Lecture 6 — April 13 2018

Prof. Emmanuel Candes

Scribe: S. Wager, E. Candes

1 Outline

Agenda: From global testing to multiple testing

1. Testing the global null vs. FWER
2. The closure principle: using global testing procedures for FWER control
   (a) Closing Bonferroni yields Holm's procedure
   (b) Closing Simes justifies Hochberg's procedure
3. Step-down vs. step-up procedures

2 Global Testing vs. Multiple Testing

In the first lectures, we focused on *global testing*: We have p-values $p_1, \ldots, p_n$ and a global null $H : p_1, \ldots, p_n \sim U([0, 1])$. Our goal is to get α-level control under this global null: $\mathbb{P}_H[\text{reject } H] \leq \alpha$. We saw several procedures that can be used to test H; as usual, we denote the ordered p-values as $0 \leq p_{(1)} \leq \ldots \leq p_{(n)} \leq 1$.

- **Bonferroni:** Reject H if $p_{(1)} \leq \alpha/n$.
- **Simes:** Reject H if $p_{(1)} \leq \alpha/n$ or $p_{(2)} \leq 2\alpha/n$ or ... or $p_{(n)} \leq \alpha$.
- **Fisher:** Reject H if $\sum_{i=1}^{n} 2 \log p_i^{-1} \geq \chi^2_{2n}(1 - \alpha)$.
- ...

Today, our focus is on *multiple testing*: We have the same p-values as before, except we now care about the individual hypotheses $H_i : p_i \sim U([0, 1])$, for $i = 1, 2, \ldots, n$ separately. If each p-value tests the association between a gene expression level and a disease, then global testing asks whether *any* gene is associated with the disease, whereas multiple testing asks whether *this* gene is associated with the disease, for each gene separately.

Our goal is to control the family-wise error rate (FWER). To do so, define $\mathcal{M}^0 = \{i : H_i \text{ is true}\}$, and $\mathcal{R} = \{i : H_i \text{ is rejected}\}$. The FWER is defined as

$$\text{FWER} = \mathbb{P}\left[\mathcal{M}^0 \cap \mathcal{R} \neq \emptyset\right].$$

We want procedures for which FWER $\leq \alpha$. Some first observations:

- Bonferroni, i.e., reject H_i if $p_i \leq \alpha/n$, controls FWER "out of the box" (last class).
- Simes, i.e., reject $H_{(i)}$ if $p_{(j)} \leq j\alpha/n$ for some $j \geq i$, does *not* control FWER (example).
- Fisher's two-step procedure, i.e., reject H_i if Fisher's test rejects the global null and $p_i \leq \alpha$, does *not* control FWER (last class).

Thus, FWER control does not follow trivially from global control.

3 Closure Principle

Consider a family of hypotheses $\{H_i\}_{i=1}^n$. For indices i and j, we can define the intersection null

$$H_{ij} = H_i \cap H_j = \{p_i, p_j \sim U([0, 1])\}.$$

More generally, define the **closure** of this family as (Marcus, Peritz, Gabriel; 1976)

$$H_I = \bigcap_{i \in I} H_i \quad \text{for all } I \subset \{1, 2, \ldots, , n\}$$

Example: Consider the case $n = 3$. The closure in this case is just

$$\begin{array}{ccc} & H_{123} & \\ H_{12} & H_{13} & H_{23} \\ H_1 & H_2 & H_3 \end{array}$$

For each I, take a α-level test φ_I for testing H_I (rejects if $\varphi_I = 1$)

$$\mathbb{P}(\varphi_I = 1 \mid H_I) \leq \alpha$$

Notice that H_I is the "global null' for the index-set I. Thus, the tests φ_I may be generated by any global testing procedure (e.g., Bonferroni, Simes, ...).

