STATS 300C: Theory of Statistics **Spring 2022**

Lecture 6 — April 14, 2022

Lecturer: Prof. Emmanuel Candès *Editor: Parth Nobel, Scribe: Sophia Lu*

Warning: *These notes may contain factual and/or typographic errors. They are based on Emmanuel Candès's course from 2018 and 2022, and scribe notes written by Feng Ruan, Rina Friedberg, Junjie Zhu, and Tim Morrison.*

6.1 Outline

Agenda: FDR Control

1. False Discovery Rate (FDR)
2. Properties of FDR
3. Procedures for Controlling FDR - Benjamini-Hochberg (BH) Procedure
4. FDR control under dependent tests

6.2 False Discovery Rate (FDR)

Recall that we have the following table to summarize the four types of outcome in a set of hypothesis tests:

	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

Figure 6.1. Potential outcomes for testing multiple null hypotheses

For instance, U is the number of null hypotheses H_0 that are true and that we accept. The family-wise error rate (FWER) is therefore FWER $= \mathbb{P}(V \geq 1)$, the probability of at least one false rejection under all configurations of hypotheses. Controlling the FWER makes a lot of sense when we are testing a small number of hypotheses; however, the burden of controlling FWER is too stringent and individual departures from the null often have little chance of being detected under FWER control. Especially in the settings where we test a large number of hypotheses and the cost of making false rejections is not too high, we would prefer a less stringent way of controlling false discoveries. For example, in the case of genome-wide association studies, and making a false discovery is not the end of the world.

Instead we would like 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. Benjamini and Hochberg (1995) instead consider the false discovery proportion:

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

That is, the FDP is the proportion of rejections that are false rejections, unless there are no rejections at all, in which case the FDP is defined to be zero. We do not actually observe the FDP, since we do not know V, but we can try to control its expectation, called the false discovery rate (FDR):

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

One important observation and a common objection to FDR is that controlling the FDR gives us security *on average* across many repetitions of an experiment, but unlike FWER FDR does not guarantee anything about a **particular** study so we cannot make a statement about the chance of making at least a false discovery about a single experiment. Hence, it is useful to think about FDR in the sense of the experiments done by the scientific community as a whole as opposed to each single experiment.

As we will see in the next section, FDR is also a weaker notion of control than FWER, which makes it a useful compromise in modern settings when the number of hypotheses is large enough that FWER control can be too stringent.

6.3 Properties of the FDR

If the global null is true, then $S = 0$ in Figure 6.1, so $V = R$. Thus, the FDP is 0 if $V = 0$ and 1 otherwise, so $FDP = \mathbf{1}\{V \geq 1\}$. Taking expectations, we get that FDR = FWER. In particular, this means that FDR control implies weak FWER control.

In general, we have

$$\mathbf{1}\{V \geq 1\} \geq \frac{V}{\max\{R,1\}}$$

since the righthand side is zero whenever the lefthand side is zero and bounded by one in general. Taking expectations, we get that FWER $\geq$ FDR. This means that any procedure that controls the FWER also controls the FDR, so FDR control is in general a weaker condition, allowing for more powerful tests.

6.4 Benjamini-Hochberg (BH) Procedure

The Benjamini-Hochberg procedure is a step-up procedure with adaptive p-values. The BH procedure at level α (BH(α)) is formally defined as follows. Given p-values $p_{(1)} \leq p_{(2)} \leq ... \leq p_{(n)}$, let i_0 be the largest i such that

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

We then reject all $H_{(i)}$ with $i \le i_0$. The BH critical values are in fact very different from those of Holmes test. Additionally, the BH threshold is adaptive in the sense the final significant threshold depends on the empirical distribution of p-values of $\{p_1, ..., p_n\}$. This procedure is also far less conservative than Hochberg's procedure. Let α_i^{BH} denote the critical value for the i-th smallest p-value. Recall that the critical value for the i-th smallest p-value α_i^{H} under Hochberg's procedure is $\alpha_i^{\text{H}} = \frac{\alpha}{n-i+1}$. Taking $i = \frac{n}{2}$, we have

$$\alpha_i^{\text{H}} = \frac{\alpha}{n-i+1} \approx \frac{2\alpha}{n}$$
$$\alpha_i^{\text{BH}} = \frac{\alpha i}{n} \approx \frac{\alpha}{2}$$

Hence we see from the above comparison that BH is far less conservative than Hochberg's procedure.

Not only does the BH procedure control the FDR under independence, we can even get an explicit expression for the FDR:

Theorem 1. *If $\{p_1, ..., p_n\}$ are mutually independent, the Benjamini-Hochberg procedure at level α satisfies*

$$\text{FDR} = \frac{n_0}{n}\alpha \le \alpha$$

where n_0 is the number of true nulls.

