



Adaptive p-value weighting with power optimality

Guillermo Durand

► To cite this version:

| Guillermo Durand. Adaptive p-value weighting with power optimality. 2017. hal-01596474v1

HAL Id: hal-01596474

<https://hal.science/hal-01596474v1>

Preprint submitted on 2 Oct 2017 (v1), last revised 12 Mar 2019 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptive p -value weighting with power optimality

Guillermo Durand

*Laboratoire de Probabilités et Modèles Aléatoires,
Université Pierre et Marie Curie,
4 place Jussieu, 75252 Paris Cedex 05
e-mail: guillermo.durand@upmc.fr*

Abstract: Weighting the p -values is a well-established strategy that improves the power of multiple testing procedures while dealing with heterogeneous data. However, how to achieve this task in an optimal way is rarely considered in the literature. This paper contributes to fill the gap in the case of group-structured null hypotheses, by introducing a new class of procedures named ADDOW (for Adaptive Data Driven Optimal Weighting) that adapts both to the alternative distribution and to the proportion of true null hypotheses. We prove the asymptotical FDR control and power optimality among all weighted procedures of ADDOW, which shows that it dominates all existing procedures in that framework. Some numerical experiments show that the proposed method preserves its optimal properties in the finite sample setting when the number of tests is moderately large.

MSC 2010 subject classifications: Primary 62J15; secondary 62G10.

Keywords and phrases: multiple testing, FDR, weighting, grouped hypotheses, adaptivity, optimality.

Contents

| | | |
|-----|--|----|
| 1 | Introduction | 2 |
| 2 | Setting | 3 |
| 2.1 | Model | 3 |
| 2.2 | $\pi_{g,0}$ estimation | 4 |
| 2.3 | Criticality | 5 |
| 2.4 | Leading example | 5 |
| 2.5 | Criteria | 5 |
| 3 | Weighting | 6 |
| 3.1 | Weighting the BH procedure | 6 |
| 3.2 | Choosing the weights | 7 |
| 4 | New procedures | 9 |
| 4.1 | ADDOW definition | 9 |
| 4.2 | Stabilization for weak signal | 9 |
| 5 | Results | 10 |
| 5.1 | Main results | 10 |
| 5.2 | Relation to IHW | 11 |
| 5.3 | Comparison to other existing procedures | 11 |
| 5.4 | Results for the stabilized version | 11 |
| 6 | Numerical experiments | 12 |
| 6.1 | Simulation setting | 12 |
| 6.2 | FDR control | 12 |
| 6.3 | Power | 14 |
| 7 | Concluding remarks | 15 |
| 8 | Proofs of Theorems 5.1 and 5.2 | 16 |
| 8.1 | Further generalization | 16 |
| 8.2 | Proof of Theorem 5.1 | 17 |
| 8.3 | Proof of Theorem 5.2 | 18 |
| | Acknowledgments | 19 |
| | References | 19 |
| A | Lemmas and proofs of Sections 2, 3 and 4 | 20 |

| | | |
|----------|---|-----------|
| B | Asymptotical weighting | 23 |
| C | Convergence lemmas | 27 |
| D | Proof of Corollary 5.3 for Pro 1 | 31 |
| E | Proof of Theorem 5.3 | 32 |

1. Introduction

Recent high-throughput technologies bring to the statistical community new type of data being increasingly large, heterogeneous and complex. Addressing significance in such context is particularly challenging because of the number of questions that could naturally come up. A popular statistical method is to adjust for multiplicity by controlling the False Discovery Rate (FDR), which is defined as the expected proportion of errors among the items declared as significant. Once the amount of possible false discoveries is controlled, the question of increasing the power, that is the amount of true discoveries, arises naturally. In the literature, it is well-known that the power can be increased by clustering the null hypotheses into homogeneous groups. The latter can be derived in several ways:

- sample size: a first example is the well-studied data set of the Adequate Yearly Progress (AYP) study ([Rogosa, 2005](#)), which compares the results in mathematics tests between socioeconomically advantaged and disadvantaged students in Californian high school. As studied by [Cai and Sun \(2009\)](#), ignoring the sizes of the schools tends to favor large schools among the detections, simply because large schools have more students and not because the effect is stronger. By grouping the schools in small, medium, and large schools, more rejections are allowed among the small schools, which increases the overall detection capability. This phenomenon also appears in more large-scale studies, as in GWAS (Genome-Wide Association Studies) by grouping hypotheses according to allelic frequencies ([Sun et al., 2006](#)) or in microarrays experiments by grouping the genes according to the DNA copy number status ([Roquain and Van De Wiel, 2009](#)).
- spatial structure: some data sets naturally involve a spatial (or temporal) structure into groups. A typical example is neuroimaging: in [Schwartzman, Dougherty and Taylor \(2005\)](#), a study compares diffusion tensor imaging brain scans on 15443 voxels of 6 normal and 6 dyslexic children. By estimating the densities under the null of the voxels of the front and back halves of the brain, some authors highlight a noteworthy difference which suggests that analysing the data by making two groups of hypotheses seems more appropriate, see [Efron \(2008\)](#) and [Cai and Sun \(2009\)](#).
- hierarchical relation: groups can be derived from previous knowledge on hierarchical structure, like pathways for genetic studies, based for example on known ontologies (see e.g. [The Gene Ontology Consortium \(2000\)](#)). Similarly, in clinical trials, the tests are usually grouped in primary and secondary endpoints, see [Dmitrienko, Offen and Westfall \(2003\)](#).

In these examples, while ignoring the group structure can lead to overly conservative procedures, this knowledge can easily be incorporated by using weights. This method can be traced back to [Holm \(1979\)](#) who presented a sequentially rejective Bonferroni procedure that controls the Family-Wise Error Rate (FWER) and added weights to the p -values. Weights can also be added to the type-I error criterion instead of the p -values, as presented in [Benjamini and Hochberg \(1997\)](#) with the so-called weighted FDR. [Blanchard and Roquain \(2008\)](#) generalized the two approaches by weighting the p -values and the criterion, with a finite positive measure to weight the criterion (see also [Ramdas et al. \(2017\)](#) for recent further generalizations). [Genovese, Roeder and Wasserman \(2006\)](#) introduced the p -value weighted BH procedure (WBH) which has been extensively used afterwards with different choices for the weights. [Roeder et al. \(2006\)](#); [Roeder and Wasserman \(2009\)](#) have built the weights upon genomic linkage, to favor regions of the genome with strong linkage. [Hu, Zhao and Zhou \(2010\)](#) calibrated the weights by estimating the proportion of true nulls inside each group (procedure named HZZ here). [Zhao and Zhang \(2014\)](#) went one step further by improving HZZ and BH with weights that maximize the number of rejections at a threshold

computed from HZZ and BH. They proposed two procedures Pro1 and Pro2 shown to control the FDR asymptotically and to have a better power than BH and HZZ.

However, the problem of finding optimal weights (in the sense of achieving maximal averaged number of rejected false nulls) has been only scarcely considered in the literature. For FWER control and Gaussian test statistics, Wasserman and Roeder (2006) designed oracle and data-driven optimal weights, while Dobriban et al. (2015) considered a Gaussian prior on the signal. For FDR control, Roquain and Van De Wiel (2009) designed oracle optimal weights by using the knowledge of the distribution under the alternative of the hypotheses. Unfortunately, this knowledge is not reachable in practice. This leads to the natural idea of estimating the oracle optimal weights by maximizing the number of rejections. This idea has been followed by Ignatiadis et al. (2016) with a procedure called IHW. While they proved that IHW controls the FDR, its power properties have not been considered. In particular, it is unclear whether maximizing the overall number of rejections is appropriate in order to maximize power.

In this paper, we present a general solution to the problem of optimal data-driven weighting of BH procedure in the case of grouped null hypotheses. The new class of procedures is called ADDOW (for Adaptive Data-Driven Optimal Weighting). With mild assumptions, we show that ADDOW asymptotically controls the FDR and has optimal power among all weighted step-up procedures. Interestingly, our study shows that the heterogeneity with respect to the proportion of true nulls should be taken into account in order to attain optimality. This fact seems to have been ignored so far: for instance we show that IHW is optimal when the true nulls are evenly distributed across groups but its performance can quickly deteriorate otherwise.

In Section 2, we present the mathematical model and assumptions. In Section 3, we define what is a weighting step-up procedure. In Section 4, we introduce ADDOW along with a stabilized version, designed to deal with the overfitting problem due to weak signal. Section 5 provides our main theoretical results. Our numerical simulations are presented in Section 6, while we conclude in Section 7 with a discussion. The proofs of the two main theorems are given in Section 8 and more technical results are deferred to appendix. Let us underline that an effort has been made to make the proofs as short and concise as possible, while keeping them as clear as possible.

In all the paper, the probabilistic space is denoted $(\Omega, \mathcal{A}, \mathbb{P})$. The notations $\xrightarrow{a.s.}$ and $\xrightarrow{\mathbb{P}}$ stand for the convergence almost surely and in probability. A "++" symbol is used to indicate that two cases (A) and (B) are simultaneously satisfied: (A)+(B).

2. Setting

2.1. Model

We consider the following stylized grouped p -value modeling: let $G \geq 2$ be the number of groups. In each group $g \in \{1, \dots, G\}$, let $(H_{g,1}, H_{g,2}, \dots)$ be some binary variables corresponding to the null hypotheses to be tested in this group, with $H_{g,i} = 0$ if it is true and $= 1$ otherwise. Consider in addition $(p_{g,1}, p_{g,2}, \dots)$ some random variables in $[0, 1]$ where each $p_{g,i}$ corresponds to the p -value testing $H_{g,i}$.

We make the following marginal distributional assumption for $p_{g,i}$: if $H_{g,i} = 0$, $p_{g,i}$ follows a uniform distribution on $[0, 1]$. We denote by $U : x \mapsto \mathbb{1}_{\{x > 0\}} \times \min(x, 1)$ its cumulative distribution function (c.d.f.). If $H_{g,i} = 1$, $p_{g,i}$ follows a common distribution corresponding to c.d.f. F_g . In particular, note that the p -values are assumed to have the same alternative distribution within each group. We make the mild assumption that F_g is strictly concave on $[0, 1]$ (and thus is also continuous on \mathbb{R} , see Lemma A.1). Furthermore, by concavity, $x \mapsto \frac{F_g(x) - F_g(0)}{x - 0}$ has a right limit in 0 that we denote by $f_g(0^+) \in [0, \infty]$, and $x \mapsto \frac{F_g(x) - F_g(1)}{x - 1}$ has a left limit in 1 that we denote by $f_g(1^-) \in [0, \infty)$.

Above, we considered an infinite set of hypotheses/ p -values because our study will be asymptotic in the number of tests m . At step m , we observe the p -values $p_{g,i}$, $g \leq G$, $i \leq m_g$ where the m_g are non-decreasing integer sequences depending on m and such that $\sum_{g=1}^G m_g = m$. Let us emphasize

that G is kept fixed with m throughout the paper. Note also $m_{g,1} = \sum_{i=1}^{m_g} H_{g,i}$ the number of false nulls and $m_{g,0} = m_g - m_{g,1}$ the number of true nulls in group g . We make the assumption that there exists $\pi_g > 0$ and $\pi_{g,0} > 0$ such that for all g , $m_g/m \rightarrow \pi_g$ and $m_{g,0}/m_g \rightarrow \pi_{g,0}$ when $m \rightarrow \infty$. For each g we also assume that $\pi_{g,1} = 1 - \pi_{g,0} > 0$. These assumptions mean that, asymptotically, no group, and no proportion of signal or sparsity, is vanishing. We denote $\pi_0 = \sum_g \pi_g \pi_{g,0}$ the mean of the $\pi_{g,0}$'s and denote the particular case where the nulls are evenly distributed in each group by (ED):

$$\pi_{g,0} = \pi_0, \quad 1 \leq g \leq G. \quad (\text{ED})$$

Let us now specify assumptions on the joint distribution of the p -values. While we make no assumption on the p -value dependence between two different groups, we assume that the p -values are weakly dependent within each group:

$$\frac{1}{m_{g,0}} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq t, H_{g,i}=0\}} \xrightarrow{a.s.} U(t), \quad t \geq 0, \quad (2.1)$$

and

$$\frac{1}{m_{g,1}} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq t, H_{g,i}=1\}} \xrightarrow{a.s.} F_g(t), \quad t \geq 0. \quad (2.2)$$

This assumption is mild and classical, see Storey, Taylor and Siegmund (2004). Note that weak dependence is trivially achieved if the p -values are independent.

2.2. $\pi_{g,0}$ estimation

For each g , let us assume we have at hand an estimator $\hat{\pi}_{g,0} \in (0, 1]$ of $m_{g,0}/m_g$ and assume that $\hat{\pi}_{g,0} \xrightarrow{\mathbb{P}} \bar{\pi}_{g,0}$ for some $\bar{\pi}_{g,0} \geq \pi_{g,0}$. Let also $\bar{\pi}_0 = \sum_g \pi_g \bar{\pi}_{g,0}$.

In our setting, this assumption can be fulfilled by using the estimators introduced in Storey, Taylor and Siegmund (2004):

$$\hat{\pi}_{g,0}(\lambda) = \frac{1 - \frac{1}{m_g} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \lambda\}} + \frac{1}{m}}{1 - \lambda}, \quad (2.3)$$

for a given parameter $\lambda \in (0, 1)$ let arbitrary (the $\frac{1}{m}$ is here just to ensure $\hat{\pi}_{g,0}(\lambda) > 0$). It is easy to deduce from (2.1) and (2.2) that $\frac{1}{m_g} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \lambda\}} \xrightarrow{a.s.} \pi_{g,0}\lambda + \pi_{g,1}F_g(\lambda)$, which provides our condition:

$$\hat{\pi}_{g,0}(\lambda) \xrightarrow{a.s.} \pi_{g,0} + \pi_{g,1} \frac{1 - F_g(\lambda)}{1 - \lambda} \geq \pi_{g,0}.$$

While $(\bar{\pi}_{g,0})_g$ is let arbitrary in our setting, some particular cases will be of interest in the sequel. First is the Evenly Estimation case (EE) one where

$$\bar{\pi}_{g,0} = \bar{\pi}_0, \quad 1 \leq g \leq G. \quad (\text{EE})$$

In that case, our estimators all share the same limit, and doing so they do not take in account the heterogeneity with respect to the proportion of true nulls. As we will see, (EE) is relevant when the proportion of true nulls is homogeneous across groups, that is, when (ED) holds. A particular subcase of (EE) is the Non Estimation case (NE) where:

$$\hat{\pi}_{g,0} = 1 \text{ which implies } \bar{\pi}_{g,0} = 1, \quad 1 \leq g \leq G. \quad (\text{NE})$$

In the latter, basically, the $\pi_{g,0}$ estimation step is skipped.

Finally, let us introduce the Consistent Estimation case (CE) for which the estimators $\hat{\pi}_{g,0}$ are assumed to be all consistent:

$$\bar{\pi}_{g,0} = \pi_{g,0}, \quad 1 \leq g \leq G. \quad (\text{CE})$$

While this corresponds to a favorable situation, this assumption can be met in classical situations, where $f_g(1^-) = 0$ and $\lambda = \lambda_m$ tends to 1 slowly enough in definition (2.3), see Lemma A.2 in Section A.

The condition $f_g(1^-) = 0$ is called "purity" in the literature. It has been introduced in [Genovese and Wasserman \(2004\)](#) and then deeply studied, along with the convergence of Storey estimators, in [Neuvial \(2013\)](#).

2.3. Criticality

To study asymptotic FDR control and power, it is convenient to focus only in situations where we some rejections are possible (the Power and FDR being converging to 0 otherwise). To this end, [Chi \(2007\)](#) introduced the notion of criticality: they defined some critical alpha level, denoted α^* , for which BH procedure has no asymptotical power if $\alpha < \alpha^*$ (see also [Neuvial \(2013\)](#) for a link between criticality and purity).

We extended this notion of criticality in our heterogeneous setting in Section A (see Definition A.1) and will focus in our results on the supercritical case $\alpha \in (\alpha^*, 1)$. Lemma A.3 states that $\alpha^* < 1$ so such an α always exists.

While the formal definition of α^* is reported to the appendix for the sake of clarity, let us emphasize that it depends on the $(F_g)_g$, $(\pi_g)_g$, $(\pi_{g,0})_g$ and, maybe less intuitively, on the $(\bar{\pi}_{g,0})_g$, which means that the choice of the estimators changes the value of α^* .

2.4. Leading example

While our framework allows a general choice for F_g , a canonical example that we have in mind is the Gaussian one-sided framework where the test statistic in group g follows $\mathcal{N}(0, 1)$ under the null, while it follows $\mathcal{N}(\mu_g, 1)$ under the alternative, for G unknown parameters $\mu_g > 0$.

Classically, this corresponds to consider p -values uniform under the null with alternative c.d.f. given by

$$F_g(\cdot) = \bar{\Phi}(\bar{\Phi}^{-1}(\cdot) - \mu_g),$$

with derivative

$$f_g(\cdot) = \exp\left(\mu_g\left(\bar{\Phi}^{-1}(\cdot) - \frac{\mu_g}{2}\right)\right),$$

where we denoted $\bar{\Phi}(z) = \mathbb{P}(Z \geq z)$ for $Z \sim \mathcal{N}(0, 1)$. Hence F_g is strictly concave and this framework fulfills the assumptions of Section 2.1.

Furthermore we easily check that $f_g(0^+) = \infty$, so $\alpha^* = 0$ and $f_g(1^-) = 0$ which means that this framework is supercritical ($\alpha^* = 0$, see Definition A.1) with purity and then achievable consistent estimation ([CE](#)).

2.5. Criteria

The set of indices corresponding to true nulls is denoted by \mathcal{H}_0 , that is $(g, i) \in \mathcal{H}_0$ if and only if $H_{g,i} = 0$, and we also denote $\mathcal{H}_1 = \mathcal{H}_0^c$.

