Adaptive p-value weighting with power optimality

Guillermo Durand

Laboratoire de Probabilités, Statistique et Modélisation
Sorbonne Université
4 place Jussieu, 75252 Paris Cedex 05
e-mail: guillermo.durand@polytechnique.org

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.

Contents

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). Common practice is generally used to build the groups from this type of covariate.

• 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 weigh 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) and Habiger (2014) designed oracle opti-

mal 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 asymptotically 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. Other recent works (Li and Barber, 2016; Ignatiadis and Huber, 2017; Lei and Fithian, 2018) suggest weighting methods (with additional steps or different threshold computing rules) but they don't address the power question theoretically either.

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). It relies on the computation of weights that maximize the number of detections at any rejection threshold, combined with the application of a step-up procedure with those weights. This is similar to IHW, however, by taking a larger weight space thanks to the use of estimators of true null proportion in each group, we allow for larger weights, hence more detections. 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 has optimality properties when the true nulls are evenly distributed across groups but we also show that its performance can quickly deteriorate otherwise with a numerical counterexample.

In Section 2, we present the mathematical model and assumptions. In Section 3, we define what is a weighting step-up procedure and discuss some procedures of the literature. In Section 4, we introduce ADDOW. Section 5 provides our main theoretical results. Our numerical simulations are presented in Section 6, while the overfitting problem is discussed in Section 7 with the introduction of a variant of ADDOW. We conclude in Section 8 with a discussion. The proofs of the two main theorems are given in Section 9 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.

2. Setting

2.1. Model

We consider the following stylized grouped p-value modeling: let $G \geq 2$ be the number of groups. Let us emphasize that G is kept fixed throughout the paper.

Because our study will be asymptotic in the number of tests m, for each m we assume that we test m_g hypotheses in group $g \in \{1, \dots, G\}$, where the m_g are non-decreasing integer sequences depending on m (the dependence is not written for conciseness) and such that $\sum_{g=1}^{G} m_g = m$. In each group $g \in \{1, \dots, G\}$, let $(H_{g,1}, \dots, H_{g,m_g})$ 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 $H_{g,i} = 1$ otherwise. Consider in addition $(p_{g,1}, \dots, p_{g,m_g})$ some random variables in $[0, 1]$ where each $p_{g,i}$ corresponds to the p-value testing $H_{g,i}$. 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 following marginal distributional assumptions for $p_{g,i}$.

Assumption 2.1. 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.).

Assumption 2.2. If $H_{g,i} = 1$, $p_{g,i}$ follows a common distribution corresponding to c.d.f. F_g, which is strictly concave on $[0, 1]$.

In particular, note that the p-values are assumed to have the same alternative distribution within each group. Note that the concavity assumption is mild (and implies continuity on $\mathbb{R}$ as proven in Lemma A.1 for completeness). 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 $f_g(1^-) \in [0, \infty)$.

Assumption 2.3. There exists $\pi_g > 0$ and $\pi_{g,0} > 0$ such that for all g, $m_g/m \to \pi_g$ and $m_{g,0}/m_g \to \pi_{g,0}$ when $m \to \infty$. Additionally, for each g, $\pi_{g,1} = 1 - \pi_{g,0} > 0$.

The above assumption means 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):

$$\pi_{g,0} = \pi_0, \ 1 \leq g \leq G. \tag{ED}$$

Let us finally specify assumptions on the joint distribution of the p-values.

Assumption 2.4. The p-values are weakly dependent within each group:

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

and

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

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, and that no assumption on the p-value dependence accross groups is made. Finally note that there is a hidden dependence in m in the joint distribution of the p-values $(p_{g,i})_{\substack{1 \leq g \leq G \\ 1 \leq i \leq m_g}}$ but that does not impact the remaining of the paper as long as (2.1) and (2.2) are satisfied.

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

Assumption 2.5. For each g, we have at hand an (over-)estimator $\hat{\pi}_{g,0} \in (0, 1]$ of $m_{g,0}/m_g$ such 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 the model of Section 2.1, this assumption can be fulfilled by using the estimators introduced in Storey, Taylor and Siegmund (2004):

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

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 $\frac{1}{m_g} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \lambda\}} \xrightarrow{\mathbb{P}} \pi_{g,0}\lambda + \pi_{g,1}F_g(\lambda)$, which provides our condition:

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

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

$$\bar{\pi}_{g,0} = \bar{\pi}_0, \ \ 1 \leq g \leq G. \tag{EE}$$

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. Case (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:

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

Case (NE) is basically the case where no estimation is intended, and the estimators are simply taken equal to 1.

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

$$\bar{\pi}_{g,0} = \pi_{g,0}, \ \ 1 \leq g \leq G. \tag{CE}$$

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).

Finally, the main case of interest is the Multiplicative Estimation case (ME) defined as the following:

$$\exists C \geq 1, \ \bar{\pi}_{g,0} = C\pi_{g,0}, \ 1 \leq g \leq G. \tag{ME}$$

Note that the constant C above cannot depend on g. Interestingly, the (ME) case covers the (CE) case (in this respect, $C = 1$) and also the case where (ED) and (EE) both hold (in this respect, $C = \frac{\bar{\pi}_0}{\pi_0}$). So the (ME) case can be viewed as a generalization of previous cases.

2.3. Criticality

Depending on the choice of α, multiple testing procedures may make no rejection at all when m tends to ∞. This case is not interesting and we should focus on the other case. To this end, Chi (2007) introduced the notion of criticality: they defined some critical alpha level, denoted α^*, for which BH procedure has no asymptotic power if $\alpha < \alpha^*$. Neuvial (2013) generalized this notion for any multiple testing procedure (see Section 2.5 therein) and also established a link between criticality and purity.

In Section A, Definition A.1, we define α^* in our heterogeneous setting and will focus in our results on the supercritical case.

Assumption 2.6. The target level α lies 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 parameters of the model, that are $(F_g)_g$, $(\pi_g)_g$ and $(\pi_{g,0})_g$, and on the parameters of the chosen estimators, that are $(\bar{\pi}_{g,0})_g$.

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 p-values are derived from Gaussian test statistics.

Formally, we assume that $p_{g,i} = \bar{\Phi}(X_{g,i})$, where $\bar{\Phi}(z) = \mathbb{P}(Z \geq z)$ for $Z \sim \mathcal{N}(0,1)$, and

$$\mathbf{X} = (X_{1,1}, \ldots, X_{1,m_1}, \ldots, X_{g,1}, \ldots, X_{g,m_g})$$

is a Gaussian vector with distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Here,

$$\boldsymbol{\mu} = (\mu_{1,1}, \ldots, \mu_{1,m_1}, \ldots, \mu_{G,1}, \ldots, \mu_{G,m_G}),$$

with $\mu_{g,i} = 0$ if $H_{g,i} = 0$ and $\mu_{g,i} = \mu_g$ if $H_{g,i} = 1$, and we assume that $\Sigma_{j,j} = 1$ for all $1 \geq j \geq m$. Hence $X_{g,i} \sim \mathcal{N}(0,1)$ under the null, and $X_{g,i} \sim \mathcal{N}(\mu_g, 1)$ under the alternative.

In this case, Assumption 2.1 is fulfilled, and

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

with derivative

$$f_g(\cdot) = \exp\left(\mu_g\left(\bar{\Phi}^{-1}(\cdot) - \frac{\mu_g}{2}\right)\right) > 0,$$

hence F_g is strictly concave and Assumption 2.2 is also fulfilled. 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 can achieve consistent estimation (CE) with additional independence assumptions.

Two particular subcases of interest arise when Σ has a particular form and can be written as

$$\begin{pmatrix} \Sigma_{(1)} & \mathbf{0} & \dots\dots\dots\dots & \mathbf{0} \\ \mathbf{0} & \Sigma_{(2)} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \Sigma_{(G-1)} & \mathbf{0} \\ \mathbf{0} & \dots\dots\dots & \mathbf{0} & \Sigma_{(G)} \end{pmatrix},$$

where $\Sigma_{(g)}$ is a square matrix of size m_g. The first subcase is when $\Sigma_{(g)}$ is the identity matrix. In this case, the p-values are all independent and Assumption 2.4 is fulfilled by the law of strong numbers. The second subcase is when $\Sigma_{(g)}$ is a Toeplitz matrix with $\left(\Sigma_{(g)}\right)_{j,k} = \frac{1}{|j-k|+1}$. In this case, Assumption 2.4 is also fulfilled (see e.g. Delattre and Roquain, 2016, Proposition 2.1, Equation (LLN-dep) and Theorem 3.1).

2.5. Criterion

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 $\mathcal{H}_1 = \mathcal{H}_0{}^{\mathsf{c}}$.

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 $\mathrm{FDP}(R)$, is defined as the number of false discoveries divided by the number of rejections if there are any, and 0 otherwise:

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

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

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

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 9.1) and does not have any repercussion

G. Durand

because the two definitions differ only by a multiplicative factor converging to $1 - \pi_0 \in (0,1)$ when $m \to \infty$.

Finally, let us emphasize that the power is the (rescaled) number of good rejections, that is, the number of rejected hypotheses that are false. The power is different from the number of total rejections, this distinction is fundamental and will be discussed all along this paper (like, for example, when discussing Heuristic 3.1, or in the simulations of Section 6.4).

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 $p_{(1)} \leq \ldots \leq p_{(m)}$ with $p_{(0)} = 0$, the classic BH procedure consists in rejecting all $p_{g,i} \leq \alpha \frac{\hat{k}}{m}$ where $\hat{k} = \max \left\{ k \geq 0 : p_{(k)} \leq \alpha \frac{k}{m} \right\}$.

