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

Lecture 8 — April 21, 2022

Lecturer: Prof. Emmanuel Candès *Editor: Parth Nobel, Scribe: Amber Hu*

Warning: *These notes may contain factual and/or typographic errors. They are based on Emmanuel Candès's course from 2018 to 2022, and scribe notes written by Julie Zhang.*

Outline

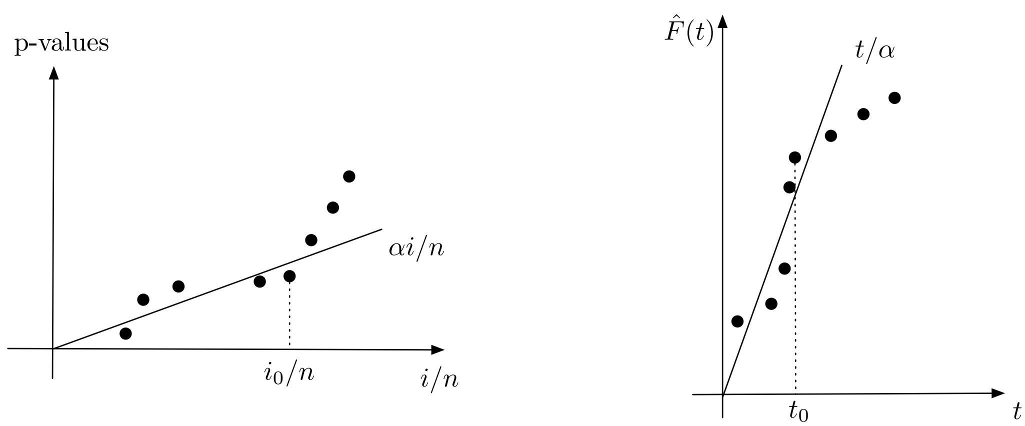
Agenda: False Discovery Rate.

1. Empirical Process viewpoint of BH.
2. Empirical Process viewpoint of FDR control.
3. Improving on BH.

Much of the material in this lecture is taken from Storey, Siegmund, and Taylor (2004) [1].

8.1 The Empirical Process Viewpoint of BH

In previous lectures, we introduced the Benjamini-Hochberg (BH) procedure by looking at the sorted p-values on the x-axis and whether they fall below a critical line. An alternative way to view BH is to flip the axes and view the sorted p-values on the y-axis. This is illustrated in the following figure.



(a) P-values on the y axis, indices on x

(b) P-values on the x axis, indices on y

Figure 8.1: Sorted p-values and BH threshold line.

This alternative view allows us to describe the BH procedure in terms of an empirical process. The coordinates on the y-axis of Figure 8.1b are the values of the empirical CDF

$$\widehat{F}_n(t) = \frac{\#\{i : p_i \leq t\}}{n}$$

evaluated at the p-values. Denote $p_{(i)}$ as the ith smallest p-value, so that

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

and let $H_{(i)}$ be the corresponding hypothesis for each $p_{(i)}$. BH is defined to reject $H_{(1)}, ..., H_{(i_0)}$ where

$$i_0 = \max\left\{i : p_{(i)} \leq \frac{qi}{n}\right\}$$

The critical p-value is $p^* = p_{(i_0)}$ and can be written as

$$\begin{aligned}
p^* &= \max\left\{p_{(i)} : p_{(i)} \leq \frac{qi}{n}\right\} \\
&= \max\left\{p_{(i)} : p_{(i)} \leq q\widehat{F}_n(p_{(i)})\right\} \\
&= \max\left\{t \in \{p_1, ..., p_n\} : t \leq q\widehat{F}_n(t)\right\}
\end{aligned}$$

If the set is empty, the convention is $p^* = q/n$. Therefore, the BH procedure is equivalent to rejecting all hypotheses with $p_i \leq \tau_{BH}$, where

$$\tau_{BH} = \max\left\{t : \frac{t}{\widehat{F}_n(t) \vee 1/n} \leq q\right\}$$

Equivalently,

$$\tau_{BH} = \max\left\{t : \hat{F}_n(t) \vee \frac{1}{n} \geq \frac{t}{q}\right\}$$

which corresponds directly to the interpretation of BH provided in Figure 8.1b. Notice that $\tau_{BH} \geq q/n$.

This formulation has a simple interpretation. Let $t \in (0, 1)$ be fixed and consider rejecting H_i iff $p_i \leq t$. We can construct the rejection/acceptance table for the hypotheses whose values depend on t.

