Lecture 12. False Discovery Rate Under Dependency

[1] *School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen)*

1. Benjamini & Hochberg's procedure under dependency

Consider the case where we have two hypotheses and assume we are under the global null. In this case the FDR and the FWER are the same. As usual, we shall assume that p-values are uniformly distributed but they may now be dependent.

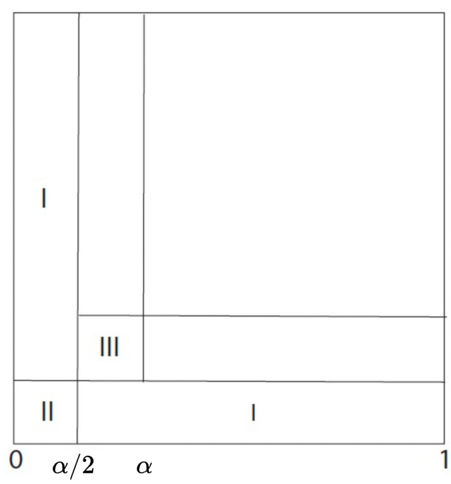


Figure 1: The BH(α) rejection region.

We refer to Figure.1 and calculate

$$\begin{aligned} \text{FDR} &= \mathbb{P}(I) + \mathbb{P}(II) + \mathbb{P}(III) = \alpha + \mathbb{P}(III) - \mathbb{P}(II) \\ &\leq \alpha + \mathbb{P}(III) \leq \alpha + \mathbb{P}(\alpha/2 \leq p_1 \leq \alpha) = 3\alpha/2 \end{aligned}$$

so Benjamini & Hochberg's procedure at least guarantees to control at level $3\alpha/2$.

However, there are configurations of p-values for which the FDR is exactly $3\alpha/2$. Consider the jointdistribution of p-values as in Figure 2. The distribution is piecewise constant with density function given by

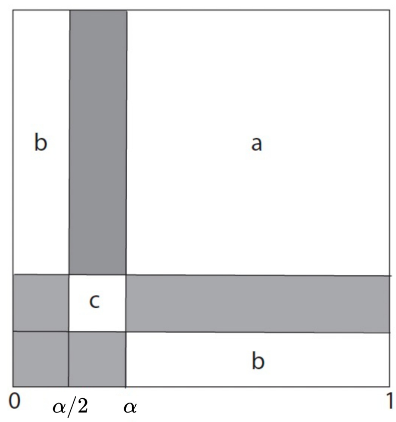


Figure 2: A piecewise constant joint distribution.

$$f(p_1, p_2 \in b) = \frac{1}{1-\alpha}, \quad f(p_1, p_2 \in c) = \frac{2}{\alpha}, \quad f(p_1, p_2 \in a) = \frac{1 - \frac{3}{2}\alpha}{(1-\alpha)^2}.$$

where the grey areas have zero probability. And one can check that the marginals are uniform. It easy to see that FDR $= 3\alpha/2$ (under the globalnull) since FDR $= \alpha + \mathbb{P}(III) - \mathbb{P}(II)$. This kind of configurations suggesting the fail of control at level α can be generalized to an arbitrary numberof hypotheses.

Theorem 1.1. *Under the global null, there are joint distributions of p-values for which the* FDR *of the BH(α) procedure is at least*

$$\alpha \cdot \big(S(n) \wedge 1\big), \tag{1.1}$$

where

$$S(n) = 1 + 1/2 + 1/3 + \cdots + 1/n \approx \log n + 0.577.$$

More specifically, consider a BH(α) type procedure where the critical values are specified by $0 \le \alpha_1 \le \alpha_2 \le \cdots \le \alpha_n \le 1$. Denote

$$\hat{k}_{\mathrm{BH}} = \max\left\{\tau \in \mathbb{N}^+ : p_{(\tau)} \le \alpha_\tau\right\}$$

and we set $\hat{k}_{\mathrm{BH}} = 0$ if the set is empty. The procedure then rejects $H_{(1)}, \cdots, H_{\hat{k}_{\mathrm{BH}}}$. This is a general step-up procedure; The BH(α) procedure we have seen uses special critical values given by $\alpha_k = \alpha \cdot k/n$. No matter the value of the level α , we have there is a joint distribution of p-values for which

$$\mathrm{FDR}(\mathrm{BH}(\boldsymbol{\alpha})) \ge \left(\sum_{k=1}^{n} \frac{n(\alpha_k - \alpha_{k-1})}{k}\right) \wedge 1.$$

When $\alpha_k = \alpha \cdot k/n$, this gives (1.1).