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

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

then $\hat{k} = m \times \mathcal{I}(\widehat{G})$. This is a particular case of Lemma A.5. Note that $\widehat{G}(u)$ is simply the number of p-values that are less than or equal to αu, divided by m.

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 $\widehat{G}$ 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 $\mathcal{I}(\widehat{G}) = 6/m$ and 6 p-values are rejected.

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

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

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

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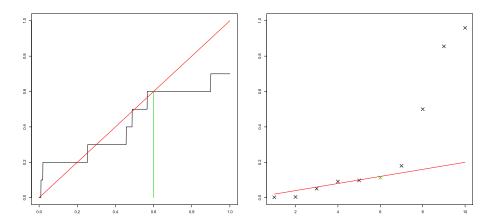
FIG 1. *The BH procedure applied to a set of 10 p-values. Right plot: the p-values and the function $k \to \alpha k/m$. Left plot: identity function and $\widehat{G}$. Each plot shows that 6 p-values are rejected.*

by introducing a weight function $W : [0, 1] \to \mathbb{R}_+^G$, which can be random, such that the following function:

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

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 MWBH(W) where, for the rest of the paper, we denote

$$\hat{u}_W = \mathcal{I}\left(\widehat{G}_W\right), \tag{3.2}$$

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 a straightforward generalization of the class of WBH procedures because for any weight vector w, w can be seen as a constant weight function $u \mapsto w$ and $\widehat{G}_w$ is nondecreasing.

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 $p_{(1)}^{[r]} \leq \ldots \leq p_{(m)}^{[r]}$ 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 (see e.g. Genovese, Roeder and Wasserman, 2006 or Zhao and Zhang, 2014), in order to achieve a valid FDR control, these procedures should be used with weights that satisfy some specific constraints. The following weight spaces will be used in the following of the paper:

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

$$\hat{K}_{\mathrm{NE}} = \left\{ w \in \mathbb{R}_+^G : \sum_g \frac{m_g}{m} w_g \leq 1 \right\}. \tag{3.4}$$

Note that $\hat{K}$ may appear unusual because it depends on the estimators $\hat{\pi}_{g,0}$, however it is completely known and usable in practice. Some intuition about the choice of $\hat{K}$ is given in next section. Note also that $\hat{K} = \hat{K}_{\mathrm{NE}}$ in the (NE) case.

Finally, for a weight function W and a rejection 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 $\widehat{G}_W(u) = m^{-1} |R_{u,W}|$ (which means that $\widehat{G}_W(u)$ is the number of rejections of $R_{u,W}$, divided by m) and that $\mathrm{MWBH}(W)$ can also be written as $R_{\hat{u}_W,W}$.

3.2. Choosing the weights

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

$$P_W^{(m)}(u) = m^{-1} \mathbb{E} \left[\sum_{g=1}^G \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha u W_g(u), H_{g,i}=1\}} \right]$$

$$= \sum_{g=1}^G \frac{m_{g,1}}{m} F_g\left(\alpha u W_g(u)\right).$$

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:

$$W_{or}^*(u) = \underset{w \in \hat{K}_{\mathrm{NE}}}{\arg\max}\, P_w^{(m)}(u). \tag{3.5}$$

Note that they defined W_{or}^* only in case (NE), but their definition easily extends to the general case as above, by replacing $\hat{K}_{\mathrm{NE}}$ by $\hat{K}$. They proved the existence and uniqueness of W_{or}^* when both (ED) and (NE) hold and that, asymptotically, $\mathrm{MWBH}(W_{or}^*)$ controls the FDR at level $\pi_0 \alpha$ and has a better power than every $\mathrm{MWBH}(w^{(m)})$ for $w^{(m)} \in \hat{K}_{\mathrm{NE}}$ 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 $u \in [0, 1]$:

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

$$= P_W^{(m)}(u) + H_W^{(m)}(u), \tag{3.6}$$

where

$$H_W^{(m)}(u) = \sum_{g=1}^{G} \frac{m_{g,0}}{m} U(\alpha u W_g(u)).$$

$G_W^{(m)}(u)$ is the mean ratio of rejections for the procedure rejecting each $p_{g,i} \leq \alpha u W_g(u)$. $P_W^{(m)}(u)$ is the rescaled mean of the number of true positives (i.e. the power) of this procedure while $H_W^{(m)}(u)$ is the rescaled mean of the number of its false positives.

Heuristic 3.1. *Maximizing $G_W^{(m)}(u)$ should be close to maximizing $P_W^{(m)}(u)$.*

Indeed, consider weight functions W such that $\sum_g \frac{m_{g,0}}{m} W_g(u) = 1$ and then replace $U(x)$ by x for all $x \in \mathbb{R}_+$ (whereas $U(x) = x$ only holds for $x \leq 1$), then $H_W^{(m)}(u)$ becomes $\alpha u \sum_g \frac{m_{g,0}}{m} W_g(u) = \alpha u$ and it does not depend on the weights. So $P_W^{(m)}(u)$ is the only term depending on W in (3.6) and maximizing $P_W^{(m)}(u)$ or $G_W^{(m)}(u)$ is the same.

Now, we can evaluate the constraint we just put 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 $\hat{K}$ 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 Heuristic 3.1 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 $\hat{K}$, 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

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

Pro2 comes from an additional step-up step compared to Pro1, hence its rejection threshold, $\hat{u}_{\widehat{W}^*(\hat{u}_M)}$, is larger than $\hat{u}_M$ and allows for more detections. 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 WBH($\hat{w}^{(2)}$), and 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 Heuristic 3.1 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 Heuristic 3.1 break down, because in general $H_W^{(m)}(u)$ can still depend on w, see remark 3.1 below. 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 Heuristic 3.1, but only in case (ME). When (ME) does not hold, we must take care that Heuristic 3.1 can fail for the same reason that it can fail with IHW. Thereby, in general, more detections do not necessarily imply more power.

Remark 3.1. In particular, we can compute numerical counterexamples where BH has larger asymptotic power 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, we can achieve a larger power with BH than with IHW. Our interpretation is that, in that case, IHW slightly favors group 2 because of its size, whereas the oracle optimal favors group 1 thanks to the knowledge of the true parameters. BH, by weighting uniformly, does not favor any group, which allows its power to end up between the power of the oracle and the power of IHW. This example is studied in Section 6.4 and illustrated in Figures 8 and 7.

3.3. Recent weighting methods

Besides IHW, there are several recent methods putting weights on p-values. We briefly discuss three of them. The first is a variation of IHW by the same authors, IHWc (Ignatiadis and Huber, 2017), where the letter 'c' stands for 'censoring'. The method bring two innovations to IHW. First, the use of cross-weighting thanks to a subdivision of the hypotheses into folds: the weights of the p-values of a fold are computed by only using the p-values of the other folds. This approach reduces overfitting since, during the step-up procedure, the information brought by a given p-value is used only once instead of twice. The second innovation is the censoring, where a threshold τ is fixed and only p-values larger than τ are used to compute the weights, while only p-values lesser than τ can be rejected during the step-up. Together, these innovations allow IHWc to control the FDR in finite sample at level α if the p-values associated to true nulls are independent. However, using only large p-values to compute the weights seems somehow counterintuitive: large p-values are likely to be associated to true nulls and to be uniform, so they won't allow the weights to properly discriminate the groups and to increase the power compared to BH. We will verify this intuition in Section 6.3. Finally, it is worth noting that IHWc allows for a kind of $\pi_{g,0}$ estimation à la Storey, with a variant called IHWc-Storey.

The censoring idea originates from the Structure Adaptive BH Algorithm (SABHA, Li and Barber, 2016), which has a group structured version with an FDR bounded by αC for a known constant $C > 1$ when the p-values are independent. Hence, applying the group structured SABHA at level α/C gives

FDR control at level α, but using a target level $< \alpha$ can induce conservatism, especially since computing the weights only with the large p-values involve the same risks that we highlighted when discussing of IHWc.

Lastly, AdaPT (Lei and Fithian, 2018) introduces threshold surfaces $s_t(x)$ that can be considered as weights and adapted to group setting. AdaPT is not a WBH procedure, its whole philosophy is totally different and relies on symmetry properties of the true null distribution of the p-values by using an estimator of the FDP, different than the one implicitly used in BH-like methods, which also relies on symmetry and allow to mask p-values during the procedure (see also Barber et al., 2015 and Arias-Castro et al., 2017 for more details on this pioneering paradigm). We won't further consider AdaPT because of its fundamental differences with WBH procedures and because we are mainly interested by optimality among said WBH procedures.

For more discussion about IHW, IHWc, SABHA and AdaPT, see Ignatiadis and Huber (2017, Section 6.2) and Lei and Fithian (2018, Section 1.4).

4. New procedure: ADDOW

We exploit Heuristic 3.1 and propose to estimate the oracle optimal weights W^*_{or} by maximizing in $w \in \hat{K}$ the empirical counterpart to $G_w^{(m)}(u)$, that is $\widehat{G}_w(u)$.

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

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

Such maximum is guaranteed to exist because $\left\{ \widehat{G}_w(u),\, w \in \hat{K} \right\}$ is a finite set. Indeed, it is a subset of $\left\{ \frac{k}{m},\, k \in [\![0, m]\!] \right\}$. However, for a given u, $\widehat{W}^*(u)$ may not be uniquely defined, hence there is no unique optimal weight function $\widehat{W}^*$ in general. So, in all the following, we fix 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 $\hat{u}_{\widehat{W}^*}$ 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, ADDOW = MWBH $\left(\widehat{W}^* \right)$.

One shall note that ADDOW is in fact a class of procedures depending on the estimators $\hat{\pi}_{g,0}$ through $\hat{K}$. Its rationale is similar to IHW in that we intend to maximize the number of rejections, but incorporating the estimators $\hat{\pi}_{g,0}$ allows for larger weights and more detections.