	H_0 accepted	H_0 rejected	Total
H_0 true	$U(t)$	$V(t)$	n_0
H_0 false	$T(t)$	$S(t)$	$n - n_0 = n_1$
	$n - R(t)$	$R(t)$	n

We define

$$FDP(t) = \frac{V(t)}{R(t) \vee 1}, \qquad FDR(t) = \mathbb{E}[FDP(t)]$$

The idea is to choose the threshold t as large as possible while controlling the FDR at level q. If we had an estimate $\widehat{FDR}$ of the FDR, we can take the threshold τ to be

$$\tau = \sup\{t \leq 1 : \widehat{FDR}(t) \leq q\}$$

and define the rejection rule to reject H_i iff $p_i \leq \tau$, where τ is a data-dependent threshold. This method defines the most liberal thresholding cutoff that controls $\widehat{FDR}(t)$, and we hope that it will also control the true $FDR(t)$ at level q. The first question is how to estimate $FDR(t)$.

Assuming $p_i \overset{iid}{\sim} \text{Unif}(0,1)$ under H_0, then $\mathbb{E}[V(t)] = n_0 t$, but n_0 is not known. Therefore, a conservative estimate of $n_0 t$ is nt, which leads to our first estimate

$$\widehat{FDR}(t) = \frac{nt}{R(t) \vee 1} = \frac{t}{\widehat{F}_n(t) \vee 1/n}$$

This leads us to exactly the BH procedure since

$$\tau_{BH} = \sup\left\{t \leq 1 : \frac{nt}{R(t) \vee 1} \leq q\right\}$$

The following theorem shows that $\widehat{FDR}(t)$ is a conservatively biased estimate. This means that our procedure that controls $\widehat{FDR}(t)$ also controls $FDR(t)$ at level q.

Theorem 1. Under independence of p-values, this FDR estimate is biased upwards:

$$\mathbb{E}[\widehat{FDR}(t)] \geq FDR(t)$$

Proof. We divide this proof into two cases: $S(t) \geq 1$ and $S(t) = 0$. First suppose $S(t) \geq 1$. Then,

$$\begin{aligned}
\mathbb{E}[\widehat{FDR}(t) - FDR(t)] &= \mathbb{E}\left[\frac{nt - V(t)}{S(t) + V(t)}\right] \\
&= \mathbb{E}\left[\frac{nt + S(t)}{S(t) + V(t)}\right] - 1 \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{nt + S(t)}{S(t) + V(t)} \mid S(t)\right]\right] - 1 \qquad (8.1)
\end{aligned}$$

Note that $\frac{nt+S(t)}{S(t)+V(t)}$ is convex in $V(t)$, so by Jensen's inequality,

$$\mathbb{E}\left[\mathbb{E}\left[\frac{nt + S(t)}{S(t) + V(t)} \mid S(t)\right]\right] \geq \mathbb{E}\left[\frac{nt + S(t)}{S(t) + \mathbb{E}[V(t) \mid S(t)]}\right] \qquad (8.2)$$

$V(t)$ and $S(t)$ are independent by assumption, so

$$\mathbb{E}[V(t) \mid S(t)] = \mathbb{E}[V(t)] = n_0 t \qquad (8.3)$$

Substituting Equations 8.2 and 8.3 into Equation 8.1, we have

$$\mathbb{E}[\widehat{FDR}(t) - FDR(t)] \geq \mathbb{E}\left[\frac{nt + S(t)}{n_0 t + S(t)}\right] - 1 \geq 0$$

For the second case, suppose $S(t) = 0$. Then,

$$\mathbb{E}[\widehat{FDR}(t) - FDR(t)] = \mathbb{E}\left[\frac{nt - V(t)}{1 \vee V(t)}\right]$$

We know that $V(t) \sim \text{Bin}(n_0, t)$. This implies that $\mathbb{E}[1 \vee V(t)] = \mathbb{P}(V(t) = 0) + n_0 t$. Applying Jensen's inequality with this identity yields

$$\begin{aligned}
\mathbb{E}\left[\frac{nt}{1 \vee V(t)}\right] &\geq \mathbb{E}\left[\frac{n_0 t}{1 \vee V(t)}\right] \\
&\geq \frac{n_0 t}{\mathbb{E}[1 \vee V(t)]} \\
&= 1 + \frac{n_0 t - \mathbb{E}[1 \vee V(t)]}{\mathbb{E}[1 \vee V(t)]} \\
&= 1 - \frac{\mathbb{P}(V(t) = 0)}{\mathbb{E}[1 \vee V(t)]} \\
&\geq 1 - \mathbb{P}(V(t) = 0) \\
&= \mathbb{P}(V(t) = 1) \qquad (8.4)
\end{aligned}$$

In addition, note that

$$\mathbb{E}\left[\frac{V(t)}{1 \vee V(t)}\right] = \mathbb{P}(V(t) = 1) \qquad (8.5)$$

Substituting Equation 8.5 into Equation 8.4 yields

$$\mathbb{E}\left[\frac{nt - V(t)}{1 \vee V(t)}\right] \geq 0$$

□

8.2 Martingale Theory and FDR Control

We can invert the estimate of FDR to prove FDR control using martingales, giving us an alternate proof of the Benjamini-Hochberg result.

