Tests multiples : généralités, problème du weighting optimal

Guillermo Durand

PhD under the supervision of E. Roquain and P. Neuvial
LPMA, Université Pierre et Marie Curie, France

24/10/2017

Table of contents

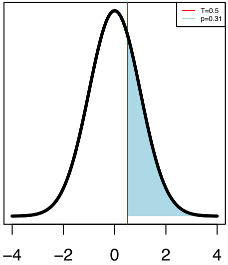
Table of contents

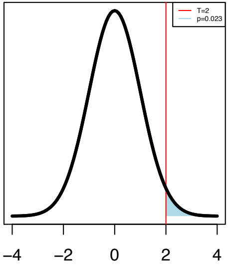
Simple testing

- Data: $X = (X_1, \ldots X_n)$ i.i.d.$\sim \mathcal{N}(\mu, 1)$, $\mu \geq 0$ unknown
- Question: is $\mu = 0$ (no signal) or > 0 (signal) ?
- Null hypothesis H_0: "$\mu = 0$" versus alternative H_1: "$\mu > 0$"
- Test statistic: $T(X) = n^{-1/2} \sum_{i=1}^{n} X_i$, under H_0 $T(X) \sim \mathcal{N}(0, 1)$
- $T(X)$ in the tail of $\mathcal{N}(0, 1) \Rightarrow$ unrealistic $\Rightarrow$ reject H_0

Test of level α

- Reject H_0 if $T(X) \geq z_{1-\alpha}$ the $1-\alpha$ quantile of $\mathcal{N}(0,1)$
- $\Leftrightarrow$ if the p-value $p(X) = \bar{\Phi}\left(T(X)\right)$ is $\leq \alpha$





Why "level α" ?

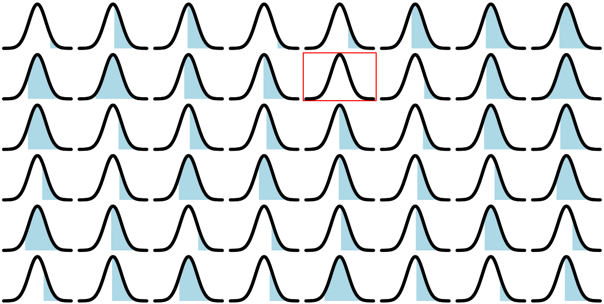
Chance to wrongly reject = make a false positive :

$$\mathbb{P}_{H_0}\left(p(X) \leq \alpha\right) = \mathbb{P}_{H_0}\left(T(X) \geq z_{1-\alpha}\right) \leq \alpha$$

$\Rightarrow$ Control of type I error

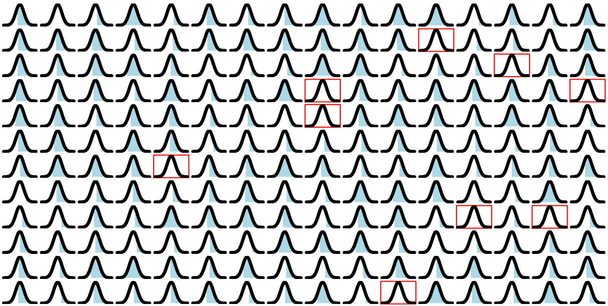
Multiple testing

- Now each X_i is a vector $(X_{i1}, \ldots, X_{im}) \sim \mathcal{N}(\mu, \mathrm{Id}_m)$ with $\mu = (\mu_1, \ldots, \mu_m) \in \mathbb{R}_+^m$
- m null hypotheses H_{0j}: "$\mu_j = 0$" versus H_{1j}: "$\mu_j > 0$"
- At least one false positive with $\mathbb{P} = 1 - (1 - \alpha)^{m_0} \underset{m_0 \to \infty}{\longrightarrow} 1$
- Example if $m = m_0 = 48$, $\alpha = 0.05$:



Multiple testing

- False positives explosion with m
- $m = m_0 = 192$, $\alpha = 0.05$:



Modern applications

- Gene expression, GWAS, fMRI...
- $m = 10^4, 10^5, 10^6$
- Too many false positives without correction