Proof. To proof this phenomenon, we shall prove the result only for the case, where

$$\sum_{k=1}^{n} \frac{n(\alpha_k - \alpha_{k-1})}{k} \le 1,$$

the other case follows with a similar argument.

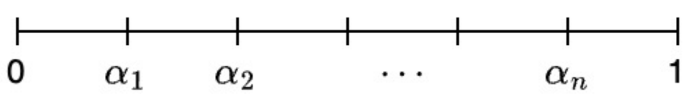
Specifically, we look at the procedure for generating p-values, for which the lower bound on the FDR is realized. We do this in a hierarchical way. First, sample K from the distribution

$$\mathbb{P}(K = k) = n \cdot \frac{\alpha_k - \alpha_{k-1}}{k}, \quad k = 1, 2, \ldots, n,$$
$$\mathbb{P}(K = 0) = 1 - \sum_{k=1}^{n} n \cdot \frac{\alpha_k - \alpha_{k-1}}{k}.$$

Second, draw a set of indices $S \subset \{1, \cdots, n\}$ of size K, uniformly at random. To sample the p-values given values of K and S, for each $i = 1, \cdots, n$,

$$p_i \sim \begin{cases} U(\alpha_{K-1}, \alpha_K) & \text{if } i \in S, \\ U(\alpha_n, 1) & \text{if } i \notin S. \end{cases}$$

In case $K = 0$, the set S would be empty, and we assume all $p_i \sim U(\alpha_n, 1)$.



We claim that

i) $p_i \sim U[0, 1]$.

ii) $\mathrm{FDR}(\mathrm{BH}(\alpha)) \ge \sum_{k=1}^{n} n \cdot \frac{\alpha_k - \alpha_{k-1}}{k}$.

To show ii), notice that we are under the global null, if $K = k > 0$, then k of the p-values lie in (α_{k-1}, α_k), which means

$$p_{(k)} \le \alpha_k \implies \hat{k}_{\mathrm{BH}} \ge k.$$

Therefore, under the global null

$$\mathrm{FDR} = \mathrm{FWER} = \mathbb{E}\big(\mathbb{1}(\hat{k}_{\mathrm{BH}} \ge 1)\big) \ge \mathbb{P}\big(K = k \ge 1\big) = \sum_{k=1}^{n} n \cdot \frac{\alpha_k - \alpha_{k-1}}{k}.$$

To show i), notice that, conditional on K and S, we have

$$p_i \sim \begin{cases} U(0,\alpha_1) & K=1, \quad i \in S, \\ U(\alpha_1,\alpha_2) & K=2, \quad i \in S, \\ \quad\vdots & \\ U(\alpha_{n-1},\alpha_n) & K=n, \quad i \in S, \\ U(\alpha_n,1) & \text{otherwise.} \end{cases}$$

Now from the distribution of K and the fact that once K is sampled, S is a randomly chosen subsetof size K, we have that

$$\mathbb{P}(i \in S \ \text{ and } \ K=k) = \alpha_k - \alpha_{k-1}$$

this gives the marginal distribution of p-values given by

$$p_i \sim \begin{cases} U(0,\alpha_1) & \text{with probability } \alpha_1, \\ U(\alpha_1,\alpha_2) & \text{with probability } \alpha_2-\alpha_1, \\ \quad\vdots & \\ U(\alpha_{n-1},\alpha_n) & \text{with probability } \alpha_n-\alpha_{n-1}, \\ U(\alpha_n,1) & \text{with probability } 1-\alpha_n. \end{cases}$$

This is precisely the $U[0,1]$ distribution. □

Theorem.1.1 gives a lower bound of the FDR under the global null and the scenario where we allow p-values to be dependent. Surprisingly, this lower bound is tight in the sence that there is a matching upper bound.

