

Lecture 7 — April 16, 2018

*Prof. Emmanuel Candes**Scribe: Feng Ruan; Edited by: Rina Friedberg, Junjie Zhu*

1 Outline

Agenda:

1. False Discovery Rate (FDR)
2. Properties of FDR
3. Procedures for Controlling FDR (BH(q))

2 False Discovery Rate

The False Discovery Rate (FDR) is an error control criterion developed in the 90's as an alternative to the FWER. The FWER makes sense when we are testing a small number of hypotheses. For example, in comparing six or ten different treatments, it is very reasonable to control the probability of returning even one ineffective treatment. The consequences of a false discovery may be severe, and even so strong a criterion as FWER control can leave us with high power. The state of data analysis, however, has changed. Now we are testing millions of hypotheses at once, for example in genome-wide association studies, and making a false discovery is not the end of the world. Moreover, FWER is so stringent a method that individual departures from the null in this setting have little chance of being detected; FWER control often returns nothing. While we do not want to waste time on too many null genes, we need a compromise: the burden of proof for FWER is too high. 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.

A new point of view advanced by Benjamini & Hochberg (1995) proposes controlling the expected proportion of errors among the rejected hypotheses.

	H_0 accepted	H_0 rejected	Total
H_0 true	U	V	n_0
H_0 false	T	S	$n - n_0$
	$n - R$	R	n

Table 1: Formal definition of FDR

Using the notation from the above table, we define the following criteria.

- Family wise error rate: $\text{FWER} = \mathbb{P}(V \geq 1)$

- False discovery proportion (FDP):

$$\text{FDP} = \frac{V}{\max(R, 1)} = \begin{cases} V/R & \text{if } R \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

If we made no rejections, then our false discovery proportion is 0.

- Notice that although we observe R , we do not observe V , and so FDP is an unobserved random variable. The criterion we propose to control is its expectation, which we refer to as the False Discovery Rate:

$$\text{FDR} = \mathbb{E}[\text{FDP}]$$

Consider what it means to control FDR: if we repeat our experiment many times, on average we control the FDP. This is not a statement about our individual experiment, and does not say much about the chance of having our FDP exceed a certain threshold—other than the very weak bound we can obtain via Markov's inequality. FWER, on the other hand, *does* control for an individual experiment. That is, 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. Moreover, although controlling FWER certainly gives FDR control, the converse does not hold. Overall, FDR is a relatively weak notion of control. Alternatives to FDR in the literature include the false exceedence rate that aims at controlling $\mathbb{P}(\text{FDP} \geq q)$ where q is a number between 0 and 1.

3 Properties of the FDR

Two properties Here we mention two key properties. The first relates the FDR to the FWER, and the second is a statement about the relative power and level of conservativeness.

1. Under the global null, the FDR is equivalent to the FWER.

To see that this is the case, note that in this setting, all rejections are false rejections, so $V = R$. Therefore $V/(R \vee 1) = 0$ for $V = 0$, and otherwise $V/(R \vee 1) = V/R = 1$. That is, $\text{FDP} = V/(R \vee 1) = \mathbb{1}_{\{V \geq 1\}}$. Taking expectations gives

$$\text{FDR} = \mathbb{E}(\text{FDP}) = \mathbb{P}(V \geq 1)$$

Conclusion: Control of FDR implies weak control of FWER.

2. $\text{FWER} \geq \text{FDR}$. Why? $\mathbb{1}_{\{V \geq 1\}} \geq \text{FDP}$, since $V \leq R$, and taking expectations gives $\mathbb{P}(V \geq 1) \geq \text{FDR}$. Therefore, any procedure that controls the FWER must also control the FDR. That is, procedures that control FDR can only be less stringent and thus more powerful.

4 FDR Controlling procedure [Benjamini, Hochberg 1995]

Setup: Begin by ordering the p -values in ascending order.

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

Fix a level $q \in [0, 1]$.

Benjamini-Hochberg procedure: 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 1. *For independent test statistics [p-values], the Benjamini-Hochberg procedure stated above controls the FDR at level q . More precisely,*

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

Remarks:

- This theorem holds for all configurations of the hypotheses.
- An interesting note is that the BH(q) threshold is adaptive; it depends on the specific values of $p_{(1)}, \dots, p_{(n)}$. So two different sets of p-values, both tested at level q , could generate quite different BH(q) thresholds, depending on their last point of crossing the line qi/n .

