Storey's Procedure

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

/ compare with $\hat{\pi}_0^{\,lk} = \dfrac{n-R(1/2)}{n/2}$ ↖ we add 1 in the numerator.

Our test reject H_1 if $p_i \le \tau$, where

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

Theorem 3

Storey's procedure controls FDR at level q.

⌈ We know that $\widehat{FDR}(\tau) = q$. Then

$$FDR(\tau) = E\left[\frac{V(\tau)}{R(\tau)\vee 1}\right] = \Big/ \frac{n\tau}{n\tau} \cdot \frac{\hat{\pi}_0}{\hat{\pi}_0}$$

$$= E\Bigg[\frac{V(\tau)}{n\tau} \cdot \underbrace{\frac{n\tau}{R(\tau)\vee 1} \cdot \frac{1+n-R(\frac{1}{2})}{n/2}}_{\widehat{FDR}(\tau)} \cdot \frac{n/2}{1+n-R(\frac{1}{2})}\Bigg] =$$

$$= E\Bigg[\underbrace{\widehat{FDR}(\tau)}_{q} \cdot \frac{V(\tau)}{n\tau} \cdot \frac{n/2}{1+n-R(\frac{1}{2})}\Bigg] = q\,E\left[\frac{V(\tau)}{\tau} \cdot \frac{1/2}{1+n-R(\frac{1}{2})}\right] \ominus$$

$\left\{ \frac{V(t)}{t} : t \in [0, \frac{1}{2}] \right\}$ is the martingale w.r.t. $\mathcal{F}_t$

τ is the stopping time w.r.t. $\mathcal{F}_t$, $t \ge \frac{1}{2}$

⌈ τ is measurable function of $R(t)$, $R(\frac{1}{2})$

$R(t)$, $R(\frac{1}{2})$ is $\mathcal{F}_t$ - measurable $\Rightarrow \{\tau \le t\} \in \mathcal{F}_t$ ⌋

$\ominus$ / Doob's Optional Stopping Theorem / =

$$= q\,E\left(\frac{V(\frac{1}{2})}{1/2} \cdot \frac{1/2}{1+n-R(\frac{1}{2})}\right) = q\,E\left[\frac{V(\frac{1}{2})}{1+n-S(\frac{1}{2})-V(\frac{1}{2})}\right] \le$$

$$\le \Big/ \begin{matrix} n = n_0 + n_1 \\ n_1 - S(\frac{1}{2}) \ge 0 \end{matrix} \Big/ \le q\,E\left(\frac{V(\frac{1}{2})}{1+n_0-\frac{1}{2}}\right)$$

Let's show that

$$E\left(\frac{V(\frac{1}{2})}{1+n_0-\frac{1}{2}}\right) \le 1.$$

$\ulcorner\; V(\tfrac{1}{2}) = \sum_{\mathcal{H}_0} \mathbb{1}\{p_i \le \tfrac{1}{2}\} \sim \text{Binomial}(n_0, \tfrac{1}{2})$

$$\Rightarrow E\left(\frac{V(\frac{1}{2})}{1+n_0-V(\frac{1}{2})}\right) \overset{\text{Law of Total Probability}}{=} \sum_{i=1}^{n_0} P(V(\tfrac{1}{2})=i)\frac{i}{1+n_0-i} =$$

$$= \sum_{i=1}^{n_0} \binom{n_0}{i} \underbrace{\left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{n_0-i}}_{=2^{-n_0}} \frac{i}{1+n_0-i} = 2^{-n_0} \sum_{i=1}^{n_0} \binom{n_0}{i} \frac{i}{1+n_0-i} =$$

$$= 2^{-n_0} \sum_{i=1}^{n_0} \frac{i\, n_0!}{i!\,(n_0-i)!\,(n_0-i+1)} = 2^{-n_0} \sum_{i=1}^{n_0} \frac{n_0!}{(i-1)!\,(n_0-i+1)!} =$$

