Multiple testing

① Refresher

- **Setting**:
  - $\{\mathbb{P}_\theta : \theta \in \Theta\}$ a collection of distributions on $\mathcal{X}$
  - **Hypotheses**: for $\Theta_0, \Theta_1 \subset \Theta$, with $\Theta_0 \cap \Theta_1 = \emptyset$, we want to test:

    $$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \Theta_1$$
  - **Data**: we observe $X(\omega)$ a realization of $X \sim \mathbb{P}_\theta$

- **Test**:
  - a random variable $\hat{\varphi} : \Omega \to \{0,1\}$, which is $\sigma(X)$-measurable
  - **Level** $= \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\hat{\varphi} = 1)$ (max probability of false rejection)
  - **Example**: for any $\hat{S} : \Omega \to \mathbb{R}$ which is $\sigma(X)$-measurable

    set
    - $T_\theta(s) = \mathbb{P}_\theta(\hat{S} \geq s)$
    - s_α such that $\sup_{\theta \in \Theta_0} T_\theta(s_\alpha) \leq \alpha$

    Then $\hat{\varphi}_\alpha := \mathbb{1}_{\hat{S} \geq s_\alpha}$ has a level at most α.

- **p-value**:
  - any random variable $\hat{p} : \Omega \to [0,1]$, which is $\sigma(X)$-measurable and fulfills

    $$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\hat{p} \leq u) \leq u, \qquad \forall u \in [0,1]$$

    (under H_0, $\hat{p}$ is stochastically larger than a $\mathcal{U}[0,1]$ variable)
  - **Remark**: $\hat{\varphi}_\alpha := \mathbb{1}_{\hat{p} \leq \alpha}$ is of level α

- Example (continued)

$\hat{p}(\omega) := \sup_{\theta \in \Theta_0} T_\theta(\hat{S}(\omega))$ is a p-value

Proof: (case with density) Set $F_\theta(s) = \mathbb{P}_\theta[\hat{S} \leq s] \geq 1 - T_\theta(s)$ (c.d.f.)

$F_\theta^{-1}(u) = \inf\{s : F_\theta(s) \geq u\}$ (right inverse)

- under $\mathbb{P}_{\theta_0}$: $\hat{S} = F_{\theta_0}^{-1}(U)$ with $U \sim \mathcal{U}[0,1]$

So $\hat{p} \underset{\theta_0 \in \Theta_0}{\geq} T_{\theta_0}(\hat{S}) \geq 1 - F_{\theta_0}(F_{\theta_0}^{-1}(U)) \overset{\text{density}}{=} 1 - U \sim \mathcal{U}[0,1]$

i.e. $\forall \theta_0 \in \Theta_0$: $\mathbb{P}_{\theta_0}[\hat{p} \leq u] \leq u$ $\square$

(2) Multiple testing

- Setting:
  - m collections of distributions: $\{\mathbb{P}_\theta : \theta \in \Theta^{(i)}\}$, $i = 1, \dots, m$
  - m tests: $H_0^{(i)} : \theta \in \Theta_0^{(i)}$ against $H_1^{(i)} : \theta \in \Theta_1^{(i)}$, $i = 1, \dots, m$
    - $\hat{p}_i$ = p-value for test n° i
    - $\hat{p}_{(1)} \leq \hat{p}_{(2)} \leq \dots \leq \hat{p}_{(m)}$: p-values ranked in increasing order.
- Multiple testing procedure:
  - $R : [0,1]^m \longrightarrow \mathcal{P}(\{1, \dots, m\})$

    $(p_1, \dots, p_m) \longrightarrow \underbrace{R(p_1, \dots, p_m)}_{\text{rejected hypotheses}} \subset \{1, \dots, m\}$
  - $\hat{R} := R(\hat{p}_1, \dots, \hat{p}_m)$

Example: $\hat{R} = \{i : \hat{p}_i \leq \alpha\}$.

- True / False Positive :

  - $I_0 = \{ i : H_0^{(i)} \text{ true} \}$, and $m_0 = |I_0|$

  - $TP = \text{Card}(\hat{R} \setminus I_0)$

    $FP = \text{Card}(\hat{R} \cap I_0)$

  - Example: (continued)

    for $\hat{R} = \{ i : \hat{p}_i \leq \alpha \}$, we have

    $\rightarrow$ individually each test $\hat{\varphi}^{(i)} = \mathbb{1}_{\hat{p}_i \leq \alpha}$ has level α

    $\rightarrow$ but

    $$\mathbb{E}[FP] = \sum_{i \in I_0} \mathbb{P}[\hat{p}_i \leq \alpha] \leq m_0 \alpha$$

    equality if $\hat{p}_i \sim \mathcal{U}(0,1)$.

