Optimal data-driven weighting procedure with grouped hypotheses and π_0-adaptation

G. Durand guillermo.durand@upmc.fr

Laboratoire de Probabilités, Statistique & Modélisation, UMR 8001 Sorbonne Université, Paris, France.

arXiv 1710.01094

Introduction

Weighting the p-values is a common strategy to improve the power of FDR controlling multiple testing procedures, see e.g. [2]. Later, π_0-adaption has been combined with weighting to gain more power [4]. However, finding an optimal procedure among all weighting strategies has only been addressed in [6], with only an oracle, non π_0-adaptive procedure. Recent works [5, 8] introduce data-driven weighting procedures for grouped hypotheses but do not solve completely the problem of optimality.

Here we present ADDOW (Adaptive Data-Driven Optimal Weighting), a new method that improves previous approaches. While ADDOW satisfies asymptotical FDR control, it satisfies a form of optimality by maximizing asymptotical power among all weighting procedures. The superiority of ADDOW is illustrated via numerical experiments.

Grouped hypotheses model

We have:

- G fixed groups of size of hypotheses $(H_{g,1}, H_{g,2}, \dots)$, $1 \le g \le G$, to test,
- Corresponding p-values $(p_{g,1}, p_{g,2}, \dots)$, $1 \le g \le G$,
- $p_{g,i} \sim \mathcal{U}([0,1])$ if $H_{g,i} = 0$ (true nulls),
- $p_{g,i} \sim F_g$ **strictly concave** if $H_{g,i} = 1$: alternatives of the **same groups** are **identically distributed**.

Asymptotic setting

- At each m, the groups have size m_g, $1 \le g \le G$, where $\sum_{g=1}^{G} m_g = m$ and $\frac{m_g}{m} \to \pi_g > 0$.
- $m_{g,1} = \sum_{i=1}^{m_g} H_{g,i}$ and $m_{g,0} = m_g - m_{g,1}$ are such that $\frac{m_{g,0}}{m_g} \to \pi_{g,0} > 0$ and $\frac{m_{g,1}}{m_g} \to \pi_{g,1} > 0$.
- Weak dependence [7] in each group: $\frac{1}{m_{g,0}} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \le t, H_{g,i}=0\}} \xrightarrow{a.s.} U(t) = t \wedge 1$, $t \ge 0$, $\frac{1}{m_{g,1}} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \le t, H_{g,i}=1\}} \xrightarrow{a.s.} F_g(t)$, $t \ge 0$.

π_0-estimation

Estimate $\pi_{g,0}$ with $\hat{\pi}_{g,0} \le 1$ such that $\hat{\pi}_{g,0} \xrightarrow{\mathbb{P}} \bar{\pi}_{g,0} \ge \pi_{g,0}$, like the Storey estimator [7]:

$$\hat{\pi}_{g,0}(\lambda) = \frac{1 - \frac{1}{m_g}\sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \le \lambda\}} + \frac{1}{m}}{1 - \lambda}, \quad \lambda \in (0,1).$$

Criticality: $\alpha > \alpha^*$ the critical alpha level (see [1]) depending on the $\bar{\pi}_{g,0}$ and the $f_g(0^+)$.

Leading example: the Gaussian one-sided framework where the p-values are derived from a test statistic $X_{g,i}$ that follows $\mathcal{N}(0,1)$ if $H_{g,i} = 0$ and $\mathcal{N}(\mu_g, 1)$, $\mu_g > 0$, if $H_{g,i} = 1$. Letting $p_{g,i} = \bar{\Phi}(X_{g,i})$ we get $F_g(\cdot) = \bar{\Phi}\left(\bar{\Phi}^{-1}(\cdot) - \mu_g\right)$ which is strictly convex, $\alpha^* = 0$ and **consistency** of Storey estimators if:

- $\lambda = \lambda_m \to 1$ slow enough,
- the $X_{g,i}$ are mutually independent.

From BH to multi-weighting

Let

$$\widehat{G} : u \mapsto m^{-1} \sum_{g=1}^{G} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \le \alpha u\}},$$

and $\hat{u} = \max\{u \in [0,1], \widehat{G}(u) \ge u\}$, then the BH procedure rejects all $p_{g,i} \le \alpha \hat{u}$, see Figure 1.