$$= 2^{-n_0} \sum_{i=1}^{n_0} \binom{n_0}{i-1} = \left| i-1 = i' \right| = 2^{-n_0} \sum_{i'=0}^{n_0-1} \binom{n_0}{i'} = 2^{-n_0}(2^{n_0}-1) =$$

$$= 1 - 2^{-n_0} \le 1 \;\lrcorner$$

$$\Rightarrow FDR(T) \le q \;\lrcorner$$

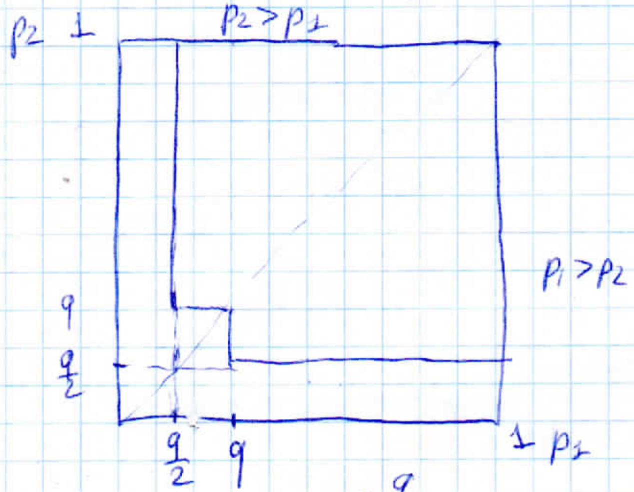
<u>BH_q under dependence.</u>

<u>Example</u>: $n = 2$ $\quad H_1, H_2$.

Assume that $H_0 = H_{01} \cap H_{02}$ is true $\Rightarrow$

$\Rightarrow$ FDR and FWER are the same.

$p_1, p_2 \sim U[0,1]$ but p_1 and p_2 may be dependent



$FDR = FWER = P(V \ge 1) =$

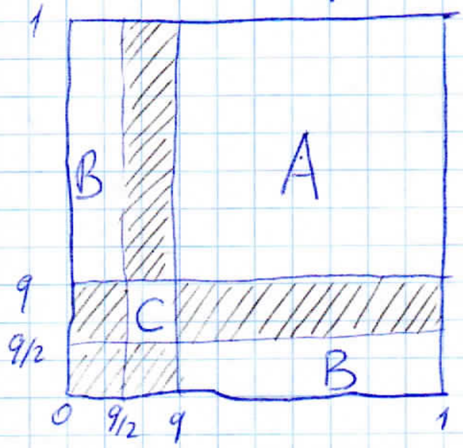
$$V \ge 1 \iff \left[\begin{array}{l} \left[\begin{array}{l} p_1 \le \frac{q}{2} \\ \frac{q}{2} < p_1 \le p_2 \le q \end{array}\right. , \quad p_1 \le p_2 \\ \left[\begin{array}{l} p_2 \le \frac{q}{2} \\ \frac{q}{2} < p_2 \le p_1 \le q \end{array}\right. \quad p_2 \le p_1 \end{array}\right.$$

$$FDR = \overset{=q}{P(p_1 \le \tfrac{q}{2})} + \overset{=q}{P(p_2 \le \tfrac{q}{2})} - P(p_1 \le \tfrac{q}{2}, p_2 \le \tfrac{q}{2}) +$$
$$+ P(p_1 \in [\tfrac{q}{2}, q],\; p_2 \in [\tfrac{q}{2}, q]) =$$

$$= q + P(p_1 \in [\tfrac{q}{2}, q], p_2 \in [\tfrac{q}{2}, q]) - P(p_1 \le \tfrac{q}{2}, p_2 \le \tfrac{q}{2}) \le$$
$$\le q + P(p_1 \in [\tfrac{q}{2}, q]) = \frac{3q}{2}$$

So, we guaranteed to control FDR at level $\frac{3q}{2}$. However, there are configurations of p-values for which $FDR = \frac{3q}{2}$.

