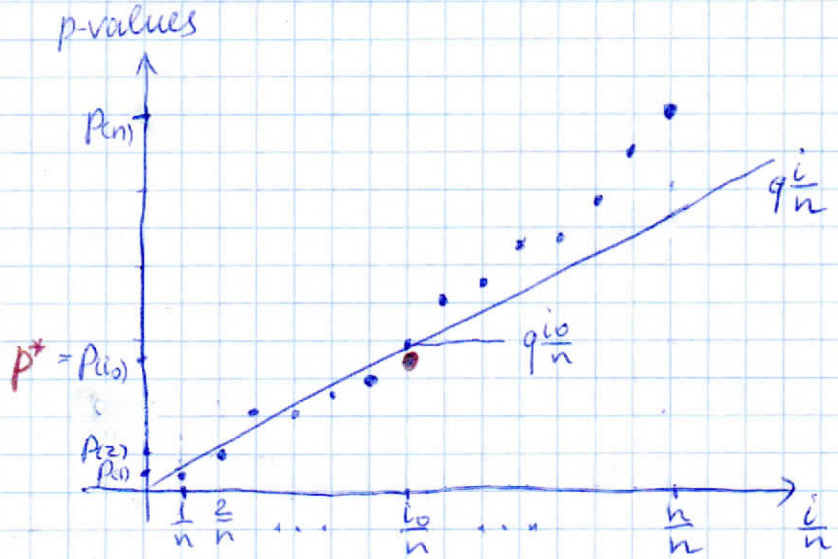
The Empirical process Viewpoint of BH_q $(q \equiv \alpha)$

BH_q : $p_{(1)} \le \dots \le p_{(n)}$
$H_{(1)} \le \dots \le H_{(n)}$

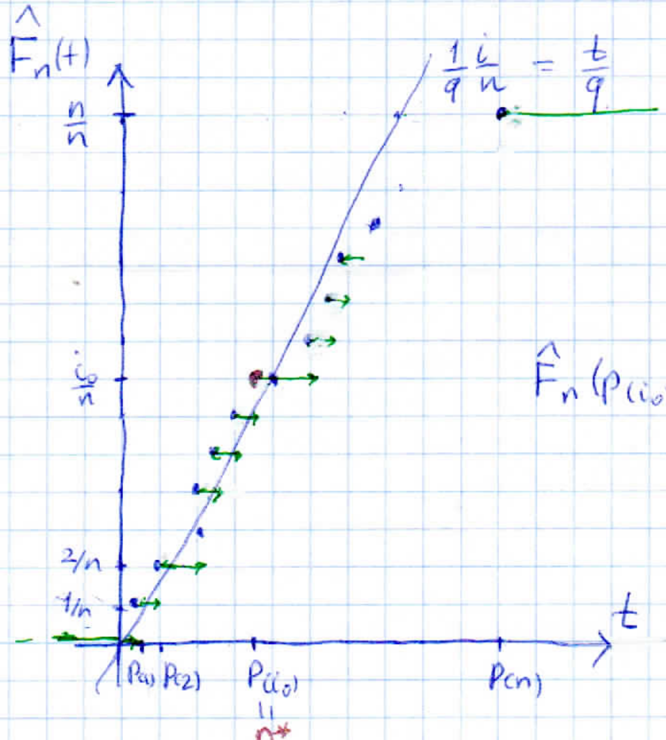
$i_0 = \max\{i : p_{(i)} \le \frac{i}{n} q\}$: $p^* = p_{(i_0)} \le \frac{i_0}{n} q$

max index for which the inequality holds



let us describe BH procedure in terms of an empirical process:

$$\hat{F}_{(n)}(t) = \frac{1}{n} \#\{i : p_i \le t\} = \begin{cases} 0, & t < p_{(1)} \\ \frac{i}{n} : & p_{(i)} \le t < p_{(i+1)} \\ 1 : & t \ge p_{(n)} \end{cases}$$



$\hat{F}_n(p_{(i_0)}) \ge \frac{p_{(i_0)}}{q}$ ← max index. $\Longleftrightarrow \hat{F}_n(t) \geq \frac{t}{q}$

$$p^* = \max\left(p_{(i)} : p_{(i)} \le q\frac{i}{n}\right) = \left| \hat{F}_n(p_{(i)}) = \frac{i}{n} \right| =$$

$$= \max\left(p_{(i)} : p_{(i)} \le q \hat{F}_n(p_{(i)})\right) = \max\left(t \in \underset{\{p_1 \dots p_n\}}{\underset{\shortparallel}{P_n}} : t \le q\hat{F}_n(t)\right)$$

$$p^* = \begin{cases} \max\left(t \in P_n : \frac{t}{q} \le \hat{F}_n(t)\right), & P_n = \{p_1 \dots p_n\} \\ \frac{q}{n} & \text{if } p_{(i)} > \frac{i}{n} q, \ \forall i = 1 \dots n \iff \\ & \iff \frac{t}{q} > \hat{F}_n(t), \ \forall t \end{cases}$$

$\iff$

BH_q procedure is equivalent to rejecting all hypothesis with

$p_i \le \tau_{BH}$, where

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

Remark: 1) $\tau_{BH} \ge \frac{q}{n}$

2) If k rejections are made $\Rightarrow$ all p-values less or equal $\frac{qk}{n}$ are rejected.