Finally, note that, in the (NE) case, ADDOW reduces to IHW.

Remark 4.1. It turns out that ADDOW is equal to a certain WBH procedure. It comes from part 2 of the proof of Theorem 5.2 and Remark 9.2. Moreover, to every MWBH procedure, corresponds a WBH procedure with power higher

or equal. This fact does not limit the interest of the MWBH class, because computing the dominating WBH procedure of a given MWBH($\widehat{W}$) procedure requires the knowledge of the step-up threshold $\hat{u}_{\widehat{W}}$ which is known by actually computing MWBH($\widehat{W}$).

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 (ME) case. The two are proven in Section 9.

Theorem 5.1. *Let us assume that Assumptions 2.1 to 2.6 are fulfilled. We have*

$$\lim_{m \to \infty} \text{FDR}\,(\text{ADDOW}) \leq \alpha. \tag{5.1}$$

Moreover, if $\alpha \leq \bar{\pi}_0$ and (ME) *holds,*

$$\lim_{m \to \infty} \text{FDR}\,(\text{ADDOW}) = \frac{\alpha}{C}. \tag{5.2}$$

Remark 5.1. Equation (5.2) means that in the (CE) case (where $C = 1$), exact asymptotic control is achieved.

Theorem 5.2. *Let us assume that Assumptions 2.1 to 2.6 are fulfilled, with the additional assumption that* (ME) *holds. For any sequence of random weight functions $(\widehat{W})_{m \geq 1}$, such that $\widehat{W} : [0,1] \to \hat{K}$ and $\widehat{G}_{\widehat{W}}$ is nondecreasing, we have*

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

5.2. Relation to **IHW**

Recall that IHW reduces ADDOW in the (NE) case, that (NE) is a subcase of (EE), and that when both (EE) and (ED) hold then (ME) is achieved. Hence, as a byproduct, we deduce from Theorems 5.1 and 5.2 the following result on IHW.

Corollary 5.1. *Let us assume that Assumptions 2.1 to 2.6 are fulfilled, with the additional assumption that* (ED) *holds. Then*

$$\lim_{m \to \infty} \text{FDR}\,(\text{IHW}) = \pi_0 \alpha, \tag{5.3}$$

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

$$\lim_{m \to \infty} \text{Pow}\,(\text{IHW}) \geq \limsup_{m \to \infty} \text{Pow}\left(\text{MWBH}\left(\widehat{W}\right)\right). \tag{5.4}$$

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 $\hat{K}_{\mathrm{NE}} \subset \hat{K}$. Hence, any MWBH procedure estimating $\frac{m_{g,0}}{m_g}$ by 1 uses a weight function valued in $\hat{K}$. This immediately yields the following corollary.

Corollary 5.2. *Let us assume that Assumptions 2.1 to 2.6 are fulfilled, with the additional assumption that* (ME) *holds. Then*

$$\lim_{m \to \infty} \mathrm{Pow}\,(\mathrm{ADDOW}) \geq \limsup_{m \to \infty} \mathrm{Pow}\,(R),$$

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

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

Corollary 5.3. *Let us assume that Assumptions 2.1 to 2.6 are fulfilled, with the additional assumption that* (ME) *holds. Then*

$$\lim_{m \to \infty} \mathrm{Pow}\,(\mathrm{ADDOW}) \geq \limsup_{m \to \infty} \mathrm{Pow}\,(R),$$

for any $R \in \{\mathrm{Pro\,1}, \mathrm{Pro\,2}, \mathrm{HZZ}, \mathrm{ABH}\}$.

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.

6. Numerical experiments

6.1. Simulation setting

FDR analysis and power analysis from Sections 6.2 and 6.3 are conducted using simulations which setting we describe here. Section 6.4 presents a counter-example using its own setting.

We consider the one-sided Gaussian framework described in Section 2.4 for $G = 2$ groups. We set $\alpha = 0.05$, $m_1 = m_2 = 4000$ (hence $m = 8000$), $m_{1,0} = 2800$ and $m_{2,0} = 3200$, such that $\pi_{1,0} = 0.7$ and $\pi_{2,0} = 0.8$. The values of μ_1

and μ_2 are defined according to a varying parameter $\bar{\mu}$, which values are in $\{0.1, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3\}$.

Our experiments have been performed by using the four following scenarios Each simulation of each scenario is replicated 1000 times.

- Scenario 1: $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$ and the p-values are independent.
- Scenario 2: $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$ and the dependence follows the Toeplitz pattern described in the end of Section 2.4.
- Scenario 3: $\mu_1 = \bar{\mu}$ and $\mu_2 = 0.01$ and the p-values are independent.
- Scenario 4: $\mu_1 = \bar{\mu}$ and $\mu_2 = 0.01$ and the dependence follows the Toeplitz pattern described in the end of Section 2.4.

In each scenario, three groups of procedures are compared. The difference between the three groups lies in the way π_0 is estimated. Group 1 corresponds to the (NE) case: $\hat{\pi}_{g,0} = 1$. Group 2 corresponds to the (CE) case, with an oracle estimator: $\hat{\pi}_{g,0} = \pi_{g,0}$. Groups 3 use the Storey estimator $\hat{\pi}_{g,0}(1/2)$ defined in Equation (2.3). We choose $\lambda = 1/2$ as it is a standard value (see e.g. Storey, 2002). The compared procedures are the following:

- ABH as defined in section 3.2 (which is BH in Group 1),
- HZZ as defined in section 3.2 (except in Group 1 where it is not defined),
- Pro2 as defined in section 3.2 (for Group 1, we only use the BH threshold),
- ADDOW (which is equal to IHW in Group 1),
- An oracle ADDOW wich is the MWBH procedure using the oracle weights W_{or}^* given by equation 3.5 (only in Groups 1 and 2),
- IHWc (only in Groups 1 and 3). The version of IHWc used in Group 3 is IHWc-Storey.

For IHWc, the censoring level chosen is the default of the IHW R package, that is α.

In the following, only plots of scenarios 1 and 3 are shown, as the situation with Toeplitz dependence is found to be similar to the independent case, up to a slight increase of the FDR of most of the procedures.

6.2. FDR control

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

In scenario 1, we can distinguish two different regimes depending on the signal strength. For $\mu \geq 1$ the signal strength is not weak in both groups (from $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$) and the FDR is controlled at level α for all procedures of Groups 2 & 3 except ADDOW and Pro2, the two procedures using the data-driven weights, that is $\widehat{W}^*$. In particular, Oracle ADDOW in Group 2 controls the FDR at level α. As the data driven weights converge to the oracle weights (see Lemma C.4), we get an illustration of Theorem 5.1 in the (CE) case. The situation is similar for Group 1 and level $\pi_0\alpha$, except for Oracle ADDOW which controls the FDR only for $\mu \geq 2$.
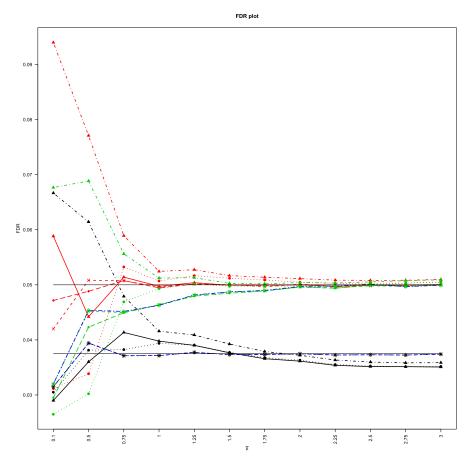
Fig 2. *FDR against $\bar{\mu}$ in scenario 1. Group 1 in black; Group 2 in red; Group 3 in green. The type of procedure depends on the shape: Oracle ADDOW (triangles and solid line); ADDOW (triangles and dashed line); Pro2 (disks); HZZ (diamonds) and finally BH/ABH (crosses). IHWc and IHWc-Storey are in blue, respectively with black and green points. Horizontal lines: α and $\pi_0\alpha$ levels. See Section 6.1.*

The situation get more confused when the signal is weak ($\mu < 1$). The FDR of ADDOW (in each group) is largely inflated. The FDR control at level α also fails sometimes for Oracle ADDOW, Pro2, ABH and HZZ (only in Group 2).

In scenario 3, one group has always weak signal. The FDR inflation of AD-DOW (in each group) and Group 2 is worse for small $\bar{\mu}$, whereas, for large $\bar{\mu}$, the situation is similar to scenario 1, up to one exception: the FDR of ABH and IHWc in Group 3 does not reach α as it did in scenario 1, which suggests some sort of conservatism.

In both scenarios, procedures of Group 2 have a larger FDR than their equivalent in Group 3, which in turn have larger FDR than in Group 1.
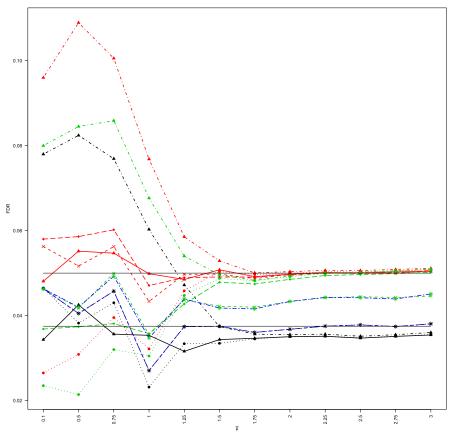
FDR plot

FIG 3. *FDR against $\bar{\mu}$ in scenario 3. Same legend as in Figure 2.*

As a side note, in both scenarios, and both Groups 1 and 3, the FDR plots of IHWc and ABH are nearly indistinguishable.