Bonferroni correction : $\hat{R}_{Bonf} := \{ i : \hat{p}_i \leq \frac{\alpha}{m} \}$

Then

$$\mathbb{P}[FP > 0] \leq \mathbb{E}[FP] \leq \sum_{i \in I_0} \sup_{\theta \in \Theta_0^{(i)}} \mathbb{P}_\theta[\hat{p}_i \leq \tfrac{\alpha}{m}] \leq m_0 \frac{\alpha}{m} \leq \alpha$$

:-) FP small

:-(TP also

- Can we imagine a less restrictive criterion ?

?

③ False discovery rate control

- False discovery proportion: $FDP = \dfrac{FP}{TP + FP}$ with $\frac{0}{0} = 0$

- False discovery rate: $FDR = \mathbb{E}[FDP]$

  our goal: construct $\hat{R}_\alpha$ such that $FDR \leq \alpha$ ← chosen in advance

- Informal discussion:
  - focus on $\hat{R} = \{ i : \hat{p}_i \leq t(\hat{p}_1, \dots, \hat{p}_m) \}$ (we keep the smallest p-values)
  - case $t(\hat{p}_1, \dots, \hat{p}_m) = \tau$:

    $\mathbb{E}[FP] \leq m_0 \tau$, so we can hope that

    $FDP = \dfrac{FP}{TP + FR}$ "$\leq$" $\dfrac{m_0 \tau}{\mathrm{Card}(\hat{R})}$ ⚠ nothing rigorous here!

  - case $t(\hat{p}_1, \dots, \hat{p}_m) = \hat{p}_{(\hat{k})}$ with $\hat{k} = \hat{k}(\hat{p}_1, \dots, \hat{p}_m)$

    ← ordered p-values.

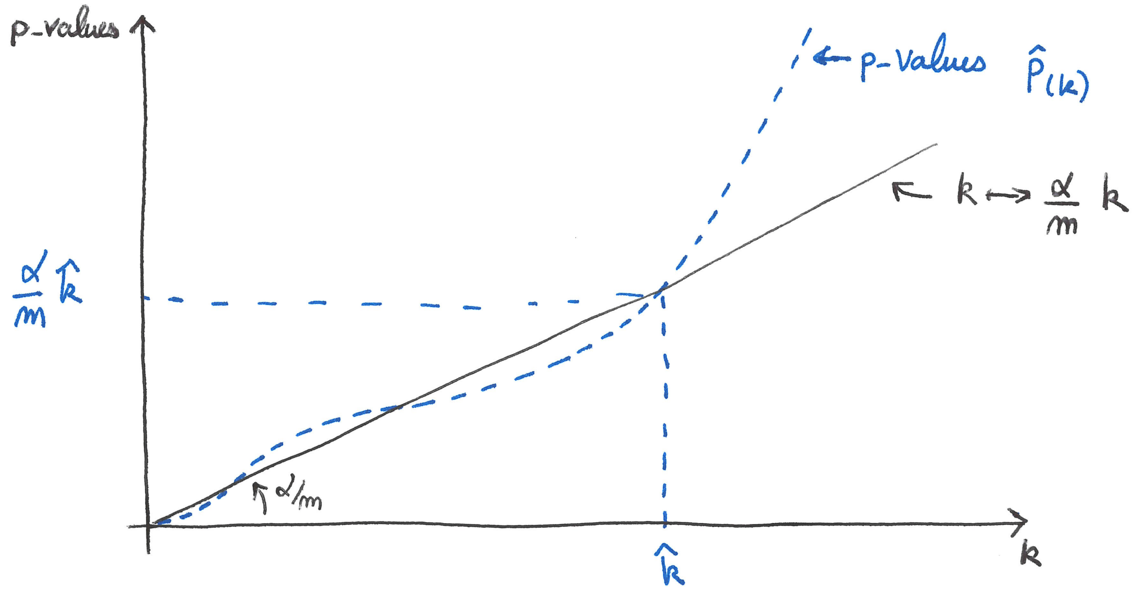
    We can hope that

    $$FDP \text{ "}\leq\text{" } \frac{m_0 \hat{p}_{(\hat{k})}}{\hat{k}} \overset{?}{\leq} \alpha$$

    If: → we choose $\hat{p}_{(\hat{k})} \leq \frac{\alpha}{m_0} \hat{k}$
    → $\hat{k}$ as large as possible

    $\Big\} \Rightarrow \hat{k} = \max \left\{ k : \hat{p}_{(k)} \leq \frac{\alpha}{m} k \right\}$

and $\hat{R} = \{ i : \hat{p}_i \leq \hat{p}_{(\hat{k})} \} = \{ i : \hat{p}_i \leq \frac{\alpha}{m} \hat{k} \}$



