Statistics 210A Lecture 29 Notes

Daniel Raban

December 9, 2021

1 Multiple Testing via Control of the False Discovery Rate

1.1 False discovery rate

In our multiple testing setup, we have data $X \sim P_\theta$, hypotheses $H_i : \theta \in \Theta_{0,i}$ for $i = 1, \dots, m$, and p-values $p_1, \dots, p_m$. We also denote the rejection set as $\mathcal{R}(X) \subseteq \{1, \dots, m\}$and the true null set as $\mathcal{H}_0 \subseteq \{1, \dots, m\}$. We have been trying to control the **familywise error rate (FWER)**,

$$\mathbb{P}_\theta(|\mathcal{H}_0 \cap \mathcal{R}| \geq 1) \leq \alpha.$$

However, if we are making several hundred rejections, it might be okay if we only have a few false alarms.

Definition 1.1. Benjamini and Hochberg (1995)[1] defined the **false discovery proportion (FDP)**

$$\mathrm{FDP} = \frac{V}{R \vee 1}, \qquad V = |\mathcal{H}_0 \cap \mathcal{R}|, \qquad R = |\mathcal{R}|.$$

This is the probability is that a randomly selected rejection is a false one, which we want to control. The maximum in the denominator is just so if $R = 0$, we don't divide by 0.

Definition 1.2. Benjamini and Hochberg also define the **false discovery rate (FDR)**

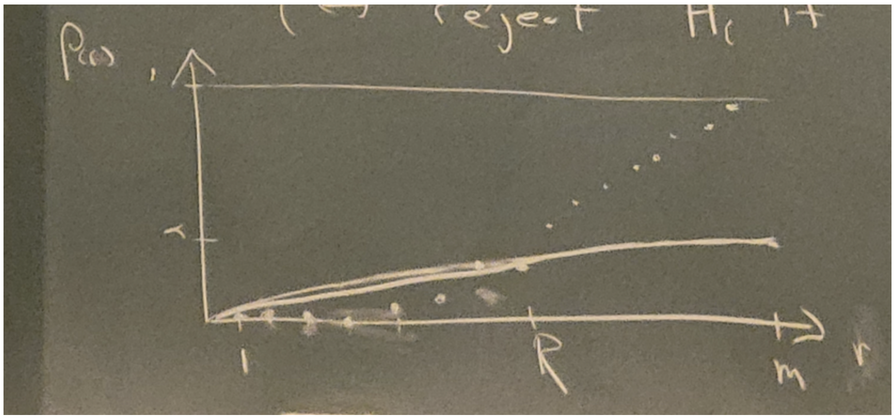
$$\mathrm{FDR} = \mathbb{E}_\theta[\mathrm{FDP}].$$

Benjamini and Hochberg didn't just introduce the FDR; they introduced a way to control it.

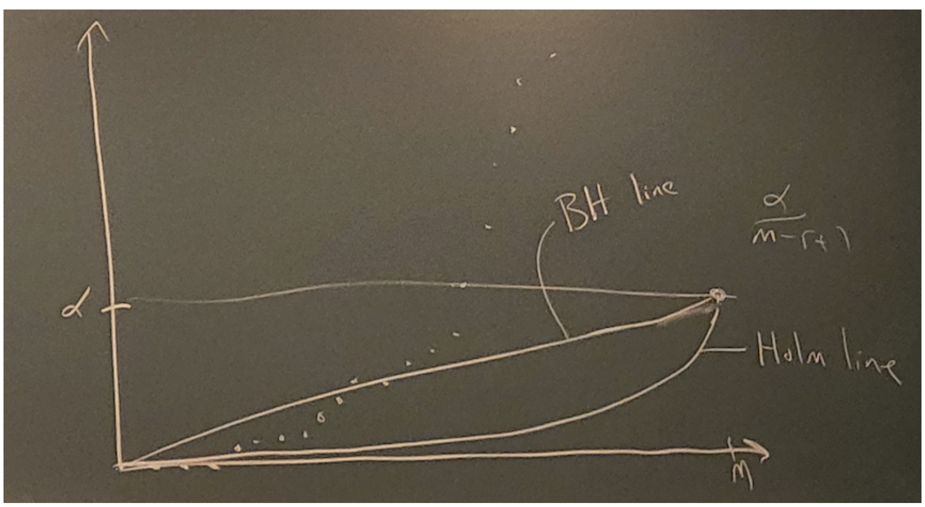
[1] They proposed this in 1988, but this radical idea of accepting some false discoveries took 7 years for any journal to accept. Professor Fithian has heard that this is the most cited paper in the entire field of statistics.

1.2 The Benjamini-Hochberg procedure

Let the p-values have order statistics $p_{(1)} \leq \cdots \leq p_{(n)}$. Then let $R^{\text{BH}} = \max\{r : p_{(r)} \leq \frac{\alpha r}{m}\}$, so the R^{BH} rejection set is $H_{(1)}, \ldots, H_{(R^{\text{BH}})}$. That is, we reject H_i if $p_i \leq \frac{\alpha R^{\text{BH}}}{m}$.