For example, consider the following distribution of (p_1, p_2)



$$f_{p_1p_2}(x_1,x_2) = \begin{cases} a = b\left(1-\frac{bq}{2}\right), & (x_1,x_2)\in A \\ b = \frac{1}{1-q}, & (x_1,x_2)\in B \\ c = \frac{2}{q}, & (x_1,x_2)\in C \end{cases}$$

a) $f_{p_1p_2}(x_1,x_2)$ is pdf

$$\ulcorner\ aS(A)+bS(B)+cS(C) =$$

$$= \underbrace{a}_{\frac{1}{1-q}\left(1-\frac{bq}{2}\right)}(1-q)^2 + \underbrace{b}_{\frac{1}{1-q}}\, q(1-q) + \underbrace{c}_{\frac{2}{q}}\cdot\frac{q^2}{4} = (1-q)\Big(1-\underbrace{\frac{bq}{2}}_{\frac{q}{2(1-q)}}\Big)+q+\frac{q}{2} =$$

$$= (1-q) - \frac{q}{2} + q + \frac{q}{2} = 1 \ \lrcorner$$

b) the marginals are uniform

$$\ulcorner\ f_{p_2}(x_2) = \int f_{p_1p_2}(x_1,x_2)\,dx_1 = (1-q)\,b\ \mathbb{1}_{\{x_2\in[0,\frac{q}{2}]\}} +$$

$$+\frac{q}{2}c\ \mathbb{1}_{\{x_2\in[\frac{q}{2},q]\}} + \underbrace{\Big(b\frac{q}{2}+a(1-q)\Big)}_{=\frac{bq}{2}+1-\frac{bq}{2}}\ \mathbb{1}_{\{x_2\in[q,1]\}} =$$

$$= \mathbb{1}_{\{x_2\in[0,\frac{q}{2}]\}} + \mathbb{1}_{\{x_2\in[\frac{q}{2},q]\}} + \mathbb{1}_{\{x_2\in[q,1]\}} = \mathbb{1}_{\{x_2\in[0,1]\}} \ \lrcorner$$

c) $FDR = q + P(C) - P(B) = q + bS(B) - cS(C) =$

$$= q + q - \frac{q}{2} = \frac{3q}{2}$$

Theorem 1

There are distributions of p-values for which the FDR of BH_q is at least

$$q(S(n) \wedge 1),$$

where $S(n) = 1 + \frac{1}{2} + \dots + \frac{1}{n} \approx \log n + 0{,}577$

Theorem 2 (Benjamini - Yekuteli (2001))

Under dependence, the BH_q procedure controls FDR at level $qS(n)$

$$FDR \le q S(n) \frac{n_0}{n}$$

Γ $FDP = \frac{V}{R \vee 1} = \sum_{\mathcal{H}_0} \frac{V_i}{R \vee 1}$, where $V_i = \mathbb{1}\{H_{0i} \text{ is rejected}\}$

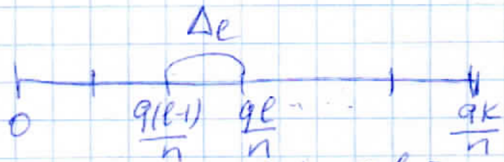
It is enough to show that for $\forall i \in \mathcal{H}_0$:

$$E \frac{V_i}{1 \vee R} \le \frac{q}{n} S(n)$$

$$\frac{V_i}{1 \vee R} = \left| \begin{array}{l} \text{Law of Total} \\ \text{probability} \end{array} \right| = \sum_{k=1}^{n} \frac{1}{k} \mathbb{1}\{H_{0i} \text{ is rejected}\} \cdot \mathbb{1}\{R = k\} =$$