In this paper, we define a multiple testing procedure R as a set of indices that are rejected: $p_{g,i}$ is rejected if and only if $(g, i) \in R$. The False Discovery Proportion (FDP) of R , denoted by $\text{FDP}(R)$, is defined as the number of false discoveries divided by the number of rejections if there are any, and 0 otherwise:

$$\text{FDP}(R) = \frac{|R \cap \mathcal{H}_0|}{|R| \vee 1}.$$

We denote $\text{FDR}(R) = \mathbb{E}[\text{FDP}(R)]$ the FDR of R . Its power, denoted $\text{Pow}(R)$, is defined as the mean number of true positives divided by m :

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

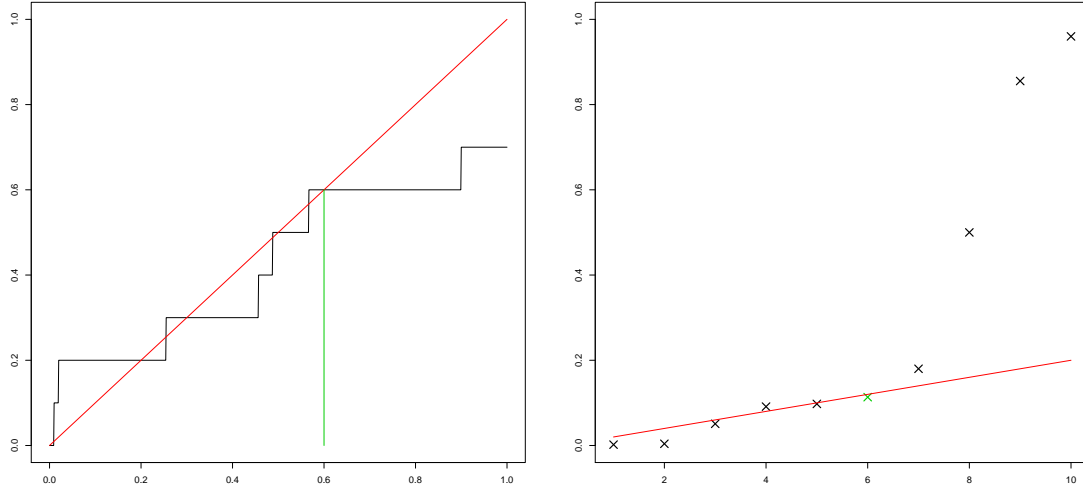


Fig 1: The BH procedure applied to a set of 10 p -values. Right plot: the p -values and the function $k \rightarrow \alpha k/m$. Left plot: identity function and \hat{G} . Each plot shows that 6 p -values are rejected.

Note that our power definition is slightly different than the usual one for which the number of true discoveries is divided by $m_1 = \sum_g m_{g,1}$ instead of m . This simplifies our expressions (see Section 8.1) and does not have any repercussion because the two definitions differ only by a multiplicative factor converging to $1 - \pi_0 \in (0, 1)$ when $m \rightarrow \infty$.

3. Weighting

3.1. Weighting the BH procedure

Say we want to control the FDR at level α . Assume that the p -values are arranged in increasing order $p_{(1)} \leq \dots \leq p_{(m)}$ with $p_{(0)} = 0$, the classic BH procedure consists in rejecting all $p_{g,i} \leq \alpha \frac{\hat{k}}{m}$ where $\hat{k} = \max \{k \geq 0 : p_{(k)} \leq \alpha \frac{k}{m}\}$.

Take a nondecreasing function h defined on $[0, 1]$ such that $h(0) = 0$ and $h(1) \leq 1$, we denote $\mathcal{I}(h) = \sup \{u \in [0, 1] : h(u) \geq u\}$. Some properties of the functional $\mathcal{I}(\cdot)$ are gathered in Lemma A.4, in particular $h(\mathcal{I}(h)) = \mathcal{I}(h)$. We now reformulate BH with the use of $\mathcal{I}(\cdot)$, because it is more convenient when dealing with asymptotics. Doing so, we follow the formalism notably used in Roquain and Van De Wiel (2009) and Neuvial (2013). Define the empirical function

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

then $\hat{k} = m \times \mathcal{I}(\hat{G})$. This is a particular case of Lemma A.5.

The graphical representation of the two points of view for BH is depicted in Figure 1 with $m = 10$. The p -values are plotted on the right part of the figure along with the function $k \mapsto \alpha k/m$ and we see that the last p -value under the line is the sixth one. On the left, the function \hat{G} corresponding to these p -values is displayed alongside the identity function, with the last crossing point being located between the sixth and seventh jumps, thus $\mathcal{I}(\hat{G}) = 6/m$ and 6 p -values are rejected.

The weighted BH (WBH) with weight vector $w \in \mathbb{R}_+^G$ is defined by computing

$$\hat{G}_w : u \mapsto m^{-1} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha u w_g\}}$$

and rejecting all $p_{g,i} \leq \alpha \mathcal{I}(G_w) w_g$. We denote it $\text{WBH}(w)$. Note that w is authorized to be random, hence it can be computed from the p -values. In particular, $\text{BH} = \text{WBH}(\mathbf{1})$ where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}_+^G$.

Following [Roquain and Van De Wiel \(2009\)](#), to deal with optimal weighting, we need to further generalize WBH into a multi-weighted BH (MWBH) procedure by introducing a weight function $W : [0, 1] \rightarrow \mathbb{R}_+^G$, which can be random, such that the following function:

$$\hat{G}_W : u \mapsto m^{-1} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha u W_g(u)\}}, \quad (3.1)$$

is nondecreasing. The resulting procedure rejects all the p -values such that $p_{g,i} \leq \alpha \hat{u}_W W_g(\hat{u}_W)$ and is denoted $\text{MWBH}(W)$ where, for the rest of the paper, we denote

$$\hat{u}_W = \mathcal{I}(\hat{G}_W), \quad (3.2)$$

and name it the step-up threshold. One different weight vector $W(u)$ is associated to each u , hence the "multi"-weighting. Note that the class of MWBH procedures is more general than the one of WBH procedures because any weight vector can be seen as a constant weight function.

Note that, there is a simple way to compute \hat{u}_W . For each r between 1 and m denote the $W(r/m)$ -weighted p -values $p_{g,i}^{[r]} = p_{g,i} / W_g(r/m)$ (with the convention $p_{g,i}/0 = \infty$), order them $p_{(1)}^{[r]} \leq \dots \leq p_{(m)}^{[r]}$ and note $p_{(0)}^{[r]} = 0$. Then $\hat{u}_W = m^{-1} \max \left\{ r \geq 0 : p_{(r)}^{[r]} \leq \alpha \frac{r}{m} \right\}$ (this is Lemma [A.5](#)).

As in previous works, in order to achieve a valid FDR control, these procedures should be used with weights that satisfy some specific relation. Here, we introduce the following weight spaces:

$$K^m = \left\{ w \in \mathbb{R}_+^G : \sum_g \frac{m}{m_g} \hat{\pi}_{g,0} w_g \leq 1 \right\}, \quad (3.3)$$

$$K_{\text{NE}}^m = \left\{ w \in \mathbb{R}_+^G : \sum_g \frac{m}{m_g} w_g \leq 1 \right\}. \quad (3.4)$$

Note that K^m may appear unusual because it depends on the estimators $\hat{\pi}_{g,0}$, however it is completely known and usable in practice. Note also that $K^m = K_{\text{NE}}^m$ in the [\(NE\)](#) case.

Finally, for a weight function W and a threshold $u \in [0, 1]$, we denote by $R_{u,W}$ the double indexed procedure rejecting the p -values less than or equal to $\alpha u W_g(u)$, that is $R_{u,W} = \{(g, i) : p_{g,i} \leq \alpha u W_g(u)\}$. By [\(3.1\)](#), note that $\hat{G}_W(u) = m^{-1} |R_{u,W}|$ and $\text{MWBH}(W) = R_{\hat{u}_W, W}$.

3.2. Choosing the weights

Take W and u , and let $P_W^{(m)}(u) = \text{Pow}(R_{u,W})$. We have

$$\begin{aligned} P_W^{(m)}(u) &= m^{-1} \mathbb{E} \left[\sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha u W_g(u), H_{g,i}=1\}} \right] \\ &= \sum_{g=1}^G \frac{m_{g,1}}{m} F_g(\alpha u W_g(u)). \end{aligned}$$

Note that these relations are valid only if W and u are deterministic. In particular, they are not valid when used a posteriori with a data-driven weighting and $u = \hat{u}_W$.

In [Roquain and Van De Wiel \(2009\)](#), the authors define the oracle optimal weight function W_{or}^* as:

$$W_{or}^*(u) = \arg \max_{w \in K_{NE}^m} P_w^{(m)}(u). \quad (3.5)$$

Note that they defined W_{or}^* only in case (NE), but their definition easily extends to the general case as above. They proved the existence and uniqueness of W_{or}^* in (ED)+(NE) case and that, asymptotically, $\text{MWBH}(W_{or}^*)$ controls the FDR at level $\pi_0\alpha$ and has a better power than every $\text{MWBH}(w^{(m)})$ for $w^{(m)} \in K_{NE}^m$ some deterministic weight vectors satisfying a convergence criterion.

However, computing W_{or}^* requires the knowledge of the F_g , not available in practice, so the idea is to estimate W_{or}^* with a data driven weight function \widehat{W}^* and then apply MWBH with this random weight function. For this, consider the functional defined by, for any (deterministic) weight function W and $u \in [0, 1]$:

$$\begin{aligned} G_W^{(m)}(u) &= \mathbb{E} \left[\widehat{G}_W(u) \right] = \sum_{g=1}^G \left(\frac{m_{g,0}}{m} U(\alpha u W_g(u)) + \frac{m_{g,1}}{m} F_g(\alpha u W_g(u)) \right) \\ &= P_W^{(m)}(u) + \sum_{g=1}^G \frac{m_{g,0}}{m} U(\alpha u W_g(u)). \end{aligned} \quad (3.6)$$

$G_W^{(m)}(u)$ is the mean ratio of rejections for the procedure rejecting each $p_{g,i} \leq \alpha u W_g(u)$. The intuitive idea is that maximizing $G_W^{(m)}(u)$ is close to maximizing $P_W^{(m)}(u)$. We justify this heuristic as follows: assuming U is the identity function, then the right term of (3.6) becomes $\alpha u \sum_g \frac{m_{g,0}}{m} W_g(u)$ and it does not depend of the weights if additionally $\sum_g \frac{m_{g,0}}{m} W_g(u) = 1$, which makes $P_W^{(m)}(u)$ the only term depending on W . Now, we can evaluate the constraint on W by estimating $\frac{m_{g,0}}{m} = \frac{m_g}{m} \frac{m_{g,0}}{m_g}$ by $\frac{m_g}{m} \hat{\pi}_{g,0}$ (which leads to the weight space K^m defined in equation (3.3)), and $G_w^{(m)}(u)$ can be easily estimated by the (unbiased) estimator $\widehat{G}_w(u)$. As a result, maximizing the latter in w should lead to good weights, not too far from $W_{or}^*(u)$.

[Zhao and Zhang \(2014\)](#) followed this heuristic by applying a two-stage approach to derive two procedures, named Pro1 and Pro2. Precisely, in the first stage they use the weight vectors $\hat{w}^{(1)} = (\frac{1}{\hat{\pi}_0}, \dots, \frac{1}{\hat{\pi}_0})$, where $\hat{\pi}_0 = \sum_g \frac{m_g}{m} \hat{\pi}_{g,0}$, and $\hat{w}^{(2)}$ defined by $\hat{w}_g^{(2)} = \frac{\hat{\pi}_{g,1}}{\hat{\pi}_{g,0}(1-\hat{\pi}_0)}$, where $\hat{\pi}_{g,1} = 1 - \hat{\pi}_{g,0}$, and let $\hat{u}_M = \max(\hat{u}_{\hat{w}^{(1)}}, \hat{u}_{\hat{w}^{(2)}})$. In the second stage, they maximize $\widehat{G}_w(\hat{u}_M)$ over K^m , which gives rise to the weight vector $\widehat{W}^*(\hat{u}_M)$ according to our notation. Then they define their procedures as the following:

$$\text{Pro 1} = R_{\hat{u}_M, \widehat{W}^*(\hat{u}_M)},$$

and

$$\text{Pro 2} = \text{WBH} \left(\widehat{W}^*(\hat{u}_M) \right).$$

The caveat of this approach is that the initial thresholding, that is the definition of \hat{u}_M , seems somewhat arbitrary, which will result in sub-optimal procedures, see Corollary 5.3. As a side remark, $\hat{w}^{(1)}$ and $\hat{w}^{(2)}$ are involved in other procedures of the literature. The HZZ procedure of [Hu, Zhao and Zhou \(2010\)](#) is $\text{WBH}(\hat{w}^{(2)})$, and $\text{WBH}(\hat{w}^{(1)})$ is the classical Adaptive BH procedure (see e.g. Lemma 2 of [Storey, Taylor and Siegmund \(2004\)](#)) denoted here as ABH.

[Ignatiadis et al. \(2016\)](#) actually used the above heuristic with multi-weighting (while their formulation differs from ours) which consists in maximizing $\widehat{G}_w(u)$ in w for each u . However, their choice of the weight space is only suitable for the case (NE) and can make the above heuristic break down, because in general the right term in (3.6) can still depend on w , see remark 3.1.

In the next section, we take the best of the two approaches to attain power optimality with data-driven weighting. Let us already mention that the crucial point is Lemma B.3, that fully justifies the heuristic (in cases (CE) and (ED)+(EE)).

Remark 3.1. We can compute numerical examples where BH has asymptotic power larger than IHW. For example, if we break (ED) by taking a small $\pi_{1,0}$ (almost pure signal) and a large $\pi_{2,0}$ (sparse signal), along with a small group and a large one (π_1 much smaller than π_2) and strong signal in both groups, IHW slightly favors group 2 whereas the oracle optimal favors group 1. BH does not favor any group thus a larger power than IHW. This example is simulated in Section 6.3 (see Figure 5).

4. New procedures

4.1. ADDOW definition

We exploit previous intuition and propose to estimate the oracle optimal weights W_{or}^* by maximizing in $w \in K^m$ the empirical counterpart to $G_w^{(m)}(u)$, that is $\hat{G}_w(u)$.

Definition 4.1. We call an adaptive data-driven optimal weight function a random function $\widehat{W}^* : [0, 1] \rightarrow K^m$ such that for all $u \in [0, 1]$:

$$\widehat{G}_{\widehat{W}^*}(u) = \max_{w \in K^m} \widehat{G}_w(u).$$

Such function always exists because $\{\widehat{G}_w(u), w \in K^m\} \subset \{\frac{k}{m}, k \in \llbracket 0, m \rrbracket\}$ is finite, but may not be unique. So in all the following, we take a certain \widehat{W}^* , and our results do not depend on the choice of \widehat{W}^* . An important fact is that $\widehat{G}_{\widehat{W}^*}$ is nondecreasing (see Lemma A.6) so $\widehat{u}_{\widehat{W}^*}$ exists and the corresponding MWBH procedure is well-defined:

Definition 4.2. The ADDOW procedure is the MWBH procedure using \widehat{W}^* as the weight function, that is, $\text{ADDOW} = \text{MWBH}(\widehat{W}^*)$.

One shall note that ADDOW is in fact a class of procedures depending on the estimators $\hat{\pi}_{g,0}$ through K^m . Note that, in the (NE) case, ADDOW reduces to IHW.

Remark 4.1. It turns out that ADDOW reduces to a WBH procedure. It comes from part 2 of the proof of Theorem 5.2 and Remark 8.2. Moreover, to every MWBH procedure, corresponds a WBH procedure with power higher or equal.

4.2. Stabilization for weak signal

Since ADDOW uses the data both through the p-values and the weights, this will result in a slight increase of the FDR, as we will see in the simulations (Section 6.2). This effect is close in spirit to the well known overfitting phenomenon. In our setting where the signal is strong enough, this drawback is proved to vanish when m is large enough, see the simulations and Theorem 5.1. However, the latter is not true for weak signal: if the data are close to be random noise, making the weight optimization can lead ADDOW to find signal only by chance, that is, to make false positives. To circumvent this concern, we propose to stabilize ADDOW by using a pre-testing phase close in spirit to the Kolmogorov-Smirnov (KS) test (Kolmogorov, 1933) to determine whether the signal is weak or not and then to apply ADDOW only if the signal is large enough (and just apply BH otherwise).

Formally, we reject the hypothesis that the signal is weak if $Z_m > q_{\beta,m}$, where

$$Z_m = \sqrt{m} \sup_{w \in K_{NE}^m} \sup_{u \in [0,1]} \left(\widehat{G}_w(u) - \alpha u \right),$$

and $q_{\beta,m}$ is the $(1 - \beta)$ -quantile of the distribution of Z_{0m} , where Z_{0m} is defined as

$$Z_{0m} = \sqrt{m} \sup_{u \in [0,1]} \left(m^{-1} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{U_{g,i} \leq \alpha u \widehat{W}_g^*(u)\}} - \alpha u \right), \quad (4.1)$$

where the $U_{g,i}$ are uniform variables over $[0, 1]$ with, for all g , $U_{g,1}, \dots, U_{g,m_g}$ independent, and

$$\widetilde{W}^*(u) \in \arg \max_{w \in K_{NE}^m} m^{-1} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{U_{g,i} \leq \alpha u w_g\}}.$$

Z_{0m} can be viewed as a copy of Z_m but under the full null model where the p -values are uniform on $[0, 1]$ and independent, and without estimating $\pi_{g,0}$. We denote the test rejecting the weak signal scenario by $\phi_\beta = \mathbb{1}_{\{Z_m > q_{\beta,m}\}}$. This gives us a stabilization procedure depending on β that we call sADDOW_β :

$$\text{sADDOW}_\beta = \begin{cases} \text{ADDOW} & \text{if } \phi_\beta = 1 \\ \text{BH} & \text{if } \phi_\beta = 0 \end{cases} \quad (4.2)$$

We expect that in the weak signal case, the stabilized procedures have better control of the FDR than ADDOW, because in that case, without estimating $\pi_{g,0}$ and if the p -values are all independent, the distribution of Z_m is close to the distribution of Z_{0m} , and we have the following approximation:

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

where $\mathbb{P}(Z_{0m} > q_{\beta,m}) \leq \beta$ by definition of $q_{\beta,m}$ and $m_0 = \sum_g m_{g,0}$ is the number of true nulls. If β is chosen small the control at level α should be achieved. This heuristic will be supported by the simulations in Section 6.2.