Comparison to Simes' procedure: Under the global null, we have shown $\text{FDR} = \text{FWER}$. Notice that in this case, $n_0 = n$ and so $\text{FWER} = \text{FDR} = qn_0/n = q$. This gives us the following chain of equalities:

$$q = \text{FDR} = \text{FWER} = \mathbb{P}(V \geq 1) = \mathbb{P}(\min(i : p_{(i)}n/i \leq q))$$

So the result $\mathbb{P}(\min(i : p_{(i)}n/i \leq q)) = q$ for Simes' procedure is an immediate consequence of Theorem 1.

Under H_0 , BH(q) controls FWER by Theorem 1. On the other hand, as shown by Hommel [1988], the BH procedure fails to control the FWER in a strong sense. Thus BH turns out to be less stringent than strong FWER controlling procedures, and thus gives us more power at the cost of controlling a different error criterion.

Comparison to Hochberg's procedure: Benjamini-Hochberg is a Step-Up procedure. It is interesting to compare it with Hochberg's procedure, which controls the FWER in a strong sense under independence.

Recall that Hochberg's procedure considers the p -values in descending order, and rejects all hypotheses after the first time that

$$p(i) \leq \frac{\alpha}{n - i + 1}$$

If we set $q = \alpha$, the ratio of the two thresholds is

$$\frac{i/n}{1/(n - i + 1)} = i \left(1 - \frac{i - 1}{n} \right)$$

Thus BH(q) is approximately i times more liberal than Hochberg's procedure (for small values of i). When i is around $n/2$, the ratio of the thresholds is about $n/4$, which is a very big difference from the viewpoint of power.

5 Proof of FDR control [E. Candes, R. Foygel Barber]

In this section, we prove Theorem 1. As the conclusion is obvious when $n_0 = 0$, we assume $n_0 \geq 1$.

By defining $V_i = \mathbb{1}_{\{H_i \text{ rejected}\}}$ for each null $i \in H_0$, we can express the FDP as

$$\text{FDP} = \sum_{i \in H_0} \frac{V_i}{R \vee 1}.$$

We claim that $\mathbb{E}[V_i/(R \vee 1)] = q/n$, based on which we immediately have

$$\text{FDR} = \mathbb{E}[\text{FDP}] = \sum_{i \in H_0} \mathbb{E} \left[\frac{V_i}{R \vee 1} \right] = \sum_{i \in H_0} \frac{q}{n} = \frac{n_0}{n} q.$$

What remains for the proof is to show that this claim is true. It is helpful to re-write

$$\frac{V_i}{R \vee 1} = \sum_{k=1}^n \frac{V_i \mathbb{1}_{\{R=k\}}}{k}$$

by summing over the possible values of the number of rejections $R = 1, \dots, n$. (Note when $R = 0$, we get $V_i/(R \vee 1) = 0$, and the equation holds). Here, we use two key observations that are based on how the Benjamini-Hochberg's procedure is defined:

- When there are k rejections, then H_i is rejected if and only if $p_i \leq qk/n$, and therefore, we have $V_i = \mathbb{1}_{\{H_i \text{ rejected}\}} = \mathbb{1}_{\{p_i \leq qk/n\}}$.
- Suppose $p_i \leq qk/n$ (or H_i is rejected), let us take p_i and set its value to 0, and denote the new number of rejections by $R(p_i \rightarrow 0)$. Then this new number of rejections is exactly R , because we are only reordering the first k p -values, all of which remain below the threshold qk/n . Figure 1 illustrates this concept. On the other hand, if $p_i > qk/n$, then we do not reject p_i , and so $V_i = 0$. Therefore, we have $V_i \mathbb{1}_{\{R=k\}} = V_i \mathbb{1}_{\{R(p_i \rightarrow 0)=k\}}$.

