Optimal weighting for false discovery rate control

Etienne Roquain, Mark van de Wiel

HAL Id: hal-00306219
https://hal.science/hal-00306219v3

Submitted on 13 Jul 2009

Optimal weighting for false discovery rate control

Etienne Roquain[*]

*University of Paris 6, LPMA,
4, Place Jussieu, 75252 Paris cedex 05, France
e-mail:* etienne.roquain@upmc.fr

and

Mark A. van de Wiel

*VU University & VU University Medical Center,
De Boelelaan 1117, 1007MB Amsterdam, The Netherlands
e-mail:* mark.vdwiel@vumc.nl

Abstract: How to weigh the Benjamini-Hochberg procedure? In the context of multiple hypothesis testing, we propose a new step-wise procedure that controls the false discovery rate (FDR) and we prove it to be more powerful than any weighted Benjamini-Hochberg procedure. Both finite-sample and asymptotic results are presented. Moreover, we illustrate good performance of our procedure in simulations and a genomics application. This work is particularly useful in the case of heterogeneous p-value distributions.

AMS 2000 subject classifications: Primary 62J15; secondary 62G10.
Keywords and phrases: False discovery rate, multiple testing, p-value weighting, power maximization.

Received June 2009.

1. Introduction

In many practical situations, e.g. functional Magnetic Resonance Imaging (fMRI) or microarray data, the problem of testing simultaneously a large number m of null hypotheses arises. Since Neyman-Pearson's approach is the most commonly used strategy for single testing, many researches have focused on generalizing this approach to the multiple testing case. First, one should choose a global type I error to be controlled, as the probability of making at least one false discovery (family-wise error rate, FWER) or more recently the mean proportion of false discoveries among all the discoveries (false discovery rate, FDR, see Benjamini and Hochberg (1995)). Second, one should build a procedure that controls the so-chosen global type I error rate. For instance, Benjamini and Hochberg (1995) proved that the linear step-up procedure (LSU) controls the FDR when the underlying tests are independent. Third, one should show that

---

[*]Research partly carried out at the VU University, Amsterdam.

the obtained procedure has good power performance, the power being generally defined as the expected number of true discoveries.

To our knowledge, while the two first points above are widely studied (e.g. building FDR controlling procedures, see e.g. Benjamini and Yekutieli (2001); Storey (2002); Sarkar (2002))), the last point is most of the time evaluated with simulations, without a full theoretical support. Only few works have studied rigorously the optimality of certain classes of multiple testing procedures (see Lehmann et al. (2005); Wasserman and Roeder (2006); Rubin et al. (2006); Storey (2007); Finner et al. (2009)).

Maximizing the power while controlling the FDR remains a difficult task, because the FDR involves a random denominator (the number of discoveries). The present paper gives a contribution to the latter maximization problem, in the simple case where the null and alternative distributions are known. This framework is natural for the power maximization of tests, as it was also used in Neyman-Pearson's lemma for single testing. Although leading to oracle procedures, it can be used in practice as soon as the null and alternative distributions are estimated or guessed reasonably accurately from independent data.

More formally, assume that each hypothesis is tested using a test statistic, that can then be transformed into a p-value p_i, and denote by F_i the alternative c.d.f. of p_i. In general, the F_i's can be possibly very different (e.g. with heterogeneous underlying data) and the p-values cannot be considered interchangeably. Therefore, a p-value weighting approach seems appropriate to improve the performance of a multiple testing procedure. This technic, that can be traced back to Holm (1979), consists in replacing in input each original p-value p_i by the weighted p-value $p'_i = p_i/w_i$ for some weight vector $(w_1, \ldots, w_m)$ summing to m. Here, we focus on the weighted version of the LSU procedure that was proposed in Genovese et al. (2006) (see also Blanchard and Roquain (2008)). In the latter paper, it was demonstrated that the weighted LSU still controls the FDR for any weighting (under independence between the p-values), and that some of these procedures could improve the power of the LSU asymptotically. In the present paper, *we aim to find the most powerful procedure among all the weighted LSU procedures*, or more precisely, *to find a procedure that mimics the best procedure among the weighted LSU procedures*. Moreover, this procedure should be computable from the p-values distributions, i.e. the F_i's.

When using the weighted version of the FWER-controlling Bonferroni procedure, Wasserman and Roeder (2006) (see also Rubin et al. (2006)) have found the optimal weighting. In Storey (2007), an optimal procedure was also proposed, maximizing the expected number of true discoveries while controlling the expected number of false discoveries. All these procedures use deterministic thresholds, which make the power maximization feasible. However, in the case of the FDR-controlling weighted LSU, the threshold depends on the final number of discoveries and a power maximization seems very difficult to make, even in the asymptotic framework where the number of p-values m tends to infinity (see Genovese et al. (2006)).

The main idea of this paper is to find the optimal weights simultaneously for all the possible rejection proportions $u \in [0, 1]$. These multi-weights are then

collected in optimal weight functions $u \mapsto W_i^\star(u)$ which in turn are sequentially integrated in a step-up procedure. While the LSU procedure uses as threshold function $u \mapsto \alpha u$, we find that the new procedure uses a threshold function $u \mapsto \alpha u W_i^\star(u)$ not necessarily linear (and depending on the $F_i's$).

The new procedure, called "optimal multi-weighted step-up procedure", will be presented in detail in Section 3. In Section 4, we show that it enjoys the following properties:

(i) FDR control for a finite number of hypotheses, up to slight modifications;
(ii) power optimality for a finite number of hypotheses, up to error terms;
(iii) power optimality without error term and FDR control without modification when the number of hypotheses m tends to infinity (in a specific asymptotic setting).