Proof. The theorem is trivial when $n_0 = 0$, since both sides are zero, so we assume $n_0 \ge 1$. Let $V_i = \mathbf{1}\{H_i \text{ is rejected}\}$. Then

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

Note that $\frac{V_i}{\max\{R,1\}}$ are identically distributed for all $i \in H_0$, since the null p-values have the same (uniform) distribution. Thus, it suffices to show that each $\frac{V_i}{\max\{R,1\}}$ in this sum has expectation $\frac{\alpha}{n}$. Note that

$$\frac{V_i}{\max\{R, 1\}} = \sum_{k=1}^{n} \frac{V_i \mathbf{1}\{R = k\}}{k}.$$

Define $R(p_i \to 0)$ to be the number of rejections we would get from BH(α) if we changed the ith p-value to zero, keeping all the rest the same. We claim that

$$\sum_{k=1}^{n} \frac{V_i \mathbf{1}\{R = k\}}{k} = \sum_{k=1}^{n} \frac{V_i \mathbf{1}\{R(p_i \to 0) = k\}}{k}.$$

If $V_i = 0$, this is trivially true. Otherwise, let $1 \le m \le n$ be such that $R = m$. Then the lefthand side of the above sum is $\frac{1}{m}$. Changing p_i to zero does not increase the number of rejections, since p_i already lies below the rejection region whenever $V_i = 1$. Then $R(p_i \to 0) = m$ as well, confirming the above equality.

Now let $\mathcal{F}_i$ be the σ-algebra generated by $\{p_1, ..., p_{i-1}, p_{i+1}, ..., p_n\}$, i.e. every p-value besides p_i. We have

$$\begin{aligned}
\mathbb{E}\left(\frac{V_i}{\max\{R,1\}} \mid \mathcal{F}_i\right) &= \sum_{k=1}^{n} \mathbb{E}\left(\frac{V_i \mathbf{1}\{R(p_i \to 0) = k\}}{k} \mid \mathcal{F}_i\right) \\
&= \sum_{k=1}^{n} \mathbb{E}\left(\frac{\mathbf{1}\{p_i \le \frac{k\alpha}{n}\}\mathbf{1}\{R(p_i \to 0) = k\}}{k} \mid \mathcal{F}_i\right) \\
&= \sum_{k=1}^{n} \frac{\mathbf{1}\{R(p_i \to 0) = k\}}{k} \mathbb{P}\left(p_i \le \frac{k\alpha}{n}\right) \\
&= \frac{\alpha}{n} \sum_{k=1}^{n} \mathbf{1}\{R(p_i \to 0) = k\} \\
&= \frac{\alpha}{n}
\end{aligned}$$

Here, we use that $R(p_i \to 0)$ is $\mathcal{F}_i$-measurable to pull it out of the expectation, then use that V_i is independent of $\mathcal{F}_i$ (because the p-values are all independent by assumption) to get that $\mathbb{P}\left(p_i \le \frac{k\alpha}{n} \mid \mathcal{F}_i\right) = \mathbb{P}\left(p_i \le \frac{k\alpha}{n}\right) = \frac{k\alpha}{n}$. Finally, note that $R(p_i \to 0) \in \{1, \dots, n\}$ can only take on one particular number $k \in \{1, \dots, n\}$, so $\mathbf{1}\{R(p_i \to 0) = k\} = 1$ for one and only one $k \in \{1, \dots, n\}$. Since $\text{FDR} = \sum_{i \in H_0} \mathbb{E}\left(\frac{V_i}{\max\{R,1\}}\right)$, the tower property implies that $\mathbb{E}[\text{FDP}] = \frac{n_0}{n}\alpha$.

Remark. In the above proof, we only used that the null p-values are mutually independent, and that the null p-values are independent of the non-null p-values.

6.5 BH under Dependence

The theorem in the previous section established FDR control when the p-values are independent, but we now turn to the case that the p-values are dependent. It turns out that we can still get a bound on the FDR for BH(α), but this bound is roughly of size $\alpha \log(n)$. In order to motivate the theorem in this section, we first illustrate in the case of two p-values (p_1, p_2), which are both marginally uniform but may not be independent. Let us define the regions I, II, and III as in the figure below:

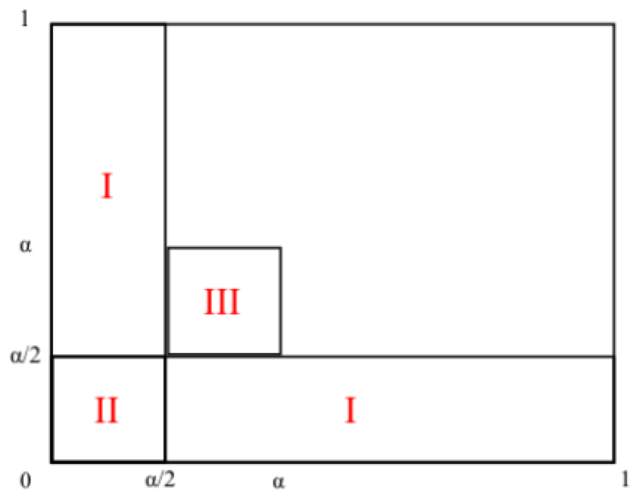

Figure 6.2. BH rejection regions. In order to find the worst-case FDR, our goal is to maximize the area of region III and make the area of region II zero in order to maximize dependence.