Table of contents

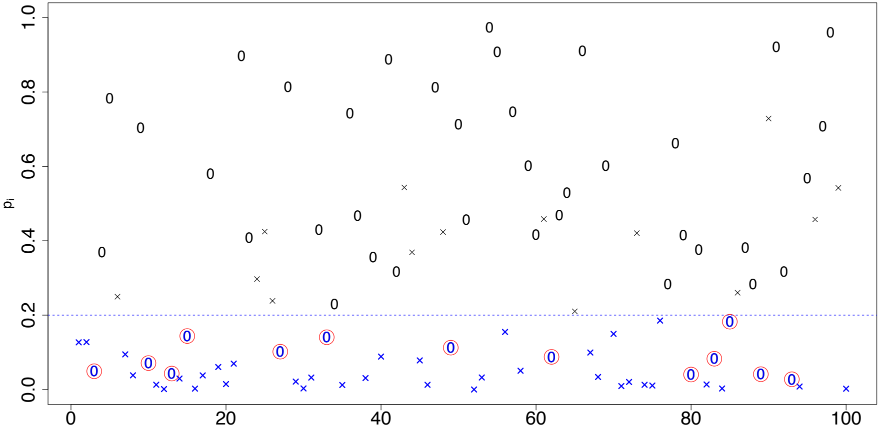
Formal setting

- Data $X \in (\mathcal{X}, \mathfrak{X})$ with $X \sim P \in \mathcal{P}$ a collection of distributions, P unknown
- m null hypotheses $H_{0,i}$ which are subsets of $\mathcal{P}$
- Label vector $\theta(P) \in \{0,1\}^m$ with $\theta_i = 0 \Leftrightarrow P \in H_{0,i}$
- m p-values $p_i = p_i(X)$ such that $p_i \succeq \mathcal{U}([0,1])$ if $\theta_i = 0$ (stochastic dominance)
  - Each p_i provides an α level test : $\mathbb{P}_{\theta_i=0}(p_i \leq \alpha) \leq \alpha$

Rejection set

Thresholding

- $R(\hat{t}) = \{i : p_i \leq \hat{t}\}$



Thresholding

Ordering p-values

- $R = \{i : p_i \leq \hat{t}\}$

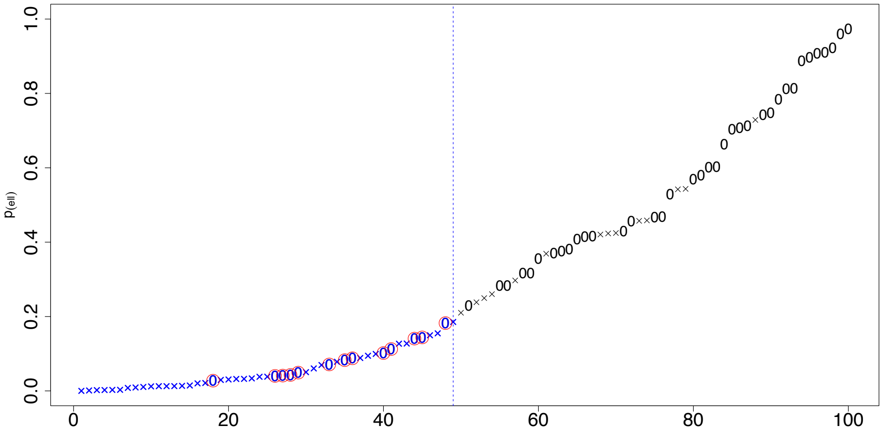


Table of contents

Family-Wise Error Rate (FWER)

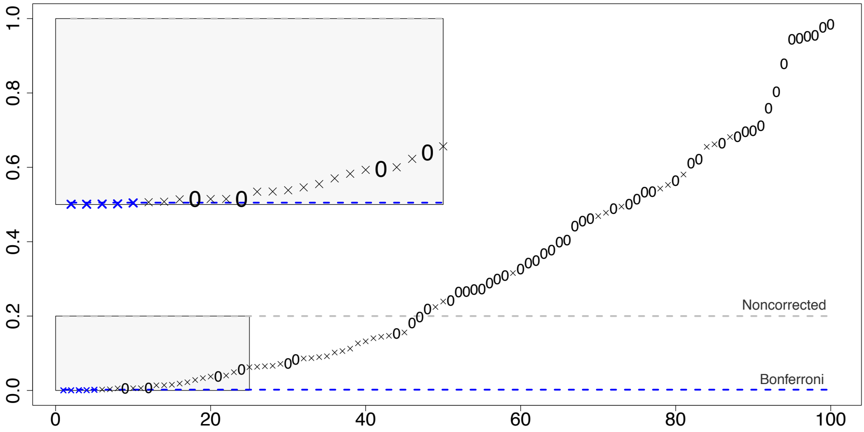
- Probability to make at least one false positive
- Let $V(R) = \sum_{i=1}^{m}(1-\theta_i)\mathbb{1}_{\{i \in R\}}$,

$$\text{FWER}(R) = \mathbb{P}(V(R) > 0)$$
$$\text{FWER}(R(\hat{t})) = \mathbb{P}(\exists i, \theta_i = 0 : p_i \leq \hat{t})$$

- Philosophy : we don't want any false positive
- Choose $\hat{t}_\alpha$ such that $\text{FWER}(R(\hat{t}_\alpha)) \leq \alpha$
- Bonferroni method: $\hat{t}_\alpha = \frac{\alpha}{m}$ is valid (union bound)
- Variant: $k\text{-FWER}(R) = \mathbb{P}(V(R) \geq k)$

Illustration of Bonferroni method

$\alpha = 0.2, m = 100$



False Discovery Rate (FDR)

FWER too stringent

Especially for exploratory experiment where:

- m is large
- we want a lot of detections,
- so we can allow some false positive

False Discovery Proportion(FDP) and FDR

$$\mathrm{FDP}(R) = \frac{V(R)}{|R| \vee 1}$$

$$\mathrm{FDR}(R) = \mathbb{E}\left[\mathrm{FDP}(R)\right]$$

- Choose a $\hat{t}_\alpha$ such that $\mathrm{FDR}(R(\hat{t}_\alpha)) \leq \alpha$?