5. Results

5.1. Main results

Now we present the two main Theorems of this paper. The two are asymptotical and justify the use of ADDOW when m is large. The first is the control of the FDR at level at most α . The second shows that ADDOW has maximum power over all MWBH procedures in the (CE) case, and also in (ED)+(EE) case. The two are proven in Section 8.

Theorem 5.1. *Let us consider the framework defined in Sections 2.1 and 2.2. We have*

$$\lim_{m \rightarrow \infty} \text{FDR}(\text{ADDOW}) \leq \alpha. \quad (5.1)$$

If $\alpha \leq \bar{\pi}_0$, we have two more accurate results: in (CE) case,

$$\lim_{m \rightarrow \infty} \text{FDR}(\text{ADDOW}) = \alpha,$$

and in (ED)+(EE) case,

$$\lim_{m \rightarrow \infty} \text{FDR}(\text{ADDOW}) = \frac{\pi_0}{\bar{\pi}_0} \alpha.$$

Theorem 5.2. *Let us consider defined in Sections 2.1 and 2.2, with the additional assumption (CE) or (ED)+(EE). For any sequence of random weight functions $(\widehat{W})_{m \geq 1}$, such that $\widehat{W} : [0, 1] \rightarrow K^m$ and $\widehat{G}_{\widehat{W}}$ is nondecreasing, we have*

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

5.2. Relation to IHW

Recall that IHW is ADDOW in the (NE) case, and that (NE) is a subcase of (EE). Hence, as a byproduct, we deduce from Theorems 5.1 and 5.2 the following result on IHW.

Corollary 5.1. *Let us consider the framework defined in Sections 2.1 and 2.2, with the additional assumption (ED). Then*

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

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

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

While equation (5.1) of Theorem 5.1 covers Theorem 4 of the supplementary material of Ignatiadis et al. (2016) (with slightly stronger assumption on the smoothness of the F_g s), the FDR controlling result of Corollary 5.1 gives a slightly sharper bound ($\pi_0 \alpha$ instead of α) in (ED) case.

The power optimality stated in Corollary 5.1 is new and was not shown in Ignatiadis et al. (2016). It thus supports the fact that IHW should be used under the assumption (ED) and when π_0 is close to 1 or not estimated.

5.3. Comparison to other existing procedures

For any estimators $\hat{\pi}_{g,0} \in [0, 1]$, any weighting satisfying $\sum_g \frac{m_g}{m} w_g \leq 1$ also satisfies $\sum_g \frac{m_g}{m} \hat{\pi}_{g,0} w_g \leq 1$, that is $K_{NE}^m \subset K^m$. Hence, any MWBH procedure estimating $\frac{m_{g,0}}{m_g}$ by 1 uses a weight function valued in K^m . This immediately yields the following corollary.

Corollary 5.2. *Let us consider the framework defined in Sections 2.1 and 2.2, with the additional assumption (CE) or (ED)+(EE). Then*

$$\lim_{m \rightarrow \infty} \text{Pow}(\text{ADDOW}) \geq \limsup_{m \rightarrow \infty} \text{Pow}(R),$$

for any $R \in \{\text{BH}, \text{IHW}\}$.

The next corollary simply states that ADDOW outperforms many procedures of the "weighting with π_0 adaptation" literature.

Corollary 5.3. *Let us consider the framework defined in Sections 2.1 and 2.2, with the additional assumption (CE) or (ED)+(EE). Then*

$$\lim_{m \rightarrow \infty} \text{Pow}(\text{ADDOW}) \geq \limsup_{m \rightarrow \infty} \text{Pow}(R),$$

for any $R \in \{\text{Pro1}, \text{Pro2}, \text{HZZ}, \text{ABH}\}$.

The results for Pro2, HZZ and ABH follow directly from Theorem 5.2 because these are MWBH procedures. The proof for Pro1 (which is not of the MWBH type) can be found in Section D.

5.4. Results for the stabilized version

Next theorem shows that, asymptotically, the procedure sADDOW_β is the same as ADDOW. Our result is true even if $\beta = \beta_m \xrightarrow{m \rightarrow \infty} 0$ provided that the convergence is not too fast. It is proven in Section E.

Theorem 5.3. *Let us consider the framework defined in Sections 2.1 and 2.2. Take a sequence $(\beta_m)_{m \geq 1}$ such that $\beta_m \geq a \exp(-bm^{1-\nu})$ for some $a \in (0, 1]$, $b > 0$ and $\nu > 0$.*

Then $\phi_{\beta_m} \rightarrow 1$ almost surely. In particular, all Theorems and Corollaries of Sections 5.1 and 5.3 hold when replacing ADDOW with sADDOW_{β_m} .

6. Numerical experiments

6.1. Simulation setting

In our experiments, additionally to BH which is not adaptive, three groups of procedures are compared:

- Group 1: some procedures not adaptive to π_0 but adaptive to the signal via optimal weights:
 - MWBH (W_{or}^*) where W_{or}^* is given by equation 3.5 in the (NE) case.
 - ADDOW in the (NE) case, that is, IHW.
 - sADDOW $_{\beta}$ in the (NE) case for some value of β .
 - Pro2 as defined in section 3.2 and in the (NE) case.
- Group 2: procedures only adaptive to $\pi_{g,0}$ and not to the signal strength, with an oracle adaptation to $\pi_{g,0}$:
 - ABH as defined in section 3.2 with $\hat{\pi}_{g,0} = \pi_{g,0}$.
 - HZZ as defined in section 3.2 with $\hat{\pi}_{g,0} = \pi_{g,0}$.
- Group 3: procedures that combine both adaptive properties, with an oracle adaptation to $\pi_{g,0}$:
 - MWBH (W_{or}^*) where W_{or}^* is given by equation 3.5 with $\hat{\pi}_{g,0} = \pi_{g,0}$.
 - ADDOW with $\hat{\pi}_{g,0} = \pi_{g,0}$.
 - sADDOW $_{\beta}$ with $\hat{\pi}_{g,0} = \pi_{g,0}$ for some value of β .
 - Pro2 as defined in section 3.2 with $\hat{\pi}_{g,0} = \pi_{g,0}$.

We consider the one-sided Gaussian framework described in Section 2.4 for $G = 2$ groups. The parameters are thus given by $m_1, m_2, m_{1,0}, m_{2,0}, \mu_1, \mu_2$, and α . For the stabilisation, $q_{\beta,m}$ is estimated with realizations of Z_{0m} (as defined in equation (4.1)), where Z_{0m} and Z_m are computed as suprema on $\{k/m, 1 \leq k \leq m\}$ instead of $[0, 1]$ for an easier computation. Our experiments have been performed by using the three following scenarios, for which the values of μ_1 and μ_2 are defined according to a parameter $\bar{\mu}$. Each simulation of each scenario is replicated 1000 times.

- Scenario 1: $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$, $\alpha = 0.05$, $\beta = 0.001$, $m_1 = m_2 = 2000$, $m_{1,0}/m_1 = 0.7$ and $m_{2,0}/m_2 = 0.8$. Furthermore the values of $\bar{\mu}$ are in $\{0.01, 0.02, 0.05\}$ and then from 0.5 to 3 with jumps of size 0.25. Here, $q_{\beta,m}$ is estimated with 10000 realizations of Z_{0m} .
- Scenario 2: $\mu_1 = 2$ and $\mu_2 = \bar{\mu}$, $\alpha = 0.7$, $m_1 = 1000$ and $m_2 = 9000$, $m_{1,0}/m_1 = 0.05$ and $m_{2,0}/m_2 = 0.85$. Furthermore $\bar{\mu} \in \{1.7, 1.8, 1.9, 2, 2.1, 2.2, 2.3\}$.
- Scenario 3: $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$, $\alpha = 0.05$, $\beta = 0.05$, $m \in \{100, 300, 500, 1000, 2000, 5000\}$, $m_1 = m_2 = m/2$, $m_{g,0}/m_g = 0.8$. Furthermore $\bar{\mu} \in \{0.01, 3\}$. Here, $q_{\beta,m}$ is estimated with 1000 realizations of Z_{0m} .

6.2. FDR control

The FDR of all above procedures are compared in Figure 2, Figure 3a and Figure 3b.

First, Figure 3b shows that the convergence of the FDR holds for moderate m . This supports the theoretical finding of Corollary 5.1 showing that the FDR shall converge to $\pi_0\alpha$ in scenario 3. This Figure also shows that when the signal is strong, sADDOW $_{\beta}$ behaves as ADDOW, which is well expected for the definition of ϕ_{β} . While Figure 2 supports the latter for large signal ($\bar{\mu} \geq 2$), we see that the FDR control of data-driven weighted procedures (ADDOW, Pro2) can deteriorate as $\bar{\mu}$ decreases. This is due to an overfitting phenomenon.

As $\bar{\mu}$ get smaller, the overfitting seems to increase and the FDR control seems to get violated. Let us underline that this does not contradict our theory because considering a small $\bar{\mu}$ might imply a smaller convergence rate while m stays ≤ 5000 in scenarios 1 and 3. Fortunately, in

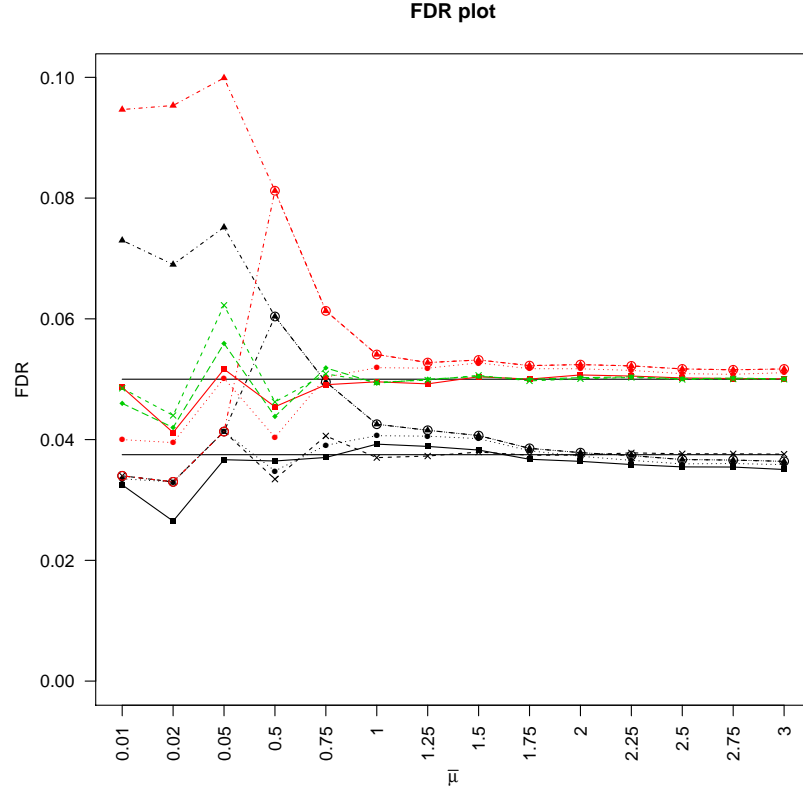


Fig 2: FDR against $\bar{\mu}$ in scenario 1. Group 1 in black; Group 2 in green; Group 3 in red. The type of procedure is MWBH (W_{or}^*) (squares); ADDOW (triangles); $sADDOW_{\beta}$ (circles); Pro2 (disks); HZZ (diamonds) and finally BH/ABH (crosses). Horizontal lines: α and $\pi_0\alpha$ levels. See Section 6.1.

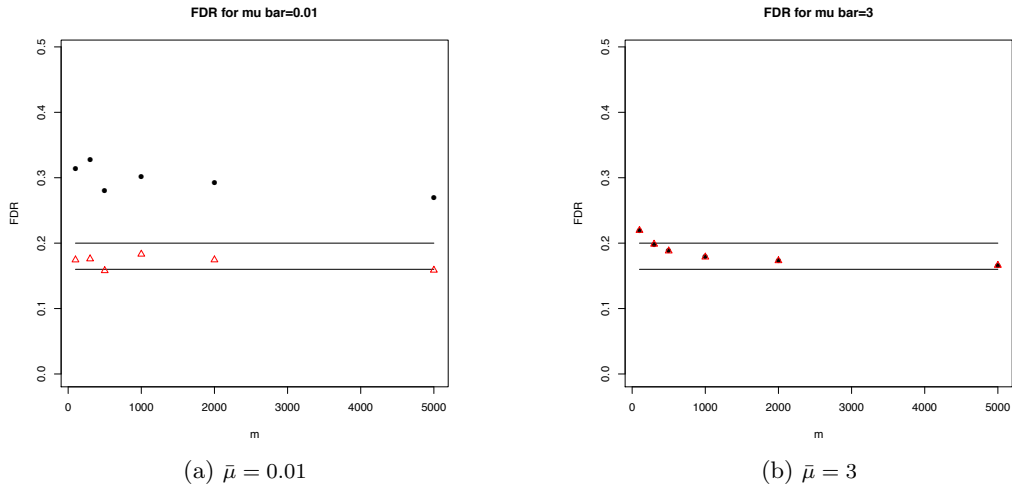


Fig 3: FDR against m in scenario 3. Group 1, ADDOW (black dots) and $sADDOW_{\beta}$ (red triangles). Horizontal lines: α and $\pi_0\alpha$ levels.

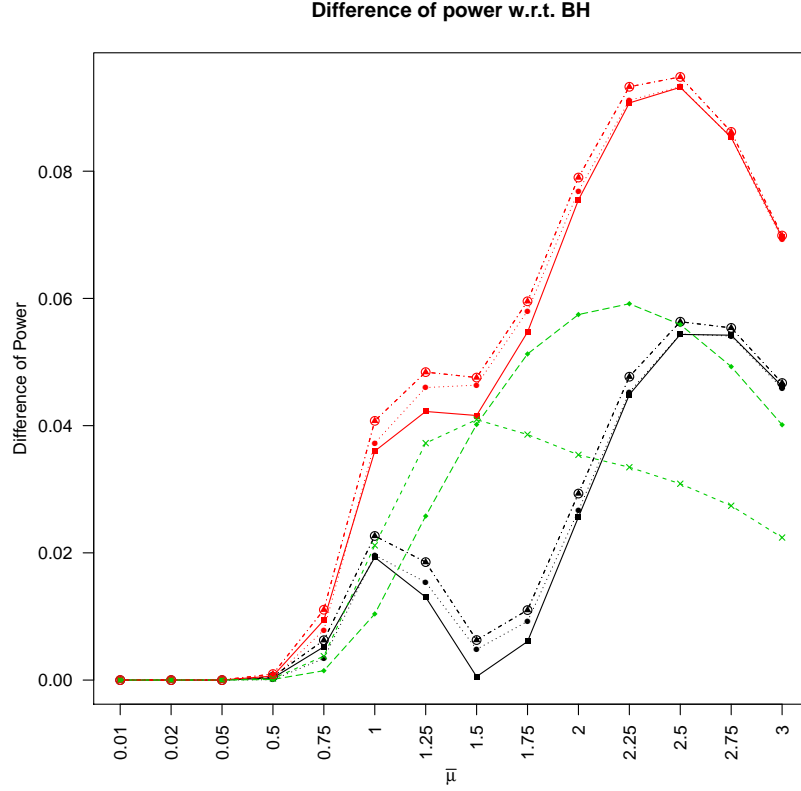


Fig 4: DiffPow against $\bar{\mu}$ in scenario 1. Same legend as Figure 2

this regime, it is apparent from Figure 2 and Figure 3a that the regularization process correctly addresses the overfitting by maintaining the FDR control holds true. Again, this is well expected because sADDOW_β simply corresponds to BH in that regime, see equation (4.2).

6.3. Power

Now that the FDR control has been studied, let us compare the procedures in terms of power. First, to better emphasize the benefit of adaptation, the power is rescaled in the following way: we define the normalized difference of power with respect to BH, or DiffPow, by

$$\text{DiffPow}(R) = \frac{m}{m_1} (\text{Pow}(R) - \text{Pow}(\text{BH})),$$

for any procedure R .

Figure 4 displays the power of all the procedures defined in Section 6.1. We can make several observations:

- We see a huge difference of behavior between the Group 1 and the Group 3. Hence, incorporating the knowledge of π_0 can lead to a large improvement of power.
- In both groups (that is in both (NE) and (CE) cases) ADDOW achieves the best power, which supports Theorem 5.2. Additionnaly, maybe surprisingly, in both cases, Pro2 behaves quite well, with a power close to the one of ADDOW and despite its theoretical sub-optimality. Hence, it seems to also be a good choice in practice.
- The comparison between the Group 2 and the Group 3 shows the benefit of adding the F_g adaptation to the π_0 adaptation: the fourth group has better power than the third for all

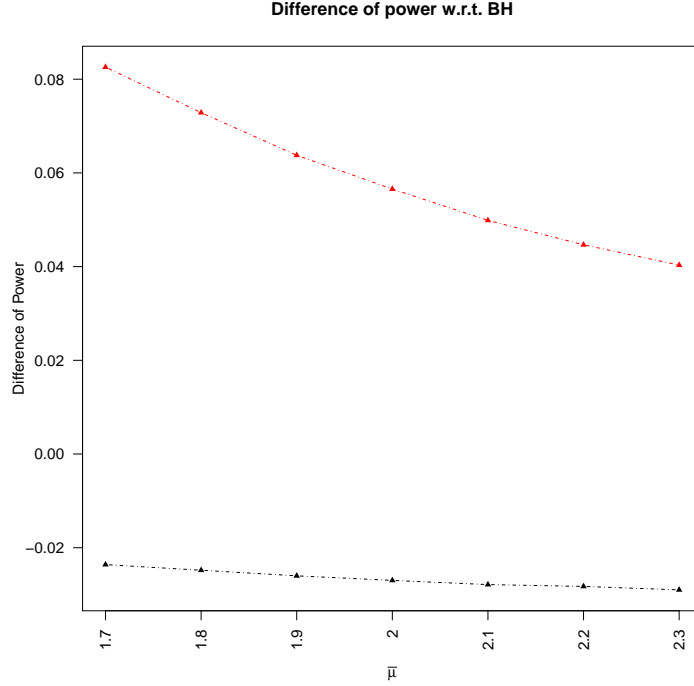


Fig 5: DiffPow of ADDOW against $\bar{\mu}$ in scenario 2. Group 1 in black and Group 3 in red.

signals. We can see a zone of moderate signal (around $\bar{\mu} = 1.5$) where the two groups are close. That is the same zone where HZZ becomes better than ABH. We deduce that in that zone the optimal weighting is the same as the uniform $\hat{w}^{(1)}$ weighting of ABH.