In this procedure, we reject all the hypotheses with p-values up until the last point which is below the line; even if a point is above the line, we reject it as long as there is a further point which is below the line. We can compare this to Holm's procedure, which has a lower line, since we are comparing $p_{(k)}$ to $\frac{\alpha}{m-k+1}$:



If $m = 10000$, then for Holm's procedure to make $R = 100$ rejections, $p_{(R)} \leq \frac{\alpha}{9901}$. But for BH to make 100 rejections, we need $p_{(R)} \leq \frac{\alpha 100}{10000} = \frac{\alpha}{100}$.

Remark 1.1. One issue with controlling the FDR instead of the FWER is that you can cheat. Suppose you have 5000 hypotheses you care about, but you can't make any rejections. Then you can throw in 10000 clearly false hypotheses and be able to make a lot more rejections.

To understand this procedure, first consider rejecting H_i iff $p_i \leq t$ for some fixed t. What is the false discovery proportion? Suppose $t = 5/m$. Then we expect about 5 rejections of null hypotheses. If we get 100 rejections, then we can say with more confidence that we must have had some correct rejections.

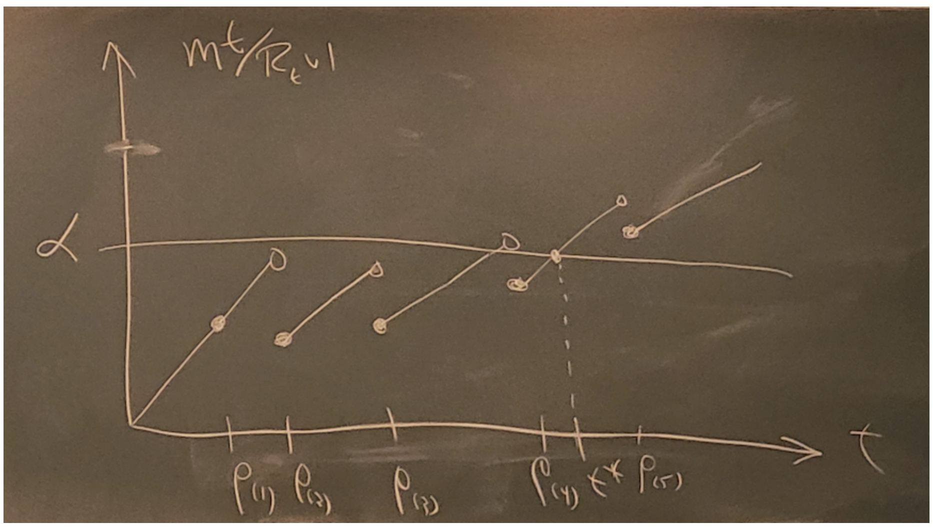
An equivalent formulation of the Benjamini-Hochberg procedure is to define

$$\widehat{\mathrm{FDP}}_t = \frac{mt}{R_t \vee 1}, \qquad R_t = \#\{i : p_i \leq t\}.$$

Then we can let

$$t^*(X) = \max\{t : \widehat{\mathrm{FDP}}_t \leq \alpha\}$$

and reject H_i if $p_i \leq t^*$.



This is equivalent because the rejection set only depends on the order statistics of the p-values and does not actually need the information of t^*; we reject $H_{(1)}, \ldots, H_{(R)}$, where

$$\begin{aligned} R &= \max\{r : \widehat{\mathrm{FDP}}_{p_{(r)}} \leq \alpha\} \\ &= \max\{r : \tfrac{mp_{(r)}}{r} \leq \alpha\} \\ &= \max\{r : p_{(r)} \leq \tfrac{\alpha r}{m}\}. \end{aligned}$$

1.3 Finite sample control of FDR using the Benjamini-Hochberg procedure

This makes sense on controlling the FDR from an asymptotic perspective (if we let the number of samples and rejections both go to infinity), but there are many interesting multiple testing problems where we only reject, say, 10 hypotheses. Asymptotic control is philosophically unsatisfactory here, but fortunately, we do have finite sample control with the Benjamini-Hochberg procedure.

Theorem 1.1. *The Benjamini-Hochberg procedure controls* FDR $\leq \alpha$*.*

Here is a celebrated proof due to Stoiey, Taylor, and Siegmund (2002) based on optional stopping of a martingale. Since we are looking at the last time the line crosses the α threshold, we need to index time backwards, starting from $t = 1$. This proof assumes that the p-values p_i are independent and that $p_i \sim U[0, 1]$ for $i \in \mathcal{H}_0$.