Benjamini-Hochberg method (BH)

[Benjamini and Hochberg (1995)]

- Sort p-values: $p_{(1)} \leq \cdots \leq p_{(m)}$
- Let $\hat{k} = \max\{k \in [\![1, m]\!] : p_{(k)} \leq \alpha \frac{k}{m}\}$ or 0 if empty set
- Choose $\hat{t}_\alpha = \hat{t}_{\text{BH}} = \alpha \frac{\hat{k}}{m}$

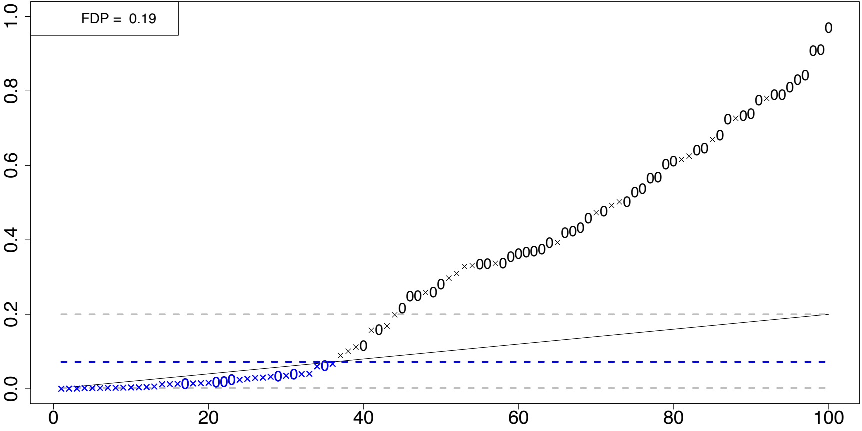
Theorem [Benjamini and Hochberg (1995)]

If the p_i are independent, then

$$\text{FDR}\left(R\left(\hat{t}_{\text{BH}}\right)\right) \leq \alpha$$

Illustration of BH method

$\alpha = 0.2, m = 100$



Proof of FDR control

Lemma

- Let i such that $p_i \leq \alpha \frac{\hat{k}}{m}$ and apply BH to all p-values except p_i, changing the critical constants in the following way:
- Let $\hat{k}^{-i} = \max\{k \in [\![1, m-1]\!] : p^{-i}_{(k)} \leq \alpha \frac{k+1}{m}\}$ or 0 if empty set
- Then $\hat{k}^{-i} = \hat{k} - 1$

Proof

- For $\hat{k} \leq k \leq m-1$, $p^{-i}_{(k)} = p_{(k+1)} > \alpha \frac{k+1}{m} \Rightarrow \hat{k}^{-i} \leq \hat{k} - 1$
- $p^{-i}_{(\hat{k}-1)} = p_{(\hat{k}-1)}$ or $p_{(\hat{k})} \leq \frac{\hat{k}-1+1}{m} \Rightarrow \hat{k}^{-i} \geq \hat{k} - 1$

Proof of FDR control

$$
\begin{aligned}
\mathrm{FDR}(\mathrm{BH}) &= \mathbb{E}\left[\frac{\sum_{i:\theta_i=0} \mathbb{1}_{\{p_i \le \alpha \hat{k}/m\}}}{\hat{k} \vee 1}\right] \\
&= \sum_{i:\theta_i=0} \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\left(p_i \le \alpha k/m, \hat{k} = k\right) \\
&\le \sum_{i:\theta_i=0} \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\left(p_i \le \alpha k/m, \hat{k}^{-i} = k-1\right) \text{ (lemma)} \\
&\le \sum_{i:\theta_i=0} \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\left(p_i \le \alpha k/m\right) \mathbb{P}\left(\hat{k}^{-i} = k-1\right) \text{ (independence)} \\
&\le \frac{\alpha}{m} \sum_{i:\theta_i=0} \sum_{k=1}^{m} \mathbb{P}\left(\hat{k}^{-i} = k-1\right) = \alpha \frac{m_0}{m} \le \alpha.
\end{aligned}
$$

Table of contents

Motivation

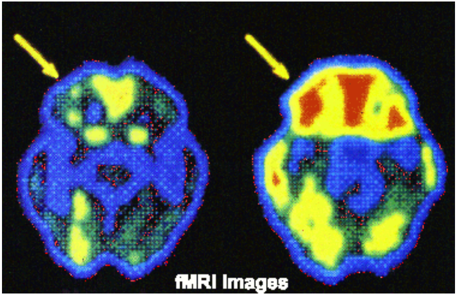
Grouped hypotheses

Context

Test hypotheses that have a group structure

Example :

- fMRI studies. 1 voxel = 1 p-value. Different groups in the brain where they have different distribution.