Theorem 1.2. *Under dependence, the BH(α) procedure controls at level $\alpha \cdot S(n)$. In fact,*

若希望控制 FDR=α,

$$\mathrm{FDR}(\mathrm{BH}(\alpha)) \le \alpha \cdot S(n) \cdot \underbrace{\frac{n_0}{n}}_{<0}. \tag{1.2}$$

则可以使用 BH at level $\alpha_n \triangleq \frac{\alpha}{\log n+1}$

$S(n) \approx \log n + 0.577$ (或 $\frac{\alpha}{\log n + 0.577}$)

Proof. Notice that

$$\mathrm{FDP} = \sum_{i \in \mathcal{H}_0} \frac{V_i}{1 \vee R},$$

Where $V_i = 1$ iff H_i is rejected. If we show that for any true null H_i,

$$\mathbb{E}\Big[\frac{V_i}{1 \vee R}\Big] \le \frac{\alpha}{n} S(n),$$

then appearently we would have (1.2) holds. Setting $\alpha_k = \alpha k/n$, we have shown in the proof of controling FDR, that when there are exactly k rejections, then H_i is rejected if and only if $p_i \le k\alpha/n$. Therefore,

$$\frac{V_i}{1 \vee R} = \sum_{k=1}^{n} \frac{\mathbb{1}\{p_i \le \alpha_k\}\mathbb{1}\{R=k\}}{k}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} \sum_{\ell=1}^{k} \frac{\mathbb{1}\{p_i \in (\alpha_{\ell-1}, \alpha_\ell)\}\mathbb{1}\{R = k\}}{k} \\
&= \sum_{\ell=1}^{n} \sum_{k \geq \ell} \frac{\mathbb{1}\{R = k\}}{k} \mathbb{1}\{p_i \in (\alpha_{\ell-1}, \alpha_\ell)\} \\
&= \sum_{\ell=1}^{n} \frac{\mathbb{1}\{R \geq \ell\}}{R} \mathbb{1}\{p_i \in (\alpha_{\ell-1}, \alpha_\ell)\} \\
&\leq \sum_{\ell=1}^{n} \frac{1}{\ell} \mathbb{1}\{p_i \in (\alpha_{\ell-1}, \alpha_\ell)\}
\end{aligned}
$$

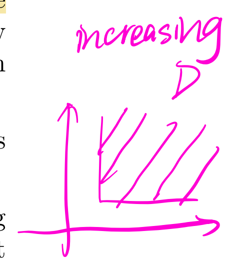
where we have set $\alpha_0 = 0$. Then taking expectation gives

$$
\mathbb{E}\Big[\frac{V_i}{1 \vee R}\Big] \leq \frac{\alpha}{n} \sum_{\ell=1}^{n} \frac{1}{\ell} = \frac{\alpha}{n} S(n).
$$

□

2. The PRDS property

So far, we have discussed FDR control via the BH(α) procedure under dependence. Inparticular, theorem.1.1 and theorem.1.2 states that under dependence, this procedure controls FDR at level $\alpha \cdot S(n)$. For large n, this is close to $\alpha \cdot (\log n + 0.577)$, which is intolerable. Natually, one would ask when would the orginal BH(α) procedure control the FDR under dependency. Here, we show that it would indeed controls the FDR under a notion of positive correlation between test statistics or p-values.



We begin by define the increasing/decreasing sets first. (Below $x \geq y$ means that $x_i \geq y_i$ for all coordinates.)

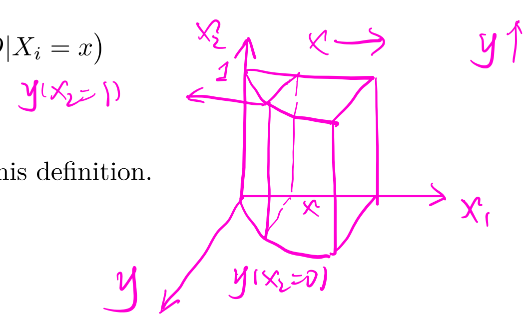
Definition 2.1 (increasing and decreasing sets). A set $D \in \mathbb{R}^n$ is called increasing if $x \in D$ and $y \geq x$ implies $y \in D$. (These sets have no boundaries in the up-right directions).