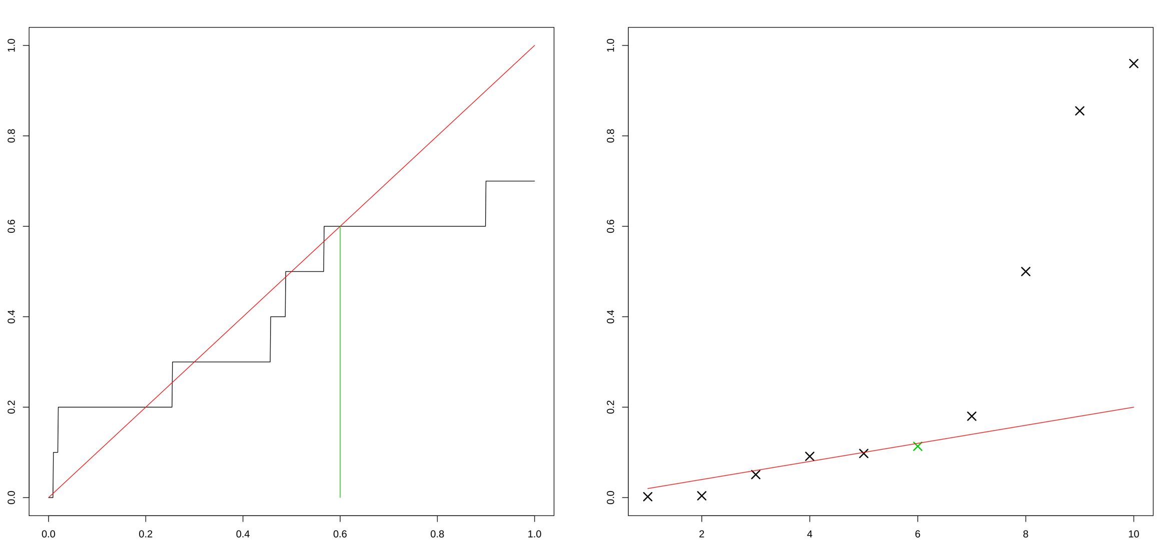

Figure 1: The BH procedure applied to a set of 10 p-values. Right plot: the p-values and the function $k \to \alpha k/m$. Left plot: identity function and $\widehat{G}$. Each plot shows that 6 p-values are rejected.

Following [6], we generalize BH into a multi-weighted BH (MWBH) procedure by introducing a weight function $W : [0,1] \to \mathbb{R}_+^G$, which can be random, such that the following:

$$\widehat{G}_W : u \mapsto m^{-1} \sum_{g=1}^{G} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \le \alpha u W_g(u)\}},$$

is nondecreasing. The $\mathrm{MWBH}(W)$ procedure rejects all $p_{g,i} \le \alpha \hat{u}_W W_g(\hat{u}_W)$, where $\hat{u}_W = \max\{u \in [0,1], \widehat{G}_W(u) \ge u\}$.

ADDOW

ADDOW is $\mathrm{MWBH}(\widehat{W}^*)$ where $\widehat{W}^*$ is an adaptive data-driven optimal weight function:

$$\forall u \in [0,1], \widehat{W}^*(u) \in \operatorname*{arg\,max}_{w \in K^m} \widehat{G}_w(u), \quad K^m = \left\{ w \in \mathbb{R}_+^G : \sum_g \frac{m_g}{m} \hat{\pi}_{g,0} w_g \le 1 \right\}.$$

MAIN IDEA: MAXIMIZE REJECTIONS ON A WELL-CHOSEN WEIGHT SPACE

Remark 1. ADDOW depends on the $\hat{\pi}_{g,0}$ which makes it a class of procedure. If $\hat{\pi}_{g,0} = 1$ we recover IHW.

Remark 2. ADDOW can be generalized by using the LCM of the e.c.d.f. instead.

Main results

Let the following assumption:

$$\exists C \ge 1, \forall g, \bar{\pi}_{g,0} = C \pi_{g,0}, \tag{1}$$

which includes the consistent case ($C = 1$).

