

## Lecture 3: Methods for false discovery rate control

May 7, 2019

Lecturer: Aldo Solari

The seminal paper in which Benjamini & Hochberg introduced the concept of FDR has changed thinking about multiple testing quite radically, showing that FWER control is not the only way to do multiple testing, and stimulating the field of multiple testing enormously.

Currently Benjamini & Hochberg (1995) paper has 50K citations. It is one of the most-cited research of all time. See <https://www.nature.com/news/the-top-100-papers-1.16224>

For a broader discussion, see [2]. The proof in Section 1 and Section 3 are taken from [1].

### 1 Benjamini-Hochberg

The Benjamini-Hochberg procedure is a step-up procedure just like the Hochberg procedure, only with higher critical values. It compares each ordered  $p$ -value  $p_{(i)}$  with the critical value  $(i\alpha)/m$  finds the largest  $j$  such that  $p_{(j)}$  is smaller than its corresponding critical value and rejects the  $j$  hypotheses with the  $j$  smallest  $p$ -values.

*Benjamini-Hochberg procedure:* If  $p_i > i\alpha/m$  for all  $i$ , reject nothing. Otherwise, let

$$i_0 = \max \left\{ i \in \{1, \dots, m\} : p_{(i)} \leq \frac{i\alpha}{m} \right\}$$

be the largest  $i$  for which  $p_{(i)} \leq \frac{i\alpha}{m}$ . Reject all  $H_{(i)}$  with  $i \leq i_0$ .

Assume for simplicity that  $p$ -values from true null hypotheses have a Uniform(0,1) distribution.

**Theorem 1.1.** *For independent  $p$ -values, the Benjamini-Hochberg procedure controls the FDR at level  $\alpha$ :*

$$\text{FDR} = \pi_0 \alpha \leq \alpha$$

*Proof.* E. Candès and R. Foygel Barber version. As the conclusion is obvious when  $m_0 = 0$ , assume  $m_0 \geq 1$ . Define  $V_i = \mathbb{1}\{H_i \text{ rejected}\}$  for each  $i \in T$  where  $T = \{i : H_i \in \mathcal{T}\}$ . We can express the FDP as

$$Q = \sum_{i \in T} \frac{V_i}{R \vee 1}$$

We claim that

$$\mathbb{E}\left(\frac{V_i}{R \vee 1}\right) = \frac{\alpha}{m}, \quad i \in T$$

based on which we have

$$Q = \sum_{i \in T} \mathbb{E}\left(\frac{V_i}{R \vee 1}\right) = \sum_{i \in T} \frac{\alpha}{m} = \pi_0 \alpha$$

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

$$\frac{V_i}{R \vee 1} = \sum_{k=1}^m \frac{V_i \mathbb{1}\{R = k\}}{k}, \quad i \in T$$

by summing over the possible values of the number of rejections  $R = 1, \dots, m$ . Note when  $R = 0$ , we get  $V_i/(R \vee 1) = 0$ , and the equality holds. Here, we use two key observations:

- When there are  $k$  rejections, then  $H_i$  is rejected if and only if  $p_i \leq (\alpha k)/m$ , and therefore, we have

$$V_i = \mathbb{1}\{p_i \leq (\alpha k)/m\}$$

- Suppose  $p_i \leq (\alpha k)/m$  (i.e.  $H_i$  is rejected). Let us take  $p_i$  and set its value to 0, and denote the new number of rejections by  $R(p_i \downarrow 0)$ . This new number of rejections is exactly  $R$ , because we have only reordering the first  $k$   $p$ -values, all of which remain below the threshold  $(\alpha k)/m$ . On the other hand, if  $p_i > (\alpha k)/m$ , then we do not reject  $H_i$ , and so  $V_i = 0$ . Therefore we have

$$V_i \mathbb{1}\{R = k\} = V_i \mathbb{1}\{R(p_i \downarrow 0) = k\}$$

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

$$\mathbb{E}\left(\frac{V_i}{R \vee 1} \middle| \mathcal{F}_i\right) = \sum_{k=1}^m \frac{\mathbb{E}(\mathbb{1}\{p_i \leq (\alpha k)/m\} \mathbb{1}\{R(p_i \downarrow 0) = k\} | \mathcal{F}_i)}{k} = \sum_{i=1}^k \frac{\mathbb{1}\{R(p_i \downarrow 0) = k\} (\alpha k)/m}{k}$$

where the second equality holds because knowing  $\mathcal{F}_i$  and  $p_i = 0$  makes  $\mathbb{1}\{R(p_i \downarrow 0)\}$  deterministic, and the fact that  $p_i \sim U(0, 1)$  and the null  $p$ -values are independent.