Following which, we may define the PRDS property.

Definition 2.2 (♣ **PRDS property**). A family of random variable $X = (X_1, \cdots, X_n)$ is PRDS (positive regression dependence on each of a subset) on I_0, if for any increasing set D and each $i \in I_0$,

$$
\mathbb{P}\big(X_1, \cdots, X_n) \in D | X_i = x\big)
$$

is increasing in x.



We make a few observations concerning this definition.

i) The PRDS property is invariant by co-monotone transformations. If $Y_i = f_i(X)$, where all the f_i's are either increasing or decreasing, then X is PRDS implies that Y is also PRDS.

ii) D is increasing if and only if D^C is decreasing. As a consequence, we have that a random vector X is PRDS if and only if for any decreasing C, $\mathbb{P}(X \in C | X_i = x_i)$ is decreasing in x.

iii) If $\{X_i\}$ is PRDS on I_0(true nulls), then $p_i = \bar{F}_{H_i}(X_i)$ [right-sided p-value] and $p_i = F_{H_i}(X_i)$ [left-sided p-value] are both PRDS. But for two-sided test, $p_i = 2\bar{F}_{H_i}(|X_i|)$. Since $|X_i|$ is not a monotone transformation, p_i may not be PRDS.

2.1. *Examples of* PRDS *distributions*

Proposition 2.3. *Take a multivariate normal Gaussian distribution $X = (X_1, \cdots, X_n) \sim \mathcal{N}(\mu, \Sigma)$. If $\Sigma_{ij} \geq 0$ for all $i \in I_0$ and all j, then $X = (X_1, \cdots, X_n)$ is PRDS over I_0. (The converse also holds).*

Remark 2.4. With Gaussian data, PRDS is equivalent to positive correlations.

Proof.

$$X = \begin{pmatrix} X_1 \\ \underbrace{X_{(-1)}}_{X_2 \cdots X_n} \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_{(-1)} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{1(-1)} \\ \Sigma_{(-1)1} & \Sigma_{(-1)(-1)} \end{pmatrix}$$

In this setup, the distribution of $X_{(1)}$ given $X_1 = x$ is given by

$$\mathcal{L}(X_{(-1)} \mid X_1 = x) = N\left(\mu_{(-1)} + \Sigma_{(-1)1}\Sigma_{11}^{-1}(x - \mu_1), \Sigma_{(-1)(-1)} - \Sigma_{(-1)1}\Sigma_{11}^{-1}\Sigma_{1(-1)}\right)$$

If $\Sigma_{(-1),1}$ is positive, then the conditional means increase with x. The covariance does not dependon x, so the conditional distribution is stochastically increasing in x. Thus for any nondecreasing f, $x \leq x'$ implies

$$\mathbb{E}(f(X|X_1 = x) \leq \mathbb{E}(f(X)|X_1 = x')$$

Taking f to be the indicator function of an increasing set verifies the PRDS property. □

2.2. *FDR control under* PRDS *property*

Theorem 2.5 (♣ **FDR control under** PRDS **property**)**.** *If the joint distribution of the statistics (or joint dist. of the p-values) is* PRDS *on the set of true nulls $\mathcal{H}_0$, then the Benjamini-Hochberg procedure BH(α) controls the FDR at level $\alpha n_0/n$. BH(α) may become conservative under positive dependence since* FDR $= \alpha n_0/n$ *under independence).*

Remark 2.6. An important feature of this theorem is that it does not make explicit assumptions on the dependence structure among the non-null hypotheses. This is good from the point of view of applications, where we typically have some knowledge of the phenomenon under the null, but know (and are willing to assume) very little of the phenomenon under the alternative. Unfortunately, we typically don't know much about how the non-nulls depend on the nulls, so it's generally not known whether the statistics arising in a particular application are PRDS.

A consequence of the PRDS property is that for $t \le t'$,

$$\mathbb{P}(D|p_i \le t) \le \mathbb{P}(D|p_i \le t')$$

if i is a null and the set D is increasing.