Deal with groups
Weighting

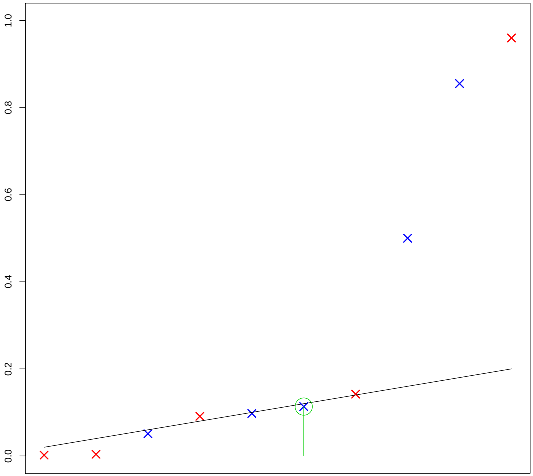
- By weighting the type I error rate [Benjamini and Hochberg (1997)]
- By weighting the p-values :
  - For FWER control [Holm (1979)]
  - For FDR control [Genovese, Roeder, et al. (2006)], [Blanchard and Roquain (2008)], [Hu, Zhao, et al. (2010)], [Zhao and Zhang (2014)], [Ignatiadis, Klaus, et al. (2016)]
- Search for optimal power [Roquain and Wiel (2009)]

The well-known BH procedure

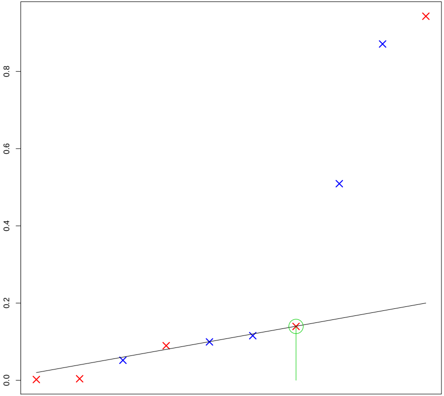
- Sort p-values $p_{(1)} \leq \cdots \leq p_{(m)}$
- Let $\hat{k} = \max\{k : p_{(k)} \leq \alpha k/m\}$
- Reject all $p_i \leq \alpha \frac{\hat{k}}{m}$
- FDR control at level $\pi_0 \alpha$ when wPRDS [Benjamini and Yekutieli (2001)]

$\mathrm{FDR}(R) = \mathbb{E}[\mathrm{FDP}(R)]$ and $\mathrm{FDP}(R) = \frac{|R \cap \mathcal{H}_0|}{|R|}$ with $\mathcal{H}_0$ the true nulls

The well-known BH procedure



An example of weighted BH



- Weights can increase detections $\Longrightarrow$ increase power ?

In this talk

- A generalization of IHW [Ignatiadis, Klaus, et al. (2016)]
- Asymptotic FDR control and power optimality
- Also a stabilization variant
- Numerical illustrations

Table of contents

Back to BH

- Order p-values : $p_{(1)} \leq \cdots \leq p_{(m)}$
- Let $\hat{k} = \max\{k : p_{(k)} \leq \alpha k/m\}$
- Reject all $p_i \leq \alpha \frac{\hat{k}}{m}$

Useful other formulation

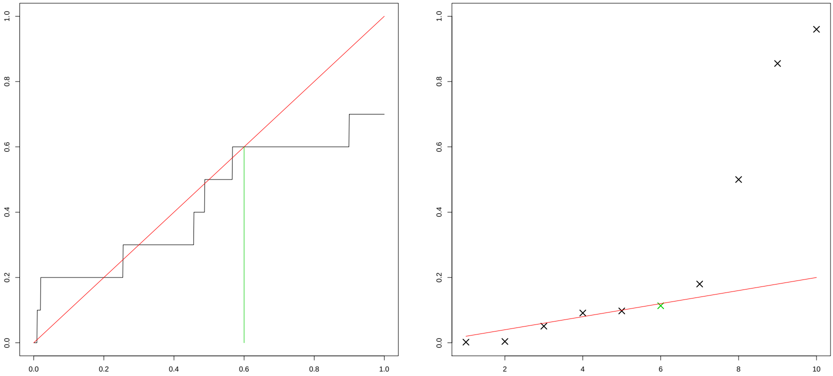
$\frac{\hat{k}}{m} := \max\left\{u : \widehat{G}(u) \geq u\right\} := \mathcal{I}\left(\widehat{G}\right)$ where

$$\widehat{G} : u \mapsto m^{-1} \sum_{i=1}^{m} \mathbb{1}_{\{p_i \leq \alpha u\}}, u \in [0,1]$$

- The only formulation that makes sense asymptotically

An illustration of $\mathcal{I}(\widehat{G})$

Last crossing point between $\widehat{G}$ and the identity



Weighted BH (WBH)

Take some weights $(w_i)_{1\le i\le m}, w_i \ge 0$, form

$$\widehat{G}_w : u \mapsto m^{-1} \sum_{i=1}^{m} \mathbb{1}_{\{p_i \le \alpha u w_i\}}$$

and reject all $p_i \le \alpha \hat{u} w_i$ with $\hat{u} = \mathcal{I}\left(\widehat{G}_w\right)$.