Theorem 2. BH (1995).
The procedure rejecting all hypotheses with $p_i \leq \tau_{BH}$ controls the FDR:

$$\mathbb{E}[FDR(\tau_{BH})] = qn_0/n$$

Proof. We let $\tau = \tau_{BH}$. Define the filtration

$$\mathcal{F}_t = \sigma(V(s), R(s) : t \le s \le 1)$$

Notice this is a backwards filtration: for $t_1 < t_2$, $\mathcal{F}_{t_2} \subset \mathcal{F}_{t_1}$. Define the reverse martingale $\{V(t)/t, 0 \le t \le 1\}$. We prove this is indeed a martingale: Let $s \le t$.

$$\begin{aligned}\mathbb{E}\left[\frac{V(s)}{s}\big|\mathcal{F}_t\right] &= \frac{1}{s}\mathbb{E}[V(s)|\mathcal{F}_t] \\ &= \frac{1}{s} \cdot \frac{s}{t} V(t) \qquad (*) \\ &= \frac{V(t)}{t}\end{aligned}$$

where in $(*)$ we used the fact that under $\mathcal{F}_t$, $V(t) = \#\{p_i : p_i \le t, H_i \text{ null}\}$ and these $p_i \sim U[0, t]$ and are independent. This proves $\{V(t)/t, 0 \le t \le 1\}$ is a martingale.

Next, notice that τ is a stopping time with respect to $\{\mathcal{F}_t\}$. This is because knowing $V(s), R(s) = n\widehat{F}_n(s)$ for $s \ge t$ will determine whether $\tau \le t$. Therefore, $\{\tau \le t\} \in \mathcal{F}_t$ and τ is a stopping time.

We are ready to apply Doob's Optional Stopping Theorem. By definition, $R(\tau) \vee 1 = n\tau/q$. Therefore,

$$\begin{aligned}FDR(\tau) &= \mathbb{E}\left[\frac{V(\tau)}{R(\tau) \vee 1}\right] \\ &= \frac{q}{n}\mathbb{E}\left[\frac{V(\tau)}{\tau}\right] \\ &= \frac{q}{n}\mathbb{E}\left[\frac{V(1)}{1}\right] \\ &= \frac{q}{n} \cdot n_0\end{aligned}$$

□

8.3 Improving on BH

In our estimate $\widehat{FDR}(t)$, we used the simple conservative bound $\pi_0 = \frac{n_0}{n} \le 1$. Here, we consider using the distribution of p-values to improve this estimate. Fix $\lambda \in [0, 1)$ and define

$$\hat{\pi}_0^\lambda = \frac{n - R(\lambda)}{(1-\lambda)n}$$

We usually will take $\lambda = 1/2$, while $\lambda = 0$ recovers the BH procedure. The motivation for this estimation is the following:

$$\hat{\pi}_0^\lambda = \frac{n_0 - V(\lambda) + n_1 - S(\lambda)}{(1-\lambda)n}$$

We would expect the non-null p-values to be small, so $n_1 - S(\lambda) \approx 0$, and hence for $\lambda = 1/2$,

$$\hat{\pi}_0^\lambda \approx \frac{n_0 - V(\lambda)}{(1-\lambda)n} \approx \frac{n_0 - (n_0/2)}{n/2} = \frac{n_0}{n}$$

For a general λ, note that

$$\mathbb{E}[\hat{\pi}_0^\lambda] = \frac{n - \mathbb{E}[R(\lambda)]}{(1-\lambda)n} \geq \frac{n - n_1 - n_0\lambda}{(1-\lambda)n} = \frac{n_0}{n} = \pi_0$$

so $\hat{\pi}_0^\lambda$ is a conservatively biased estimate of π_0. Our estimate for the false discovery rate is

$$\widehat{FDR}^\lambda(t) = \hat{\pi}_0^\lambda \cdot \frac{nt}{R(t) \vee 1}$$

and the natural test would be to reject H_i iff $p_i \leq \tau$,

$$\tau = \sup\{t \leq 1 : \widehat{FDR}^\lambda(t) \leq q\}$$

In cases where $\hat{\pi}_0^\lambda$ is smaller than 1, say 0.8, we may get more powerful results than BH because we have a significant proportion of non-nulls.