In both settings regarding $\bar{\mu}$ (large or small), procedures based on $\widehat{W}^*$ suffer from some sort of overfitting causing a loss of FDR control. This is discussed in Section 7 with an attempt to stabilize the weights. Let us underline that this does not contradict Theorem 5.1 because a small μ_g might imply a smaller convergence rate while m stays $< 10^4$ in our setting.

6.3. Power analysis

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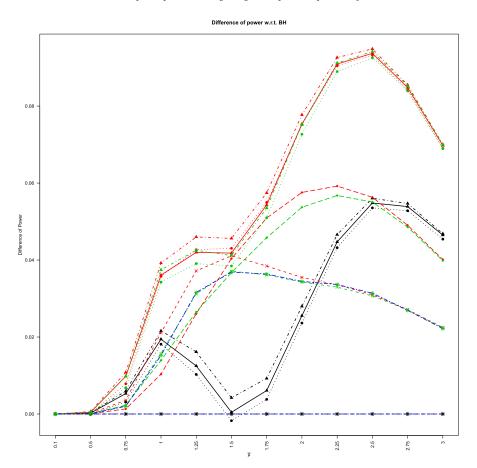
FIG 4. *DiffPow against $\bar{\mu}$ in scenario 1. Same legend as Figure 2.*

with respect to BH, or DiffPow, by

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

for any procedure R.

Figures 4 and 5 display the power of all the procedures defined in Section 6.1. Figures 6a and 6b display only a subset of them in Scenario 1, for clarity. We can make several observations:

- In both scenarios, procedures of Group 2 are more powerful than their equivalent in Group 3, which are better than in Group 1 (up to one exception, see next point), see e.g. Figure 6a. In particular, the difference between Group 2 and Group 1 is huge. This illustrates the importance of incorporating the knowledge of π_0 to improve power.
- In scenario 2, HZZ is largely better in Group 3 than in Group 2. Our interpretation is that the signal is so weak in the second group of p-values
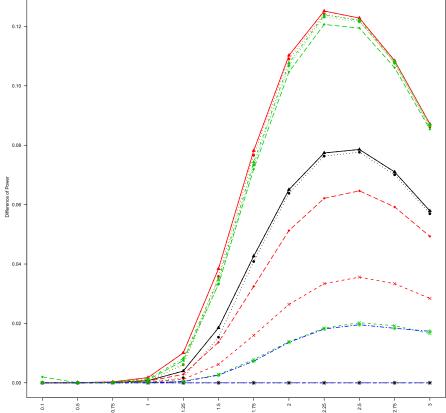
Fig 5. *DiffPow against $\bar{\mu}$ in scenario 3. Same legend as Figure 2.*

that the estimator $\hat{\pi}_{2,0}(1/2)$ is close to one, while $\hat{\pi}_{1,0}(1/2)$ stays close to $\pi_{1,0}$. Hence $\hat{w}_1^{(2)}$ in Group 3 is larger than $\hat{w}_1^{(2)}$ in Group 2 which allows for more good discoveries. The drawback of having $\hat{w}_2^{(2)}$ in Group 3 smaller than $\hat{w}_2^{(2)}$ in Group 2 is not a real one since the signal is so small that it is impossible to detect no matter the weight. Recall that $\hat{w}^{(2)}$ is defined in Section 3.2.

- In every Group (that is for any choice of $\hat{\pi}_{g,0}$), and for both scenarios, ADDOW achieves the best power (see e.g. Figure 6b), which supports Theorem 5.2. Additionnaly, maybe surprisingly, Pro2 behaves quite well, with a power close to the one of ADDOW (sometimes larger than Oracle ADDOW) and despite its theoretical sub-optimality.
- Inside Group 2 or Group 3, and for both scenarios, comparing ABH and HZZ to ADDOW and Pro2 shows the benefit of adding the F_g adaptation
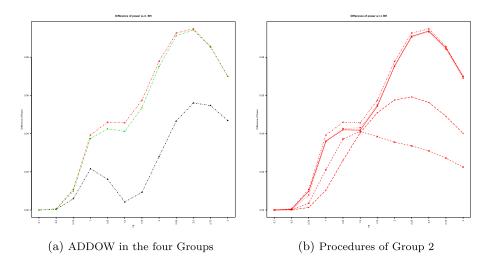
(a) ADDOW in the four Groups   (b) Procedures of Group 2

FIG 6. *Details of Figure 4 where only a subset of procedures is plotted.*

to the π_0 adaptation: the ADDOW and Pro2 have better power than ABH and HZZ for all signals (see e.g. Figure 6b).

In scenario 1, for Groups 2 and 3, we can see a zone of moderate signal (around $\bar{\mu} = 1.5$) where the two categories of procedures 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, on the one hand, IHW and, on the other hand, ABH or HZZ from Group 2, in Figure 4, shows the difference between adapting only to the F_g's versus adapting only 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.

- In scenario 1, for all signals, methods of Group 3 are close to their equivalent of Group 2, which indicates that using $\lambda = 1/2$ gives a good estimate of $\pi_{g,0}$ in practice (see e.g. Figure 6a). Furthermore, the larger the signal is, the more methods of Group 3 get closer to Group 2.

- In both scenarios, once again IHWc and ABH are nearly indistinguishable, which confirms the intuition given in Section 3.3 that IHWc performs badly in terms of power due to using only large p-values to compute the weights. See in particular how the power of IHW is larger than the power of IHWc (and even than the power of IHWc-Storey) in Figure 4.

6.4. Importance of (ME) *for optimality results*

We provide here a setting and a simulation where Corollary 5.1 fails because (ED) does not hold, to illustrate the importance of (ME) in Theorem 5.2 and in

Theorem 5.1 (to get (5.2)). The setting is chosen according to what we sketched in Remark 3.1 and is the following.

We consider again the one-sided Gaussian framework described in Section 2.4 for $G = 2$ groups and independent p-values. The parameters are the same as in Section 6.1 and each simulation of each scenario is replicated 1000 times. We choose a large value for α ($\alpha = 0.7$) which is unlikely to appear in practice but allows us to get our counterexample. We set $m_1 = 1000$ and $m_2 = 9000$, $m_{1,0}/m_1 = 0.05$ and $m_{2,0}/m_2 = 0.85$. So group 1 is small and has a lot of signal, while group 2 is large but has not much signal. The signal strength is given by $\mu_1 = 2$ and $\mu_2 = \bar{\mu}$, and $\bar{\mu} \in \{1.7, 1.8, 1.9, 2, 2.1, 2.2, 2.3\}$, so the signal is strong and almost equal in both groups.

We compare only BH, ADDOW in the (CE) case (with $\hat{\pi}_{g,0} = \pi_{g,0}$) and ADDOW in the (NE) case (that is, IHW, with $\hat{\pi}_{g,0} = 1$). The simulation is illustrated with an FDR plot in Figure 7 and a DiffPow plot in Figure 8.

In Figure 7, the FDR of BH is $\pi_0 \alpha$ as expected, and we see that the FDR of IHW is above that level, hence Equation (5.3) is violated. On a side note, we see that, thanks to a large m (10^4) and a rather strong signal, ADDOW in (CE) does not overfit and we get an illustration of Equation (5.2) with $C = 1$.

Figure 8 is rather unequivocal and shows that our parameter choice implies that IHW has a power smaller than BH (ADDOW in (CE) case stays better as expected), hence Equation (5.4) is violated. Let us recall our interpretation proposed in Remark 3.1: 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 a WBH procedure with weights (1,1), BH does not favor any group. Figure 8 demonstrates the limitation of Heuristic 3.1 by providing a direct counterexample, and underlines the necessity of estimating the $\pi_{g,0}$ when nothing lets us think that (ED) may be met.

7. Stabilization for overfitting

7.1. Overfitting phenomena

Since ADDOW uses the data both through the p-values and the weights, it suffers from an overfitting phenomena where the FDR in finite samples is above the target level α, as we saw in Section 6.2. In our setting, if 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 leads ADDOW to assign its weighting budget at random, and giving large weights to the wrong groups increases the FDP.

As said before, our intuition is that the overfitting is at least partly due to using each p-value twice in the step-up procedure of ADDOW: in the expression $\mathbb{1}_{\left\{p_{g,i} \leq \alpha u \widehat{W}_g^*(u)\right\}}$, $p_{g,i}$ appears in both sides of the inequality because it is used
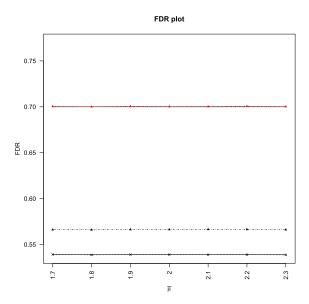
FDR plot

FIG 7. *FDR of ADDOW and BH against $\bar{\mu}$ in the simulation of Section 6.4. The two solid lines are the α and $\pi_0\alpha$ levels, the FDR of BH is confounded with the $\pi_0\alpha$ level. ADDOW in the* (NE) *case is given by the black triangles and ADDOW in the* (CE) *case is given by the red triangles.*
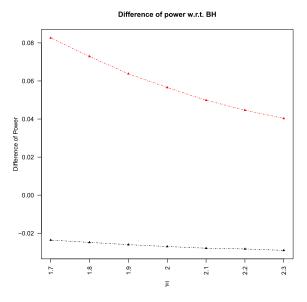
Difference of power w.r.t. BH

FIG 8. *DiffPow of ADDOW against $\bar{\mu}$ in the simulation of Section 6.4. ADDOW in the* (NE) *case is given by the black triangles and ADDOW in the* (CE) *case is given by the red triangles.*

to compute $\widehat{W}_g^*(u)$. Following this, we propose a variation of ADDOW that uses the same cross-weighting trick as IHWc.