- The comparison of the DiffPow between the Group 1 and Group 2 in Figure 4 shows the difference between adapting to the F_g 's versus adapting to π_0 . No method is generally better than the other: as we see in the plot, it depends on the signal strength. We also see that neither ABH nor HZZ is better than the other.

Finally, let us discuss Figure 5. Here, the scenario 3 entails that IHW favors the large and sparse second group of hypotheses whereas the optimal power is achieved by favoring the small first group of hypotheses which contains almost only signal, as expected in remark 3.1. As a WBH procedure with weights (1,1), BH does not favor any group. Hence, IHW has a power smaller than BH. This demonstrates the limitation of the heuristic upon which IHW is built, and underlines the necessity of estimating the $\pi_{g,0}$ when nothing lets us think that (ED) may be met.

7. Concluding remarks

In this paper we presented a new class of data-driven step-up procedure, ADDOW, that generalizes IHW by incorporating $\pi_{g,0}$ estimators in each group. We showed that while this procedure asymptotically controls the FDR at the targeted level, it has the best power among all MWBH procedures when the π_0 estimation can be made consistently. In particular it dominates all the existing procedures of the weighting literature and solves the p -values weighting issue in a group-structured multiple testing problem. As a by-product, our work established the optimality of IHW in the case of homogeneous π_0 structure. Finally we proposed a stabilization variant designed to deal with the case where only few discoveries can be made (very small signal strength or sparsity). Some numerical simulations illustrated that our properties are also valid in a finite sample framework, provided that the number of tests is large enough.

Assumptions Our assumptions are rather mild: basically we only added the concavity of the F_g to the assumptions of Ignatiadis et al. (2016). Notably we dropped the other regularity assumptions on F_g that were made in Roquain and Van De Wiel (2009) while keeping all the useful properties on W^* in the (NE) case. Note that the criticality assumption is often made in the literature, see Ignatiadis et al. (2016) (assumption 5 of the supplementary material), Zhao and Zhang (2014) (assumption A.1), or the assumption of Theorem 4 in Hu, Zhao and Zhou (2010). Finally, the weak dependence assumption is extensively used in our paper. An interesting direction could be to extend our result to some strong dependent cases, for instance by assuming the PRDS (positive regression dependence), as some previous work already studied properties of MWBH procedures under that assumption, see Roquain and Van De Wiel (2008).

Computational aspects The actual maximization problem of ADDOW is difficult, it involves a mixed integer linear programming that may take a long time to resolve. Some regularization variant may be needed for applications. To this end, we can think to use the least concave majorant (LCM) instead of the empirical c.d.f. in equation (3.1) (as proposed in modification (E1) of IHW in Ignatiadis et al. (2016)). As we show in Section 8, ADDOW can be extended to that case (see especially Section 8.1) and our results are still valid for this new regularized version of ADDOW.

Toward nonasymptotic results Interesting direction for future research can be to investigate the convergence rate in our asymptotic results. One possible direction can be to use the work of Neuvial (2008). However, it would require to compute the Hadamard derivative of the functional involved in our analysis, which might be very challenging. Finally, another interesting future work could be to develop other versions of ADDOW that ensure finite sample FDR control property: this certainly requires to use a different optimization process, which will make the power optimality difficult to maintain.

8. Proofs of Theorems 5.1 and 5.2

8.1. Further generalization

Define for any u and W

$$\hat{H}_W(u) = m^{-1} |R_{u,W} \cap \mathcal{H}_0| = m^{-1} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha u W_g(u), H_{g,i}=0\}}$$

and

$$\hat{P}_W(u) = m^{-1} |R_{u,W} \cap \mathcal{H}_1| = \hat{G}_W(u) - \hat{H}_W(u),$$

so that $\text{FDP}(R_{u,W}) = \frac{\hat{H}_W(u)}{\hat{G}_W(u) \vee m^{-1}}$ and $\text{Pow}(R_{u,W}) = \mathbb{E}[\hat{P}_W(u)]$ (recall that $\text{MWBH}(W) = R_{\tilde{u}_W, W}$). Also define $\hat{D}_g(t) = m_g^{-1} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq t\}}$ so that $\hat{G}_W(u) = \sum_g \frac{m_g}{m} \hat{D}_g(\alpha u W_g(u))$.

For the sake of generality \hat{D}_g is not the only estimator of D_g (defined in equation (B.1)) that we will use to prove our results (for example, we can use the LCM of \hat{D}_g , denoted $\text{LCM}(\hat{D}_g)$, see Section 7). So let us increase slightly the scope of the MWBH class by defining $\tilde{G}_W(u) = \sum_g \frac{m_g}{m} \tilde{D}_g(\alpha u W_g(u))$ for any estimator \tilde{D}_g such that \tilde{D}_g is nondecreasing, $\tilde{D}_g(0) = 0$, $\tilde{D}_g(1) = 1$ and $\|\tilde{D}_g - D_g\| \xrightarrow{\mathbb{P}} 0$, where $\|\cdot\|$ is the sup norm for the bounded functions on their definition domain. Note that at least $(D_g)_g$, $(\hat{D}_g)_g$ (by Lemma C.1), and $(\text{LCM}(\hat{D}_g))_g$ (by Lemma C.6) are eligible.

If W is such that \tilde{G}_W is nondecreasing, we then define the generalized MWBH as

$$\text{GMWBH}\left((\tilde{D}_g)_g, W\right) = R_{\tilde{u}_W, W} \text{ where } \tilde{u}_W = \mathcal{I}\left(\tilde{G}_W\right).$$

If $(\tilde{D}_g)_g$ is such that we can define, for all $u \in [0, 1]$,

$$\tilde{W}^*(u) \in \arg \max_{w \in K^m} \tilde{G}_w(u), \quad (8.1)$$

we define the generalized ADDOW by

$$\text{GADDOW} \left((\tilde{D}_g)_g \right) = \text{GMWBH} \left((\tilde{D}_g)_g, \tilde{W}^* \right),$$

the latter being well defined because $\tilde{G}_{\tilde{W}^*}$ is nondecreasing (by a proof similar to the one of Lemma A.6). Note that for any continuous \tilde{D}_g , such as $\text{LCM}(\hat{D}_g)$ or D_g itself, the $\arg \max$ in (8.1) is non empty and GADDOW can then be defined.

What we show below are more general theorems, valid for any $\text{GADDOW} \left((\tilde{D}_g)_g \right)$. Our proofs combined several technical lemmas deferred to Sections B and C, which are based on the previous work of Roquain and Van De Wiel (2009); Hu, Zhao and Zhou (2010); Zhao and Zhang (2014).

Remark 8.1. $\text{GADDOW} \left((\tilde{D}_g)_g \right)$ when $\tilde{D}_g = \text{LCM}(\hat{D}_g)$ and $\hat{\pi}_{g,0} = 1$ is not exactly the same as IHW with modification (E1). In our notation, their procedure is $\text{WBH} \left(\tilde{W}^* \left(\mathcal{I} \left(\tilde{G}_{\tilde{W}^*} \right) \right) \right)$.

8.2. Proof of Theorem 5.1

We have

$$\text{FDP} \left(\text{GMWBH} \left((\tilde{D}_g)_g, \tilde{W}^* \right) \right) = \frac{\hat{H}_{\tilde{W}^*}(\tilde{u})}{\hat{G}_{\tilde{W}^*}(\tilde{u}) \vee m^{-1}} \in [0, 1],$$

where \tilde{u} is defined as in (C.6) so by Lemma C.5 we deduce that

$$\text{FDP} \left(\text{GADDOW} \left((\tilde{D}_g)_g \right) \right) \xrightarrow{m \rightarrow \infty} \frac{H_{W^*}^\infty(u^*)}{G_{W^*}^\infty(u^*)} = \frac{H_{W^*}^\infty(u^*)}{u^*},$$

and then

$$\lim_{m \rightarrow \infty} \text{FDR} \left(\text{GADDOW} \left((\tilde{D}_g)_g \right) \right) = u^{*-1} H_{W^*}^\infty(u^*),$$

where $G_{W^*}^\infty$, $H_{W^*}^\infty$ and u^* are defined in Section B.

If $\alpha \geq \bar{\pi}_0$, $u^* = 1$ by Lemma B.2 and $\alpha u^* W_g^*(u^*) \geq 1$ by Lemma B.1 so $u^{*-1} H_{W^*}^\infty(u^*) = \pi_0 \leq \bar{\pi}_0 \leq \alpha$.

If $\alpha \leq \bar{\pi}_0$, $\alpha u^* W_g^*(u^*) \leq 1$ by Lemma B.1 so $U(\alpha u^* W_g^*(u^*)) = \alpha u^* W_g^*(u^*)$ for all g and then

$$\begin{aligned} u^{*-1} H_{W^*}^\infty(u^*) &= \alpha \sum_g \pi_g \pi_{g,0} W_g^*(u^*) \\ &\leq \alpha \sum_g \pi_g \bar{\pi}_{g,0} W_g^*(u^*) = \alpha. \end{aligned} \quad (8.2)$$

Moreover if we are in (CE) case (that is $\bar{\pi}_{g,0} = \pi_{g,0}$) the inequality above becomes an equality. Finally if we are in (ED)+(EE) case (that is $\pi_{g,0} = \pi_0$ and $\bar{\pi}_{g,0} = \bar{\pi}_0$) we write

$$\begin{aligned} u^{*-1} H_{W^*}^\infty(u^*) &= \alpha \sum_g \pi_g \pi_0 W_g^*(u^*) \\ &= \frac{\pi_0}{\bar{\pi}_0} \alpha \sum_g \pi_g \bar{\pi}_0 W_g^*(u^*) \\ &= \frac{\pi_0}{\bar{\pi}_0} \alpha. \end{aligned} \quad (8.3)$$

The equalities in (8.2) and (8.3) are due to $\sum_g \pi_g \bar{\pi}_{g,0} W_g^*(u^*) = 1$ (by Lemma B.1).

8.3. Proof of Theorem 5.2

First, in any case,

$$\hat{P}_{\widehat{W}^*}(\tilde{u}) = \hat{G}_{\widehat{W}^*}(\tilde{u}) - \hat{H}_{\widehat{W}^*}(\tilde{u}) \xrightarrow{a.s.} G_{\widehat{W}^*}^\infty(u^*) - H_{\widehat{W}^*}^\infty(u^*) = P_{\widehat{W}^*}^\infty(u^*)$$

by Lemma C.5, where $P_{\widehat{W}^*}^\infty$ is defined in Section B.. Hence the limit of $\text{Pow}\left(\text{GADDOW}\left((\tilde{D}_g)_g\right)\right)$ is $P_{\widehat{W}^*}^\infty(u^*)$.

For the rest of the proof, we assume we are in case (CE) or (ED)+(EE), which implies by Lemma B.3 that $W^*(u) \in \arg \max_{w \in K^\infty} P_w^\infty(u)$ for all u , and that $P_{\widehat{W}^*}^\infty$ is nondecreasing. We also split the proof in two parts. For the first part we assume that for all m , \widehat{W} is a weight vector $\hat{w} \in K^m$ therefore not depending on u . In the second part we will conclude with a general sequence of weight functions.

Part 1 $\widehat{W} = \hat{w} \in K^m$ for all m . Let $\ell = \limsup \text{Pow}(\text{MWBH}(\hat{w}))$. Up to extracting a subsequence, we can assume that $\ell = \lim \mathbb{E} \left[\hat{P}_{\hat{w}}(\hat{u}_{\hat{w}}) \right]$ and $\hat{\pi}_{g,0} \xrightarrow{a.s.} \bar{\pi}_{g,0}$ for all g . Define the event

$$\tilde{\Omega} = \left\{ \begin{array}{l} \forall g, \hat{\pi}_{g,0} \longrightarrow \bar{\pi}_{g,0} \\ \sup_{w \in \mathbb{R}_+^G} \left\| \hat{P}_w - P_w^\infty \right\| \longrightarrow 0 \\ \sup_{w \in \mathbb{R}_+^G} \left\| \hat{G}_w - G_w^\infty \right\| \longrightarrow 0 \end{array} \right\}$$

then $\mathbb{P}(\tilde{\Omega}) = 1$ (by Lemma C.1), $\ell = \lim \mathbb{E} \left[\hat{P}_{\hat{w}}(\hat{u}_{\hat{w}}) \mathbf{1}_{\tilde{\Omega}} \right]$ and by reverse Fatou Lemma $\ell \leq \mathbb{E} \left[\limsup \hat{P}_{\hat{w}}(\hat{u}_{\hat{w}}) \mathbf{1}_{\tilde{\Omega}} \right]$.

Now consider that $\tilde{\Omega}$ occurs and fix a realization of it, the following of this part 1 is deterministic. Let $\ell' = \limsup \hat{P}_{\hat{w}}(\hat{u}_{\hat{w}})$. The sequences $\left(\frac{m}{m_g \hat{\pi}_{g,0}} \right)$ are converging and then bounded, hence the sequence (\hat{w}) is also bounded. By compactity, once again up to extracting a subsequence, we can assume that $\ell' = \lim \hat{P}_{\hat{w}}(\hat{u}_{\hat{w}})$ and that \hat{w} converges to a given w^{cv} . By taking $m \rightarrow \infty$ in relation $\sum \frac{m_g}{m} \hat{\pi}_{g,0} \hat{w}_g \leq 1$, it appears that w^{cv} belongs to K^∞ . $\|\hat{G}_{\hat{w}} - G_{w^{cv}}^\infty\| \leq \sup_w \|\hat{G}_w - G_w^\infty\| + \|G_{\hat{w}}^\infty - G_{w^{cv}}^\infty\| \rightarrow 0$ so by Remark B.2 $\hat{u}_{\hat{w}} \rightarrow u_{w^{cv}}^\infty$ and finally

$$\begin{aligned} \left| \hat{P}_{\hat{w}}(\hat{u}_{\hat{w}}) - P_{w^{cv}}^\infty(u_{w^{cv}}^\infty) \right| &\leq \sup_{w \in \mathbb{R}_+^G} \left\| \hat{P}_w - P_w^\infty \right\| + |P_{\hat{w}}^\infty(\hat{u}_{\hat{w}}) - P_{w^{cv}}^\infty(u_{w^{cv}}^\infty)| \\ &\longrightarrow 0, \end{aligned}$$

by continuity of F_g and because $\omega \in \tilde{\Omega}$. So $\ell' = P_{w^{cv}}^\infty(u_{w^{cv}}^\infty) \leq P_{\widehat{W}^*}^\infty(u_{w^{cv}}^\infty)$ by maximality. Note also that $G_{w^{cv}}^\infty(\cdot) \leq G_{\widehat{W}^*}^\infty(\cdot)$ which implies that $u_{w^{cv}}^\infty \leq u_{\widehat{W}^*}^\infty = u^*$ so $\ell' \leq P_{\widehat{W}^*}^\infty(u^*)$ because $P_{\widehat{W}^*}^\infty$ is nondecreasing. Finally $\limsup \hat{P}_{\hat{w}}(\hat{u}_{\hat{w}}) \mathbf{1}_{\tilde{\Omega}} \leq P_{\widehat{W}^*}^\infty(u^*)$ for any realization of $\tilde{\Omega}$, by integrating we get that $\ell \leq P_{\widehat{W}^*}^\infty(u^*)$ which concludes that part 1.

Part 2 Now consider the case where \widehat{W} is a weight function $u \mapsto \widehat{W}(u)$. Observe that

$$\hat{u}_{\widehat{W}} = \hat{G}_{\widehat{W}}(\hat{u}_{\widehat{W}}) = \hat{G}_{\widehat{W}(\hat{u}_{\widehat{W}})}(\hat{u}_{\widehat{W}}),$$

so by definition of $\mathcal{I}(\cdot)$, $\hat{u}_{\widehat{W}} \leq \hat{u}_{\widehat{W}(\hat{u}_{\widehat{W}})}$, and then

$$\hat{P}_{\widehat{W}}(\hat{u}_{\widehat{W}}) = \hat{P}_{\widehat{W}(\hat{u}_{\widehat{W}})}(\hat{u}_{\widehat{W}}) \leq \hat{P}_{\widehat{W}(\hat{u}_{\widehat{W}})}(\hat{u}_{\widehat{W}(\hat{u}_{\widehat{W}})}).$$

As a consequence, $\text{Pow}(\text{MWBH}(\widehat{W})) \leq \text{Pow}(\text{MWBH}(\widehat{W}(\hat{u}_{\widehat{W}})))$. Finally, apply part 1 to the weight vector sequence $(\widehat{W}(\hat{u}_{\widehat{W}}))$ to conclude.

Remark 8.2. We just showed that for every MWBH procedure, there is a corresponding WBH procedure with better power. In particular, by defining $\hat{u} = u_{\widehat{W}^*}$ the ADDOW threshold, we showed that $\hat{u} \leq \hat{u}_{\widehat{W}^*(\hat{u})}$. But $\widehat{G}_{\widehat{W}^*} \geq \widehat{G}_{\hat{u}}$ and then $\hat{u} \geq u_{\hat{u}}$ for any \hat{u} . Hence $\hat{u} = \hat{u}_{\widehat{W}^*(\hat{u})}$ and ADDOW is the WBH procedure associated to the weight vector $\widehat{W}^*(\hat{u})$.