- Choose the right weight space for FDR control, such as

$$\left\{ w : \sum_i w_i \le m \right\}.$$

- BH is a WBH procedure with $w_i = 1\ \forall i$.

Multi-Weighting

A generalization to non-linear weight functions [Roquain and Wiel (2009)]

Take a weight function $u \mapsto W(u)$ such that

$$\widehat{G}_W : u \mapsto m^{-1} \sum_{i=1}^{m} \mathbb{1}_{\{p_i \leq \alpha u W_i(u)\}} \text{ is nondecreasing,}$$

then $\mathrm{MWBH}\,(W) = \{i : p_i \leq \alpha \hat{u} W_i\,(\hat{u})\}$ with $\hat{u} = \mathcal{I}\left(\widehat{G}_W\right)$.

A WBH(w) is a MWBH with constant weight function $u \mapsto w$.

Optimal weighting

- Consider the procedure $R_{u,w}$ rejecting p_i if $p_i \leq \alpha u w_i$
- Maximize its power for all u on a given weight space $\mathcal{W}$:

Definition of optimal weights [Roquain and Wiel (2009)]

$$W^*_{or}(u) = \arg\max_{w \in \mathcal{W}} \mathrm{Pow}\left(R_{u,w}\right)$$

Power definition

$$\mathrm{Pow}\left(R\right) = m^{-1}\mathbb{E}\left[\left|R \cap \mathcal{H}_1\right|\right]$$

- Differs from the usual definition with m_1 by a multiplicative factor
- Depends on the F_i's, the c.d.f. under $\mathcal{H}_1$

Optimal weighting

Existence, uniqueness and asymptotics

- Assume regularity properties of the F_i (concavity), fulfilled in the gaussian 1-sided framework
- $\mathcal{W} = \{w : \sum_i w_i \leq m\}$

Theorem [Roquain and Wiel (2009)]

We have existence, uniqueness of W^*_{or} (and other nice properties).

Theorem [Roquain and Wiel (2009)]

Moreover, MWBH (W^*_{or}) asymptotically enjoys FDR control at level $\pi_0\alpha$ and power optimality among all WBH procedures.

Optimal weighting
Motivation

- F_i unknown under the alternative ! So is W^*_{or}
- $\{w : \sum_i w_i \leq m\} \Longrightarrow \pi_0\alpha$-FDR control $\Longrightarrow$ conservativeness

Goal

- Estimate the oracle optimal weights
- enlarge $\mathcal{W}$ in a way that incorporates π_0 estimation
- obtain asymptotical results on FDR control and power optimality

$\Longrightarrow$ Adaptive Data-Driven Optimal Weighting (ADDOW)

Table of contents

ADDOW

Model

- G groups of sizes m_g, hypotheses $H_{g,i}$, p-values $p_{g,i}$
- $p_{g,i} | H_{g,i} = 0 \sim \mathcal{U}([0,1])$
- $p_{g,i} | H_{g,i} = 1 \sim F_g$ with F_g **strictly concave**
- p-values follow *weak dependency* [Storey, Taylor, et al. (2004)]
- $\alpha >$ some critical α^* (0 in Gaussian 1-sided) [Chi (2007)], [Neuvial (2013)]

ADDOW

π_0 estimation

- G estimators $\hat{\pi}_{g,0}$ such that $\hat{\pi}_{g,0} \xrightarrow{\mathbb{P}} \tilde{\pi}_{g,0} \geq \pi_{g,0}$
- Non Estimation (NE) : $\hat{\pi}_{g,0} = \tilde{\pi}_{g,0} = 1 \ \forall g$
- Evenly Estimation (EE) : $\tilde{\pi}_{g,0} = \tilde{\pi}_0 \ \forall g$
- Consistent Estimation (CE) : $\tilde{\pi}_{g,0} = \pi_{g,0} \ \forall g$
- $\mathcal{W} = \left\{ w : \sum_g \frac{m_g}{m} \hat{\pi}_{g,0} w_g \leq 1 \right\}$ allows larger weights !