7.2. The crADDOW variant

The main idea is to split the p-values into F folds, where F is some fixed integer ≥ 2, and to use only p-values of the remaining $F - 1$ folds to compute the weights assigned to the p-values of a given fold. The resulting procedure can be seen as a WBH procedure using $F \times G$ groups.

Formally, for each m we have a random function $\mathbb{F}_m : (g, i) \mapsto \mathbb{F}_m(g, i) \in \{1, \ldots, F\}$ such that, for each $f \in \{1, \ldots, F\}$ and each $g \in \{1, \ldots, G\}$, $|\{1 \leq i \leq m_g : \mathbb{F}_m(g, i) = f\}| \geq \lfloor \frac{m_g}{F} \rfloor$, which simply means that the p-values of each group g are evenly distributed between the F folds. Some dependence assumptions are required:

Assumption 7.1. The σ-algebra generated by $(\mathbb{F}_m)_m$ and the σ-algebra generated by $\big((p_{g,i})_{(g,i)}\big)_m$ are independent.

Assumption 7.2. Conditionally to $(\mathbb{F}_m)_m$, we have weak dependence (as in Assumption 2.4) inside each fold.

For each fold $f \in \{1, \ldots, F\}$, we compute ADDOW_{-f}, that is ADDOW but using only p-values for the folds in $\{1, \ldots, F\} \setminus \{f\}$. This is done by constructing the empirical function

$$\widehat{G}_w^{-f} : u \mapsto |\{(g, i) : \mathbb{F}_m(g, i) \neq f\}|^{-1} \sum_{\substack{(g,i): \\ \mathbb{F}_m(g,i) \neq f}} \mathbb{1}_{\{p_{g,i} \leq \alpha u w_g\}},$$

and then maximizing it in $w \in \hat{K}^{-f}$ for each $u \in [0, 1]$, where:

$$\hat{K}^{-f} = \left\{ w \in \mathbb{R}_+^G : \sum_g \frac{|\{1 \leq i \leq m_g : \mathbb{F}_m(g, i) \neq f\}|}{|\{(g, i) : \mathbb{F}_m(g, i) \neq f\}|} \hat{\pi}_{g,0} w_g \leq 1 \right\}.$$

While this expression seems complicated, note that if F divides each m_g, then $|\{1 \leq i \leq m_g : \mathbb{F}_m(g, i) = f\}| = \frac{m_g}{F}$ and $\hat{K}^{-f} = \hat{K}$. The maximization provides a weight function $\widehat{W}^{*-f}$ and the MWBH procedure provides a step-up threshold $\hat{u}_{\widehat{W}^{*-f}} = \mathcal{I}\left(\widehat{G}_{\widehat{W}^{*-f}}^{-f}\right)$. To lighten notation, let $w_{g,f}^* = \widehat{W}^{*-f}(\hat{u}_{\widehat{W}^{*-f}})$.

Our ADDOW variant, named crADDOW for cross-ADDOW, is the WBH procedure which assigns the weight $w_{g,f}^*$ to all p-values $p_{g,i}$ such that $\mathbb{F}_m(g, i) = f$. Now, in $\mathbb{1}_{\{p_{g,i} \leq \alpha u w_{g,f}^*\}}$, $p_{g,i}$ is only used once. While we don't have a finite-sample result about crADDOW, we expect it to have a lesser FDR than AD-DOW, especially for weak signal. We expect crADDOW to act like a stabilization of ADDOW and to not lose the good performances of ADDOW when the signal is not weak. Those intuitions are verified in the simulations of Section 7.3. Still, crADDOW has the nice property of being asymptotically equivalent to ADDOW.
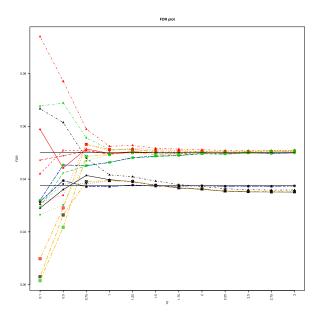
FIG 9. *FDR against $\bar{\mu}$ in scenario 1. Same legend as in Figure 9, with the addition of* crADDOW *(yellow lines). The color of the points (black, red, green) indicates the Group (respectively, 1, 2 and 3).*

Theorem 7.1. *Let us assume that Assumptions 2.1 to 2.6, 7.1 and 7.2 are fulfilled. Assume also that $\alpha \leq \bar{\pi}_0$. We have*

$$\lim_{m \to \infty} \text{FDR}\,(\text{crADDOW}) = \lim_{m \to \infty} \text{FDR}\,(\text{ADDOW}), \tag{7.1}$$

and

$$\lim_{m \to \infty} \text{Pow}\,(\text{crADDOW}) = \lim_{m \to \infty} \text{Pow}\,(\text{ADDOW}). \tag{7.2}$$

This Theorem is proved in Section E.

7.3. *Simulations with* **crADDOW**

The simulations presented here are the same as the simulations depicted in Section 6.1, with the addition of crADDOW in each Group.

From the FDR plots, we see that the FDR is hugely deflated and is now controlled at level α for weak $\bar{\mu}$ in each scenario, while for large $\bar{\mu}$ we are still slightly above the target level but with a small improvement over ADDOW. In scenario 1 there is a small window between large and small $\bar{\mu}$, around $\bar{\mu} = 0.75$, where crADDOW in Group 2 overfits more than for really large $\bar{\mu}$, but even there we see a large improvement over ADDOW.

As for the power, we see that crADDOW is less powerful than ADDOW, as expected since we reject less hypotheses, but we see that in most Groups
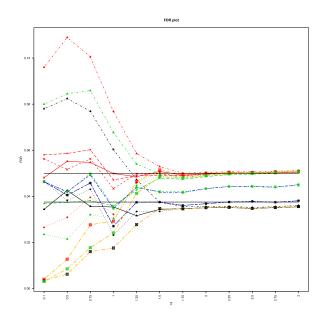
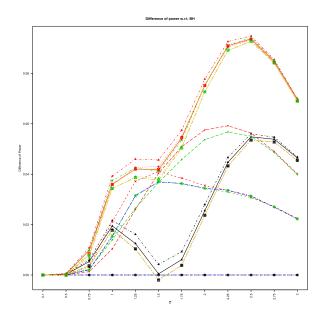FIG 10. *FDR against $\bar{\mu}$ in scenario 3. Same legend as in Figure 9.*

FIG 11. *DiffPow against $\bar{\mu}$ in scenario 1. Same legend as Figure 9.*
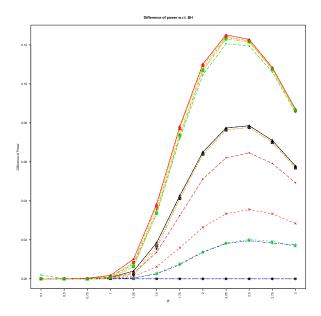
Fig 12. *DiffPow against $\bar{\mu}$ in scenario 3. Same legend as Figure 9.*

and scenarios the loss of power is almost negligible and crADDOW remains even as powerful as Oracle ADDOW (with the exception of Group 1 in scenario 1). The difference of power between crADDOW and Pro2 is even smaller and crADDOW is better in most configurations, with the exception of Groups 2 and 3 around $\bar{\mu} = 1.5$, which is the zone that we identified in Section 6.3 as the zone where the optimal weights are given by the uniform $\hat{w}^{(1)}$ weighting of ABH.

The simulations hence confirm our intuitions about the stabilization properties of crADDOW especially for weak signal where ADDOW was totally unreliable. Studying the finite sample properties of crADDOW, especially its FDR, is an interesting direction for future works.

8. Concluding remarks

In this paper we presented a new class of data-driven step-up procedures, 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 and the signal strength are large enough. We also introduced crADDOW, a variant of ADDOW that uses cross-weighting to reduce the overfitting while having the exact same asymptotic properties.

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 9, ADDOW can be extended to that case (see especially Section 9.1) 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. A possible such variation is crADDOW, whose FDR in finite sample has yet to be investigated.

9. Proofs of Theorems 5.1 and 5.2

9.1. Further generalization

Define, for any u and W,

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

and

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

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

$$\widehat{G}_W(u) = \sum_g \frac{m_g}{m} \widehat{D}_g(\alpha u W_g(u)).$$

For the sake of generality $\widehat{D}_g$ 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 $\widehat{D}_g$, denoted $\text{LCM}(\widehat{D}_g)$, see Section 8). So let us increase slightly the scope of the MWBH class by defining $\widetilde{G}_W(u) = \sum_g \frac{m_g}{m} \widetilde{D}_g(\alpha u W_g(u))$ for any estimator $\widetilde{D}_g$ such that $\widetilde{D}_g$ is nondecreasing, $\widetilde{D}_g(0) = 0$, $\widetilde{D}_g(1) = 1$ and $\left\|\widetilde{D}_g - D_g\right\| \xrightarrow{\mathbb{P}} 0$, where $\|\cdot\|$ is the sup norm for the bounded functions on their definition domain. Note that at least $(D_g)_g$, $(\widehat{D}_g)_g$ (by Lemma C.1), and $\left(\text{LCM}(\widehat{D}_g)\right)_g$ (by Lemma C.6) are eligible.

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

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

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

$$\widetilde{W}^*(u) \in \underset{w \in \hat{K}}{\arg\max} \, \widetilde{G}_w(u), \tag{9.1}$$

we define the generalized ADDOW by

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

the latter being well defined because $\widetilde{G}_{\widetilde{W}^*}$ is nondecreasing (by a proof similar to the one of Lemma A.6). Note that for any continuous $\widetilde{D}_g$, such as $\text{LCM}(\widehat{D}_g)$ or D_g itself, the arg max in (9.1) is non empty and GADDOW can then be defined.