Proof. We know that

$$\mathrm{FDR} = \mathbb{E}\Big(\sum_{i\in\mathcal{H}_0} \frac{V_i}{1\vee R}\Big), \quad \text{where } V_i = \mathbb{1}\big(H_i \text{ is reject}\big)$$

Recall that in the independent case, $\mathbb{E}V_i/(1\vee R) = \alpha/n$ (no matter the null the expectation is always α/n). Now we want to show that $\mathbb{E}V_i/(1\vee R) \le \alpha/n$. This will then imply that FDR is at most $\alpha n_0/n$. Set $\alpha_k = \alpha k/n$ and notice that

$$\begin{aligned}
\frac{V_i}{1\vee R} &= \sum_{k\ge 1} \frac{\mathbb{1}(p_i \le \alpha_k)\mathbb{1}(R=k)}{k} \\
&= \sum_{k\ge 1} \frac{\mathbb{1}(p_i \le \alpha_k)\big[\mathbb{1}(R\le k) - \mathbb{1}(R \le k-1)\big]}{k} \\
&= \sum_{k=1}^{n-1} \left[\frac{\mathbb{1}(p_i\le\alpha_k)}{k} - \frac{\mathbb{1}(p_i \le \alpha_{k+1})}{k+1}\right]\mathbb{1}(R\le k) + \frac{\mathbb{1}(R\le n)\mathbb{1}(p_i\le\alpha)}{n}.
\end{aligned}$$

(idea 类似 integrate by part)

Note that

always$=1$

$$\mathbb{E}\left(\frac{\mathbb{1}(R\le n)\mathbb{1}(p_i\le\alpha)}{n}\right) = \frac{\alpha}{n}$$

always holds, since $R \le n$ all the time and $p_i \sim U(0,1)$ under the null. If we can prove

$$\mathbb{E}\Bigg(\sum_{k=1}^{n-1}\left[\frac{\mathbb{1}(p_i\le\alpha_k)}{k} - \frac{\mathbb{1}(p_i\le\alpha_{k+1})}{k+1}\right]\mathbb{1}(R\le k)\Bigg) \le 0,$$

we are done. For each k, we have

$$\begin{aligned}
&\mathbb{E}\Bigg(\left[\frac{\mathbb{1}(p_i\le\alpha_k)}{k} - \frac{\mathbb{1}(p_i\le\alpha_{k+1})}{k+1}\right]\mathbb{1}(R\le k)\Bigg) \\
&= \frac{\mathbb{P}(p_i\le\alpha_k, R\le k)}{k} - \frac{\mathbb{P}(p_i \le \alpha_{k+1}, R\le k)}{k+1}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mathbb{P}(R \le k | p_i \le \alpha_k)\mathbb{P}(p_i \le \alpha_k)}{k} - \frac{\mathbb{P}(R \le k | p_i \le \alpha_{k+1})\mathbb{P}(p_i \le \alpha_{k+1})}{k+1} \\
&\le \frac{\mathbb{P}(R \le k | p_i \le \overset{\text{increasing}}{\underset{\downarrow\,\text{PRDS}}{\alpha_{k+1}}})\underbrace{\mathbb{P}(p_i \le \alpha_k)}_{\frac{k}{n}}}{k} - \frac{\mathbb{P}(R \le k | p_i \le \alpha_{k+1})\mathbb{P}(p_i \le \alpha_{k+1})}{k+1} \\
&= 0
\end{aligned}
$$

where the inequality follows from the PRDS property. Indeed $\{R \le k\}$ is an increasing set since when the p_i's increase for each i, R decreases (we make fewer rejections). $\square$

- *Example* 2.7. 1. Suppose $X \sim N(\mu, \Sigma)$. If we wish to test

$$H_{0i} : \mu_i = 0 \;\; v.s. \;\; H_{1i} : \mu_i > 0, \;\; i = 1, \cdots, n$$

in the setup of the claim above, our test statistics X_i are PRDS.

2. The same conclusion holds if we wish to test

$$H_{0i} : \mu_i = 0 \;\; v.s. \;\; H_{1i} : \mu_i < 0, \;\; i = 1, \cdots, n$$

3. However, for testing

$$H_{0i} : \mu_i = 0 \;\; v.s. \;\; H_{1i} : \mu_i \neq 0, \;\; i = 1, \cdots, n$$

the test statistics $|X_i|$ no longer have the PRDS property.