Acknowledgments

Thanks to my advisors Etienne Roquain and Pierre Neuvial for the constant and deep improvements they enabled, in many areas of the paper. Also thanks to Christophe Giraud and Patricia Reynaud-Bouret for useful discussions. This work has been supported by CNRS (PEPS FaSciDo) and ANR-16-CE40-0019.

References

- BENJAMINI, Y. and HOCHBERG, Y. (1997). Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics* **24** 407–418.
- BLANCHARD, G. and ROQUAIN, E. (2008). Two simple sufficient conditions for FDR control. *Electronic journal of Statistics* **2** 963–992.
- CAI, T. T. and SUN, W. (2009). Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks. *Journal of the American Statistical Association* **104** 1467–1481.
- CAROLAN, C. A. (2002). The least concave majorant of the empirical distribution function. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* 317–328.
- CHI, Z. (2007). On the performance of FDR control: constraints and a partial solution. *The Annals of Statistics* **35** 1409–1431.
- DMITRIENKO, A., OFFEN, W. W. and WESTFALL, P. H. (2003). Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Statistics in medicine* **22** 2387–2400.
- DOBRIBAN, E., FORTNEY, K., KIM, S. K. and OWEN, A. B. (2015). Optimal multiple testing under a Gaussian prior on the effect sizes. *Biometrika* **102** 753–766.
- DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics* 642–669.
- EFRON, B. (2008). Simultaneous inference: When should hypothesis testing problems be combined? *The annals of applied statistics* 197–223.
- GENOVESE, C. R., ROEDER, K. and WASSERMAN, L. (2006). False discovery control with p-value weighting. *Biometrika* 509–524.
- GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Annals of Statistics* 1035–1061.
- HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* 65–70.
- HU, J. X., ZHAO, H. and ZHOU, H. H. (2010). False discovery rate control with groups. *Journal of the American Statistical Association* **105**.
- IGNATIADIS, N., KLAUS, B., ZAUGG, J. B. and HUBER, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature methods* **13** 577–580.
- KOLMOGOROV, A. N. (1933). Sulla Determinazione Empirica di una Legge di Distribuzione. *Giornale dell'Istituto Italiano degli Attuari* **4** 83–91.
- MASSART, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability* 1269–1283.
- NEUVIAL, P. (2008). Asymptotic properties of false discovery rate controlling procedures under independence. *Electronic journal of statistics* **2** 1065–1110.
- NEUVIAL, P. (2013). Asymptotic results on adaptive false discovery rate controlling procedures based on kernel estimators. *The Journal of Machine Learning Research* **14** 1423–1459.

- RAMDAS, A., BARBER, R. F., WAINWRIGHT, M. J. and JORDAN, M. I. (2017). A Unified Treatment of Multiple Testing with Prior Knowledge. *arXiv preprint arXiv:1703.06222*.
- ROEDER, K. and WASSERMAN, L. (2009). Genome-wide significance levels and weighted hypothesis testing. *Statistical science: a review journal of the Institute of Mathematical Statistics* **24** 398.
- ROEDER, K., BACANU, S.-A., WASSERMAN, L. and DEVLIN, B. (2006). Using linkage genome scans to improve power of association in genome scans. *The American Journal of Human Genetics* **78** 243–252.
- ROGOSA, D. (2005). Accuracy of API index and school base report elements: 2003 Academic Performance Index, California Department of Education.
- ROQUAIN, E. and VAN DE WIEL, M. (2008). Multi-weighting for FDR control. *arXiv:0807.4081v1*.
- ROQUAIN, E. and VAN DE WIEL, M. A. (2009). Optimal weighting for false discovery rate control. *Electronic Journal of Statistics* **3** 678–711.
- SCHWARTZMAN, A., DOUGHERTY, R. F. and TAYLOR, J. E. (2005). Cross-subject comparison of principal diffusion direction maps. *Magnetic Resonance in Medicine* **53** 1423–1431.
- STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). 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** 187–205.
- SUN, L., CRAIU, R. V., PATERSON, A. D. and BULL, S. B. (2006). Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genetic epidemiology* **30** 519–530.
- THE GENE ONTOLOGY CONSORTIUM (2000). Gene Ontology: tool for the unification of biology. *Nature genetics* **25** 25–29.
- WASSERMAN, L. and ROEDER, K. (2006). Weighted hypothesis testing. *arXiv preprint math/0604172*.
- ZHAO, H. and ZHANG, J. (2014). Weighted p-value procedures for controlling FDR of grouped hypotheses. *Journal of Statistical Planning and Inference* **151** 90–106.

Appendix A: Lemmas and proofs of Sections 2, 3 and 4

Lemma A.1. *For all g , F_g is continuous.*

Proof. $F_g(x) = 0$ for all $x \leq 0$ and $F_g(x) = 1$ for all $x \geq 1$. F_g is concave so it is continuous over $(0, 1)$. F_g is càdlàg (as a c.d.f.) so it is continuous in 0. Write $F_g\left(\frac{t+1}{2}\right) \geq \frac{F_g(t)+F_g(1)}{2}$ and make t converge to 1^- to get that $F_g(1^-) \geq F_g(1)$ but we also have $F_g(1^-) \leq F_g(1)$ because F_g is nondecreasing. \square

Lemma A.2. *Take a real valued sequence (λ_m) with $\lambda_m \in (0, 1)$, converging to 1, such that $\frac{1}{\sqrt{m}} = o(1 - \lambda_m)$ and $\frac{m_{g,0}}{m_g} = \pi_{g,0} + o(1 - \lambda_m)$ for all g . If $f_g(1^-) = 0$ for all g and the p-values inside each group are mutually independent, then*

$$\forall g \in \{1, \dots, G\}, \hat{\pi}_{g,0}(\lambda_m) \xrightarrow{\mathbb{P}} \pi_{g,0}.$$

Proof. First note that $\frac{m_{g,1}}{m_g} - \pi_{g,1} = \pi_{g,0} - \frac{m_{g,0}}{m_g} = o(1 - \lambda_m)$.

Thus we have

$$\begin{aligned}
|\hat{\pi}_{g,0}(\lambda_m) - \pi_{g,0}| &= \left| \frac{1 - \frac{1}{m_g} \sum_i \mathbb{1}_{\{p_{g,i} \leq \lambda_m\}} + \frac{1}{m}}{1 - \lambda_m} - \pi_{g,0} \right| \\
&\leq \frac{\lambda_m \left| \pi_{g,0} - \frac{m_{g,0}}{m_g} \right| + \frac{m_{g,0}}{m_g} \left| \lambda_m - \frac{1}{m_{g,0}} \sum_i \mathbb{1}_{\{p_{g,i} \leq \lambda_m, H_{g,i}=0\}} \right|}{1 - \lambda_m} \\
&\quad + \frac{\left| \pi_{g,1} - \frac{m_{g,1}}{m_g} \right| + \frac{m_{g,1}}{m_g} \left| F_g(\lambda_m) - \frac{1}{m_{g,1}} \sum_i \mathbb{1}_{\{p_{g,i} \leq \lambda_m, H_{g,i}=1\}} \right|}{1 - \lambda_m} \\
&\quad + \frac{m_{g,1}}{m_g} \frac{1 - F_g(\lambda_m)}{1 - \lambda_m} + \frac{1}{m(1 - \lambda_m)} \\
&\leq \frac{m_{g,0}}{m_g} \frac{\sup_{x \in [0,1]} \left| x - \frac{1}{m_{g,0}} \sum_i \mathbb{1}_{\{p_{g,i} \leq x, H_{g,i}=0\}} \right|}{1 - \lambda_m} \\
&\quad + \frac{m_{g,1}}{m_g} \frac{\sup_{x \in [0,1]} \left| F_g(x) - \frac{1}{m_{g,1}} \sum_i \mathbb{1}_{\{p_{g,i} \leq x, H_{g,i}=1\}} \right|}{1 - \lambda_m} + o(1).
\end{aligned}$$

The two suprema of the last display, when multiplied by \sqrt{m} , converge in distribution (by Kolmogorov-Smirnov's theorem). So when divided by $1 - \lambda_m$ they converge to 0 in distribution and then in probability (because $\frac{1}{1 - \lambda_m} = o(\sqrt{m})$). \square

Definition A.1. The critical alpha value is

$$\alpha^* = \inf_{w \in K^\infty} \frac{1}{\sum_g \pi_g w_g (\pi_{g,0} + \pi_{g,1} f_g(0^+))},$$

where $K^\infty = \{w \in \mathbb{R}_+^G : \sum_g \pi_g \bar{\pi}_{g,0} w_g \leq 1\}$.

Lemma A.3. α^* is always such that $\alpha^* < 1$.

Proof. We only need to show that for one $w \in K^\infty$, we have

$$\sum_g \pi_g w_g (\pi_{g,0} + \pi_{g,1} f_g(0^+)) > 1.$$

Let us show that this is true for every $w \in K^\infty$ such that $\sum_g \pi_g \bar{\pi}_{g,0} w_g = 1$, e.g. the w defined by $w_g = \frac{1}{\bar{\pi}_{g,0}}$ for all g . We use the fact that $f_g(0^+) > \frac{F_g(1) - F_g(0)}{1 - 0} = 1$ by the strict concavity of F_g . Then $\pi_{g,0} + \pi_{g,1} f_g(0^+) > 1$ and

$$\sum_g \pi_g w_g (\pi_{g,0} + \pi_{g,1} f_g(0^+)) > \sum_g \pi_g w_g \geq \sum_g \pi_g \bar{\pi}_{g,0} w_g = 1. \quad \square$$

Recall that $\mathcal{I}(\cdot)$ is defined as $\mathcal{I}(h) = \sup \{u \in [0, 1] : h(u) \geq u\}$ on the function space:

$$\mathcal{F} = \{h : [0, 1] \rightarrow [0, 1] : h(0) = 0, h \text{ is nondecreasing}, \} \quad (\text{A.1})$$

which has the natural order $h_1 \leq h_2 \iff h_1(u) \leq h_2(u) \forall u \in [0, 1]$. \mathcal{F} is also normed with the sup norm $\|\cdot\|$.

Lemma A.4. For all $h \in \mathcal{F}$, $\mathcal{I}(h)$ is a maximum and $h(\mathcal{I}(h)) = \mathcal{I}(h)$. Moreover, $\mathcal{I}(\cdot)$, seen as a map on \mathcal{F} , is nondecreasing and continuous on each continuous $h_0 \in \mathcal{F}$ such that either $u \mapsto h_0(u)/u$ is decreasing over $(0, 1]$, or $\mathcal{I}(h_0) = 0$.

Proof. $\mathcal{I}(h)$ is a maximum because there exists $\epsilon_n \rightarrow 0$ such that

$$h(\mathcal{I}(h)) \geq h(\mathcal{I}(h) - \epsilon_n) \geq \mathcal{I}(h) - \epsilon_n \rightarrow \mathcal{I}(h).$$

So $h(\mathcal{I}(h)) \geq \mathcal{I}(h)$. Then $h(h(\mathcal{I}(h))) \geq h(\mathcal{I}(h))$ thus $h(\mathcal{I}(h)) \leq \mathcal{I}(h)$ by the definition of $\mathcal{I}(h)$ as a supremum.

Next, if $h_1 \leq h_2$, $\mathcal{I}(h_1) = h_1(\mathcal{I}(h_1)) \leq h_2(\mathcal{I}(h_1))$ so $\mathcal{I}(h_1) \leq \mathcal{I}(h_2)$ by definition of $\mathcal{I}(h_2)$.

Now take a continuous $h_0 \in \mathcal{F}$ such that either $u \mapsto h_0(u)/u$ is decreasing or $\mathcal{I}(h_0) = 0$, and h any element of \mathcal{F} . Let $\gamma > 0$, let $u_- = \mathcal{I}(h_0) - \gamma$ and $u_+ = \mathcal{I}(h_0) + \gamma$. We want to prove that there exists an η_γ such that $\|h - h_0\| \leq \eta_\gamma$ implies $u_- \leq \mathcal{I}(h) \leq u_+$.

If $u_+ > 1$ then obviously $\mathcal{I}(h) \leq u_+$. If not, let $s_\gamma = \max_{u' \in [u_+, 1]} (h_0(u') - u')$. It is a maximum by continuity over a compact and is such that $s_\gamma < 0$, because $s_\gamma \geq 0$ would contradict the maximality of $\mathcal{I}(h_0)$.

Then, for all $u' \in [u_+, 1]$,

$$h(u') - u' \leq h_0(u') - u' + \|h - h_0\|,$$

and then

$$\sup_{u' \in [u_+, 1]} (h(u') - u') \leq s_\gamma + \|h - h_0\|.$$

Hence, as soon as $\|h - h_0\| \leq \frac{1}{2}|s_\gamma|$, $\sup_{u' \in [u_+, 1]} (h(u') - u') < 0$ and $\mathcal{I}(h) < u_+$.

If $u_- \leq 0$, which is always the case if $\mathcal{I}(h_0) = 0$, then $\mathcal{I}(h) \geq u_-$. If $u_- > 0$, $u \mapsto h_0(u)/u$ is decreasing and

$$\frac{h_0(u_-)}{u_-} > \frac{h_0(\mathcal{I}(h_0))}{\mathcal{I}(h_0)} = 1,$$

so $h_0(u_-) > u_-$. We can then write the following:

$$h(u_-) - u_- \geq h_0(u_-) - u_- - \|h - h_0\| > 0,$$

as soon as $\|h - h_0\| \leq \frac{1}{2}(h_0(u_-) - u_-)$. This implies $\mathcal{I}(h) > u_-$. Taking

$$\eta_\gamma = \frac{1}{2} \min(|s_\gamma| \mathbb{1}_{\{u_+ \leq 1\}} + \mathbb{1}_{\{u_+ > 1\}}, (h_0(u_-) - u_-) \mathbb{1}_{\{u_- > 0\}} + \mathbb{1}_{\{u_- \leq 0\}})$$

completes the proof. \square

Lemma A.5. Let a weight function $W : [0, 1] \rightarrow \mathbb{R}_+^G$. For each r between 1 and m denote the $W(r/m)$ -weighted p-values $p_{g,i}^{[r]} = p_{g,i}/W_g(r/m)$ (with the convention $p_{g,i}/0 = \infty$), order them $p_{(1)}^{[r]} \leq \dots \leq p_{(m)}^{[r]}$ and note $p_{(0)}^{[r]} = 0$.

Then $\hat{u}_W = m^{-1} \max \left\{ r \geq 0 : p_{(r)}^{[r]} \leq \alpha \frac{r}{m} \right\}$.

Proof. Let us denote $\hat{r} = \max \left\{ r \geq 0 : p_{(r)}^{[r]} \leq \alpha \frac{r}{m} \right\}$ and show $\hat{u}_W = \hat{r}/m$ by double inequality. First, we have

$$\begin{aligned} \hat{G}_W \left(\frac{\hat{r}}{m} \right) &= m^{-1} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha \frac{\hat{r}}{m} W_g(\frac{\hat{r}}{m})\}} \\ &= m^{-1} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i}^{[r]} \leq \alpha \frac{\hat{r}}{m}\}} \\ &= m^{-1} \sum_{r=1}^m \mathbb{1}_{\{p_{(r)}^{[r]} \leq \alpha \frac{\hat{r}}{m}\}} \geq \hat{r}/m, \end{aligned}$$

because $p_{(1)}^{[\hat{r}]}, \dots, p_{(\hat{r})}^{[\hat{r}]} \leq \alpha \frac{\hat{r}}{m}$. Then $\hat{r}/m \leq \hat{u}_W$ by definition of \hat{u}_W . Second, we know that \hat{u}_W can be written as $\hat{\kappa}/m$ because $\hat{u}_W = \hat{G}_W(\hat{u}_W)$, so we want to show that $\hat{\kappa} \leq \hat{r}$ which is implied by $\hat{r}, p_{(\hat{\kappa})}^{[\hat{\kappa}]} \leq \alpha \frac{\hat{\kappa}}{m}$. The latter is true because

$$\sum_{r=1}^m \mathbb{1}_{\{p_{(r)}^{[\hat{\kappa}]} \leq \alpha \frac{\hat{\kappa}}{m}\}} = m \hat{G}_W\left(\frac{\hat{\kappa}}{m}\right) = m \hat{G}_W(\hat{u}_W) \geq \hat{\kappa}.$$

□

Lemma A.6. $\hat{G}_{\hat{W}^*}$ is nondecreasing.

Proof. Let $u \leq u'$. $\hat{G}_{\hat{W}^*}(u') = \max_{w \in K^m} \hat{G}_w(u')$ so by denoting $w = \hat{W}^*(u)$ we have $\hat{G}_{\hat{W}^*}(u') \geq \hat{G}_w(u')$. Furthermore,

$$\hat{G}_w(u') = \frac{1}{m} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha u' w_g\}} \geq \frac{1}{m} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha u w_g\}} = \hat{G}_{\hat{W}^*}(u),$$

which entails $\hat{G}_{\hat{W}^*}(u') \geq \hat{G}_{\hat{W}^*}(u)$. □

Appendix B: Asymptotical weighting

Define, for a weight function $W : [0, 1] \rightarrow \mathbb{R}_+^G$, possibly random,

$$P_W^\infty : u \mapsto \sum_{g=1}^G \pi_g \pi_{g,1} F_g(\alpha u W_g(u));$$

$$G_W^\infty : u \mapsto \sum_{g=1}^G \pi_g D_g(\alpha u W_g(u));$$

and

$$H_W^\infty(u) = G_W^\infty(u) - P_W^\infty(u),$$

where

$$D_g : t \mapsto \pi_{g,0} U(t) + \pi_{g,1} F_g(t) \tag{B.1}$$

is strictly concave on $[0, 1]$ because F_g is and $\pi_{g,1} > 0$. Note that, if W is a fixed deterministic weight function, P_W^∞ and G_W^∞ are the uniform limits of $P_W^{(m)}$ and $G_W^{(m)}$ when $m \rightarrow \infty$. If W is such that G_W^∞ is nondecreasing, we also define

$$u_W^\infty = \mathcal{I}(G_W^\infty). \tag{B.2}$$

Recall that $K^\infty = \{w \in \mathbb{R}_+^G : \sum_g \pi_g \bar{\pi}_{g,0} w_g \leq 1\}$. It is the asymptotic version of K^m . We now define oracle optimal weights over K^∞ for $G^\infty(u)$ and $P^\infty(u)$, for all $u > 0$.