The FDR of BH(α) is

$$\mathrm{FDR} = \mathbb{P}(\mathrm{I}) + \mathbb{P}(\mathrm{II}) + \mathbb{P}(\mathrm{III}) = \alpha + \mathbb{P}(\mathrm{III}) - \mathbb{P}(\mathrm{II})$$

where we use that p_1 and p_2 are marginally uniform to conclude that $\mathbb{P}(\mathrm{I}) + 2\mathbb{P}(\mathrm{II}) = \alpha$. The largest that the FDR can be is $\alpha + \mathbb{P}(\mathrm{III})$, and the largest that $\mathbb{P}(\mathrm{III})$ can be is $\frac{\alpha}{2}$, since

$$\mathbb{P}\left(\frac{\alpha}{2} \le p_1 \le \alpha, \frac{\alpha}{2} \le p_2 \le \alpha\right) \le \mathbb{P}\left(\frac{\alpha}{2} \le p_1 \le \alpha\right) = \frac{\alpha}{2}.$$

Thus, the FDR is at most $\frac{3\alpha}{2}$ and attains this upper bound when $\mathbb{P}(\mathrm{II}) = 0$ and $\mathbb{P}(\mathrm{III}) = \frac{\alpha}{2}$. For general n, Guo and Rao (2008) proved the following theorem:

Theorem. (Tightness of inflation) There are joint distributions of p-values for which FDR of the BH procedure is at least $\min\{\alpha S(n), 1\}$, with

$$S(n) = 1 + \frac{1}{2} + ... + \frac{1}{n} \approx \log(n) + \gamma \approx \log(n) + 0.577$$

where γ is the Euler–Mascheroni constant.

Finally, we have the following theorem from Benjamini and Yekuteli (2001).

Theorem. *Under dependence, BH(α) controls the FDR at level $\alpha S(n)$, and specifically*

$$\mathrm{FDR} \le \alpha S(n) \frac{n_0}{n}.$$

Proof. This proof is due to Barber and Candès. As in the previous theorem, we write

$$\mathrm{FDP} = \sum_{i \in H_0} \frac{V_i}{\max\{R, 1\}}$$

so it suffices to show that $\mathbb{E}\left(\frac{V_i}{\max\{R,1\}}\right) \leq \frac{\alpha}{n}S(n)$. Letting $\alpha_k = \frac{k\alpha}{n}$, we have

$$\begin{aligned}
\frac{V_i}{\max\{R,1\}} &= \sum_{k=1}^{n} \frac{\mathbf{1}\{p_i \leq \alpha_k\}\mathbf{1}\{R=k\}}{k} \\
&= \sum_{k=1}^{n}\sum_{\ell=1}^{k} \frac{\mathbf{1}\{\alpha_{\ell-1} < p_i \leq \alpha_\ell\}\mathbf{1}\{R=k\}}{k} \\
&= \sum_{\ell=1}^{n}\sum_{k\geq\ell} \frac{\mathbf{1}\{\alpha_{\ell-1} < p_i \leq \alpha_\ell\}\mathbf{1}\{R=k\}}{k} \\
&= \sum_{\ell=1}^{n} \frac{\mathbf{1}\{R \geq \ell\}}{R}\mathbf{1}\{p_i \in (\alpha_{\ell-1}, \alpha_\ell]\} \\
&\leq \sum_{\ell=1}^{n} \frac{1}{\ell}\mathbf{1}\{p_i \in (\alpha_{\ell-1}, \alpha_\ell]\}.
\end{aligned}$$

The interval $(\alpha_{\ell-1}, \alpha_\ell]$ is of length $\frac{\alpha}{n}$, so the event $p_i \in (\alpha_{\ell-1}, \alpha_\ell]$ has probability $\frac{\alpha}{n}$ (since p_i is uniform). Taking expectations then gives

$$\begin{aligned}
\mathbb{E}\left(\frac{V_i}{\max\{R,1\}}\right) &= \frac{\alpha}{n}\sum_{\ell=1}^{n}\frac{1}{\ell} \\
&= \frac{\alpha}{n}S(n).
\end{aligned}$$

This implies that $\mathrm{FDR} \leq \alpha S(n)\frac{n_0}{n}$, completing the proof.