The Closure procedure: Reject H_I iff for all $J \supseteq I$, H_J is rejected at level α; that is, $T_I = \min_{J \supseteq I} \phi_J$

Example Consider the case of 4 hypotheses. Suppose the underlined hypotheses are rejected at the α level.

$$\begin{array}{cccccc} & & \underline{H_{1234}} & & & \\ & \underline{H_{123}} & \underline{H_{124}} & \underline{H_{134}} & H_{234} & \\ \underline{H_{12}} & \underline{H_{13}} & \underline{H_{14}} & H_{23} & H_{24} & H_{34} \\ & \underline{H_1} & \underline{H_2} & H_3 & H_4 & \end{array}$$

- In this example, only H_1 is rejected by the closed test procedure.

- H_2 is not rejected, although it is marginally significant (because not all intersections are found to be significant).
- **Question:** What would Fisher's two-step procedure do? What is the corresponding picture?

Theorem 1. *The closure principle controls the FWER strongly.*

Proof. Let $\mathcal{M}^0$ be the set of true nulls, and assume it's non-empty (otherwise no rejection can be false).

$$\begin{aligned} A &: \{\text{at least one false rejection}\} \\ B &: \{\text{reject } H_{\mathcal{M}^0}\} \end{aligned}$$

For the closure procedure, $A \cap B = A$. So

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B) \leq \alpha.$$

$\square$

Thus, the closure procedure provides us with a **generic recipe** for translating global testing procedures into valid FWER controlling procedures. The next question is whether the closure principle yields **powerful** multiple testing procedures. As it turns out, it often does (but not always).

4 Example 1: Closing Bonferroni

We use Bonferroni's procedure to generate the φ_I:

$$\varphi_I = 1 \quad \textbf{iff} \quad \inf\{p_i : i \in I\} \leq \frac{\alpha}{|I|}.$$

We know that applying the closure principle to these test statistics will yield a valid FWER-controlling procedure. However, evaluating the closure tests explicitly requires computing 2^n test statistics; ideally we would like to have an analytic short-cut for evaluating the closure. It turns out that the closure of the Bonferroni procedure has a simple closed-form solution.

To derive a concise description of the procedure, note the following:

1. We consider an index set I for which $p_{(j)} = \min\{p_i : i \in I\}$, where (j) denote the index of the j-th smallest p-value. Let $I_{(j)+}$ denote the index set corresponding to the $n - j + 1$ largest p-values $\{(j), (j+1), ..., (n)\}$. Then:

   $$\phi_{I^+_{(j)}} = 1 \quad \text{implies that} \quad \phi_I = 1.$$

   - *Proof:* Because $\phi_{I^+_{(j)}} = 1$, we know that $p_{(j)} \leq \alpha/(n - j + 1)$. Now, because $p_{(j)}$ is the smallest p-value with index in I, we know that $|I| \leq n - j + 1$. Thus, $p_{(j)} = \min\{p_i : i \in I\} \leq \alpha/|I|$.

2. Given the first observation, we notice that

$$\begin{aligned} H_{(j)} \text{ is rejected } &\iff \phi_{I^+_{(1)}} = 1 \quad \textbf{and} \quad \textbf{...} \quad \textbf{and} \quad \phi_{I^+_{(j)}} = 1 \\ &\iff p_{(1)} \leq \frac{\alpha}{n} \quad \textbf{and} \quad \textbf{...} \quad \textbf{and} \quad p_{(j)} \leq \frac{\alpha}{n-j+1}. \end{aligned}$$

Procedure *Closed Bonferroni*

```
j = 0
while p_(j+1) ≤ α/(n − j) do
    j = j + 1
end
Reject H_(1), ..., H_(j)
```

We can also describe the closed Bonferroni procedure procedurally, as above. Closing Bonferroni has thus led us back to a procedure we already knew: **Holm's procedure**. This is encouraging, in that the closure principle appears to yield reasonable procedures.