Lemma B.1. Fix an $u \in [0, 1]$. Then $\arg \max_{w \in K^\infty} G_w^\infty(u)$ is non empty.

If $0 < \alpha u \leq \bar{\pi}_0$ it is a singleton. Its only element w^* belongs to $[0, \frac{1}{\alpha u}]^G$ and checks $\sum_g \pi_g \bar{\pi}_{g,0} w_g^* = 1$. If $\alpha u \geq \bar{\pi}_0$ it is included in $[\frac{1}{\alpha u}, \infty)^G$. Finally $\max_{w \in K^\infty} G_w^\infty(u) = 1$ if and only if $\alpha u \geq \bar{\pi}_0$.

The same statements are true for P^∞ , except that $\max_{w \in K^\infty} P_w^\infty(u) = 1 - \pi_0$ if and only if $\alpha u \geq \bar{\pi}_0$.

Proof. The function $w \mapsto G_w^\infty(u)$ is continuous over the compact K^∞ so it has a maximum. Note that $\max_{w \in K^\infty} G_w^\infty(0) = 0$ and $\arg \max_{w \in K^\infty} G_w^\infty(0) = K^\infty$. For the rest of the proof u is greater than 0.

First we show that any $w^* \in \arg \max_{w \in K^\infty} G_w^\infty(u)$ belongs to $[0, \frac{1}{\alpha u}]^G$ or $[\frac{1}{\alpha u}, \infty)^G$. If not, there is $w^* \in \arg \max_{w \in K^\infty} G_w^\infty(u)$ such that $\alpha u w_{g_1}^* > 1$ and $\alpha u w_{g_2}^* < 1$ for some $g_1, g_2 \leq G$. Now then we define \tilde{w} such that $\tilde{w}_g = w_g^*$ for all $g \notin \{g_1, g_2\}$, $\tilde{w}_{g_1} = \frac{1}{\alpha u}$ and

$$\tilde{w}_{g_2} = w_{g_2}^* + \left(w_{g_1}^* - \frac{1}{\alpha u} \right) \frac{\pi_{g_1} \bar{\pi}_{g_1,0}}{\pi_{g_2} \bar{\pi}_{g_2,0}} > w_{g_2}^*.$$

So \tilde{w} belongs to K^∞ and satisfies

$$\begin{aligned} G_{\tilde{w}}^\infty(u) &= \sum_{g \neq g_1, g_2} \pi_g D_g(\alpha u w_g^*) + \pi_{g_1} + \pi_{g_2} D_{g_2}(\alpha u \tilde{w}_{g_2}) \\ &> \sum_{g \neq g_1, g_2} \pi_g D_g(\alpha u w_g^*) + \pi_{g_1} + \pi_{g_2} D_{g_2}(\alpha u w_{g_2}^*) = G_{w^*}^\infty(u), \end{aligned}$$

because D_g is increasing over $[0, 1]$ and then constant equal to 1. This contradicts the definition of w^* so is impossible.

Next we distinct three cases.

(i) $\alpha u = \bar{\pi}_0$. Then $w_0 = (\frac{1}{\alpha u}, \dots, \frac{1}{\alpha u}) = (\frac{1}{\bar{\pi}_0}, \dots, \frac{1}{\bar{\pi}_0})$ is obviously an element of $\arg \max_{w \in K^\infty} G_w^\infty(u)$ because

$$G_{w_0}^\infty(u) = \sum_{g=1}^G \pi_g D_g(1) = 1,$$

and we easily check that $\sum_g \pi_g \bar{\pi}_{g,0}(w_0)_g = 1$. Thus for every $w \in K^\infty$ distinct from w_0 , there must exist a $g_1 \in \{1, \dots, G\}$ such that $\alpha u w_{g_1} < 1$, so $D_{g_1}(\alpha u w_{g_1}) < 1$ and $G_w^\infty(u) < \sum_g \pi_g = 1$: w_0 is the only element of $\arg \max_{w \in K^\infty} G_w^\infty(u)$.

(ii) $\alpha u < \bar{\pi}_0$. If a $w^* \in \arg \max_{w \in K^\infty} G_w^\infty(u)$ exists in $[\frac{1}{\alpha u}, \infty)^G$, then $w_g^* \geq \frac{1}{\alpha u} > \frac{1}{\bar{\pi}_0}$ and $\sum_g \pi_g \bar{\pi}_{g,0} w_g^* > 1$ which is impossible. So

$$\arg \max_{w \in K^\infty} G_w^\infty(u) = \arg \max_{w \in K^\infty \cap [0, \frac{1}{\alpha u}]^G} G_w^\infty(u).$$

The function $w \mapsto G_w^\infty(u)$ is strictly concave over the convex set $K^\infty \cap [0, \frac{1}{\alpha u}]^G$ because $\pi_{g,1} > 0$ and D_g is strictly concave over $[0, 1]$ for all g , hence the maximum is unique.

We showed that the only $w^* \in \arg \max_{w \in K^\infty} G_w^\infty(u)$ is not in $[\frac{1}{\alpha u}, \infty)^G$ so there exists $g_1 \leq G$ such that $\alpha u w_{g_1}^* < 1$ thus $G_{w^*}^\infty(u) < 1$.

Furthermore $\sum_g \pi_g \bar{\pi}_{g,0} w_g^* = 1$: if not there is a \tilde{w} with $\tilde{w}_{g_1} > w_{g_1}^*$ and $\tilde{w}_g = w_g^*$ for all $g \neq g_1$ such that $\tilde{w} \in K^\infty$ and $G_{\tilde{w}}^\infty(u) > G_{w^*}^\infty(u)$ which is impossible.

(iii) $\alpha u > \bar{\pi}_0$. So $u > \frac{\bar{\pi}_0}{\alpha}$ and obviously

$$\max_{w \in K^\infty} G_w^\infty(u) \geq \max_{w \in K^\infty} G_w^\infty\left(\frac{\bar{\pi}_0}{\alpha}\right) = G_{w_0}^\infty\left(\frac{\bar{\pi}_0}{\alpha}\right) = 1,$$

as stated in case (i). So $\max_{w \in K^\infty} G_w^\infty(u) = 1$ and then the elements w^* of $\arg \max_{w \in K^\infty} G_w^\infty(u)$ are the ones fulfilling $D_g(\alpha u w_g^*) = 1$ for all g that is $w^* \in [\frac{1}{\alpha u}, \infty)^G$.

The proof is similar for P^∞ , by replacing D_g by $\pi_{g,1} F_g$. \square

From now on, $W^*(u)$ denotes an element of $\arg \max_{w \in K^\infty} G_w^\infty(u)$ (just like we write $\widehat{W}^*(u)$ as an element of $\arg \max_{w \in K^m} \widehat{G}_w(u)$), our results will not depend on the chosen element of the argmax. Next Lemma gives some properties on the function $G_{W^*}^\infty$, among them $G_{W^*}^\infty$ is nondecreasing which allow us to define

$$u^* = u_{W^*}^\infty = \mathcal{I}(G_{W^*}^\infty). \quad (\text{B.3})$$

Lemma B.2. $G_{W^*}^\infty$ is nondecreasing and $u^* > 0$. $G_{W^*}^\infty$ is strictly concave over $[0, \frac{\bar{\pi}_0}{\alpha} \wedge 1]$ and, if $\alpha \geq \bar{\pi}_0$, constant equal to 1 over $[\frac{\bar{\pi}_0}{\alpha}, 1]$.

In particular, (i) $u^* = 1$ if and only if $\alpha \geq \bar{\pi}_0$ (ii) the function $u \mapsto G_{W^*}^\infty(u)/u$ is decreasing over $(0, 1]$ (iii) $G_{W^*}^\infty$ is continuous over $[0, 1]$.

Proof. $G_{W^*}^\infty$ is nondecreasing by exactly the same argument as in the proof of Lemma A.6. The result can be strengthened thanks to Lemma B.1, by writing, for $u < u' \leq \frac{\bar{\pi}_0}{\alpha} \wedge 1$, that $G_{W^*(u)}^\infty(u') > G_{W^*(u)}^\infty(u)$ because $1 > G_{W^*}^\infty(u)$. So $G_{W^*}^\infty$ is increasing on $[0, \frac{\bar{\pi}_0}{\alpha} \wedge 1]$.

To prove that $u^* > 0$, take some $w \in K^\infty$ such that

$$\alpha > \frac{1}{\sum_g \pi_g w_g (\pi_{g,0} + \pi_{g,1} f_g(0^+))} \geq \alpha^*.$$

Because the expression above is continuous of the w_g , they can always be chosen nonzero. We have $u^* \geq u_w^\infty$ because $G_{W^*}^\infty \geq G_w^\infty$. Then we have, for $x > 0$, $x \rightarrow 0^+$,

$$\begin{aligned} \frac{G_w^\infty(x) - G_w^\infty(0)}{x - 0} &= \frac{G_w^\infty(x)}{x} = \sum_g \pi_g \pi_{g,0} \alpha w_g + \sum_g \pi_g \pi_{g,1} \alpha w_g \frac{F_g(\alpha x w_g)}{\alpha x w_g} \\ &\rightarrow \alpha \sum_g \pi_g w_g (\pi_{g,0} + \pi_{g,1} f_g(0^+)) > 1, \end{aligned}$$

so $G_w^\infty(u) > u$ in the neighborhood of 0^+ , which entails $u_w^\infty > 0$.

Now take $a, b \in [0, \frac{\bar{\pi}_0}{\alpha} \wedge 1]$ with $a < b$ and $\lambda \in (0, 1)$, by Lemma B.1 $\alpha a W_g^*(a), \alpha b W_g^*(b) \leq 1$ and then, for all g :

$$D_g(\lambda \alpha a W_g^*(a) + (1 - \lambda) \alpha b W_g^*(b)) \geq \lambda D_g(\alpha a W_g^*(a)) + (1 - \lambda) D_g(\alpha b W_g^*(b)).$$

Moreover, because $G_{W^*}^\infty(a) < G_{W^*}^\infty(b)$, for at least one g_1 we have $a W_{g_1}^*(a) \neq b W_{g_1}^*(b)$ and by strict concavity of D_{g_1} the inequality above is strict for g_1 . Then define $\tilde{w}_g = \frac{\lambda a W_g^*(a) + (1 - \lambda) b W_g^*(b)}{\lambda a + (1 - \lambda) b}$. We have $\tilde{w} \in K^\infty$ and then for all g :

$$\pi_g D_g(\alpha(\lambda a + (1 - \lambda) b) \tilde{w}_g) \geq \lambda \pi_g D_g(\alpha a W_g^*(a)) + (1 - \lambda) \pi_g D_g(\alpha b W_g^*(b)),$$

the inequality being strict for g_1 . Finally by summing:

$$G_{W^*}^\infty(\lambda a + (1 - \lambda) b) \geq G_w^\infty(\lambda a + (1 - \lambda) b) > \lambda G_{W^*}^\infty(a) + (1 - \lambda) G_{W^*}^\infty(b).$$

Additionally, $G_{W^*}^\infty(u) = 1$ for $\alpha u \geq \bar{\pi}_0$ comes from Lemma B.1. The fact that $u^* = 1 \iff \alpha \geq \bar{\pi}_0$ follows directly from the previous statements and Lemma B.1. The decreasingness of $u \mapsto G_{W^*}^\infty(u)/u$ is straightforward from strict concavity properties because it is the slope of the line between the origin and the graph of $G_{W^*}^\infty$ at abscissa $u > 0$. Previous statements imply that $G_{W^*}^\infty$ is continuous at least over $(0, \frac{\bar{\pi}_0}{\alpha} \wedge 1)$ and, if $\alpha \geq \bar{\pi}_0$, over $[\frac{\bar{\pi}_0}{\alpha}, 1]$. K^∞ is bounded, let B such that $|w_g| \leq B$ for all $w \in K^\infty$, then $G_{W^*}^\infty(u) \leq \sum_g \frac{m_g}{m} D_g(\alpha u B) \rightarrow 0$ when $u \rightarrow 0$ which gives the continuity in 0. As in the proof of Lemma A.1, the continuity in $\frac{\bar{\pi}_0}{\alpha} \wedge 1$ is given by the combination of concavity and nondecrease. \square

Remark B.1. The case $\alpha \geq \bar{\pi}_0$ is rarely met in practice because α is chosen small and the signal is assumed to be sparse (so $\bar{\pi}_0$ is large) but it is kept to cover all situations. It confirms the intuitive idea that in this situation the best strategy is to reject all hypotheses because then the FDP is equal to $\pi_0 \leq \bar{\pi}_0 \leq \alpha$.

Remark B.2. For a weight vector $w \in \mathbb{R}_+^G$, G_w^∞ is obviously continuous. Moreover if $w \neq 0$, let $M = \max_{0 \leq u \leq 1} G_w^\infty(u) \leq 1$ and $u^\diamond = \min\{u : G_w^\infty(u) = M\} > 0$, then G_w^∞ is strictly concave over $[0, u^\diamond]$ and constant equal to M on $[u^\diamond, 1]$, hence $u \mapsto G_w^\infty(u)/u$ is decreasing. So whether $w = 0$ or not, $\mathcal{I}(\cdot)$ is continuous in G_w^∞ by Lemma A.4.

Remark B.3. The proof of the strict concavity of $G_{W^*}^\infty$ can easily be adapted to show the (non necessary strict) concavity of $\tilde{G}_{\tilde{W}^*}$ when $\tilde{D}_g = \text{LCM}(\hat{D}_g)$.

Figure 6 illustrates all the properties stated in Lemma B.2, with the two cases $\alpha \geq \bar{\pi}_0$ and $\alpha < \bar{\pi}_0$.

The next Lemma justifies the intuitive idea that maximizing the rejections and the power is the same thing (as exposed in Section 3.2), but only under (CE) or (ED)+(EE).

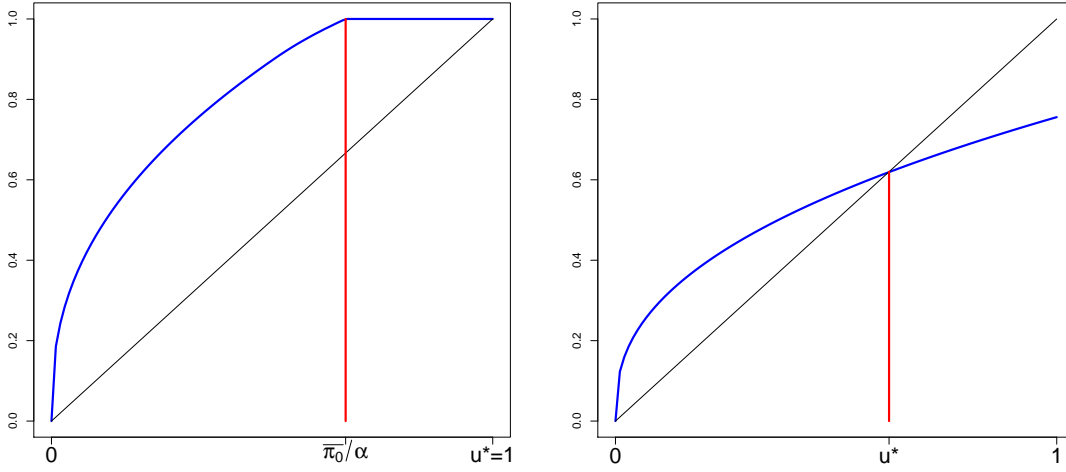


Fig 6: Plot of $u \mapsto G_{W*}^\infty(u)$ when $\alpha \geq \bar{\pi}_0$ (left panel) and $\alpha < \bar{\pi}_0$ (right panel).

Lemma B.3. In (CE) or (ED)+(EE) case, for all $u \in [0, 1]$,

$$\arg \max_{w \in K^\infty} G_w^\infty(u) = \arg \max_{w \in K^\infty} P_w^\infty(u).$$

In particular, P_{W*}^∞ is continuous nondecreasing.

Proof. First, $\arg \max_{w \in K^\infty} G_w^\infty(0) = \arg \max_{w \in K^\infty} P_w^\infty(0) = K^\infty$, so assume $u > 0$. If $\alpha u \geq \bar{\pi}_0$, $\max_{w \in K^\infty} G_w^\infty(u) = 1$ and $\max_{w \in K^\infty} P_w^\infty(u) = 1 - \pi_0$ by Lemma B.1, thus $\arg \max_{w \in K^\infty} G_w^\infty(u)$ and $\arg \max_{w \in K^\infty} P_w^\infty(u)$ are both equal to the set of weights $w \in K^\infty$ such that $\alpha u w_g \geq 1$ for all g .

Now if $\alpha u \leq \bar{\pi}_0$, both $\arg \max$ are singletons. Take w^* the only element of $\arg \max_{w \in K^\infty} P_w^\infty(u)$ and write, for all $w \in K^\infty$,

$$\begin{aligned} G_w^\infty(u) &= \sum_g \pi_g \pi_{g,0} U(\alpha u w_g) + P_w^\infty(u) \\ &\leq \alpha u \sum_g \pi_g \pi_{g,0} w_g^* + P_{w^*}^\infty(u). \end{aligned}$$

In case (CE), $\pi_{g,0} = \bar{\pi}_{g,0}$ so $\sum_g \pi_g \pi_{g,0} w_g \leq 1$ and

$$\begin{aligned} G_w^\infty(u) &\leq \alpha u + P_{w^*}^\infty(u) = \alpha u \sum_g \pi_g \pi_{g,0} w_g^* + P_{w^*}^\infty(u) \\ &= \sum_g \pi_g \pi_{g,0} U(\alpha u w_g^*) + P_{w^*}^\infty(u) = G_{w^*}^\infty(u), \end{aligned}$$

because $\sum_g \pi_g \pi_{g,0} w_g^* = 1$ and $\alpha u w_g^* \leq 1$ for all g , by Lemma B.1. This means that w^* is also the unique element of $\arg \max_{w \in K^\infty} G_w^\infty(u)$.