ADDOW

Definition

$$\text{ADDOW} = \text{MWBH}\left(\widehat{W}^*\right)$$

where

$$\forall u,\ \widehat{W}^*(u) = \arg\max_{w \in \mathcal{W}} \widehat{G}_w(u)$$

$\Longrightarrow$ $\widehat{W}^*$ maximizes the *rejections*

Key idea

Under (CE) or (ED)+(EE), maximize the rejections is the same as maximizing the power

- (ED) : Evenly Distribution, $\pi_{g,0} = \pi_0\ \forall g$

Remark : under (NE), ADDOW=IHW [Ignatiadis, Klaus, et al. (2016)]

sADDOW$_\beta$

Stabilization for weak signal

- ADDOW overfits so FDR control lost with weak signal in finite sample
- We should prefer BH then
- $\Longrightarrow$ test if there is signal before choosing the procedure, like KS tests

Definition

$$\text{sADDOW}_\beta = \begin{cases} \text{ADDOW} & \text{if } \phi_\beta = \mathbb{1}_{\{Z_m > q_{\beta,m}\}} = 1 \\ \text{BH} & \text{if } \phi_\beta = \mathbb{1}_{\{Z_m > q_{\beta,m}\}} = 0 \end{cases}$$

with $Z_m = \sqrt{m} \sup_{u\in[0,1]} \left(\widehat{G}_{\widehat{W}^*}(u) - \alpha u \right)$ and $q_{\beta,m}$ the $(1-\beta)$ quantile of Z_{0m} (independent copy of Z_m under full null, (NE), and independence).

Table of contents

Asymptotic FDR control

Theorem

$$\lim_{m\to\infty} \text{FDR}\,(\text{ADDOW}) \leq \alpha$$

Moreover if $\alpha \leq \tilde{\pi}_0$:

$$\lim_{m\to\infty} \text{FDR}\,(\text{ADDOW}) = \frac{\pi_0}{\tilde{\pi}_0}\alpha \text{ if (ED)+(EE)}, = \alpha \text{ if (CE)}$$

Proofs inspired by [Roquain and Wiel (2009)], [Hu, Zhao, et al. (2010)] and [Zhao and Zhang (2014)].

Corollary

In (ED) case :

$$\lim_{m\to\infty} \text{FDR}\,(\text{IHW}) = \pi_0\alpha.$$

Power optimality

Theorem

In (CE) or (ED)+(EE) case,

$$\lim_{m\to\infty} \text{Pow}\,(\text{ADDOW}) \geq \limsup_{m\to\infty} \text{Pow}\left(\text{MWBH}\left(\widehat{W}\right)\right)$$

for any weight function such that $\sum_g \frac{m_g}{m} \hat{\pi}_{g,0} \widehat{W}_g(u) \leq 1$.

Corollary

In (ED) case,

$$\lim_{m\to\infty} \text{Pow}\,(\text{IHW}) \geq \limsup_{m\to\infty} \text{Pow}\left(\text{MWBH}\left(\widehat{W}\right)\right)$$

for any weight function such that $\sum_g \frac{m_g}{m} \widehat{W}_g(u) \leq 1$.

sADDOW$_\beta$ equivalent to ADDOW

Theorem

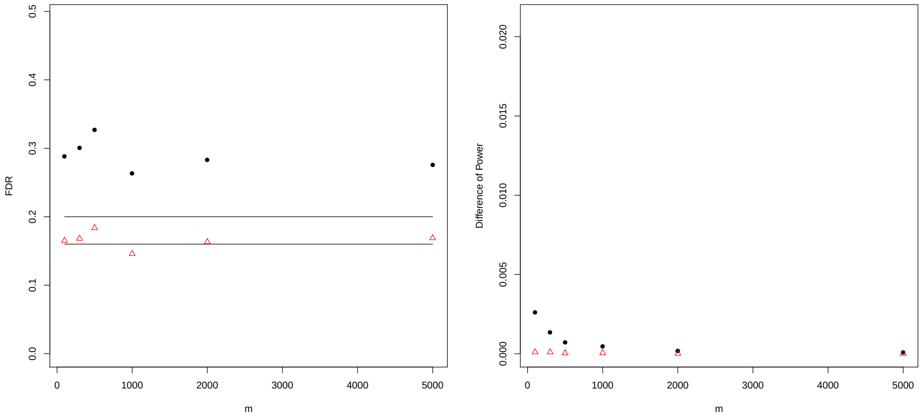
sADDOW$_\beta$ is asymptotically equivalent to ADDOW because $\phi_\beta \xrightarrow{a.s.} 1$ when $m \to \infty$, even if $\beta = \beta_m \to 0$ not too slowly ($\beta_m \geq \exp\left(-m^{1-\nu}\right), \nu > 0$).