Below, we prove more general theorems, valid for any $\text{GADDOW}\left((\widetilde{D}_g)_g\right)$. 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 9.1. $\text{GADDOW}\left((\widetilde{D}_g)_g\right)$ when $\widetilde{D}_g = \text{LCM}(\widehat{D}_g)$ and $\hat{\pi}_{g,0} = 1$ is exactly the same as IHW with modification (E1) defined in the supplementary material of Ignatiadis et al. (2016). In our notation, the latter is $\text{WBH}\left(\widetilde{W}^*\left(\tilde{u}_{\widetilde{W}^*}\right)\right)$, which is the same as $\text{GADDOW}\left((\widetilde{D}_g)_g\right)$ because $\tilde{u}_{\widetilde{W}^*} = \tilde{u}_{\widetilde{W}^*\left(\tilde{u}_{\widetilde{W}^*}\right)}$ (same proof as in Remark 9.2).

9.2. Proof of Theorem 5.1

We have

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

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

$$\text{FDP}\left(\text{GADDOW}\left((\widetilde{D}_g)_g\right)\right) \xrightarrow[m \to \infty]{\mathbb{P}} \frac{H_{W^*}^\infty(u^*)}{G_{W^*}^\infty(u^*)} = \frac{H_{W^*}^\infty(u^*)}{u^*},$$

and then

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

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

$$u^{*-1} H_{W^*}^\infty(u^*) = \alpha \sum_g \pi_g \pi_{g,0} W_g^*(u^*)$$

$$\leq \alpha \sum_g \pi_g \bar{\pi}_{g,0} W_g^*(u^*) = \alpha. \tag{9.2}$$

Moreover if (ME) holds (that is, there exists $C \geq 1$ such that $\bar{\pi}_{g,0} = C\pi_{g,0}$ for all g), we write

$$u^{*-1} H_{W^*}^\infty(u^*) = \alpha \sum_g \pi_g \pi_{g,0} W_g^*(u^*)$$

$$= \frac{\alpha}{C} \sum_g \pi_g \bar{\pi}_{g,0} W_g^*(u^*) = \frac{\alpha}{C}. \tag{9.3}$$

The equalities in (9.2) and (9.3) are due to the fact that $\sum_g \pi_g \bar{\pi}_{g,0} W_g^*(u^*) = 1$ (by Lemma B.1).

9.3. Proof of Theorem 5.2

First, in any case,

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

by Lemma C.5, where $P_{W^*}^\infty$ is defined in Section B. Hence:

$$\lim_{m \to \infty} \text{Pow}\left(\text{GADDOW}\left((\widetilde{D}_g)_g\right)\right) = P_{W^*}^\infty(u^*).$$

For the rest of the proof, we assume we are in case (ME), 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 $P_{W^*}^\infty$ 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 $\hat{w} \in \hat{K}$ therefore not depending on u. In the second part we will conclude with a general sequence of weight functions.

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

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

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

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

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

by continuity of F_g and because $\widetilde{\Omega}$ is realized. So $\ell' = P_{w^{cv}}^\infty(u_{w^{cv}}^\infty) \leq P_{W^*}^\infty(u_{w^{cv}}^\infty)$ by maximality. Note also that $G_{w^{cv}}^\infty(\cdot) \leq G_{W^*}^\infty(\cdot)$ which implies that $u_{w^{cv}}^\infty \leq u_{W^*}^\infty = u^*$ so $\ell' \leq P_{W^*}^\infty(u^*)$ because $P_{W^*}^\infty$ is nondecreasing. Finally,

$$\limsup \widehat{P}_{\hat{w}}(\hat{u}_{\hat{w}})\mathbb{1}_{\widetilde{\Omega}} \leq P_{W^*}^\infty(u^*)$$

for any realization of Ω, by integrating we get that $\ell \leq P_{W^*}^\infty(u^*)$ 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

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

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

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

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

Remark 9.2. 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 $\widehat{G}_{\widehat{W}^*} \geq \widehat{G}_{\hat{w}}$ and then $\hat{u} \geq u_{\hat{w}}$ for any $\hat{w}$. Hence $\hat{u} = \hat{u}_{\widehat{W}^*(\hat{u})}$ and ADDOW is equal to the WBH procedure associated to the weight vector $\widehat{W}^*(\hat{u})$.

Remark 9.3. We actually proved a stronger result, as we can replace the statement $\widehat{W} : [0,1] \rightarrow \hat{K}$ by $\widehat{W} : [0,1] \rightarrow \hat{K}^{\text{alt}}$, where

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

and each $\hat{\pi}_{g,0}^{\text{alt}}$ is such that $\hat{\pi}_{g,0}^{\text{alt}} \xrightarrow{\mathbb{P}} \bar{\pi}_{g,0}^{\text{alt}}$ for some $\bar{\pi}_{g,0}^{\text{alt}} \geq \bar{\pi}_{g,0}$. That is, the weight space $\widehat{W}$ belongs to does not have to be the same weight space where we apply ADDOW, as long as it uses over-estimators of the limits of the over-estimators used in $\hat{K}$.

Appendix A: Lemmas and proofs of Section 2

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

Proof. F_g is concave so it is continuous over $\mathbb{R} \setminus \{0,1\}$. F_g is continuous in 0 because it is càdlàg. F_g is continuous in 1 by concavity and monotonicity. $\square$

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*

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

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

$$|\hat{\pi}_{g,0}(\lambda_m) - \pi_{g,0}| = \left| \frac{1 - \frac{1}{m_g} \sum_i \mathbb{1}_{\{p_{g,i} \leq \lambda_m\}} + \frac{1}{m}}{1 - \lambda_m} - \pi_{g,0} \right|$$

$$\leq \frac{\lambda_m \left| \pi_{g,0} - \frac{m_{g,0}}{m_g} \right| + \frac{m_{g,0}}{m_g} \left| \lambda_m - \frac{1}{m_{g,0}} \sum_i \mathbb{1}_{\{p_{g,i} \leq \lambda_m, H_{g,i}=0\}} \right|}{1 - \lambda_m}$$

$$+ \frac{\left| \pi_{g,1} - \frac{m_{g,1}}{m_g} \right| + \frac{m_{g,1}}{m_g} \left| F_g(\lambda_m) - \frac{1}{m_{g,1}} \sum_i \mathbb{1}_{\{p_{g,i} \leq \lambda_m, H_{g,i}=1\}} \right|}{1 - \lambda_m}$$

$$+ \frac{m_{g,1}}{m_g} \frac{1 - F_g(\lambda_m)}{1 - \lambda_m} + \frac{1}{m(1 - \lambda_m)}$$

$$\leq \frac{m_{g,0}}{m_g} \frac{\sup_{x \in [0,1]} \left| x - \frac{1}{m_{g,0}} \sum_i \mathbb{1}_{\{p_{g,i} \leq x, H_{g,i}=0\}} \right|}{1 - \lambda_m}$$

$$+ \frac{m_{g,1}}{m_g} \frac{\sup_{x \in [0,1]} \left| F_g(x) - \frac{1}{m_{g,1}} \sum_i \mathbb{1}_{\{p_{g,i} \leq x, H_{g,i}=1\}} \right|}{1 - \lambda_m} + o(1).$$

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 $\frac{1}{1-\lambda_m} = o(\sqrt{m})$). $\qquad \square$

Definition A.1. The critical alpha value is

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

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

Lemma A.3. α^* *always satisfies* $\alpha^* < 1$.

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

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

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

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

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

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

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 continous 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 $\epsilon_n \to 0$ such that

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

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 defintion 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 $s_\gamma = \max\limits_{u' \in [u_+, 1]} (h_0(u') - u')$. 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

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

completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Lemma A.5. *Let a weight function* $W : [0,1] \to \mathbb{R}_+^G$. *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* $p_{(1)}^{[r]} \leq \ldots \leq p_{(m)}^{[r]}$ *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

$$\widehat{G}_W\left(\frac{\hat{r}}{m}\right) = m^{-1} \sum_{g=1}^{G} \sum_{i=1}^{m_g} \mathbb{1}_{\left\{p_{g,i} \leq \alpha \frac{\hat{r}}{m} W_g\left(\frac{\hat{r}}{m}\right)\right\}}$$

$$= m^{-1} \sum_{g=1}^{G} \sum_{i=1}^{m_g} \mathbb{1}_{\left\{p_{g,i}^{[\hat{r}]} \leq \alpha \frac{\hat{r}}{m}\right\}}$$

$$= m^{-1} \sum_{r=1}^{m} \mathbb{1}_{\left\{ p_{(r)}^{[\hat{r}]} \leq \alpha \frac{\hat{r}}{m} \right\}} \geq \hat{r}/m,$$

because $p_{(1)}^{[\hat{r}]}, \ldots, p_{(\hat{r})}^{[\hat{r}]} \leq \alpha \frac{\hat{r}}{m}$. 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 $\hat{u}_W = \widehat{G}_W(\hat{u}_W)$, 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

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

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

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

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

which entails $\widehat{G}_{\widehat{W}^*}(u') \geq \widehat{G}_{\widehat{W}^*}(u)$. $\qquad \square$

Appendix B: Asymptotical weighting

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

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

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

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 $m \to \infty$. If W is such that G_W^∞ is nondecreasing, we also define

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

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 $\hat{K}$. We now define oracle optimal weights over K^∞ for $G_\cdot^\infty(u)$ and $P_\cdot^\infty(u)$, 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. In this case, its only element w^* belongs to $[0, \frac{1}{\alpha u}]^G$ and satisfies $\sum_g \pi_g \bar{\pi}_{g,0} w_g^* = 1$. If $\alpha u \geq \bar{\pi}_0$ it is included in $[\frac{1}{\alpha u}, \infty)^G$.
Finally, $\max_{w \in K^\infty} G_w^\infty(u) \leq 1$ with equality if and only if $\alpha u \geq \bar{\pi}_0$.
The same statements are true for P^∞, except that $\max_{w \in K^\infty} P_w^\infty(u) \leq 1 - \pi_0$ with equality 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 $w_0 = (\frac{1}{\alpha u}, \ldots, \frac{1}{\alpha u}) = (\frac{1}{\bar{\pi}_0}, \ldots, \frac{1}{\bar{\pi}_0})$ 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 $g_1 \in \{1, \ldots, G\}$ 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 $g_1 \le G$ 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 would exist a $\tilde{w}$ with $\tilde{w}_{g_1} > w_{g_1}^*$ (for the same g_1 as in previous sentence) and $\tilde{w}_g = w_g^*$ for all $g \ne g_1$ 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

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

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

The proof is similar for $P_\cdot^\infty$, by replacing D_g by $\pi_{g,1} F_g$. $\qquad\square$

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 $\arg\max_{w \in \hat{K}} \widehat{G}_w(u)$), 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

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

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* $\alpha \ge \bar{\pi}_0$, *constant equal to 1 over* $[\frac{\bar{\pi}_0}{\alpha}, 1]$.