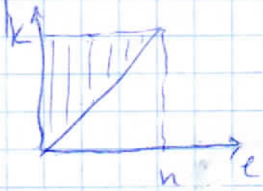
$$= \left| \begin{array}{l} \text{If } \{R = k\} \Rightarrow \text{we made } k \text{ rejections} \\ \Rightarrow \text{we reject } H_{0i} \text{ if } p_i \le \frac{qk}{n} \end{array} \right| =$$

$$= \sum_{k=1}^{n} \frac{1}{k} \mathbb{1}\{p_i \le \frac{qk}{n}\} \mathbb{1}\{R = k\} =$$



$$\Delta_\ell = \left(q\frac{(\ell-1)}{n};\ q\frac{\ell}{n}\right]$$

$$= \sum_{k=1}^{n} \sum_{\ell=1}^{k} \frac{1}{k} \mathbb{1}\{p_i \in \Delta_\ell\} \mathbb{1}\{R = k\} = \left| \begin{array}{l} 1 \le \ell \le k \\ 1 \le k \le n \end{array} \Rightarrow \begin{array}{l} k \ge \ell \\ 1 \le \ell \le n \end{array} \right|$$



$$= \sum_{\ell=1}^{n} \sum_{k \ge \ell} \frac{1}{k} \mathbb{1}\{p_i \in \Delta_\ell\} \mathbb{1}\{R = k\} = \sum_{\ell=1}^{n} \mathbb{1}\{p_i \in \Delta_\ell\} \frac{\overbrace{\mathbb{1}\{R \ge \ell\}}^{\le 1}}{\ell} \le$$

$$\le \sum_{\ell=1}^{n} \frac{1}{\ell} \mathbb{1}\{p_i \in \Delta_\ell\}$$

$$\Rightarrow E\frac{V_i}{1 \vee R} \le \sum_{l=1}^{n} \frac{1}{l} \underbrace{P\{p_i \in \Delta_l\}}_{= \frac{q}{n}} = \frac{q}{n}\sum_{l=1}^{n}\frac{1}{l} = \frac{q}{n} S(n)$$

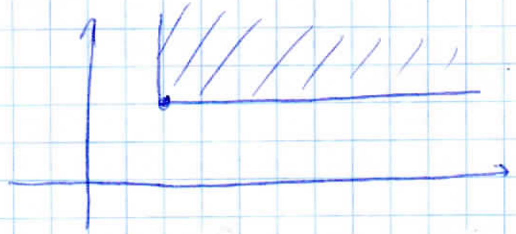
The PRDS property

We show that BH_q procedure controls the FDR under the assumption of positive correlation between test statistics or p-values.

$X \geq y \iff x_i \geq y_i$ for all coordinates

Df 1 A set $D \in \mathbb{R}^n$ is called increasing if
$x \in D$ and $y \geq x \Rightarrow y \in D$
(D have no boundaries in the North-East direction)

$\mathbb{R}^2$:



Df 2 A random vector $X = (X_1 \dots X_n)$ is PRDS (positive regression dependence on each of subset) on I_0 if
$\forall D$ - increasing set and $\forall i \in I_0$:
$P((X_1 \dots X_n) \in D / X_i = x)$ is increasing in x.

Properties:

1) PRDS property is invariant by co-monotone transformations
$\iff y_i = f_i(X)$, where $\forall f_i$ are either $\uparrow\uparrow$ or $\downarrow\downarrow$
$\Rightarrow$ if X - PRDS $\Rightarrow y$ is also PRDS
2) r.vector X is PRDS $\iff \forall$ decreasing C: $P(X \in C / X_i = x_i) \downarrow\downarrow$ in x_i
3) if X_i is PRDS $\Rightarrow p_i$ (for right-sided or left sided tests) are both PRDS (for two-sided test it may not be true)

Example of PRDS distribution:

$X = (X_1, \dots X_n) \sim N(\mu, \Sigma)$

if $\Sigma_{ij} \geq 0$ for $\forall i,j \in I_0 \Rightarrow X$ is PRDS over I_0.

$\Longleftrightarrow$ With gaussian data PRDS is equivalent to positive correlations.

FDR control under PRDS property:

Theorem 3 (Benjamini, Yekutieli (2001)):

If the joint distribution of the statistics (or joint disribution of p-values) is PRDS on the set of true nulls $\mathcal{H}_0 \Rightarrow BH_q$ controls FDR at level $q\frac{n_0}{n}$

Remark: a consequence of the PRDS property:

$\forall t_1 \leq t_2$: $P(D / p_i \leq t_1) \leq P(D / p_i \leq t_2)$,

for all null i and increasing D.