Note The closure principle does not always yield such nice procedures. For example, try closing Fisher's procedure. What does it take to reject $H_{(1)}$? When will $\phi_{(1)(n)} = 1$, $\phi_{(1)(n-1)(n)} = 1$, etc.?

5 Example 2: Closing Simes

Suppose we now apply the closure principle with Simes test statistics

$$\varphi_I = 1(\{p_{(1,\,I)} \leq \alpha/|I|\}) \ \ \textbf{or} \ \ 1(\{p_{(2,\,I)} \leq 2\alpha/|I|\}) \ \ \textbf{or} \ \ \textbf{...} \ \ \textbf{or} \ \ 1(\{p_{(|I|,\,I)} \leq \alpha\}),$$

where $p_{(j,\,I)}$ denotes the j-th p-value with index in I. Because Simes is strictly more powerful than Bonferroni, the closure of Simes' procedure will be strictly more powerful than Holm's procedure. However, just like Simes, the closure of Simes requires the p-values to be independent (or for a similar weaker condition to hold).

This closure can be evaluated, although it is a little complicated. Here, we will derive a simple procedure that is **strictly more conservative** than the closure of Simes. Thus, because the closure of Simes controls FWER under independence, our procedure will also control FWER under independence.

Lemma 2. *For a set I, suppose that* **(a)** *the index (j) corresponding to the j-th smallest p-value is in I, and* **(b)** *that there exists a $j' \geq j$ such that $p_{(j')} \leq \frac{\alpha}{n-j'+1}$. Then $\varphi_I = 1$ for the test statistic φ_I given by Simes' procedure.*

Proof. Let k be the index given by

$$k = \sup\{i : (i) \in I \text{ and } i \leq j'\};$$

by hypothesis, we know that the set over which we are maximizing is non-empty. Then

$$
\begin{aligned}
p_{(k)} &\le p_{(j')} && \text{by choice of } k \\
&\le \frac{\alpha}{n - j' + 1} && \text{by assumption } \textbf{(b)} \\
&\le \frac{\alpha}{1 + |\{(j'+1), ..., (n)\} \cap I|} && ... \\
&= \frac{\alpha}{|\{(k), ..., (n)\} \cap I|} && \text{by definition of } k \\
&\le \frac{|\{(1), ..., (k)\} \cap I|}{|I|}\,\alpha && \text{by algebraic manipulation}
\end{aligned}
$$

Thus, by definition of Simes' procedure, $\varphi_I = 1$. □

By closing Simes, we have verified that the following procedure controls FWER:

$$\text{Reject } H_{(j)} \text{ if there is an index } j' \ge j \text{ such that } p_{(j')} \le \frac{\alpha}{n - j' + 1}.$$

This procedure is known as **Hochberg's procedure**.

6 Step-Down vs. Step-Up Procedures

We can write Holm's and Hochberg's procedures side-by-side:

Procedure *Holm*

```
j = 0
while p_(j+1) ≤ α/(n − j) do
    j = j + 1
end
Reject H_(1), ..., H_(j)
```

Procedure *Hochberg*

```
j = n
while p_(j) > α/(n − j + 1) do
    j = j − 1
end
Reject H_(1), ..., H_(j)
```

The two procedures have the same thresholds, i.e., $p_{(j)}$ is compared to $\alpha/(n - j + 1)$. However,

- Holm's scans forward, and stops as soon as a p-value fails to clear its threshold. This pessimistic approach is called a **step-down** procedure (think stepping downwards on the χ^2-statistics).
- Hochberg's scans backwards, and stops as soon as a p-value succeeds in passing its threshold. This optimistic approach is called a **step-up** procedure (think stepping upwards on the χ^2-statistics).

In general, **step-up** procedures can be substantially more powerful than **step-down** procedures. In the extreme case where $p_1 = ... = p_n = \alpha$, Holm's rejects nothing whereas Hochberg's rejects everything. A perhaps less extreme example is shown in the following figure; here, by scanning forward, Holm's can only reject 3 hypotheses whereas Hochberg's can reject 8 hypotheses because $p_{(8)} \le \alpha/(n - 8 + 1)$.

