Part III False discovery rate control

(1) FDR and Benjamini Hochberg procedure

Consider a general multiple testing framework with $X, (\mathcal{X}, \mathfrak{X}), P \in \mathcal{P}, \theta \in \mathbb{R}^m, \mathcal{H}_0(P)$

$m_0(P) = |\mathcal{H}_0(P)|$ number of 'non signal'

Let R being a multiple testing procedure

The FDR of R is $\mathrm{FDR}(R, P) = \mathbb{E}\left[\mathrm{FDP}(R, P)\right]$

where $\mathrm{FDP}(R, P) = \dfrac{|\mathcal{H}_0(P) \cap R|}{|R| \vee 1}$ proportion of errors among the positives

Controlling FDR at level α means :

for some $\alpha \in (0,1)$, find $R = R_\alpha$ with $\forall P \in \mathcal{P}$, $\mathrm{FDR}(R, P) \leq \alpha$ (and as many rejections as possible)

Consider a family of p-values $p_1, \dots, p_m$ satisfying $(*)$

BH procedure : step-up stopping rule $\hat{\ell} = \max\left\{ \ell \in \{0, .., m\} : p_{(\ell)} \leq \alpha \frac{\ell}{m} \right\}$

rejects the nulls corresponding to $p_{(1)} \cdots p_{(\hat{\ell})}$,

that is, $R^{BH} = \left\{ 1 \leq j \leq m : p_j \leq \dfrac{\alpha(\hat{\ell} \vee 1)}{m} \right\}$ (threshold $\hat{t} = \dfrac{\alpha(\hat{\ell} \vee 1)}{m}$)

Proposition : in a general multiple testing setting consider p-values satisfying $(*)$ and consider $P \in \mathcal{P}$ such that $(p_j, 1 \leq j \leq m)$ are mutually independent. Then we have

(i) $\mathrm{FDR}(R^{BH}, P) \leq \alpha \dfrac{m_0(P)}{m} \leq \alpha$ [Benjamini and Hochberg (1995)]

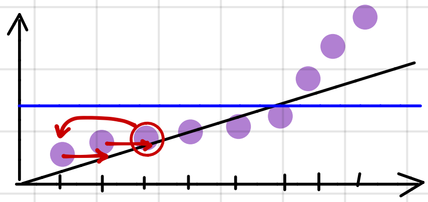
(ii) Moreover if each p-value is uniform (exactly) under the null

$\mathrm{FDR}(R^{BH}, P) = \alpha \dfrac{m_0(P)}{m} \leq \alpha$ [Benjamini and Yekutieli (2001)]

Proof: (i) $\mathrm{FDR}(R^{BH}, P) = \sum_{j=1}^{m} (1-\theta_j)\, \mathbb{E}\left[\frac{\mathbb{1}\{p_j \leq \alpha(\hat{\ell} \vee 1)/m\}}{\hat{\ell} \vee 1} \right]$ [Ferreira and Zwinderman (2006)]

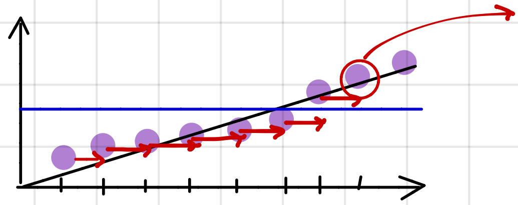
we have $p_j \leq \frac{\alpha}{m}(\hat{\ell} \vee 1) \Leftrightarrow \hat{\ell} = \hat{\ell}^{(-j)} + 1 \Leftrightarrow p_j \leq \frac{\alpha}{m}(\hat{\ell}^{(-j)} + 1)$

where $\hat{\ell}^{(-j)}$ is the number of rejections of SU aplied to the $m-1$ pvalues $(p_{j'},\ j' \neq j)$ and with modified critical values $\left(\frac{\alpha(\ell+1)}{m},\ 1 \leq \ell \leq m-1\right)$



if p_j rejected, the number of rejections is as if p_j were in first position

hence $\hat{\ell} = \hat{\ell}^{(-j)} + 1$



if p_j not rejected some $p_{j'}$ shifted $\hat{\ell}^{(-j)} \geq \hat{\ell}$

hence $\hat{\ell}^{(-j)} + 1 > \hat{\ell}$

Hence
$$\mathrm{FDR}(R^{BH}, P) = \sum_{j=1}^{m} (1-\theta_j)\, \mathbb{E}\left[\frac{\mathbb{1}\{p_j \leq \alpha(\hat{\ell}^{(-j)}+1)/m\}}{\hat{\ell}^{(-j)}+1} \right]$$
indep
$$= \sum_{j=1}^{m} (1-\theta_j)\, \mathbb{E}\left[\frac{1}{\hat{\ell}^{(-j)}+1}\ \underbrace{\mathbb{P}\left[p_j \leq \alpha(\hat{\ell}^{(-j)}+1)/m \;\middle|\; \hat{\ell}^{(-j)} \right]}_{\leq \alpha(\hat{\ell}^{(-j)}+1)/m} \right] \leq \alpha \frac{m_0}{m}$$

(ii) Above, this is an equality if p_j is uniform $\square$

Can we remove the independence assumption?

In fact, no, there is in general an additionnal factor $\gamma_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$

Proposition: consider p_values satisfying (*) in a general multiple testing setting

(i) $\forall P \in \mathcal{P}$, $\mathrm{FDR}(R^{BH}, P) \leq \alpha \frac{m_0(P)}{m} \gamma_m \leq \alpha \gamma_m$ [Benjamini & Yekutieli (2001)]

(ii) if any distribution in $[0,1]^m$ with uniform marginals corresponds to the distribution of the p-value family for some $P_0 \in \mathcal{P}$ with $\theta(P_0)_j = 0$ for all j ← under the 'full null'

$$\sup_{P \in \mathcal{P}} \{\mathrm{FDR}(R^{BH}, P)\} = (\alpha \gamma_m) \wedge 1$$ [Hommel 1986, Metrika] [Lehmann & Romano 2005, AoS]