In particular, (i) $u^* = 1$ *if and only if* $\alpha \ge \bar{\pi}_0$, *(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 $u < u' \le \frac{\bar{\pi}_0}{\alpha} \wedge 1$, 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

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

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

$$\frac{G_w^\infty(x) - G_w^\infty(0)}{x - 0} = \frac{G_w^\infty(x)}{x} = \sum_g \pi_g \pi_{g,0} \alpha w_g + \sum_g \pi_g \pi_{g,1} \alpha w_g \frac{F_g(\alpha x w_g)}{\alpha x w_g}$$

$$\to \alpha \sum_g \pi_g w_g \left(\pi_{g,0} + \pi_{g,1} f_g(0^+)\right) > 1,$$

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, we have that $\alpha a W_g^*(a), \alpha b W_g^*(b) \le 1$ and then, for all g:

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

Moreover, because $G_{W^*}^\infty(a) < G_{W^*}^\infty(b)$, for at least one g_1 we have $aW_{g_1}^*(a) \neq bW_{g_1}^*(b)$ 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:

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

the inequality being strict for g_1. Finally by summing:

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

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 $G_{W^*}^\infty(u) \leq \sum_g \frac{m_g}{m} D_g(\alpha u B) \to 0$ when $u \to 0$ 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 nondecreasingness. □

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 $\widetilde{G}_{\widetilde{W}^*}$ when $\widetilde{D}_g = \text{LCM}\left(\widehat{D}_g\right)$.

Figure 13 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 (ME).

Lemma B.3. *If* (ME) *holds, for all* $u \in [0, 1]$,

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

In particular, $P_{W^*}^\infty$ 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
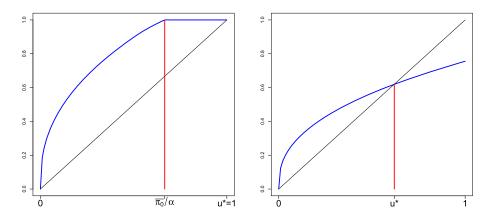
FIG 13. *Plot of $u \mapsto G_{W^*}^{\infty}(u)$ when $\alpha \geq \bar{\pi}_0$ (left panel) and $\alpha < \bar{\pi}_0$ (right panel).*

Lemma B.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 $\arg\max_{w \in K^{\infty}} P_w^{\infty}(u)$. Recall that there exists $C \geq 1$ such that, for all $1 \leq g \leq G$, $\bar{\pi}_{g,0} = C\pi_{g,0}$, and write, for all $w \in K^{\infty}$,

$$
\begin{aligned}
G_w^{\infty}(u) &= \sum_g \pi_g \pi_{g,0} U(\alpha u w_g) + P_w^{\infty}(u) \\
&\leq \alpha u \sum_g \pi_g \pi_{g,0} w_g + P_{w^*}^{\infty}(u) \\
&= \frac{\alpha u}{C} \sum_g \pi_g \bar{\pi}_{g,0} w_g + P_{w^*}^{\infty}(u) \\
&\leq \frac{\alpha u}{C} \times 1 + P_{w^*}^{\infty}(u) \\
&= \sum_g \pi_g \pi_{g,0} U(\alpha u w_g^*) + P_{w^*}^{\infty}(u) = G_{w^*}^{\infty}(u),
\end{aligned}
$$

because $\sum_g \pi_g \bar{\pi}_{g,0} w_g^* = 1$ and $\alpha u w_g^* \leq 1$ for all g, by Lemma B.1. This means that w^* is also the unique element of $\arg\max_{w \in K^{\infty}} G_w^{\infty}(u)$. Finally the properties on $P_{W^*}^{\infty}$ 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 $\alpha u \leq \bar{\pi}_0$,

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

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

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

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

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

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 $\alpha u \geq \bar{\pi}_0$, $M_u = \{w \in K^\infty : \alpha u w_g \geq 1 \forall g\}$ by Lemma B.1, hence (B.5). □

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 in probability:*

- $\sup_{w \in \mathbb{R}_+^G} \left\| \widehat{H}_w - H_w^\infty \right\|$,
- $\sup_{w \in \mathbb{R}_+^G} \left\| \widehat{P}_w - P_w^\infty \right\|$,
- $\sup_{w \in \mathbb{R}_+^G} \left\| \widehat{G}_w - G_w^\infty \right\|$,
- *for all $g \in \{1, \ldots, G\}$, $\left\| \widehat{D}_g - D_g \right\|$.*

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

$$\sup_{w \in \mathbb{R}_+^G} \left\| \widetilde{G}_w - G_w^\infty \right\| \xrightarrow{\mathbb{P}} 0. \tag{C.1}$$

Proof. By using the same proof as the one of the Glivenko-Cantelli theorem (which adapts trivially to convergence in probability instead of almost surely), we get from (2.1) and (2.2) that, for all g,

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

and

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

Next, we write that

$$\left\|\frac{1}{m_g}\sum_{i=1}^{m_g}\mathbb{1}_{\{p_{g,i}\leq\cdot,H_{g,i}=0\}}-\pi_{g,0}U\right\|\leq\left|\frac{m_{g,0}}{m_g}-\pi_{g,0}\right|$$
$$+\pi_{g,0}\left\|\frac{1}{m_{g,0}}\sum_{i=1}^{m_g}\mathbb{1}_{\{p_{g,i}\leq\cdot,H_{g,i}=0\}}-U\right\|$$
$$\xrightarrow{\mathbb{P}}0,$$

and similarly $\left\|\frac{1}{m_g}\sum_{i=1}^{m_g}\mathbb{1}_{\{p_{g,i}\leq\cdot,H_{g,i}=1\}}-\pi_{g,1}F_g\right\|\xrightarrow{\mathbb{P}}0.$

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

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

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

Finally,

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

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

G. Durand

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

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

which in turn implies that

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

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

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

and

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

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

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

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

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

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

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

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

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

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

$$\sup_u \left(G^\infty_{W^*}(u) - G^\infty_{\widetilde{W}^\dagger}(u) \right) \le \sup_u \left(G^\infty_{W^*}(u) - \widetilde{G}_{W^*}(u) \right)$$
$$+ \sup_u \left(\widetilde{G}_{W^*}(u) - \widetilde{G}_{\widetilde{W}^\dagger}(u) \right)$$
$$+ \sup_u \left(\widetilde{G}_{\widetilde{W}^\dagger}(u) - G^\infty_{\widetilde{W}^\dagger}(u) \right),$$

with $\sup_u \left(G^\infty_{W^*}(u) - \widetilde{G}_{W^*}(u) \right) \overset{\mathbb{P}}{\to} 0$ and $\sup_u \left(\widetilde{G}_{\widetilde{W}^\dagger}(u) - G^\infty_{\widetilde{W}^\dagger}(u) \right) \overset{\mathbb{P}}{\to} 0$ by (C.1).

Finally,

$$\sup_u \left(\widetilde{G}_{W^*}(u) - \widetilde{G}_{\widetilde{W}^\dagger}(u) \right) \le \sup_u \left(\widetilde{G}_{W^*}(u) - \widetilde{G}_{W^\dagger}(u) \right)$$
$$+ \sup_u \left(\widetilde{G}_{W^\dagger}(u) - \widetilde{G}_{\widetilde{W}^*}(u) \right)$$
$$+ \sup_u \left(\widetilde{G}_{\widetilde{W}^*}(u) - \widetilde{G}_{\widetilde{W}^\dagger}(u) \right),$$

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

$$\left\| \widetilde{G}_{\widetilde{W}^*} - G^\infty_{W^*} \right\| \le \sup_u \left(\widetilde{G}_{W^\dagger}(u) - \widetilde{G}_{\widetilde{W}^*}(u) \right) + V_m,$$

but $\widetilde{G}_{W^\dagger}(u) - \widetilde{G}_{\widetilde{W}^*}(u) \le 0$ by maximality of $\widetilde{G}_{\widetilde{W}^*}(u)$ over $\hat{K}$, so

$$\left\| \widetilde{G}_{\widetilde{W}^*} - G^\infty_{W^*} \right\| \le V_m \overset{\mathbb{P}}{\to} 0. \qquad \square$$

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

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

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

$$\begin{cases} \tilde{u} & \overset{\mathbb{P}}{\longrightarrow} & u^* \\ \widetilde{G}_{\widetilde{W}^*}(\tilde{u}) & \overset{\mathbb{P}}{\longrightarrow} & G^\infty_{W^*}(u^*). \end{cases}$$