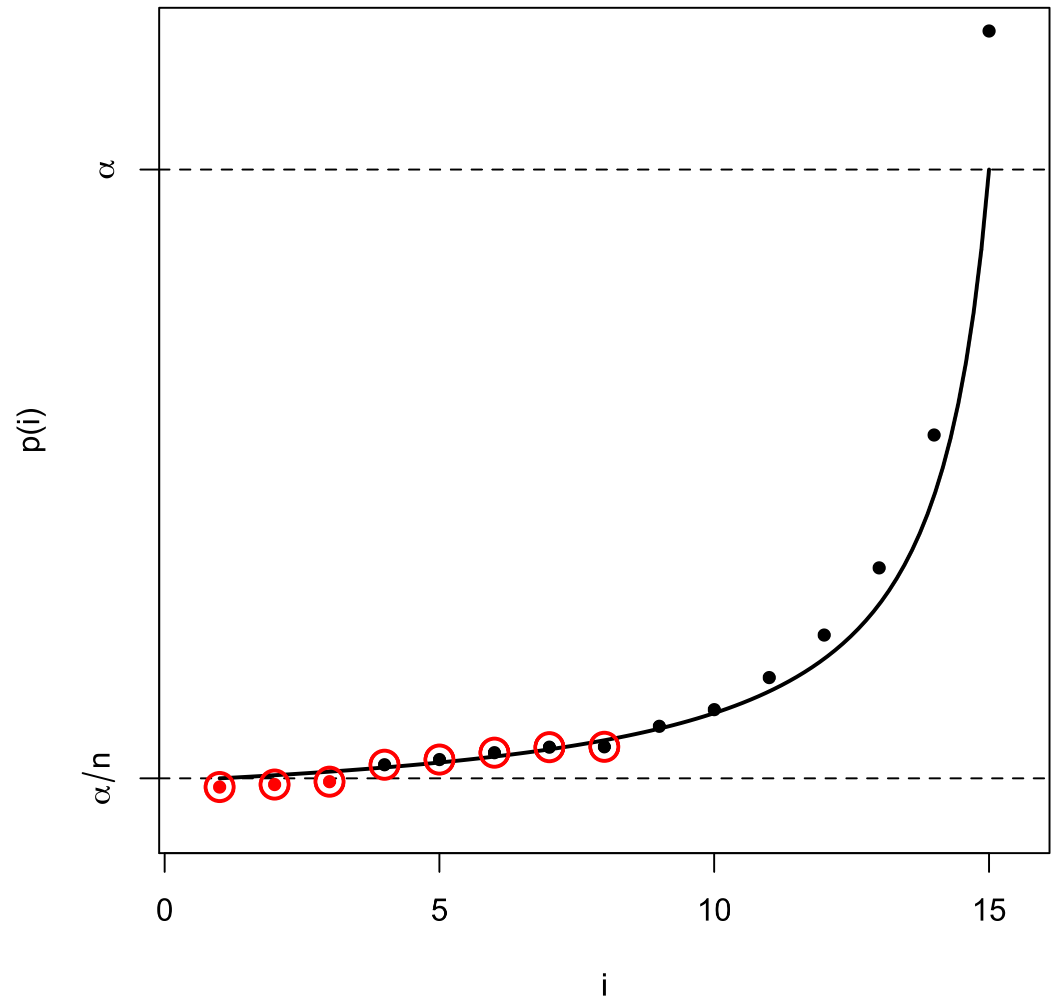

Figure 1: Comparison of Holm's Step-Down procedure and Hochberg's Step-Up procedure. Both control the FWER, but Hochberg's procedure is more powerful (makes more rejections). Solid red points are hypotheses rejected by Holm's procedure (3 rejections). Points circled in red are rejected by Hochberg's procedure (8 rejections).