Combining the observations above and taking the expectation conditional on all p -values except for p_i , i.e., $\mathcal{F}_i = \{p_1, \dots, p_{i-1}, \dots, p_{i+1}, p_n\}$, we have

$$\mathbb{E} \left[\frac{V_i}{R \vee 1} \middle| \mathcal{F}_i \right] = \sum_{k=1}^n \frac{\mathbb{E} [\mathbb{1}_{\{p_i \leq qk/n\}} \mathbb{1}_{\{R(p_i \rightarrow 0)=k\}} | \mathcal{F}_i]}{k} = \sum_{k=1}^n \frac{\mathbb{1}_{\{R(p_i \rightarrow 0)=k\}} qk/n}{k},$$

where the second equality holds because knowing \mathcal{F}_i and $p_i = 0$ makes $\mathbb{1}_{\{R(p_i \rightarrow 0)=k\}}$ deterministic. Here, we also used the fact that $p_i \sim \text{Uniform}[0, 1]$ and the nulls are independent. Next, we have

$$\mathbb{E} \left[\frac{V_i}{R \vee 1} \middle| \mathcal{F}_i \right] = \frac{q}{n} \sum_{k=1}^n \mathbb{1}_{\{R(p_i \rightarrow 0)=k\}} = \frac{q}{n}.$$

after noticing $\sum_{k=1}^n \mathbb{1}_{\{R(p_i \rightarrow 0)=k\}} = 1$. Since we have set p_i to 0, we must make at least one rejection – we will always reject H_i . Therefore $R(p_i \rightarrow 0) \geq 1$, and $R(p_i \rightarrow 0)$ must take a value between 1 and n . The tower property verifies that $\mathbb{E} V_i/(R \vee 1) = q/n$.

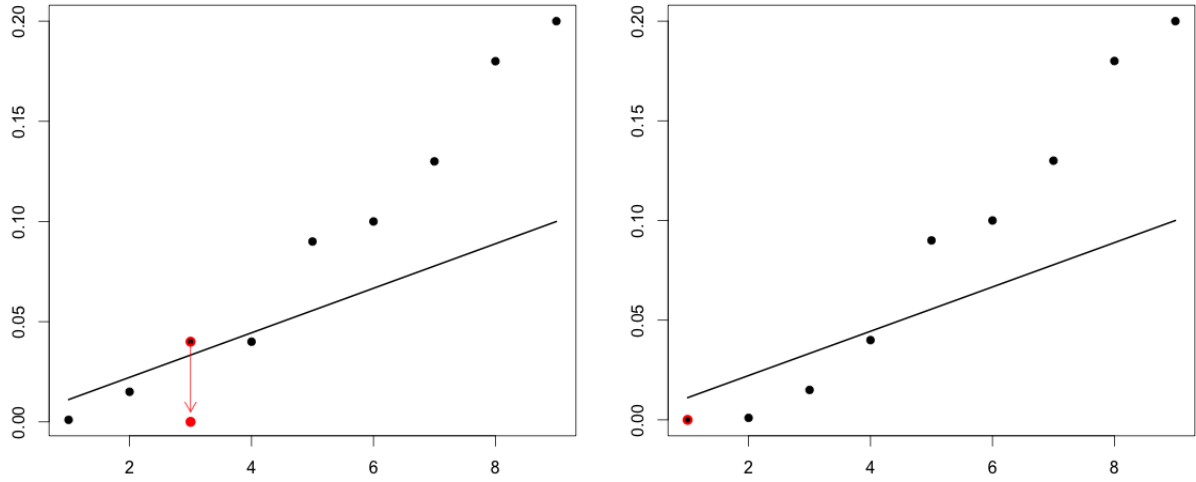


Figure 1: Demonstration of $R(p_i \rightarrow 0)$. Here we have plotted ordered p -values and let $i = 3$.

Some remarks:

- The assumptions we need here are only the independence of the null p -values p_1, \dots, p_{n_0} among themselves and from the non-nulls. We do not require the independence between the non-null p -values.
- Benjamini-Hochberg's procedure finds the smallest j such that that $p_{(n-j)}$ falls below $\frac{(n-j)q}{n}$. Such definition is a 'stopping time' with respect to some filtration, and we will see later how to apply the theory of martingales to prove this theorem again.