Interpretation: let $t \in (0,1)$ be fixed and reject H_i: $p_i \le t$.

$\Rightarrow$ we construct rejection/acceptance table:

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

$$FDP(t) = \frac{V(t)}{R(t) \vee 1} \quad ; \quad FDR(t) = E(FDP(t))$$

We would like to find the threshold t as large as possible while controlling the FDR at level q. $\widehat{FDR}$ is the estimate of the FDR.

$\Rightarrow \quad \tau = \sup\left(t \le 1 : \widehat{FDR}(t) \le q\right)$

and define the rejection rule: reject H_i: $p_i \le \tau$.

How to estimate $FDR(t)$?

$E\,V(t) = n_0 t$, but n_0 is unknown.

$n_0 t \le n t$ – conservative estimate

$$\Rightarrow \widehat{FDR}(t) = E\,\widehat{\frac{V(t)}{R(t) \vee 1}} = \frac{n_0 t}{R(t) \vee 1} \le \frac{nt}{R(t) \vee 1} \overset{\circledast}{=} \frac{t}{\hat F_n(t) \vee \frac{1}{n}}$$

($V(t)$ – r.v.; $R(t)$ – is known)

$\circledast$ if $R(t) = k \Rightarrow \#\{i : p_i \le t\} = K \Rightarrow$
$$\Rightarrow \frac{\#\{i : p_i \le t\}}{n} = \frac{k}{n} \Rightarrow \hat F_n(t) = \frac{1}{n} R(t)$$

$$\Rightarrow \tau = \sup\left(t \le 1 : \frac{nt}{R(t) \vee 1} \le q\right) =$$
$$= \sup\left(t \le 1 : \frac{t}{\hat F_n(t) \vee \frac{1}{n}} \le q\right) = \tau_{BH}.$$

<u>Theorem 1</u> Under independence, this FDR estimate is biased upwards:
$$E(\widehat{FDR}(t)) \ge FDR(t)$$

Martingale theory and FDR control:

<u>Theorem 2</u> The procedure rejecting all $H_i : p_i \le \tau_{BH}$ controls the FDR:
$$E(FDR(\tau_{BH})) = \frac{q n_0}{n} \le q$$

$\vdash$ $\tau := \tau_{BH}$

Define the filtration:
$$\mathcal{F}_t = \sigma\{V(s), R(s) : s \in [t, 1)\}$$

Remark: $\mathcal{F}_{t_2} \subset \mathcal{F}_{t_1}$ for $t_1 < t_2$

$V(t) := \frac{1}{t} V(t)$, $t \in [0, 1]$

Let us prove, that $V(t)$ is the martingale w.r.t. $\mathcal{F}_t$.

It is need to show that $E(V(s)/\mathcal{F}_t) = V(t), \forall s \le t$.

$\ulcorner$ $E(V(s)/\mathcal{F}_t) = E\left(\frac{V(s)}{s}/\mathcal{F}_t\right) = \frac{1}{s} E(V(s)/\mathcal{F}_t) =$

$$\begin{cases} V_i(s) = \mathbb{1}_{\{H_{0i} \text{ is rejected}\}} = \mathbb{1}_{\{p_i \le s\}} \\ V(s) = \sum_{\mathcal{H}_0} \mathbb{1}_{\{p_i \le s\}} \end{cases}$$

$= \frac{1}{s} \sum_{\mathcal{H}_0} E(\mathbb{1}_{\{p_i \le s\}} / \mathcal{F}_t) \overset{(*)}{=} \frac{1}{s} \cdot \frac{s}{t} \sum_{\mathcal{H}_0} \mathbb{1}_{\{p_i \le t\}} = V(t)$

$(*)$: $\ulcorner$ $E\left(\mathbb{1}_{\{p_i \le s\}} \underset{\text{test function}}{f(\mathbb{1}_{\{p_i \le t\}})}\right) =$