$FDR = E\frac{V}{R \vee 1} = E\sum_{i \in \mathcal{H}_0}\left(\frac{V_i}{1 \vee R}\right)$; where $V_i = \mathbb{1}_{\{H_{0i} \text{ is rejected}\}}$

If p_i are independent $\Rightarrow E\frac{V_i}{1 \vee R} = \frac{q}{n}$

here we will show that $E\frac{V_i}{1 \wedge R} \leq \frac{q}{n} \Rightarrow$

$\Rightarrow FDR \leq \frac{q n_0}{n}$

Let $q_k = \frac{qk}{n} \Rightarrow$

$$\frac{V_i}{1 \vee R} = \left| \text{Law of Total probability} \right| = \sum_{k=1}^{n} \frac{1}{k} \mathbb{1}_{\{p_i \leq q_k\}} \mathbb{1}_{\{R = k\}} =$$

$$= \sum_{k=1}^{n} \frac{1}{k} \mathbb{1}_{\{p_i \leq q_k\}} \left(\mathbb{1}_{\{R \leq k\}} - \mathbb{1}_{\{R \leq k-1\}} \right) =$$

$$\sum_{k=1}^{n} \frac{1}{k} \mathbb{1}_{\{p_i \leq q_k\}} \mathbb{1}_{\{R \leq k-1\}} = \sum_{k=2}^{n} \frac{1}{k} \mathbb{1}_{\{p_i \leq q_k\}} \mathbb{1}_{\{R \leq k-1\}} = \left| k-1 = k' \right| =$$

$$= \sum_{k'=1}^{n-1} \frac{1}{k'+1} \mathbb{1}_{\{p_i \leq q_{k'+1}\}} \mathbb{1}_{\{R \leq k'\}}$$

$$= \sum_{k=1}^{n-1} \mathbb{1}_{\{R \le k\}} \left(\frac{\mathbb{1}_{\{p_i \le q_k\}}}{k} - \frac{\mathbb{1}_{\{p_i \le q_{k+1}\}}}{k+1} \right) + \frac{\overbrace{\mathbb{1}_{\{R \le n\}}}^{=1} \mathbb{1}_{\{p_i \le q\}}}{n}$$

$$\Big/ \text{Note that: } E\, \frac{\mathbb{1}_{\{R \le n\}} \mathbb{1}_{\{p_i \le q\}}}{n} = \frac{E\, \mathbb{1}_{\{p_i \le q\}}}{n} \overset{=q,\ p_i \sim U[0,1]}{=} \frac{q}{n} \Big/$$

So, it is enough to prove that

$$E \sum_{k=1}^{n-1} \mathbb{1}_{\{R \le k\}} \left(\frac{\mathbb{1}_{\{p_i \le q_k\}}}{k} - \frac{\mathbb{1}_{\{p_i \le q_{k+1}\}}}{k+1} \right) \le 0$$

$\ulcorner$ for each k

$$E\, \mathbb{1}_{\{R \le k\}} \left(\frac{\mathbb{1}_{\{p_i \le q_k\}}}{k} - \frac{\mathbb{1}_{\{p_i \le q_{k+1}\}}}{k+1} \right) =$$

$$= \frac{1}{k} P(R \le k,\ p_i \le q_k) - \frac{1}{k+1} P(R \le k,\ p_i \le q_{k+1}) =$$

$$= \big/ P(A \cap B) = P(A)\, P(B/A) \big/ =$$

$$= \frac{1}{k} \underbrace{P(R \le k / p_i \le q_k)}_{\underset{\text{PRDS}}{\le} P(R \le k / p_i \le q_{k+1})} \underbrace{P(p_i \le q_k)}_{q_k} - \frac{1}{k+1} P(R \le k / p_i \le q_{k+1}) \underbrace{P(p_i \le q_{k+1})}_{q_{k+1}}$$

$$\le P(R \le k / p_i \le q_{k+1}) \underbrace{\left(\frac{1}{k} q_k - \frac{1}{k+1} q_{k+1} \right)}_{=0} = 0. \lrcorner$$