Proof. Then define $V_t = \#\{i \in \mathcal{H}_0 : p_i \leq t\} \leq R_t$. Then we estimate

$$\mathrm{FDP}_t = \frac{V_t}{R \vee 1}$$

by

$$\widehat{\mathrm{FDP}}_t = \frac{mt}{R_t \vee 1}.$$

This gives

$$\mathrm{FDP}_t = \widehat{\mathrm{FDP}}_t \cdot \underbrace{\frac{V_t}{mt}}_{:=Q_t}.$$

This quotient Q_t is what we will apply the optional stopping argument to. This gives

$$\begin{aligned} \mathrm{FDR} &= \mathbb{E}[\mathrm{FDP}_{t^*}] \\ &= \mathbb{E}[\widehat{\mathrm{FDP}}_{t^*} \cdot Q_{t^*}] \\ &= \alpha\, \mathbb{E}[Q_{t^*}] \end{aligned}$$

Using the optional stopping theorem,

$$\begin{aligned} &= \alpha\, \mathbb{E}[Q_1] \\ &= \alpha \frac{m_0}{m}. \end{aligned}$$

It now remains to show that Q_t is a martingale and t^* is a stopping time with respect to the filtration $\mathcal{F}_t = \sigma(p_i \vee t, i = 1, \dots, m)$; we could alternatively use $\mathcal{F}_t = \sigma(V_s : s \geq t)$. Conditional on $\mathcal{F}_t$, we know V_s and $\widehat{\mathrm{FDP}}_s$ for all $s \geq t$. As a result, t^* is a stopping time ($\mathbb{1}_{\{t^* \geq s\}}$ is $\mathcal{F}_t$-measurable). To check that this is a martingale, we have for $s < t$ that

$$\mathbb{E}[V_s \mid V_t = v] = v \frac{s}{t}.$$

(More precisely, we have that $\mathbb{E}[V_s \mid \mathcal{F}_t] = V_t \frac{s}{t}$.) So

$$\mathbb{E}\left[\frac{V_s}{ms} \mid \frac{V_t}{mt} = q\right] = \frac{1}{ms} \cdot (qmt) \cdot \frac{s}{t} = q.$$

(More precisely, we have $\mathbb{E}[Q_s \mid \mathcal{F}_t] = Q_t$.) $\square$

Here is another proof:

Proof. Define $B_i = \mathbb{1}_{\{H_i \text{ rejected}\}}$. The we can decompose

$$\frac{V}{R \vee 1} = \sum_{i \in \mathcal{H}_0} \frac{V_i}{R \vee 1}.$$

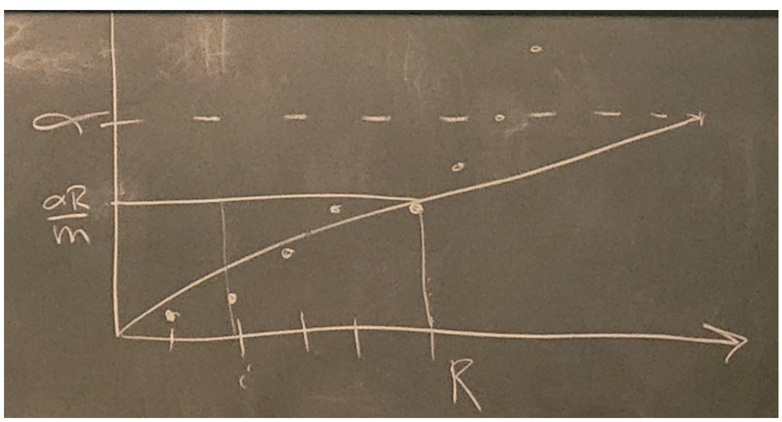
By the linearity of expectation, we can say that

$$\text{FDR} = \sum_{i \in \mathcal{H}_0} \underbrace{\mathbb{E}\left[\frac{\mathbb{1}_{\{i \text{ rejected}\}}}{R \vee 1}\right]}_{\text{want to show } \leq \alpha/m}.$$

Assume that $p_1, \ldots, p_m$ are independent. Then condition on p_{-i}. We will be in good shape if we can show that

$$\mathbb{E}\left[\frac{\mathbb{1}_{\{i \text{ rejected}\}}}{R \vee 1} \mid p_{-i}\right] \leq \frac{\alpha}{m}.$$

Rewrite the indicator as $\mathbb{1}_{\{p_i \leq \alpha R/m\}}$. We would like to pull out R, but R is not a deterministic function of p_{-i}. The key observation (which is generalizable) is that if p_i were already being rejected and we send it to 0, then it is still rejected:



Define $R^{(i)} = R(p_{-i}, 0)$. We claim that on the event $\{p_i \leq \frac{\alpha R}{m}\}$, $R^{(i)} = R$. So we can look at

$$\mathbb{E}\left[\frac{\mathbb{1}_{\{p_i \leq \frac{\alpha R}{m}\}}}{R \vee 1} \mid p_{-i}\right] = \mathbb{E}\left[\frac{\mathbb{1}_{\{p_i \leq \frac{\alpha R^{(i)}}{m}\}}}{R^{(i)}} \mid p_{-i}\right]$$

$$
\begin{aligned}
&= \frac{1}{R^{(i)}} \mathbb{P}\left(p_i \leq \frac{\alpha R^{(i)}}{m} \mid p_{-i}\right) \\
&= \frac{1}{R^{(i)}} \frac{\alpha R^{(i)}}{m} \\
&= \frac{\alpha}{m}.
\end{aligned}
$$

$\square$

Professor Fithian and a collaborator were able to generalize this proof to non-independent p_i by conditioning on something other than p_{-i}.