Henceforth, we focus on $\hat{R}_\beta$ defined by:

Algorithm:

- $\hat{p}_{(1)} \le \hat{p}_{(2)} \le \cdots \le \hat{p}_{(m)}$: ordered p-values
- $\beta : \{1, \ldots, m\} \to \mathbb{R}^+$, increasing
- $\hat{R}_\beta := \left\{ i : \hat{p}_i \le \frac{\alpha}{m} \beta(\hat{k}) \right\}$ where $\hat{k} = \max\left\{ k : \hat{p}_{(k)} \le \frac{\alpha}{m} \beta(k) \right\}$

Theorem 10.2

$$\mathrm{FDR}(\hat{R}_\beta) \le \alpha \frac{m_0}{m} \sum_{j \ge 1} \frac{\beta(j \wedge m)}{j(j+1)}$$

Proof:

$$\mathrm{FDR} = \mathbb{E}\left[\frac{\mathrm{Card}\left\{ i \in I_0 : \hat{p}_i \le \frac{\alpha}{m} \beta(\hat{k}) \right\}}{\hat{k}} \mathbb{1}_{\hat{k} \ge 1} \right]$$

$$= \sum_{i \in I_0} \mathbb{E}\left[\mathbb{1}_{\hat{p}_i \le \frac{\alpha}{m} \beta(\hat{k})} \underbrace{\frac{\mathbb{1}_{\hat{k} \ge 1}}{\hat{k}}}_{= \sum_{j \ge 1} \frac{1}{j(j+1)} \mathbb{1}_{j \ge \hat{k} \ge 1}} \right]$$

$$\overset{\text{Fubini}}{=} \sum_{i \in I_0} \sum_{j \ge 1} \frac{1}{j(j+1)} \underbrace{\mathbb{P}\left[\hat{p}_i \le \frac{\alpha}{m} \beta(\hat{k}) \text{ and } j \ge \hat{k} \ge 1 \right]}_{\overset{\beta \text{ increasing}}{\le} \mathbb{P}\left[\hat{p}_i \le \frac{\alpha}{m} \beta(j \wedge m) \right]}$$

$$\underset{\text{p-value}}{\le} \alpha \frac{m_0}{m} \sum_{j \ge 1} \frac{\beta(j \wedge m)}{j(j+1)}$$

$\square$

Remark: there exist (pathological) p-values, for which equality holds in Theorem 10.2.

Choice of β? if we want $\beta(j) = \gamma j$, then

$$\sum_{j\geq 1} \frac{\beta(j\wedge m)}{j(j+1)} = \gamma \sum_{j=1}^{m-1} \frac{1}{j+1} + \gamma \underbrace{\sum_{j\geq m} \frac{m}{j(j+1)}}_{=\frac{m}{m}=1}$$

$$= \gamma H_m \text{ where } H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}.$$

To have $FDR \leq \alpha$, we must take $\gamma = \frac{1}{H_m} \underset{m\to\infty}{\sim} \frac{1}{\log m}$.

The choice $\beta(j) = j/H_m$ corresponds to the Benjamini-Yekutieli procedure.

Better than Bonferroni?

$$\hat{t}_{BY} = \frac{\alpha}{m} \times \frac{\hat{k}}{H_m} \overset{?}{\geq} \hat{t}_{Bonf} = \frac{\alpha}{m}$$

only if $\hat{k} \geq H_m$!

If not, it is worst than Bonferroni :(

Can we sometimes use $\beta(j) = j$? ← Benjamini-Hochberg procedure

only under some hypotheses.

Let us try to bound the FDR for $\beta(j)=j$ in a different way.

$$\mathrm{FDR} = \sum_{i\in I_0} \mathbb{E}\left[\mathbb{1}_{\hat{p}_i \le \frac{\alpha \hat{k}}{m}} \frac{\mathbb{1}_{\hat{k}\ge 1}}{\hat{k}} \right]$$

$$= \sum_{i\in I_0} \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\Big[\underbrace{\hat{p}_i \le \frac{\alpha k}{m}}_{\text{easy to handle}} \text{ and } \underbrace{\hat{k}=k}_{\text{tricky}} \Big]$$

$$= \sum_{i\in I_0} \sum_{k=1}^{m} \frac{1}{k} \underbrace{\mathbb{P}\Big[\hat{p}_i \le \frac{\alpha k}{m} \Big]}_{\le \frac{\alpha k}{m}} \mathbb{P}\Big[\hat{k}=k \,\Big|\, \hat{p}_i \le \frac{\alpha k}{m} \Big]$$

