expression levels for each gene among healthy individuals and those with prostate cancer.
  - k healthy patients: $X_{ij}^{(0)}$, $1 \leq j \leq k$, expression level of gene i.
  - l patients with prostate cancer: $X_{ij}^{(1)}$, $1 \leq j \leq l$, expression level of gene i.
  - The i-th null hypothesis, denoted $H_{0,i}$, would state that the mean expression level of the i-th gene is the same in both groups of patients:

$$H_{0,i} : \mathbb{E}X_i^{(0)} = \mathbb{E}X_i^{(1)}$$

  - Resolution: Global testing.

Global Testing

- Global null:

$$H_0 = \bigcap_{i=1}^{n} H_{0,i}.$$

- The global null states that all of the individual nulls are true.
- Suppose that for each hypothesis $H_{0,i}$, we already have a test statistic, and hence, a p-value p_i.
- For simplicity, we assume that $p_i \sim Unif[0, 1]$. We would like to combine $p_1, \ldots, p_n$ to test H_0.

Bonferroni's Method

Bonferroni's global test:
Given a desired level α, we can test the global null by simply testing each $H_{0,i}$ at level α/n and rejecting H_0 whenever any of the $H_{0,i}$ is rejected, i.e. reject H_0 whenever

$$\min_i p_i \leq \alpha/n.$$

- What is the size of this test?
- Most suited for situations where we expect at least one of the p-values to be very significant.
- In the biological example, we might apply this test if we expect one (or a few) of the genes to be very significantly linked to prostate cancer.

Simes Test

As before we start with n p-values and order them

$$p_{(1)} \leq \cdots \leq p_{(n)}.$$

The Simes statistic is

$$T_n = \min_i \left\{ p_{(i)} \frac{n}{i} \right\}.$$

Assume that p_i's are independent, then the Simes test reject H_0 if $T_n \leq \alpha$.

Note: The Simes procedure is strictly less conservative than Bonferroni.

Suppose in the prostate cancer example, we have $n = 6033$ tests of hypotheses.

- We collect p-values $p_1, \ldots, p_n$ for each test.
- We have four types of outcomes in multiple testing:

	accepted	rejected	total
true	U	V	n_0
false	T	S	$n - n_0$
total	$n - R$	R	n

- U, V, S, T are unobserved random variables. R is an observed random variable.
- The two quantities of primary interest to us are the size of V (the number of false discoveries) and the size of R (the total number of rejected hypotheses).

Familywise Error Rate (FWER)

Classical multiple comparison procedures (MCPs) aim to control

$$FWER = \mathbb{P}(V \geq 1),$$

in a strong sense; that is, under all configurations of true and false hypotheses.

- Bonferroni's method controls FWER at level α:

$$FWER \leq \frac{n_0 \alpha}{n} \leq \alpha$$

False Discovery Rate (FDR)

- The FWER makes sense when we are testing a small number of hypotheses.
- FWER is so stringent a method that individual departures from the null in this setting have little chance of being detected.
- We prefer to return some false positives along with many potentially interesting genes, because this enables scientists to follow these leads and to distinguish the important genes from the false discoveries.

False Discovery Rate (FDR)

- False discovery proportion (FDP) is defined as:

$$FDP = \frac{V}{\max\{R, 1\}}$$

- If we made no rejections, then our FDP is 0.
- Although we observe R, we do not observe V, and so FDP is an unobserved random variable.
- $FDR := \mathbb{E}(FDP)$.

Remarks:

- With FWER control, we have managed our false discoveries unless we are very unlucky.
- With FDR control, on average our test will control FDP, but this time we may not have done a very good job.
- Under the global null, the FDR is equivalent to the FWER.
- $FWER \geq FDR$.

FDR Controlling procedure

Benjamini & Hochberg (1995)
Fix a level $q \in [0, 1]$. Let i_0 be the largest i for which

$$p_{(i)} \leq \frac{i}{n} q$$

Reject all $H_{(i)}$ with $i \leq i_0$.

Theorem

*For **independent** test statistics [p-values], the Benjamini-Hochberg procedure stated above controls the FDR at level q. More precisely,*

$$FDR = \frac{n_0}{n} q \leq q.$$