In case (ED)+(EE), we write, as in the last part of the proof of Theorem 5.1,

$$G_w^\infty(u) \leq \alpha u \frac{\pi_0}{\bar{\pi}_0} + P_{w^*}^\infty(u) = \alpha u \sum_g \pi_g \pi_0 w_g^* + P_{w^*}^\infty(u) = G_{w^*}^\infty(u),$$

and we get the same result. Finally the properties on P_{W*}^∞ are obtained by the same proof as Lemma B.2. \square

The next lemma is only a deterministic tool used in the proof of Lemma C.4. Define the distance d of a weight vector w to a subset S of \mathbb{R}_+^G by $d(w, S) = \inf_{\bar{w} \in S} \max_g |w_g - \bar{w}_g|$. Let $M_u = \arg \max_{w \in K^\infty} G_w^\infty(u)$ to lighten notations.

Lemma B.4. *Take some $u \in (0, 1]$. Then we have:*

$$\forall \epsilon > 0, \exists \xi > 0, \forall w \in K^\infty, |G_w^\infty(u) - G_{W^*}^\infty(u)| \leq \xi \Rightarrow d(w, M_u) < \epsilon.$$

In particular, if $\alpha u \leq \bar{\pi}_0$,

$$\forall \epsilon > 0, \exists \xi > 0, \forall w \in K^\infty, |G_w^\infty(u) - G_{W^*}^\infty(u)| \leq \xi \Rightarrow \max_g |w_g - W_g^*(u)| < \epsilon, \quad (\text{B.4})$$

and if $\alpha u \geq \bar{\pi}_0$,

$$\forall \epsilon > 0, \exists \xi > 0, \forall w \in K^\infty, |G_w^\infty(u) - G_{W^*}^\infty(u)| \leq \xi \Rightarrow (\forall g, \alpha u w_g > 1 - \epsilon). \quad (\text{B.5})$$

Proof. If the statement is false, there exists some $\epsilon > 0$ and some sequence $(w_n)_{n \geq 1}$ converging to a w^ℓ in K^∞ (because K^∞ is compact), such that $d(w_n, M_u) \geq \epsilon$ and

$$|G_{w_n}^\infty(u) - G_{W^*}^\infty(u)| \rightarrow 0.$$

By continuity of D_g , $G_{w^\ell}^\infty(u) = G_{W^*}^\infty(u)$ so $w^\ell \in M_u$ which contradicts $d(w^\ell, M_u) \geq \epsilon$. If $\alpha u \leq \bar{\pi}_0$, M_u is a singleton by Lemma B.1, hence (B.4). However, if $\alpha u \geq \bar{\pi}_0$, $M_u = \{w \in \mathbb{R}_+^G : \alpha u w_g \geq 1 \forall g\}$ by Lemma B.1, hence (B.5). \square

Appendix C: Convergence lemmas

Recall that $\|\cdot\|$ is the sup norm for the bounded functions on their definition domain: $\|f\| = \sup_{u \in [0,1]} |f(u)|$ or $\|f\| = \sup_{t \in \mathbb{R}} |f(t)|$.

Lemma C.1. *The following quantities converge to 0 almost surely:*

$$\sup_{w \in \mathbb{R}_+^G} \left\| \hat{H}_w - H_w^\infty \right\|, \sup_{w \in \mathbb{R}_+^G} \left\| \hat{P}_w - P_w^\infty \right\|, \sup_{w \in \mathbb{R}_+^G} \left\| \hat{G}_w - G_w^\infty \right\|, \text{ and, for all } g, \left\| \hat{D}_g - D_g \right\|.$$

Furthermore, for any $(\tilde{D}_g)_g$ such that $\left\| \tilde{D}_g - D_g \right\| \xrightarrow{\mathbb{P}} 0$,

$$\sup_{w \in \mathbb{R}_+^G} \left\| \tilde{G}_w - G_w^\infty \right\| \xrightarrow{\mathbb{P}} 0. \quad (\text{C.1})$$

Proof. By using the same proof as the one of the Glivenko-Cantelli theorem, we get from (2.1) and (2.2) that, for all g ,

$$\left\| \frac{1}{m_{g,0}} \sum_{i=1}^{m_g} \mathbf{1}_{\{p_{g,i} \leq \cdot, H_{g,i}=0\}} - U \right\| \xrightarrow{a.s.} 0,$$

and

$$\left\| \frac{1}{m_{g,1}} \sum_{i=1}^{m_g} \mathbf{1}_{\{p_{g,i} \leq \cdot, H_{g,i}=1\}} - F_g \right\| \xrightarrow{a.s.} 0.$$

Next, we write that

$$\begin{aligned} \left\| \frac{1}{m_g} \sum_{i=1}^{m_g} \mathbf{1}_{\{p_{g,i} \leq \cdot, H_{g,i}=0\}} - \pi_{g,0} U \right\| &\leq \left| \frac{m_{g,0}}{m_g} - \pi_{g,0} \right| \\ &\quad + \pi_{g,0} \left\| \frac{1}{m_{g,0}} \sum_{i=1}^{m_g} \mathbf{1}_{\{p_{g,i} \leq \cdot, H_{g,i}=0\}} - U \right\| \\ &\xrightarrow{a.s.} 0, \end{aligned}$$

and similarly $\left\| \frac{1}{m_g} \sum_{i=1}^{m_g} \mathbf{1}_{\{p_{g,i} \leq \cdot, H_{g,i}=1\}} - \pi_{g,1} F_g \right\| \xrightarrow{a.s.} 0$. So by summing, $\left\| \widehat{D}_g - D_g \right\| \xrightarrow{a.s.} 0$. Apply the triangular inequality once again to get that $\left\| \frac{1}{m} \sum_{i=1}^{m_g} \mathbf{1}_{\{p_{g,i} \leq \cdot, H_{g,i}=0\}} - \pi_g \pi_{g,0} U \right\| \xrightarrow{a.s.} 0$ which implies

$$\sup_{w \in \mathbb{R}_+^G} \left\| \widehat{H}_w - H_w^\infty \right\| \leq \sum_{g=1}^G \left\| \frac{1}{m} \sum_{i=1}^{m_g} \mathbf{1}_{\{p_{g,i} \leq \cdot, H_{g,i}=0\}} - \pi_g \pi_{g,0} U \right\| \xrightarrow{a.s.} 0.$$

Similarly $\sup_{w \in \mathbb{R}_+^G} \left\| \widehat{P}_w - P_w^\infty \right\| \xrightarrow{a.s.} 0$ and $\sup_{w \in \mathbb{R}_+^G} \left\| \widehat{G}_w - G_w^\infty \right\| \xrightarrow{a.s.} 0$ by sum.

Finally,

$$\sup_{w \in \mathbb{R}_+^G} \left\| \widetilde{G}_w - G_w^\infty \right\| \leq \sum_g \left(\left| \frac{m_g}{m} - \pi_g \right| + \pi_g \left\| \widetilde{D}_g - D_g \right\| \right) \xrightarrow{\mathbb{P}} 0. \quad \square$$

From now on \widetilde{D}_g is assumed to converge uniformly to D_g in probability and that $\widetilde{W}^*(u) \in \arg \max_{w \in K^m} \widetilde{G}_w(u)$ exists for all u .

Next Lemma is the main technical one (with the longest proof).

Lemma C.2. *We have the following convergence in probability:*

$$\left\| \widetilde{G}_{\widetilde{W}^*} - G_{\widetilde{W}^*}^\infty \right\| \xrightarrow{\mathbb{P}} 0.$$

Proof. First,

$$\left\| \widetilde{G}_{\widetilde{W}^*} - G_{\widetilde{W}^*}^\infty \right\| \leq \sup_{w \in \mathbb{R}_+^G} \left\| \widetilde{G}_w - G_w^\infty \right\| + \left\| G_{\widetilde{W}^*}^\infty - G_{\widetilde{W}^*}^\infty \right\|,$$

where the first term tends to 0 by (C.1), so we work on the second term.

The main idea is to use the maximality of $\widetilde{G}_w(u)$ in $\widetilde{W}^*(u)$ and the maximality of $G_w^\infty(u)$ in $W^*(u)$. The problem is that one is a maximum over K^m and the other is over K^∞ . The solution consists in defining small variations of $\widetilde{W}^*(u)$ and $W^*(u)$ to place them respectively in K^∞ and K^m .

Let $\widetilde{W}_g^\dagger(u) = \frac{m_g \widehat{\pi}_{g,0}}{m \pi_g \bar{\pi}_{g,0}} \widetilde{W}_g^*(u)$. Then $\widetilde{W}^\dagger(u) \in K^\infty$ and

$$\begin{aligned} \left\| \widetilde{W}_g^\dagger - \widetilde{W}_g^* \right\| &= \left| \frac{m_g \widehat{\pi}_{g,0}}{m \pi_g \bar{\pi}_{g,0}} - 1 \right| \left\| \widetilde{W}_g^* \right\| \\ &\leq \left| \frac{m_g \widehat{\pi}_{g,0}}{m \pi_g \bar{\pi}_{g,0}} - 1 \right| \frac{m}{m_g \widehat{\pi}_{g,0}} \xrightarrow{\mathbb{P}} 0 \text{ because } \frac{m_g}{m} \widehat{\pi}_{g,0} \xrightarrow{\mathbb{P}} \pi_g \bar{\pi}_{g,0}, \end{aligned}$$

which in turn implies that

$$\begin{aligned} \left\| G_{\widetilde{W}^\dagger}^\infty - G_{\widetilde{W}^*}^\infty \right\| &\leq \sum_g \pi_g \sup_u \left| D_g \left(\alpha u \widetilde{W}_g^\dagger(u) \right) - D_g \left(\alpha u \widetilde{W}_g^*(u) \right) \right| \\ &\xrightarrow{\mathbb{P}} 0, \end{aligned} \quad (\text{C.2})$$

because D_g is uniformly continuous over \mathbb{R}_+ . Likewise, we define $W_g^\dagger(u) = \frac{m \pi_g \bar{\pi}_{g,0}}{m_g \widehat{\pi}_{g,0}} W_g^*(u)$. Therefore $W^\dagger(u) \in K^m$,

$$\left\| W_g^\dagger - W_g^* \right\| \leq \left| \frac{m \pi_g \bar{\pi}_{g,0}}{m_g \widehat{\pi}_{g,0}} - 1 \right| \frac{1}{\pi_g \bar{\pi}_{g,0}} \xrightarrow{\mathbb{P}} 0,$$

and

$$\left\| G_{W^\dagger}^\infty - G_{W^*}^\infty \right\| \leq \sum_g \pi_g \sup_u \left| D_g \left(\alpha u W_g^\dagger(u) \right) - D_g \left(\alpha u W_g^*(u) \right) \right| \xrightarrow{\mathbb{P}} 0. \quad (\text{C.3})$$

With (C.1) and (C.2), we deduce that

$$\begin{aligned} \left\| \tilde{G}_{\tilde{W}^\dagger} - \tilde{G}_{\tilde{W}^*} \right\| &\leq \left\| \tilde{G}_{\tilde{W}^\dagger} - G_{\tilde{W}^\dagger}^\infty \right\| + \left\| G_{\tilde{W}^\dagger}^\infty - G_{\tilde{W}^*}^\infty \right\| \\ &\quad + \left\| G_{\tilde{W}^*}^\infty - \tilde{G}_{\tilde{W}^*} \right\| \\ &\xrightarrow{\mathbb{P}} 0, \end{aligned} \tag{C.4}$$

and likewise with (C.1) and (C.3) we have

$$\left\| \tilde{G}_{W^\dagger} - \tilde{G}_{W^*} \right\| \xrightarrow{\mathbb{P}} 0. \tag{C.5}$$

Combining (C.1), (C.2), (C.4), (C.5), and the maximalities of $\tilde{G}_{\tilde{W}^*}(u)$ and $G_{W^*}^\infty(u)$ will finish the proof. As a start, write

$$\left\| G_{\tilde{W}^*}^\infty - G_{W^*}^\infty \right\| \leq \left\| G_{\tilde{W}^*}^\infty - G_{\tilde{W}^\dagger}^\infty \right\| + \left\| G_{\tilde{W}^\dagger}^\infty - G_{W^*}^\infty \right\|,$$

with $\left\| G_{\tilde{W}^*}^\infty - G_{\tilde{W}^\dagger}^\infty \right\| \xrightarrow{\mathbb{P}} 0$ by (C.2), and, for all u ,

$$\left| G_{\tilde{W}^\dagger}^\infty(u) - G_{W^*}^\infty(u) \right| = G_{W^*}^\infty(u) - G_{\tilde{W}^\dagger}^\infty(u),$$

by maximality of $G_{W^*}^\infty(u)$ over K^∞ . Then

$$\begin{aligned} \sup_u \left(G_{W^*}^\infty(u) - G_{\tilde{W}^\dagger}^\infty(u) \right) &\leq \sup_u \left(G_{W^*}^\infty(u) - \tilde{G}_{W^*}(u) \right) \\ &\quad + \sup_u \left(\tilde{G}_{W^*}(u) - \tilde{G}_{\tilde{W}^\dagger}(u) \right) \\ &\quad + \sup_u \left(\tilde{G}_{\tilde{W}^\dagger}(u) - G_{\tilde{W}^\dagger}^\infty(u) \right), \end{aligned}$$

with $\sup_u \left(G_{W^*}^\infty(u) - \tilde{G}_{W^*}(u) \right) \xrightarrow{\mathbb{P}} 0$ and $\sup_u \left(\tilde{G}_{\tilde{W}^\dagger}(u) - G_{\tilde{W}^\dagger}^\infty(u) \right) \xrightarrow{\mathbb{P}} 0$ by (C.1).

Finally,

$$\begin{aligned} \sup_u \left(\tilde{G}_{W^*}(u) - \tilde{G}_{\tilde{W}^\dagger}(u) \right) &\leq \sup_u \left(\tilde{G}_{W^*}(u) - \tilde{G}_{W^\dagger}(u) \right) \\ &\quad + \sup_u \left(\tilde{G}_{W^\dagger}(u) - \tilde{G}_{\tilde{W}^*}(u) \right) \\ &\quad + \sup_u \left(\tilde{G}_{\tilde{W}^*}(u) - \tilde{G}_{\tilde{W}^\dagger}(u) \right), \end{aligned}$$

with $\sup_u \left(\tilde{G}_{W^*}(u) - \tilde{G}_{W^\dagger}(u) \right) \xrightarrow{\mathbb{P}} 0$ by (C.5) and $\sup_u \left(\tilde{G}_{\tilde{W}^*}(u) - \tilde{G}_{\tilde{W}^\dagger}(u) \right) \xrightarrow{\mathbb{P}} 0$ by (C.4). As a consequence there exists a random variable $V_m \xrightarrow{\mathbb{P}} 0$ such that

$$\left\| \tilde{G}_{\tilde{W}^*} - G_{W^*}^\infty \right\| \leq \sup_u \left(\tilde{G}_{W^\dagger}(u) - \tilde{G}_{\tilde{W}^*}(u) \right) + V_m,$$

but $\tilde{G}_{W^\dagger}(u) - \tilde{G}_{\tilde{W}^*}(u) \leq 0$ by maximality of $\tilde{G}_{\tilde{W}^*}(u)$ over K^m , so

$$\left\| \tilde{G}_{\tilde{W}^*} - G_{W^*}^\infty \right\| \leq V_m \xrightarrow{\mathbb{P}} 0. \quad \square$$

Next Lemma is a direct application of Lemma A.4. Recall that $u^* = u_{W^*}^\infty$ (see (B.3)) and let

$$\tilde{u} = \tilde{u}_{\tilde{W}^*} = \mathcal{I} \left(\tilde{G}_{\tilde{W}^*} \right) \tag{C.6}$$

Lemma C.3. *We have the following convergences in probability:*

$$\begin{cases} \tilde{u} & \xrightarrow{\mathbb{P}} u^* \\ \tilde{G}_{\tilde{W}^*}(\tilde{u}) & \xrightarrow{\mathbb{P}} G_{W^*}^\infty(u^*). \end{cases}$$

Proof. $u \mapsto G_{W^*}^\infty(u)/u$ is nondecreasing and $G_{W^*}^\infty$ is continuous by Lemma B.2 so by Lemma A.4 $\mathcal{I}(\cdot)$ is continuous in $G_{W^*}^\infty$: let $\gamma > 0$ and η_γ as in the proof of Lemma A.4, then

$$\mathbb{P}(|\tilde{u} - u^*| \leq \gamma) \geq \mathbb{P}\left(\|\tilde{G}_{\tilde{W}^*} - G_{W^*}^\infty\| \leq \eta_\gamma\right) \xrightarrow{\text{Lemma C.2}} 1.$$

Second result follows immediately because $\tilde{G}_{\tilde{W}^*}(\tilde{u}) = \tilde{u}$ and $G_{W^*}^\infty(u^*) = u^*$ by Lemma A.4. \square

Lemma C.4.

(i) *If $\alpha \leq \bar{\pi}_0$, $\tilde{W}^*(\tilde{u}) \xrightarrow{\mathbb{P}} W^*(u^*)$.*

(ii) *If $\alpha \geq \bar{\pi}_0$, the inferior limit in probability of $\alpha \tilde{u} \tilde{W}_g^*(\tilde{u})$ is greater than or equal to 1, uniformly in g , which reads formally:*

$$\forall \epsilon > 0, \mathbb{P}\left(\forall g, \alpha \tilde{u} \tilde{W}_g^*(\tilde{u}) > 1 - \epsilon\right) \longrightarrow 1.$$

Proof. First, we use the same trick as in the proof of Lemma C.2: let $\tilde{W}_g^\dagger(u) = \frac{m_g \bar{\pi}_{g,0}}{m \pi_g \bar{\pi}_{g,0}} \tilde{W}_g^*(u)$ such that $\tilde{W}^\dagger(u) \in K^\infty$ and $\|\tilde{W}_g^* - \tilde{W}_g^\dagger\| \xrightarrow{\mathbb{P}} 0$.