$$\le \frac{\alpha}{m} \sum_{i\in I_0} \sum_{k=1}^{m} \Big(\underbrace{\mathbb{P}\Big[\hat{k}\le k \,\Big|\, \hat{p}_i \le \frac{\alpha k}{m} \Big]}_{(*)\ \overset{?}{\le}\ \mathbb{P}\left[\hat{k}\le k \,\middle|\, \hat{p}_i \le \alpha\frac{k+1}{m} \right]} - \mathbb{P}\Big[\hat{k}\le k-1 \,\Big|\, \hat{p}_i \le \frac{\alpha k}{m} \Big] \Big)$$

if (*) holds

$$\le \frac{\alpha}{m} \sum_{i\in I_0} \underbrace{\mathbb{P}\Big[\hat{k}\le m \,\Big|\, \hat{p}_i \le \frac{\alpha(m+1)}{m} \Big]}_{=1} \le \frac{m_0}{m}\alpha \quad :\text{-)}$$

<u>Under which conditions does (*) hold?</u>

$$\{\hat{k}\le k\} = \Big\{ \underbrace{\max\Big\{ j : \hat{p}_{(j)} \le \frac{\alpha j}{m} \Big\}}_{\text{decreases if } \hat{p}_i \text{ increases}} \le k \Big\} \text{ is increasing with } (\hat{p}_1, \dots, \hat{p}_m)$$

So (*) will hold if

Weak Positive Regression Dependency (wPRDS)

The distribution of $(\hat{p}_1, \dots, \hat{p}_m)$ fulfills the wPRDS property

if $\quad \cdot\ \forall g : [0,1]^m \to \mathbb{R}^+$ non-decreasing (coordinate wise)

$\quad\quad \cdot\ \forall i \in I_0$

$$u \to \mathbb{E}[g(\hat{p}_1, \dots, \hat{p}_m) \mid \hat{p}_i \leq u] \text{ is non-decreasing.}$$

Theorem 10.5:

If the distribution of $(\hat{p}_1, \dots, \hat{p}_m)$ fulfills the wPRDS property,

then $\mathrm{FDR}(\hat{R}_{BH}) \leq \alpha \frac{m_0}{m} \leq \alpha$:)

When does wPRDS hold?

- Example 1: $(\hat{p}_i)_{i \in I_0}$ ⫫ and independent from $(\hat{p}_i)_{i \notin I_0}$

Proof: take $i \in I_0$, with no loss of generality $i = 1$. For $g \nearrow$

$$\mathbb{E}[g(\hat{p}_1, \dots, \hat{p}_m) \mid \hat{p}_1 \leq u] \overset{⫫}{=} \int_{x_{-1} \in [0,1]^{m-1}} \mathbb{E}[\underbrace{g(\hat{p}_1, x_{-1})}_{=: g_1(\hat{p}_1) \nearrow} \mid \hat{p}_1 \leq u]\, \mathbb{P}[\hat{p}_{-1} \in dx_{-1}]$$

$$\mathbb{E}[g_1(\hat{p}_1) \mid \hat{p}_1 \leq u] = \int_0^{+\infty} \mathbb{P}[g_1(\hat{p}_1) \geq t \mid \hat{p}_1 \leq u]\, dt$$

$$= \int_0^{\infty} \underbrace{\mathbb{P}[\hat{p}_1 \overset{\text{or} >}{\geq} g_1^{-1}(t) \mid \hat{p}_1 \leq u]}_{} \, dt$$

$$= \left(1 - \frac{\mathbb{P}[\hat{p}_1 \overset{\text{or} \leq}{<} g_1^{-1}(t)]}{\mathbb{P}[\hat{p}_1 \leq u]}\right)_+ \nearrow \text{ with } u \qquad \square$$

- Example 2 : . assume that $(\hat{S}_1, \dots, \hat{S}_m) \sim \mathcal{N}(\mu, \Sigma)$
with $\Sigma_{ij} \geq 0 \quad \forall i,j = 1, \dots, m$.
  - Set $\hat{P}_i = T_i(\hat{S}_i)$ where $T_i(s) = \mathbb{P}[\mathcal{N}(0, \Sigma_{ii}) \geq s]$

Then, $(\hat{P}_1, \dots, \hat{P}_m)$ fulfills wPRDS

proof: exercise 10.6.3 $\square$

These test statistics are useful to test $\mu_i = 0$ against $\mu_i > 0$.

③ Take Home Message

- Keeping FP low induces a strong loss in power
- controlling FDR can be less conservative, but
  - → Benjamini - Yekutieli can be more conservative than Bonferroni
  - → Benjamini - Hochberg is less conservative, but FDR control only under wPRDS, which is hard to check in practice (from data)
- Benjamini - Hochberg is frequently used in practice.
- controlling FDR in coordinate sparse regression is a challenging problem.

  ⇝ look at Exercise 5.5.9 on Slope estimator