Theorem 1 (Asymptotic FDR control)**.**

$$\lim_{m \to \infty} \mathrm{FDR}(\mathrm{ADDOW}) \le \alpha,$$

and, under (1)*,*

$$\lim_{m \to \infty} \mathrm{FDR}(\mathrm{ADDOW}) = \frac{\alpha}{C}.$$

Theorem 2 (Asymptotic Power optimality)**.** *Under* (1)*, for any sequence of random weight functions* $(\widehat{W})_{m \ge 1}$*, such that* $\widehat{W} : [0,1] \to K^m$ *and* $\widehat{G}_{\widehat{W}}$ *is nondecreasing,*

$$\lim_{m \to \infty} \mathrm{Pow}(\mathrm{ADDOW}) \ge \limsup_{m \to \infty} \mathrm{Pow}\left(\mathrm{MWBH}\left(\widehat{W}\right)\right).$$

Corollary 1 (IHW)**.** *Assume that* $\pi_{g,0}$ *do not depend on* g*:* $\pi_{g,0} = \pi_0$, $\forall g$*. Then,*

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

and for any sequence of random weight functions $(\widehat{W})_{m \ge 1}$ *such that* $\widehat{W} : [0,1] \to K^m_{NE}$ *and* $\widehat{G}_{\widehat{W}}$ *is nondecreasing, we have*

$$\lim_{m \to \infty} \mathrm{Pow}(\mathrm{IHW}) \ge \limsup_{m \to \infty} \mathrm{Pow}\left(\mathrm{MWBH}\left(\widehat{W}\right)\right).$$

Numerical experiments

2 groups in one-sided Gaussian framework, $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$, $m_1 = m_2 = 2000$, $m_{1,0}/m_1 = 0.7$ and $m_{2,0}/m_2 = 0.8$. No π_0-adaptation in Group 1: $\hat{\pi}_{g,0} = 1$. Oracle π_0-adaptation in Groups 2 and 3: $\hat{\pi}_{g,0} = \pi_{g,0}$.

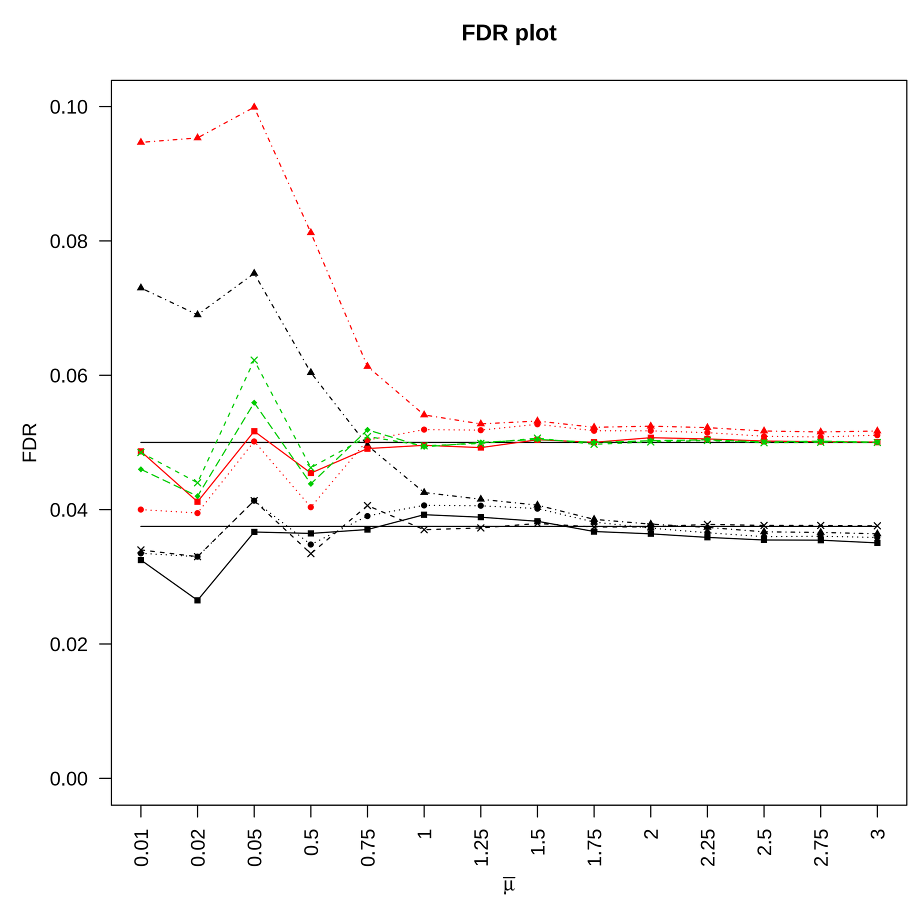

Figure 2: FDR against $\bar{\mu}$. Group 1 in black; Group 2 in green; Group 3 in red. The type of procedure is $\mathrm{MWBH}(W^*_{or})$ (squares); ADDOW (triangles); Pro2 (disks); HZZ (diamonds) and finally BH/ABH (crosses). Horizontal lines: α and $\pi_0\alpha$ levels.