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

$$\mathbb{P}\left(|\tilde{u} - u^*| \le \gamma \right) \ge \mathbb{P}\left(\left\| \widetilde{G}_{\widetilde{W}^*} - G^\infty_{W^*} \right\| \le \eta_\gamma \right) \underset{\text{Lemma C.2}}{\longrightarrow} 1.$$

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

Lemma C.4. (i) *If $\alpha \leq \bar{\pi}_0$, $\widetilde{W}^*(\tilde{u}) \overset{\mathbb{P}}{\longrightarrow} W^*(u^*)$.*

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

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

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

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

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

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

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

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), $\widetilde{W}^\dagger(\tilde{u}) \overset{\mathbb{P}}{\longrightarrow} W^*(u^*)$. But for all g

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

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

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

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

By equation (C.7) we also have

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

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

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

$$\widehat{G}_{\widetilde{W}^*}(\tilde{u}) \overset{\mathbb{P}}{\longrightarrow} G_{W^*}^\infty(u^*),$$
$$\widehat{H}_{\widetilde{W}^*}(\tilde{u}) \overset{\mathbb{P}}{\longrightarrow} H_{W^*}^\infty(u^*).$$

Proof. We have

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

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

(i) If $\alpha \leq \bar{\pi}_0$, $\tilde{u} \xrightarrow{\mathbb{P}} u^*$ and $\widetilde{W}^*(\tilde{u}) \xrightarrow{\mathbb{P}} W^*(u^*)$ by Lemma C.4. Then $\alpha \tilde{u} \widetilde{W}^*(\hat{u}) \xrightarrow{\mathbb{P}} \alpha u^* W^*(u^*)$. We get the desired convergence by D_g's continuity.

(ii) If $\alpha \geq \bar{\pi}_0$, $u^* = 1$ and $\alpha u^* W_g^*(u^*) \geq 1$ for all g so $G_{W^*}^\infty(u^*) = 1$. Then by Lemma C.4 $D_g\left(\alpha \tilde{u} \widetilde{W}_g^*(\hat{u})\right) \xrightarrow{\mathbb{P}} 1$ which means that $G_{\widetilde{W}^*}^\infty(\tilde{u}) \xrightarrow{\mathbb{P}} \sum_g \pi_g 1 = 1$.

The proof for $\hat{H}$ is similar, just replace D_g by $\pi_{g,0} U$. □

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

Lemma C.6. *Assume that* $\widetilde{D}_g = \mathrm{LCM}(\widehat{D}_g)$. *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

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

Then,

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

because $\widehat{D}_g$ is non decreasing.

Furthermore, C is concave. The inequality $C(\lambda x + (1-\lambda)y) \geq \lambda C(x) + (1-\lambda)C(y)$ for $x, y, \lambda \in [0,1]$, $x \leq y$, is trivial except in the case where $\lambda x + (1-\lambda)y \leq A \leq y$, where $A = 1 - (b-a)$. In this case, let λ' such that $\lambda x + (1-\lambda)y = \lambda'x + (1-\lambda')A$. Then

$$\begin{aligned} C(\lambda x + (1-\lambda)y) &= C(\lambda'x + (1-\lambda')A) \\ &\geq \lambda'C(x) + (1-\lambda')C(A) \\ &= \lambda'C(x) + (1-\lambda') \\ &\geq \lambda C(x) + (1-\lambda) = \lambda C(x) + (1-\lambda)C(y) \end{aligned}$$

where the last inequality is true because $\lambda' \leq \lambda$ and $C(x) \leq 1$. Indeed, we have

$$\begin{aligned} \lambda' &= \frac{\lambda x + (1-\lambda)y - A}{x - A} \\ &= \frac{\lambda(y-x) - (y-A)}{A - x} \end{aligned}$$

$$= \frac{\lambda(A-x) + \lambda(y-A) - (y-A)}{A-x}$$

$$= \lambda - (1-\lambda)\frac{y-A}{A-x} \leq \lambda.$$

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). □

Appendix D: Proof of Corollary 5.3 for Pro 1

First, $\hat{w}^{(1)} \xrightarrow{\mathbb{P}} w^{(1)}$ where $w^{(1)} = \left(\frac{1}{\bar{\pi}_0}, \ldots, \frac{1}{\bar{\pi}_0}\right)$ and $\hat{w}^{(2)} \xrightarrow{\mathbb{P}} w^{(2)}$ where, for all g, $w_g^{(2)} = \frac{\bar{\pi}_{g,1}}{\bar{\pi}_{g,0}(1-\bar{\pi}_0)}$. 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 $\widehat{W}_g^\dagger(u) = \frac{m_g \hat{\pi}_{g,0}}{m\pi_g \bar{\pi}_{g,0}} \widehat{W}_g^*(u)$ and note that the power of Pro1 is $\mathbb{E}\left[\widehat{P}_{\widehat{W}^*}(\hat{u}_M)\right]$. We have

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

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

$$\xrightarrow{\mathbb{P}} P_{W^*}^\infty(u_M),$$

because $P_{W^*}^\infty$ 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 7.1

First, note that, by the independence provided by Assumption 7.1, we can work conditionally to $(\mathbb{F}_m)_m$ and consider this sequence fixed and deterministic. Second, note that $m^{-1}|\{1 \leq i \leq m_g : \mathbb{F}_m(g,i) = f\}| \xrightarrow[m \to \infty]{} \frac{\pi_g}{F}$. Finally, note that the empirical function used for crADDOW is defined by

$$\widehat{G}^{\text{cross}} : u \mapsto m^{-1} \sum_{g=1}^G \sum_{f=1}^F \sum_{\substack{1 \leq i \leq m_g: \\ \mathbb{F}_m(g,i)=f}} \mathbb{1}_{\{p_{g,i} \leq \alpha u w_{g,f}^*\}}. \tag{E.1}$$

By using Assumption 7.2, we can apply Lemmas C.1, C.2, C.3 and C.4 to ADDOW$_{-f}$ for each f, so we get that $w_{g,f}^* \xrightarrow{\mathbb{P}} W_g^*(u^*)$. Using again Lemma C.1 for each fold, and then using the continuity of G^∞, we get that

$$\left\| \widehat{G}^{\text{cross}} - G_{W^*(u^*)}^\infty \right\| \xrightarrow{\mathbb{P}} 0.$$

Then, by Remark B.2,

$$\hat{u}^{cross} \xrightarrow{\mathbb{P}} u_{W^*(u^*)}^\infty,$$

where $\hat{u}^{cross} = \mathcal{I}\left(\widehat{G}^{\text{cross}}\right)$ is the step-up threshold of crADDOW. Now, by using the asymptotic counterpart of Remark 9.2, we have that $u_{W^*(u^*)}^\infty = u^*$. Then we can proceed as in the proofs of Lemma C.5 and Theorems 5.1 and 5.2 to finally obtain that

$$\lim_{m \to \infty} \text{FDR}\left(\text{crADDOW}\right) = \frac{H_{W^*}^\infty(u^*)}{u^*}$$

and

$$\lim_{m \to \infty} \text{Pow}\left(\text{crADDOW}\right) = P_{W^*}^\infty(u^*),$$

which concludes.

Remark E.1. To keep notation light, we did not introduce the generalization of Section 9.1 but we could use it to get a Theorem with a generalized crADDOW.

Remark E.2. It is easy to see that crADDOW is a WBH procedure with $G \times F$ groups when using the expression given by Equation (E.1).

Acknowledgments

References

ARIAS-CASTRO, E., CHEN, S. et al. (2017). Distribution-free multiple testing. *Electronic Journal of Statistics* **11** 1983–2001. MR3651021

Barber, R. F., Candès, E. J. et al. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics* **43** 2055–2085. MR3375876

Benjamini, Y. and Hochberg, Y. (1997). Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics* **24** 407–418. MR1481424

Blanchard, G. and Roquain, E. (2008). Two simple sufficient conditions for FDR control. *Electronic Journal of Statistics* **2** 963–992. MR2448601

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. MR2597000

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. MR1926068

Chi, Z. (2007). On the performance of FDR control: constraints and a partial solution. *The Annals of Statistics* **35** 1409–1431. MR2351091

Delattre, S. and Roquain, E. (2016). On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing. *Bernoulli* **22** 302–324. MR3449784

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. MR3431551

Efron, B. (2008). Simultaneous inference: When should hypothesis testing problems be combined? *The Annals of Applied Statistics* 197–223. MR2415600

Genovese, C. R., Roeder, K. and Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika* 509–524. MR2261439

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Annals of Statistics* 1035–1061. MR2065197

Habiger, J. D. (2014). Weighted Adaptive Multiple Decision Functions for False Discovery Rate Control. *arXiv preprint* arXiv:1412.0645.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 65–70. MR0538597

Hu, J. X., Zhao, H. and Zhou, H. H. (2010). False discovery rate control with groups. *Journal of the American Statistical Association* **105**. MR2752616

Ignatiadis, N. and Huber, W. (2017). Covariate powered cross-weighted multiple testing. *arXiv preprint* arXiv:1701.05179.

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.

Lei, L. and Fithian, W. (2018). AdaPT: an interactive procedure for multiple testing with side information. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 649–679. MR3849338

Li, A. and Barber, R. F. (2016). Multiple testing with the structure adaptive Benjamini-Hochberg algorithm. *arXiv preprint* arXiv:1606.07926.

MR3904779

Neuvial, P. (2008). Asymptotic properties of false discovery rate controlling procedures under independence. *Electronic Journal of Statistics* **2** 1065–1110. MR2460858

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. MR3081930

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. MR3988773

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. MR2779334

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 preprint* 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. MR2521216

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. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64** 479–498. MR1924302

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. MR2035766

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* arXiv: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. MR3216640