Let us show that $\left|G_{\tilde{W}^\dagger(\tilde{u})}^\infty(u^*) - G_{W^*}^\infty(u^*)\right| \xrightarrow{\mathbb{P}} 0$ to apply then Lemma B.4 (always possible because $u^* > 0$). We have

$$\begin{aligned} \left|G_{\tilde{W}^\dagger(\tilde{u})}^\infty(u^*) - G_{W^*}^\infty(u^*)\right| &\leq \left|G_{\tilde{W}^\dagger(\tilde{u})}^\infty(u^*) - G_{\tilde{W}^*}^\infty(\tilde{u})\right| \\ &\quad + \left|G_{\tilde{W}^*}^\infty(\tilde{u}) - \tilde{G}_{\tilde{W}^*}(\tilde{u})\right| \\ &\quad + \left|\tilde{G}_{\tilde{W}^*}(\tilde{u}) - G_{W^*}^\infty(u^*)\right|. \end{aligned}$$

First term converges to 0 because for all g , D_g is uniformly continuous and

$$\begin{aligned} \left|\alpha u^* \tilde{W}_g^\dagger(\tilde{u}) - \alpha \tilde{u} \tilde{W}_g^*(\tilde{u})\right| &\leq \left|\alpha u^* \tilde{W}_g^\dagger(\tilde{u}) - \alpha u^* \tilde{W}_g^*(\tilde{u})\right| + \left|\alpha u^* \tilde{W}_g^*(\tilde{u}) - \alpha \tilde{u} \tilde{W}_g^*(\tilde{u})\right| \\ &\leq \|\tilde{W}_g^\dagger - \tilde{W}_g^*\| + |u^* - \tilde{u}| \frac{m}{m_g \hat{\pi}_{g,0}} \xrightarrow{\mathbb{P}} 0. \end{aligned} \tag{C.7}$$

Apply (C.1) to the second term and Lemma C.3 to the third.

(i) If $\alpha \leq \bar{\pi}_0$, then $\alpha u^* \leq \bar{\pi}_0$ and by equation (B.4), $\tilde{W}^\dagger(\tilde{u}) \xrightarrow{\mathbb{P}} W^*(u^*)$. But for all g

$$\left|\tilde{W}_g^*(\tilde{u}) - W_g^*(u^*)\right| \leq \|\tilde{W}_g^* - \tilde{W}_g^\dagger\| + \left|\tilde{W}_g^\dagger(\tilde{u}) - W_g^*(u^*)\right|,$$

and then $\tilde{W}^*(\tilde{u}) \xrightarrow{\mathbb{P}} W^*(u^*)$.

(ii) If $\alpha \geq \bar{\pi}_0$, $u^* = 1$ by Lemma B.2 and by equation (B.5),

$$\forall \epsilon > 0, \mathbb{P}\left(\forall g, \alpha u^* \tilde{W}_g^\dagger(\tilde{u}) > 1 - \frac{\epsilon}{2}\right) \longrightarrow 1.$$

By equation (C.7) we also have

$$\forall \epsilon > 0, \mathbb{P}\left(\forall g, \left|\alpha u^* \tilde{W}_g^\dagger(\tilde{u}) - \alpha \tilde{u} \tilde{W}_g^*(\tilde{u})\right| \leq \frac{\epsilon}{2}\right) \longrightarrow 1,$$

and by combining the two we get the desired result. \square

Lemma C.5. *We have the following convergences in probability:*

$$\widehat{G}_{\widetilde{W}^*}(\tilde{u}) \xrightarrow{\mathbb{P}} G_{\widetilde{W}^*}^\infty(u^*),$$

$$\widehat{H}_{\widetilde{W}^*}(\tilde{u}) \xrightarrow{\mathbb{P}} H_{\widetilde{W}^*}^\infty(u^*).$$

Proof. We have

$$\left| \widehat{G}_{\widetilde{W}^*}(\tilde{u}) - G_{\widetilde{W}^*}^\infty(u^*) \right| \leq \sup_{w \in \mathbb{R}_+^G} \left\| \widehat{G}_w - G_w^\infty \right\| + \left| G_{\widetilde{W}^*}^\infty(\tilde{u}) - G_{\widetilde{W}^*}^\infty(u^*) \right|.$$

Hence, by Lemma C.1, we only need to show that $H_{\widetilde{W}^*}^\infty(\tilde{u}) \xrightarrow{\mathbb{P}} H_{\widetilde{W}^*}^\infty(u^*)$.

(i) If $\alpha \leq \pi_0$, $\tilde{u} \xrightarrow{\mathbb{P}} u^*$ and $\widetilde{W}^*(\tilde{u}) \xrightarrow{\mathbb{P}} W^*(u^*)$ by Lemma C.4. Then $\alpha \tilde{u} \widetilde{W}^*(\tilde{u}) \xrightarrow{\mathbb{P}} \alpha u^* W^*(u^*)$. We get the desired convergence by D_g 's continuity.

(ii) If $\alpha \geq \pi_0$, $u^* = 1$ and $\alpha u^* W_g^*(u^*) \geq 1$ for all g so $G_{\widetilde{W}^*}^\infty(u^*) = 1$. Then by Lemma C.4 $D_g(\alpha \tilde{u} \widetilde{W}_g^*(\tilde{u})) \xrightarrow{\mathbb{P}} 1$ which means that $G_{\widetilde{W}^*}^\infty(\tilde{u}) \xrightarrow{\mathbb{P}} \sum_g \pi_g 1 = 1$.

The proof for \widehat{H} is similar, just replace D_g by $\pi_{g,0}U$ (and 1 by $\pi_{g,0}$). \square

The last lemma states that $\text{LCM}(\widehat{D}_g)$ is a valid estimator of D_g to use in GADDOW.

Lemma C.6. *Assume that $\widetilde{D}_g = \text{LCM}(\widehat{D}_g)$. Then \widetilde{D}_g is nondecreasing, $\widetilde{D}_g(0) = 0$, $\widetilde{D}_g(1) = 1$ and $\left\| \widetilde{D}_g - D_g \right\| \xrightarrow{\mathbb{P}} 0$.*

Proof. $\widetilde{D}_g(0) = \widehat{D}_g(0) = 0$ and $\widetilde{D}_g(1) = \widehat{D}_g(1) = 1$ from the closed form given in Lemma 1 in Carolan (2002). Let $a, b \in [0, 1]$, $a < b$, and let

$$C(t) = \begin{cases} \widetilde{D}_g(t+b-a) & \text{if } t+b-a \leq 1 \\ 1 & \text{if } t+b-a \geq 1. \end{cases}$$

Then C is concave, and

$$C(t) \geq \begin{cases} \widehat{D}_g(t+b-a) & \geq \widehat{D}_g(t) & \text{if } t+b-a \leq 1 \\ 1 & \geq \widehat{D}_g(t) & \text{if } t+b-a \geq 1, \end{cases}$$

because \widehat{D}_g is non decreasing. So by definition of the LCM, $C(t) \geq \widetilde{D}_g(t)$ for all $t \in [0, 1]$. In particular,

$$\widetilde{D}_g(b) = C(a) \geq \widetilde{D}_g(a),$$

and \widetilde{D}_g is nondecreasing. Finally, the convergence comes from $\|\widetilde{D}_g - D_g\| \leq \|\widehat{D}_g - D_g\|$, see also Carolan (2002). \square

Appendix D: Proof of Corollary 5.3 for Pro 1

First, $\hat{w}^{(1)} \xrightarrow{\mathbb{P}} w^{(1)}$ where $w^{(1)} = \left(\frac{1}{\pi_0}, \dots, \frac{1}{\pi_0} \right)$ and $\hat{w}^{(2)} \xrightarrow{\mathbb{P}} w^{(2)}$ where, for all g , $w_g^{(2)} = \frac{\pi_{g,1}}{\pi_{g,0}(1-\pi_0)}$. By using Lemma C.1 and the continuity of D_g , we get that $\|\widehat{G}_{\hat{w}^{(1)}} - G_{w^{(1)}}^\infty\| \xrightarrow{\mathbb{P}} 0$ and $\|\widehat{G}_{\hat{w}^{(2)}} - G_{w^{(2)}}^\infty\| \xrightarrow{\mathbb{P}} 0$ and then by Lemma A.4 we get that $\hat{u}_{\hat{w}^{(1)}} \xrightarrow{\mathbb{P}} u_{w^{(1)}}^\infty$ and $\hat{u}_{\hat{w}^{(2)}} \xrightarrow{\mathbb{P}} u_{w^{(2)}}^\infty$ so $\hat{u}_M \xrightarrow{\mathbb{P}} u_M$ where $u_M = \max(u_{w^{(1)}}^\infty, u_{w^{(2)}}^\infty)$.

Define again $\widehat{W}_g^\dagger(u) = \frac{m_g \hat{\pi}_{g,0}}{m \pi_g \pi_{g,0}} \widehat{W}_g^*(u)$ and note that the power of Pro1 is $\mathbb{E} \left[\widehat{P}_{\widehat{W}^*}(\hat{u}_M) \right]$. We have

$$\begin{aligned} \widehat{P}_{\widehat{W}^*}(\hat{u}_M) &\leq \sup_{w \in \mathbb{R}_+^G} \left\| \widehat{P}_w - P_w^\infty \right\| + \left\| P_{\widehat{W}^*}^\infty - P_{\widehat{W}^\dagger}^\infty \right\| + P_{\widehat{W}^\dagger}^\infty(\hat{u}_M) \\ &\leq \sup_{w \in \mathbb{R}_+^G} \left\| \widehat{P}_w - P_w^\infty \right\| + \left\| P_{\widehat{W}^*}^\infty - P_{\widehat{W}^\dagger}^\infty \right\| + P_{\widehat{W}^*}^\infty(\hat{u}_M) \\ &\xrightarrow{\mathbb{P}} P_{\widehat{W}^*}^\infty(u_M), \end{aligned}$$

because $P_{\widehat{W}^*}^\infty$ is continuous by Lemma B.3.

Note that $u^* \geq u_M$ (because $G_{W^*}^\infty \geq G_{w^{(1)}}^\infty$ and $G_{W^*}^\infty \geq G_{w^{(2)}}^\infty$) to conclude.

Appendix E: Proof of Theorem 5.3

First note that $\mathbf{1} = (1, \dots, 1) \in K_{\text{NE}}^m$. Fix a given $u_0 \in (0, 1)$, say $u_0 = 1/2$. Then, by Lemma C.1,

$$\sup_{w \in K_{\text{NE}}^m} \sup_{u \in [0,1]} \left(\widehat{G}_w(u) - \alpha u \right) \geq \widehat{G}_1(u_0) - \alpha u_0 \xrightarrow{a.s.} G_1^\infty(u_0) - \alpha u_0. \quad (\text{E.1})$$

Then, denoting $K_1 = G_1^\infty(u_0) - \alpha u_0$, we have

$$\begin{aligned} K_1 &= \sum_{g=1}^G \pi_g (\pi_{g,0} \alpha u_0 + \pi_{g,1} F_g(\alpha u_0) - \alpha u_0) \\ &= \sum_{g=1}^G \pi_g \pi_{g,1} (F_g(\alpha u_0) - \alpha u_0) > 0, \end{aligned}$$

because $\pi_{g,1} > 0$ and, by strict concavity, for any $x \in (0, 1)$,

$$\frac{F_g(x)}{x} = \frac{F_g(x) - F_g(0)}{x - 0} > \frac{F_g(1) - F_g(0)}{1 - 0} = 1.$$

By multiplying both terms of (E.1) by \sqrt{m} we get that $Z_m \geq \sqrt{m} Y_m$ for some variable Y_m checking $Y_m \xrightarrow[m \rightarrow \infty]{a.s.} K_1 > 0$.

Next, recall that

$$Z_{0m} = \sqrt{m} \sup_{u \in [0,1]} \left(m^{-1} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{U_{g,i} \leq \alpha u \widetilde{W}_g^*(u)\}} - \alpha u \right),$$

where the $U_{g,i}$ are uniform variables over $[0, 1]$ with, for all g , $U_{g,1}, \dots, U_{g,m_g}$ independent, and

$$\widetilde{W}^*(u) \in \arg \max_{w \in K_{\text{NE}}^m} m^{-1} \sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{U_{g,i} \leq \alpha u w_g\}}.$$

Define also $\widehat{U}_{g,m_g} : t \mapsto m_g^{-1} \sum_{i=1}^{m_g} \mathbb{1}_{\{U_{g,i} \leq t\}}$. We then have

$$\begin{aligned} Z_{0m} &= \sqrt{m} \sup_{u \in [0,1]} \left(\sum_{g=1}^G \frac{m_g}{m} \widehat{U}_{g,m_g} \left(\alpha u \widetilde{W}_g^*(u) \right) - \alpha u \right) \\ &= \sqrt{m} \sup_{u \in [0,1]} \left(\sum_{g=1}^G \frac{m_g}{m} \left(\widehat{U}_{g,m_g} \left(\alpha u \widetilde{W}_g^*(u) \right) - \alpha u \widetilde{W}_g^*(u) \right) \right) \end{aligned} \quad (\text{E.2})$$

$$\begin{aligned} &\leq \sqrt{m} \sum_{g=1}^G \frac{m_g}{m} \sup_{t \in \mathbb{R}^+} \left(\widehat{U}_{g,m_g}(t) - t \right) \\ &\leq \sqrt{m} \sum_{g=1}^G \frac{m_g}{m} \sup_{t \in [0,1]} \left(\widehat{U}_{g,m_g}(t) - U(t) \right), \end{aligned} \quad (\text{E.3})$$

where the equality in (E.2) comes from $\sum_g \frac{m_g}{m} \widetilde{W}_g^*(u) = 1$, and the inequality in (E.3) comes from $U(t) = \min(1, t) \leq t$. Therefore

$$Z_{0m} \leq |Z_{0m}| \leq \sqrt{m} \sum_{g=1}^G \frac{m_g}{m} \left\| \widehat{U}_{g,m_g} - U \right\| = \sum_{g=1}^G \sqrt{\frac{m_g}{m}} \sqrt{m_g} \left\| \widehat{U}_{g,m_g} - U \right\|.$$

Therefore, for all $c > 0$,

$$\begin{aligned} \mathbb{P}(Z_{0m} > c) &\leq \mathbb{P} \left(\sum_{g=1}^G \sqrt{\frac{m_g}{m}} \sqrt{m_g} \left\| \widehat{U}_{g,m_g} - U \right\| > c \right) \\ &\leq \mathbb{P} \left(\exists g : \sqrt{m_g} \left\| \widehat{U}_{g,m_g} - U \right\| > \sqrt{\frac{m}{m_g}} \frac{c}{G} \right) \\ &\leq \sum_{g=1}^G \mathbb{P} \left(\sqrt{m_g} \left\| \widehat{U}_{g,m_g} - U \right\| > \sqrt{\frac{m}{m_g}} \frac{c}{G} \right). \end{aligned}$$

Now, applying G times the Dvoretzky-Kiefer-Wolfowitz-Massart inequality (Dvoretzky, Kiefer and Wolfowitz, 1956; Massart, 1990), we derive

$$\begin{aligned} \mathbb{P}(Z_{0m} > c) &\leq \sum_{g=1}^G 2 \exp \left(-2 \frac{m}{m_g} \frac{c^2}{G^2} \right) \\ &\leq 2G \exp \left(-2 \min_{1 \leq g \leq G} \left(\frac{m}{m_g} \right) \frac{c^2}{G^2} \right). \end{aligned}$$

Now define

$$c_m = \frac{G}{\sqrt{2}} \sqrt{\max_{1 \leq g \leq G} \left(\frac{m_g}{m} \right) \log \left(\frac{2G}{\beta_m} \right)} \leq K_2 \sqrt{K_3 - \log(\beta_m)} \text{ for some } K_2, K_3 > 0.$$

From above, $\mathbb{P}(Z_{0m} > c_m) \leq \beta_m$ which in turn implies $q_{\beta_m, m} \leq c_m$, because by definition

$$q_{\beta_m, m} = \min \{x : \mathbb{P}(Z_{0m} > x) \leq \beta_m\}.$$

Finally, $\phi_{\beta_m} = \mathbb{1}_{\{Z_m > q_{\beta_m, m}\}} \geq \mathbb{1}_{\{\sqrt{m} Y_m > c_m\}}$, and $c_m \xrightarrow{m \rightarrow \infty} o(\sqrt{m})$ because $\beta_m \geq a \exp(-bm^{1-\nu})$, which proves that $\phi_{\beta_m} \rightarrow 1$ almost surely.

Now showing that sADDOW_{β_m} has same asymptotical FDR and power as ADDOW is easy, because on one hand,

$$\begin{aligned} |\text{FDP}(\text{sADDOW}_{\beta_m}) - \text{FDP}(\text{ADDOW})| &= |(\phi_{\beta_m} - 1) \text{FDP}(\text{ADDOW}) \\ &\quad + (1 - \phi_{\beta_m}) \text{FDP}(\text{BH})| \\ &\leq 2 |1 - \phi_{\beta_m}| \xrightarrow{a.s.} 0, \end{aligned}$$

and on the other hand

$$\text{Pow}(\text{sADDOW}_{\beta_m}) = \mathbb{E} \left[\phi_{\beta_m} \widehat{P}_{\widehat{W}^*}(\hat{u}) + (1 - \phi_{\beta_m}) \widehat{P}_{\mathbf{1}}(u_{\mathbf{1}}) \right],$$

with

$$\left| \phi_{\beta_m} \widehat{P}_{\widehat{W}^*}(\hat{u}) + (1 - \phi_{\beta_m}) \widehat{P}_{\mathbf{1}}(u_{\mathbf{1}}) - \widehat{P}_{\widehat{W}^*}(\hat{u}) \right| \leq 2 |1 - \phi_{\beta_m}| \xrightarrow{a.s.} 0.$$