Stats 300C: Theory of Statistics Spring 2016

Lecture 6 — April 8 2016

Prof. Emmanuel Candes

Scribe: E. Candes

1 Outline

Agenda: Mutiple testing/comparison problems

1. The problem of multiple testing
2. Procedure for controlling FWER
3. Weak/strong control
4. Holm's procedure

2 Mutiple Testing / Comparison Problems

Until now, we have been considering tests of the global null $H_0 = \bigcap_i H_{0,i}$. For some testing problems, however, our goal is to accept or reject each individual $H_{0,i}$.

Motivation: DNA microarrays measure expression levels of tens of thousands of genes. The data consist of levels of mRNA, which are thought to measure how much of a protein the gene produces. A larger number implies a more active gene.

$$X_{ij} = \text{expression level of gene } i \text{ in patient } j.$$

Efron's book considers an example with 6033 genes measured on 102 patients (50 controls and 52 with prostate cancer).

Here we have $n = 6033$ tests of hypotheses:

$$\begin{aligned} H_{0,i} &: \quad \text{gene } i \text{ is "null"} \\ H_{1,i} &: \quad \text{gene } i \text{ is "non-null"} \end{aligned}$$

and we are interested in calling *each* hypothesis i. In this case the global null could still be interesting if we were trying to discover whether prostate cancer has a genetic component.

We collect p-values $p_1, \ldots, p_n$ for each test. In our example,

$$T_i = \frac{\text{Ave[Cancer] - Ave[Control]}}{\text{Est. Std. Error}}.$$

Under $H_{0,i}$, $T_i \sim t_{100}$ and the corresponding p-values are $p_i = \mathbb{P}(|t_{100}| > |t_i|)$.

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 and the size of R.

Familywise Error Rate: Classical multiple comparison procedures (MCPs) aim to control

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

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

3 Procedures for Controlling FWER

Bonferroni's method: First, we can apply Bonferroni's method by rejecting all $H_{0,i}$ for which $p_i \leq \alpha/n$.

Theorem 1. *Bonferroni's method controls FWER at level α in the strong sense. In fact,*

$$\mathbb{E}V \leq \frac{n_0}{n}\alpha$$

Proof. We introduce the random variables

$$V_i = \begin{cases} 1 & \text{if } H_{0,i} \text{ is rejected} \\ 0 & \text{otherwise} \end{cases}.$$

Denote the set of nulls by $\mathcal{H}_0$. Then $V = \sum_{i \in \mathcal{H}_0} V_i$ and

$$\mathbb{E}V = \sum_{i \in \mathcal{H}_0} \mathbb{E}(V_i) = \sum_{i \in \mathcal{H}_0} \mathbb{P}(V_i = 1) = \sum_{i \in \mathcal{H}_0} \frac{\alpha}{n} = \frac{n_0}{n}\alpha.$$

Since

$$\mathbb{P}(V \geq 1) \leq \mathbb{P}(V \geq 1) + \mathbb{P}(V \geq 2) + \ldots + \mathbb{P}(V \geq n) = \mathbb{E}V,$$

this concludes the proof. □

Sidak's procedure: Under independence, test each hypothesis at level α_n

$$\mathbb{P}(V \geq 1) = 1 - \mathbb{P}(V = 0) = 1 - (1 - \alpha_n)^{n_0}$$

If all are null,

$$\mathbb{P}(V \geq 1) = 1 - (1 - \alpha_n)^n = 1 - 1 + n\alpha_n - \binom{n}{2}\alpha_n^2 + \cdots \leq \alpha.$$

We thus want

$$n\alpha_n \left[1 - \frac{n-1}{2}\alpha_n\right] \leq \alpha = 0.05$$

So for n large ($n \geq 10$, say), this is about the same as

$$n\alpha_n \left[1 - \frac{n\alpha_n}{2}\right] \leq 0.05$$

Rearranging, we obtain

$$\alpha_n \leq \frac{0.05}{n}\left[\frac{1}{1 - 0.025}\right] \approx \frac{0.05}{n}$$

Thus for even moderate values of n, using Sidak's procedure is effectively the same as using Bonferroni. It is not surprising that we cannot do better using Sidak's procedure for large n, since we proved in week 1 that Bonferroni is optimal in a minimax sense for large samples.

3.1 Weak Control

Two-step procedure [Fisher 1934]

1. Global test for $H_0 = \bigcap_{i=1}^n H_{0,i}$
2. Test each hypothesis at level α

For example, consider the testing procedure

1. Reject H_0 if $\min_i p_i \leq \alpha/n$
2. If H_0 rejected, reject $H_{0,i}$ if $p_i \leq \alpha$

This procedure controls the FWER only weakly

Definition 2. *A testing procedure controls the FWER weakly if it controls the FWER under the global null (i.e., when all hypotheses are null).*

!! This procedure DOES NOT control FWER in a strong sense.

Example:

- $X_i \sim N(\mu_i, 1)$ and independent
- $H_{0,i}: \ \mu_i = 0$

Imagine now that μ_1 is very large, so we reject H_0. We then apply α-level test to all the others. What happens?

Well, we make a lot of false rejections; we make approximately $\alpha \cdot n_0$ false rejections

4 Holm's procedure

This is a step-down procedure operating as follows: first, order the p-values

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n)}$$

and let $H_{(1)}, H_{(2)}, \ldots, H_{(n)}$ be the corresponding hypotheses. Then examine the p-values in order

> **Step 1:** If $p_{(1)} \leq \alpha/n$ reject $H_{(1)}$ and go to Step 2.
> Otherwise, accept $H_{(1)}, H_{(2)}, \ldots, H_{(n)}$ and stop.
>
> **Step i:** If $p_{(i)} \leq \alpha/(n-i+1)$ reject $H_{(i)}$ and go to step $i+1$
> Otherwise, accept $H_{(i)}, H_{(i+1)}, \ldots, H_{(n)}$ and stop.
>
> **Step n:** If $p_{(n)} \leq \alpha$, reject $H_{(n)}$.
> Otherwise, accept $H_{(n)}$.

Hence the procedure stops the first time $p_{(i)}$ exceeds the critical value $\alpha_i = \alpha/(n-i+1)$.

Holm's procedure is not as conservative as Bonferroni; we typically make more rejections (have more power). Also Holm's procedure can always be used instead of Bonferroni, due to the following result:

Theorem 3. *Holm's procedure controls the FWER strongly.*

Proof. Let i_0 be the rank of the smallest null p-value so that in Holm's procedure, the first true null is encountered at step i_0. Obviously,

$$i_0 \leq n - n_0 + 1$$

i.e., the first true null can be preceded by at most $n - n_0$ other hypotheses (all the false ones). Holm's commits a false rejection if and only if

$$p_{(1)} \leq \frac{\alpha}{n}, \, p_{(2)} \leq \frac{\alpha}{n-1}, \ldots, p_{(i_0)} \leq \frac{\alpha}{n-i_0+1},$$

which implies

$$p_{(i_0)} \leq \frac{\alpha}{n-i_0+1} \leq \frac{\alpha}{n_0}.$$

Therefore the probability of false rejection is bounded above by

$$\mathbb{P}\left(\min_{i \in \mathcal{H}_0} p_i \leq \frac{\alpha}{n_0}\right) \leq \sum_{i \in \mathcal{H}_0} \mathbb{P}(p_i \leq \alpha/n_0) = \alpha.$$

$\square$