Next, we have

$$\mathbb{E}\left(\frac{V_i}{R \vee 1} \middle| \mathcal{F}_i\right) = \frac{\alpha}{m} \sum_{k=1}^m \mathbb{1}\{R(p_i \downarrow 0) = k\} = \frac{\alpha}{m}$$

after noticing that  $\sum_{k=1}^m \mathbb{1}\{R(p_i \downarrow 0) = k\} = 1$ . Since we have set  $p_i$  to 0, we must make at least one rejection - we will always reject  $H_i$ . Therefore  $R(p_i \downarrow 0) \geq 1$ , and  $R(p_i \downarrow 0)$  must take a value between 1 and  $m$ . □

Some remarks:

- the assumptions we need here are only the independence of null  $p$ -values among themselves and from the non-nulls. We do not require the independence between non-null  $p$ -values
- Benjamini-Hochberg procedure finds the smallest  $i$  such that  $p_{m-i}$  falls below  $[(m-i)\alpha]/m$ . Such definition is a stopping time with respect to some filtration, and it is possible to apply the theory of martingales to prove this theorem.

**Theorem 1.2.** *For PDS  $p$ -values, the Benjamini-Hochberg procedure controls the FDR at level  $\alpha$ :*

$$\text{FDR} \leq \pi_0 \alpha \leq \alpha$$

Proof omitted.

## 1.1 Comparison to Hochberg's procedure

Benjamini-Hochberg and Hochberg are both step-up procedures. The ratio of Benjamini-Hochberg and Hochberg thresholds is

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

thus Benjamini-Hochberg is approximately  $i$  times more liberal than Hochberg for small values of  $i$ . When  $i$  is around  $m/2$ , the ratio of the thresholds is about  $m/4$ .

## 2 Adaptive Benjamini & Hochberg

The Benjamini & Hochberg method, like Bonferroni, controls its error rate at level  $\pi_0 \alpha$ , rather than at  $\alpha$ . This suggests the possibility of an alternative, more powerful Benjamini & Hochberg procedure that uses critical values

$$\frac{i\alpha}{\hat{\pi}_0 m}$$

rather than  $(i\alpha)/m$  if a good estimate  $\hat{\pi}_0$  of the proportion of true hypotheses  $\pi_0$  would be available.

Such procedures are called *adaptive* procedures, and many have been proposed on the basis of various estimates of  $\pi_0$ . A problem with the adaptive approach, however, is that estimates of  $\pi_0$  can have high variance, especially if  $p$ -values are strongly correlated. Naive plug-in procedures, in which this variance is not taken into account, will therefore generally not have FDR control. More sophisticated methods are needed that do take the estimation error of  $\pi_0$  into account.

### 2.1 Adaptive BKY procedure

The procedure by Benjamini, Krieger and Yekutieli is a two-step procedure:

*BKY procedure:*

- Estimate  $\pi_0$  by

$$\hat{\pi}_0 = \frac{m - R_0}{m}$$

where  $R_0$  is the number of rejections of the Benjamini-Hochberg procedure at the slightly reduced level

$$\alpha^* = \frac{\alpha}{1 - \alpha}$$

- Perform Benjamini-Hochberg procedure at level

$$\frac{\alpha^*}{\hat{\pi}_0}$$

Note that this procedure is not guaranteed to give more rejections than the regular, nonadaptive Benjamini-Hochberg procedure because  $\alpha^*/\hat{\pi}_0$  may be smaller than  $\alpha$ .

FDR control for the adaptive BKY procedure has only yet been proven under independence, although simulations suggest FDR control under positive dependence as well.

## 2.2 Adaptive Storey procedure

Storey's procedure is based on the conservative estimate of  $m_0$

$$\hat{m}_0(\lambda) = \frac{\sum_{i=1}^m \mathbb{1}\{p_i > \lambda\}}{1 - \lambda}$$

To understand this estimator, note that  $p$ -values of true hypotheses follow a uniform distribution, and a proportion  $1 - \lambda$  is expected to be above  $\lambda$ :

$$\mathbb{E}\left(\sum_{i=1}^m \mathbb{1}\{p_i > \lambda\}\right) \geq \mathbb{E}\left(\sum_{i \in T} \mathbb{1}\{q_i > \lambda\}\right) = m_0(1 - \lambda)$$

thus  $\mathbb{E}(\hat{m}_0) \geq m_0$ .

*Storey procedure:*

- Choose  $\lambda \in (0, 1)$ . Estimate  $\pi_0$  by

$$\hat{\pi}_0(\lambda) = \frac{\sum_{i=1}^m \mathbb{1}\{p_i > \lambda\} + 1}{(1 - \lambda)m}$$

- Perform Benjamini-Hochberg procedure at level

$$\frac{\alpha}{\hat{\pi}_0(\lambda)}$$

The addition of 1 to the numerator makes sure that  $1/\hat{\pi}_0$  is always well-defined. The value of  $\lambda$  is typically  $1/2$ , although  $\lambda = \alpha$  has also been advocated. Storey procedure controls FDR under independence but not under positive dependence.