There are several drawbacks to this approach. One drawback is that we may have $\hat{\pi}_0^\lambda > 1$, in which we are being even more conservative in our estimation. More importantly, the threshold τ may not even control the FDR at level q. To resolve this issue of FDR control, we introduce a modified version called Storey's procedure.

8.4 Storey's Procedure

Storey's procedure involves a simple modification of to the estimate of π_0 defined in the previous section. Define

$$\hat{\pi}_0 = \frac{1 + n - R(1/2)}{n/2}$$

The only difference between $\hat{\pi}_0$ and $\hat{\pi}_0^{1/2}$ is the added 1 in the numerator. Our test now rejects H_i iff $p_i \leq \tau$,

$$\tau = \sup\left\{t \leq \frac{1}{2} : \widehat{FDR}(t) = \frac{1 + n - R(1/2)}{n/2} \cdot \frac{nt}{R(t) \vee 1} \leq q\right\}$$

Notice that we only take the supremum over $t \leq \frac{1}{2}$, which is necessary because the estimate of π_0 used the information of the p-values $> 1/2$.

Theorem 3. Storey's procedure controls FDR at level q.

Proof. We use martingale theory in a proof similar to the proof of Theorem 2. We know that $\widehat{FDR}(\tau) = q$. Then

$$\begin{aligned}
FDR(\tau) &= \mathbb{E}\left[\frac{V(\tau)}{R(\tau) \vee 1}\right] \\
&= \mathbb{E}\left[\frac{V(\tau)}{n\tau} \cdot \frac{n\tau}{R(\tau) \vee 1} \cdot \frac{1+n-R(1/2)}{n/2} \cdot \frac{n/2}{1+n-R(1/2)}\right] \\
&= \mathbb{E}\left[\widehat{FDR}(\tau) \cdot \frac{V(\tau)}{n\tau} \cdot \frac{n/2}{1+n-R(1/2)}\right] \\
&= q \cdot \mathbb{E}\left[\frac{V(\tau)}{\tau} \cdot \frac{1/2}{1+n-R(1/2)}\right]
\end{aligned}$$

Applying Doob's Optional Stopping Theorem to the martingale $\{V(t)/t : t \in [0, 1/2]\}$ and stopping time τ, we have

$$\begin{aligned}
FDR(\tau) &= q \cdot \mathbb{E}\left[\frac{V(1/2)}{1/2} \cdot \frac{1/2}{1+n-R(1/2)}\right] \\
&= q \cdot \mathbb{E}\left[\frac{V(1/2)}{1+n-S(1/2)-V(1/2)}\right] \\
&\le q \cdot \mathbb{E}\left[\frac{V(1/2)}{1+n_0-V(1/2)}\right]
\end{aligned}$$

where the last inequality holds because $n_1 - S(1/2) \ge 0$.

We directly calculate $\mathbb{E}[\frac{V(1/2)}{1+n_0-V(1/2)}] \le 1$. We know that $V(1/2) \sim Bin(n_0, 1/2)$. Then

$$\begin{aligned}
\mathbb{E}\left[\frac{V(1/2)}{1+n_0-V(1/2)}\right] &= \sum_{i=1}^{n_0} \mathbb{P}(V(1/2) = i)) \cdot \frac{i}{1+n_0-i} \\
&= 2^{-n_0} \sum_{i=1}^{n_0} \binom{n_0}{i} \cdot \frac{i}{1+n_0-i} \\
&= 2^{-n_0} \sum_{i=1}^{n_0} \frac{i \cdot n_0!}{(n_0-i+1) \cdot (n-i)! \cdot i!} \\
&= 2^{-n_0} \sum_{i=1}^{n_0} \frac{n_0!}{(n_0-i+1)!(i-1)!} \\
&= 2^{-n_0} \sum_{j=0}^{n_0-1} \binom{n_0}{j} \\
&= 2^{-n_0}(2^{n_0}-1) \\
&= 1 - 2^{-n_0} \\
&\le 1
\end{aligned}$$

Therefore, $FDR(\tau) \le q$ and this concludes the proof. $\square$

Bibliography

[1] John D Storey, Jonathan E Taylor, and David Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):187–205, 2004.