$= E(\underbrace{\mathbb{1}_{\{p_i \le s\}} f(0) \cdot \mathbb{1}_{\{p_i > t\}}}_{=0}) + E(\underbrace{\mathbb{1}_{\{p_i \le s\}} f(1) \mathbb{1}_{\{p_i \le t\}}}_{= \mathbb{1}_{\{p_i \le s\}} \cdot f(1)}) \overset{s \le t}{=}$

$= f(1) E \mathbb{1}_{\{p_i \le s\}} = f(1) \cdot s = \frac{s}{t} \cdot t f(1) =$

$= \frac{s}{t} E \mathbb{1}_{\{p_i \le t\}} f(\mathbb{1}_{\{p_i \le t\}})$ $\lrcorner$ $\lrcorner$

τ_{BH} is a stopping time w.r.t. $\{\mathcal{F}_t\} \Leftrightarrow$
$\Leftrightarrow \{\tau_{BH} \le t\} \in \mathcal{F}_t$, $\forall t \in (0,1)$

$\ulcorner$ $\tau_{BH} = \sup\{t \le 1 : \frac{nt}{R(t) \vee 1} \le q\}$ – measurable function of $R(t)$.
$R(t)$ is $\mathcal{F}_t$ – measurable $\Rightarrow \{\tau_{BH} \le t\} \in \mathcal{F}_t$ $\lrcorner$

$\Rightarrow$ Doob's Optional Stopping Theorem:

X_t – martingale w.r.t. $\mathcal{F}_t$,
σ, τ are stopping times w.r.t. $\mathcal{F}_t$: $\sigma \le \tau \le 1$
$E(X_\tau / \mathcal{F}_\sigma) = X_\sigma$

$$\Rightarrow FDR(\tau) = E\left(\frac{V(\tau)}{R(\tau)\vee 1}\right) = \left/ \begin{matrix} R(\tau)\vee 1 = \frac{n\tau}{q} \text{ by} \\ \text{definition of } \tau \end{matrix} \right/ =$$

$$= E\left(\frac{V(\tau)}{n\tau} q\right) = \frac{q}{n} E \frac{V(\tau)}{\tau} \overset{\text{Doob's th.}}{=} \frac{q}{n} E \frac{V(1)}{1} =$$

$$= \frac{q}{n} \cdot n_0 \cdot 1 = \frac{q n_0}{n} \quad \blacksquare$$

Improving on BH_q

When we estimate $FDR(t)$:

$$\widehat{FDR}(t) = \frac{n_0 t}{R(t)\vee 1} = \pi_0 \frac{nt}{R(t)\vee 1} \le \frac{nt}{R(t)\vee 1}, \text{ where } \pi_0 = \frac{n_0}{n}$$

we use inequality $n_0 \le n$. Can the distribution of p-values be used to improve the conservative estimate of π_0?

Fix $\lambda \in [0, 1)$ and define

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

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

<u>Remark</u>: $\lambda = 0 \Rightarrow R(0) = 0 \Rightarrow \hat{\pi}_0^0 = 1 \Rightarrow BH_q$ is recovered.

For general λ:

$$\hat{\pi}_0^\lambda = \frac{n - R(\lambda)}{(1-\lambda)\, n} = \left| \begin{matrix} n = \underset{\substack{\uparrow \\ \#\{H_{0i} \text{ is true}\}}}{n_0} + n_1 \\ R(\lambda) = V(\lambda) + S(\lambda) \end{matrix} \right| = \frac{n_0 - V(\lambda) + n_1 - S(\lambda)}{(1-\lambda)\, n} \geq$$

$$\geq \frac{n_0 - V(\lambda)}{(1-\lambda)\, n} \Rightarrow E\hat{\pi}_0^\lambda \geq \frac{n_0}{n} = \pi_0.$$

For example, for $\lambda = \frac{1}{2}$:

$$\hat{\pi}_0^{1/2} = \frac{n_0 - \overbrace{V(\tfrac{1}{2})}^{\approx n_0/2} + n_1 - S(\tfrac{1}{2})}{n/2} \approx$$

we would expect that non-null p-values to be small so $n_1 - S(\lambda) = T(\lambda) \approx 0$

$$\approx \frac{n_0 - n_0/2}{n/2} = \frac{n_0}{n}$$

Then our estimate for the FDR is

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

and we reject H_i if $p_i \le \tau$, where

$$\tau = \sup\{t \le 1, \hat{FDR}^{\lambda}(t) \le q\}$$

If $\hat{\pi}_0^{\lambda}$ is smaller than 1 we may get more powerful results than BH_q but the thresholt τ may not control the FDR.
Also we may have $\hat{\pi}_0^{\lambda} > 1$.
So, we introduce modified version called Storey's procedure.