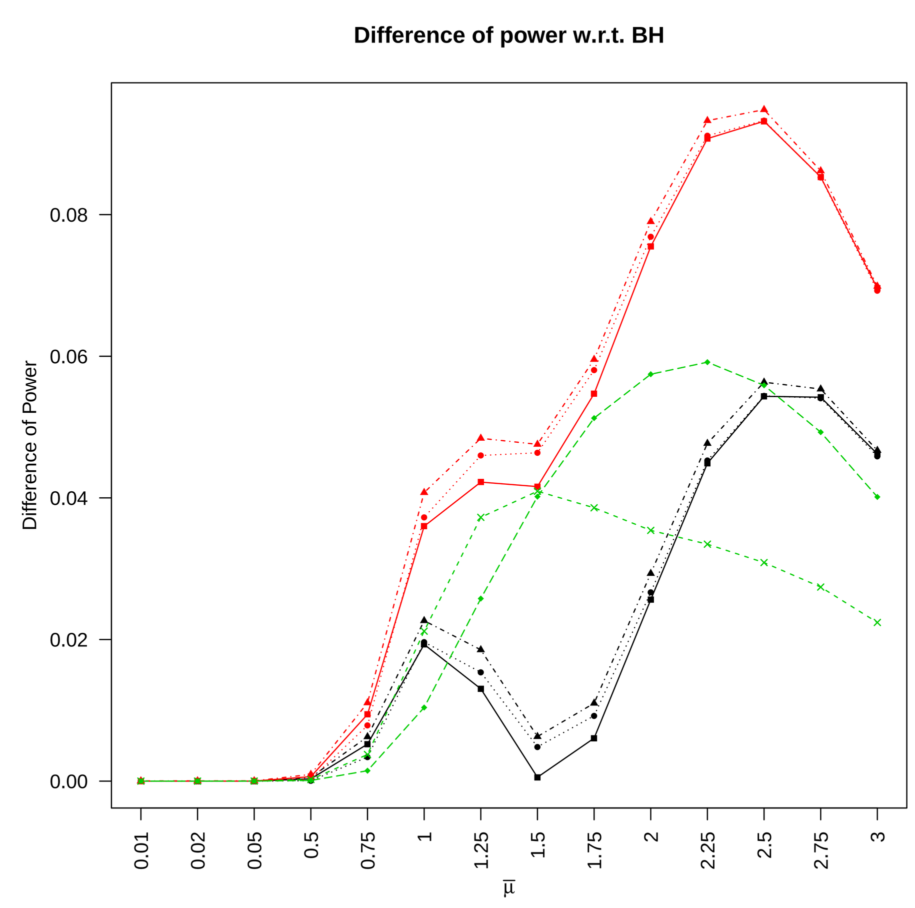

Figure 3: $\mathrm{Pow}(\cdot) - \mathrm{Pow}(\mathrm{BH})$ against $\bar{\mu}$. Same legend as Figure 2

References

[1] Z. Chi. On the performance of fdr control: constraints and a partial solution. *The Annals of Statistics*, 35(4):1409–1431, 2007.

[2] C. R. Genovese, K. Roeder, and L. Wasserman. False discovery control with p-value weighting. *Biometrika*, pages 509–524, 2006.

[3] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[4] J. X. Hu, H. Zhao, and H. H. Zhou. False discovery rate control with groups. *Journal of the American Statistical Association*, 105(491), September 2010.

[5] N. Ignatiadis, B. Klaus, J. B. Zaugg, and W. Huber. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature methods*, 13(7):577–580, 2016.

[6] E. Roquain and M. A. Van De Wiel. Optimal weighting for false discovery rate control. *Electronic Journal of Statistics*, 3:678–711, 2009.

[7] J. D. Storey, J. E. Taylor, and D. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):187–205, 2004.

[8] H. Zhao and J. Zhang. Weighted p-value procedures for controlling fdr of grouped hypotheses. *Journal of Statistical Planning and Inference*, 151:90–106, 2014.