## 2.3 Minimally adaptive BH procedure

This adaptive procedure for FDR control is uniformly more powerful than the procedure of Benjamini and Hochberg [3].

*MABH procedure:* If  $p_i > i\alpha/m$  for all  $i$ , reject nothing, otherwise set

$$i_0 = \max \left\{ i \in \{1, \dots, m\} : p_{(i)} \leq \frac{i\alpha}{m-1} \right\}$$

Reject all  $H_{(i)}$  with  $i \leq i_0$ .

This adaptive procedure, unlike the procedures mentioned above, is uniformly more powerful than the original BH procedure, and has FDR control under exactly the same assumptions as the BH procedure. As a uniform improvement, the new procedure may always replace the old one.

It is a minimally adaptive procedure that only admits the estimates  $\hat{\pi}_0 = 1$  and  $\hat{\pi}_0 = (m-1)/m$ . Because of this limitation, the gain in power relative to the BH procedure is minimal, even negligible if  $m$  is large, but it can be appreciable in smaller problems.

## 3 Benjamini-Yekutieli

It has been proved that the Benjamini-Hochberg procedure controls FDR at level  $\alpha$  also under the PDS assumption. If the PDS assumption is not valid, an alternative is the procedure of Benjamini & Yekutieli, which is valid under general dependence.

Consider the case where we have two hypotheses, both true. The two  $p$ -values are marginally uniformly distributed, but they may be dependent.

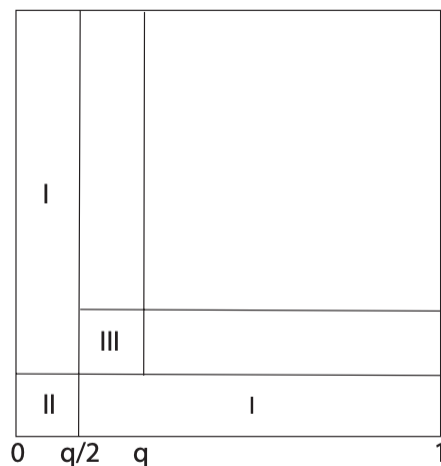


Figure 1: The BH<sub>q</sub> rejection region.

We refer to Figure 1 and calculate

$$\text{FDR} = P(I) + P(II) + P(III)$$

Consider the joint distribution of  $p$ -values as in Figure 2.

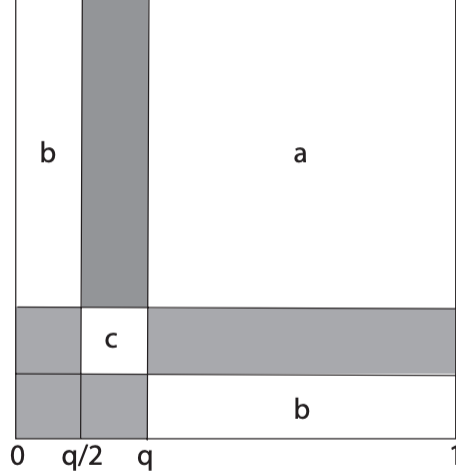


Figure 2: A piecewise constant joint distribution.

The distribution is piecewise constant with

- $b = 1/(1 - \alpha)$
- $c = 2/\alpha$
- $a = b(1 - b\alpha/2)$
- grey areas have 0 probability

One can check that marginals are uniform, and that

$$\text{FDR} = \frac{3\alpha}{2}$$

More generally, there are joint distributions of  $p$ -values for which the FDR of the BH procedure at level  $\alpha$  is

$$\alpha \sum_{j=1}^m \frac{1}{j}$$

*Benjamini-Yekutieli procedure:* If  $p_i > (i\alpha)/(m \sum_{j=1}^m j^{-1})$  for all  $i$ , reject nothing. Otherwise, let

$$i_0 = \max \left\{ i \in \{1, \dots, m\} : p_{(i)} \leq \frac{i\alpha}{m \sum_{j=1}^m j^{-1}} \right\}$$

be the largest  $i$  for which  $p_{(i)} \leq \frac{i\alpha}{m \sum_{j=1}^m j^{-1}}$ . Reject all  $H_{(i)}$  with  $i \leq i_0$ .

**Theorem 3.1.** *The Benjamini-Yekutieli procedure controls the FDR at level  $\alpha$ :*

$$\text{FDR} \leq \pi_0 \alpha \leq \alpha$$

Proof omitted.

Note that it may happen that Holm's procedure rejects more hypotheses than Benjamini-Yekutieli.

## References

- [1] E. Candes et al. Stats 300c: Theory of statistics. *Lecture notes*, 2018.
- [2] J. J. Goeman and A. Solari. Multiple hypothesis testing in genomics. *Statistics in medicine*, 33(11):1946–1978, 2014.
- [3] A. Solari and J. J. Goeman. Minimally adaptive bh: A tiny but uniform improvement of the procedure of benjamini and hochberg. *Biometrical Journal*, 59(4):776–780, 2017.