Proof relies on the DKWM inequality [Massart (1990)]

Table of contents

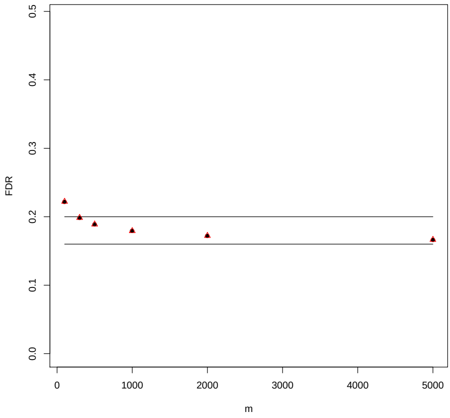
Stabilization for weak signal : $\bar{\mu} = 0.01$

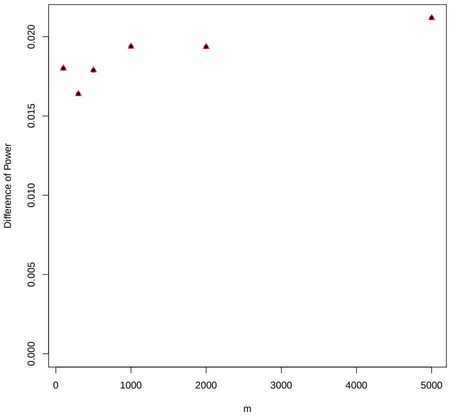
$\pi_1 = \pi_2 = 0.5$, $\pi_0 = 0.8$, $\mu_1 = \bar{\mu}$, $\mu_2 = 2\bar{\mu}$, 1000 replications



Stabilization for strong signal : $\bar{\mu} = 3$

$\pi_1 = \pi_2 = 0.5$, $\mu_1 = \bar{\mu}$, $\mu_2 = 2\bar{\mu}$, 1000 replications



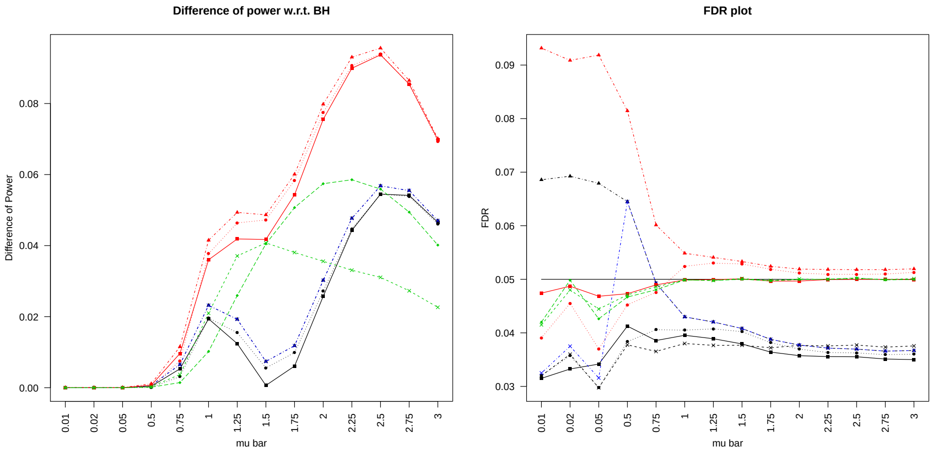


Comparison with other methods

$\alpha = 0.05$, $\pi_{1,0} = 0.7$, $\pi_{2,0} = 0.8$, $m_1 = m_2 = 1500$, $\beta = 0.025$, $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$, 2000 replications

- ADDOW, the oracle and ZZ in (NE) and (CE) cases [Roquain and Wiel (2009)], [Zhao and Zhang (2014)]
- Varying signal $\bar{\mu}$
- Also ABH and HZZ : only adaptation to π_0 [Hu, Zhao, et al. (2010)]
- sADDOW$_\beta$ in (NE) case

Comparison with other methods

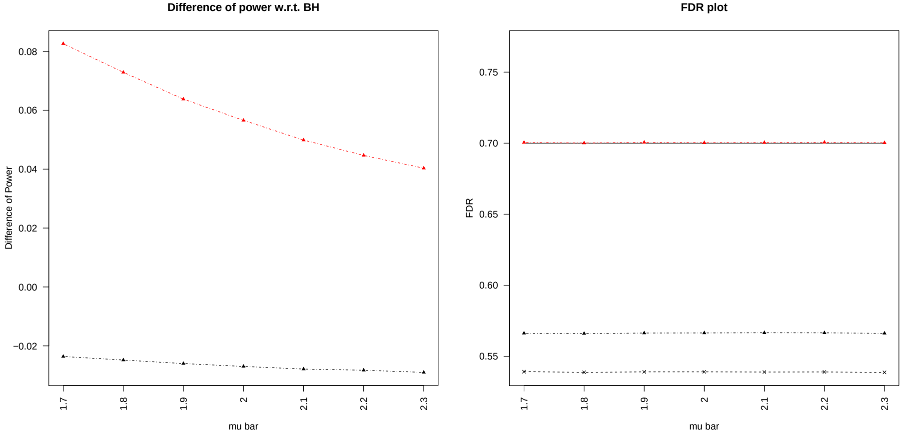


- Overfitting in both cases
- Benefit of π_0-adaptation
- α level FDR control ?

BH better than IHW

Possible outside of (ED) case !

$\alpha = 0.7$, $\pi_{1,0} = 0.05$, $\pi_{2,0} = 0.85$, $m_1 = 1000$, $m_2 = 9000$, $\mu_1 = 2$ and $\mu_2 = \bar{\mu}$, 1000 replications



Conclusion

- Optimal asymptotical properties but with strong assumptions and possible overfitting
- Positive dependence ?
- Use a better estimator of the rejections than $\widehat{G}_w$?
- FDR bound in finite sample ?
- Convergence speed ? With more regularity assumptions on F_g ?

Preprint available: arXiv:1710.01094

Bibliography I

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the royal statistical society. Series B (Methodological)*, pp. 289–300.

— (1997). "Multiple hypotheses testing with weights". In: *Scandinavian Journal of Statistics* 24.3, pp. 407–418.

Benjamini, Yoav and Daniel Yekutieli (2001). "The control of the false discovery rate in multiple testing under dependency". In: *Annals of statistics*, pp. 1165–1188.

Blanchard, Gilles and Etienne Roquain (2008). "Two simple sufficient conditions for FDR control". In: *Electronic journal of Statistics* 2, pp. 963–992.

Chi, Zhiyi (2007). "On the performance of FDR control: constraints and a partial solution". In: *The Annals of Statistics* 35.4, pp. 1409–1431.

Genovese, Christopher R., Kathryn Roeder, and Larry Wasserman (2006). "False discovery control with p-value weighting". In: *Biometrika*, pp. 509–524.

Holm, Sture (1979). "A simple sequentially rejective multiple test procedure". In: *Scandinavian journal of statistics*, pp. 65–70.

Hu, James X., Hongyu Zhao, and Harrison H. Zhou (2010). "False discovery rate control with groups". In: *Journal of the American Statistical Association* 105.491.

Ignatiadis, Nikolaos et al. (2016). "Data-driven hypothesis weighting increases detection power in genome-scale multiple testing". In: *Nature methods* 13.7, pp. 577–580.

Bibliography II

Massart, Pascal (1990). "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality". In: *The Annals of Probability*, pp. 1269–1283.

Neuvial, Pierre (2013). "Asymptotic results on adaptive false discovery rate controlling procedures based on kernel estimators". In: *The Journal of Machine Learning Research* 14.1, pp. 1423–1459.

Roquain, Etienne and Mark A. van de Wiel (2009). "Optimal weighting for false discovery rate control". In: *Electronic Journal of Statistics* 3, pp. 678–711.

Storey, John D, Jonathan E Taylor, and David Siegmund (2004). "Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66.1, pp. 187–205.

Zhao, Haibing and Jiajia Zhang (2014). "Weighted p-value procedures for controlling FDR of grouped hypotheses". In: *Journal of Statistical Planning and Inference* 151, pp. 90–106.

Multi-Weighting

A practical way to compute $\mathcal{I}\left(\widehat{G}_W\right)$

No need to compute $W(u)$ for each u !

$\forall k \in [\![1, m]\!]$, compute all $\frac{p_i}{W_i(\frac{k}{m})}$ and take q_k the k-th smallest.
Let $q_0 = 0$.
Then $\mathcal{I}\left(\widehat{G}_W\right) = m^{-1} \max\{k \in [\![0, m]\!] : q_k \leq \alpha \frac{k}{m}\}$.

Stabilization variant

Main idea (under independence)

Weak signal $\Longrightarrow Z_m$ close to Z_{0m} in distribution, and

$$\begin{aligned}
\mathrm{FDR}\left(\mathrm{sADDOW}_\beta\right) &= \mathbb{E}\left[\phi_\beta \,\mathrm{FDP}\left(\mathrm{ADDOW}\right) + (1-\phi_\beta)\,\mathrm{FDP}\left(\mathrm{BH}\right)\right] \\
&\leq \mathbb{E}\left[\phi_\beta + \mathrm{FDP}\left(\mathrm{BH}\right)\right] \\
&\leq \mathbb{P}\left(Z_m > q_{\beta,m}\right) + \frac{m_0}{m}\alpha \\
&\lesssim \mathbb{P}\left(Z_{0m} > q_{\beta,m}\right) + \frac{m_0}{m}\alpha \\
&\leq \beta + \frac{m_0}{m}\alpha
\end{aligned}$$

About the computation of $\widehat{W}^*$

Key ideas

- Compute only $\widehat{W}^*(u)$ for $u = \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m}, 1$
- Fixing u, $w \mapsto \widehat{G}_w(u)$ only jumps at the $\frac{p_{g,i}}{\alpha u} \Longrightarrow$ let $\widehat{W}_g^*(u) = \frac{p_{g,i_g}}{\alpha u}$ such that $\sum m_g \frac{p_{g,i_g}}{\alpha u} \leq m$ and $\max \sum_g i_g$
- $\widehat{G}_w(u)$ nondecreasing in u AND w : try to reject 1 hyp, then 2, then 3... for $u = \frac{1}{m}$, when fail at k hyp, try to reject k hyp for $u = \frac{2}{m}$, ...