These results are established in two different (classical) models of p-values, both assuming independence between the p-values. The results (ii) and (iii) additionally use that the F_i's are strictly concave functions and that the maximization of the power at any rejection proportion is feasible, which remain quite mild assumptions.

In Section 5, we present a simulation study which exhibits the behavior of the new procedure when the F_i's are correctly specified or misspecified. Section 5 discusses some applications and our conclusions are given in Section 7. All our results are proved in Section 8, while some technical parts are gathered in Appendix. Our proofs mainly use the "self-consistency condition" introduced in Blanchard and Roquain (2008) (see also Finner et al. (2009)) and Hoeffding's inequality (see Hoeffding (1963)).

2. Preliminaries

2.1. Models for the p-values

Let us first define the two different models for the p-values that will be used throughout the paper.

We consider a finite set of m null hypotheses on a probability space and we let $H_i := 0$ (resp. 1) if the i-th null hypothesis is true (resp. false). Letting $\mathbf{H} := (H_i)_{1 \le i \le m} \in \{0,1\}^m$, we denote by $\mathcal{H}_0 := \{i \in \{1, \ldots, m\} \mid H_i = 0\}$ the set corresponding to the true null hypotheses and by $m_0 := |\mathcal{H}_0|$ its cardinal. Analogously, we define $\mathcal{H}_1 := \{i \in \{1, \ldots, m\} \mid H_i = 1\}$ and $m_1 := |\mathcal{H}_1|$ for the alternative hypotheses. Since $\mathcal{H}_1$ is the complement of $\mathcal{H}_0$ in $\{1, \ldots, m\}$, we have $m_1 = m - m_0$. The proportion of true nulls (resp. false nulls) is denoted by $\pi_0 := m_0/m$ (resp. $\pi_1 := m_1/m$) as usual. We suppose that for the i-th null hypothesis it is given a p-value p_i i.e. a measurable function from the observation space into $[0,1]$ such that the distribution of p_i is uniform on $[0,1]$ when the i-th null hypothesis is true:

$$\forall i \in \mathcal{H}_0, \ \forall t \in [0,1], \ \mathbb{P}(p_i \le t) = t. \tag{1}$$

Under the alternative, we denote by F_i the cumulative distribution function of p_i: $\forall i \in \mathcal{H}_1$, $\forall t \in [0,1]$, $F_i(t) := \mathbb{P}(p_i \leq t)$. In our setting, the F_i's are allowed to be different and we denote $\mathbf{F} := (F_i)_{i \in \mathcal{H}_1}$ the family of alternative c.d.f.'s. The p-values are assumed mutually independent. The latter model has parameters $(\mathbf{H}, \mathbf{F})$ and will be referred troughout the paper as the *conditional model* (because it uses a fixed vector $\mathbf{H}$).

Additionally, we will consider the so-called random effects model (see e.g. Efron et al. (2001); Storey (2003); Genovese and Wasserman (2004)). In this model, $\mathbf{H}$ is generated independently from all other random variables, from m i.i.d. Bernoulli priors. The probability for a null to be true (resp. false) is denoted by $\pi_0 := \mathbb{P}(H_i = 0) \in (0,1)$ (resp. $\pi_1 := 1 - \pi_0$). Then, the p-values are assumed to follow the conditional model conditionally to $\mathbf{H}$: the p-values are mutually independent conditional to $\mathbf{H}$ and each p_i is uniform conditional to $H_i = 0$ (i.e. satisfies (1) conditional to $H_i = 0$) and has for alternative c.d.f. F_i conditional to $H_i = 1$. As a consequence, unconditionally, the p-values are independent and for $i = 1, \ldots, m$, the c.d.f. of each p-value p_i is $t \mapsto \pi_0 t + \pi_1 F_i(t)$. This model has for parameters $(\pi_0, \mathbf{F})$ where $\mathbf{F} = (F_i)_{1 \leq i \leq m}$ is the family of alternative c.d.f.'s. The latter model will be referred trough the paper as the *unconditional model*.

2.2. Assumptions and notation

We introduce the following possible regularity assumptions on the parameter $\mathbf{F}$ of each model, the derivative of F_i being denoted by f_i:

$$\text{the } F_i\text{'s are continuous, strictly concave functions on } [0,1]; \qquad (A1)$$

$$\text{the } F_i\text{'s are twice differentiable on } (0,1); \qquad (A2)$$

$$\text{the functions } i \mapsto f_i(0^+) \text{ and } i \mapsto f_i(1^-) \text{ are constant} \qquad (A3)$$

$$\text{for each } i, j, \lim_{y \to f_i(0^+)} f_j^{-1}(y)/f_i^{-1}(y) \text{ exists in } [0, +\infty]. \qquad (A4)$$

As illustration, the assumptions $(A1)$–$(A4)$ are all satisfied in the one-sided testing Gaussian case where we test for any i the null "$\mu_i = 0$" against "$\mu_i > 0$" from a Gaussian test statistic of mean μ_i and variance 1. In that case, we have

$$F_i(x) = \overline{\Phi}\big(\overline{\Phi}^{-1}(x) - \mu_i\big) \quad \text{and} \quad f_i(x) = \exp\left\{\mu_i\left(\overline{\Phi}^{-1}(x) - \frac{\mu_i}{2}\right)\right\}, \qquad (2)$$

where we denoted $\overline{\Phi}(z) := \mathbb{P}\left[Z \geq z\right]$ for $Z \sim \mathcal{N}(0,1)$.

Finally, for any non-decreasing function $F : [0,1] \to [0,1]$ we denote

$$\mathcal{I}(F) := \sup\{u \in [0,1] \mid F(u) \geq u\},$$
$$\mathcal{J}(F) := \sup\{u \in [0,1] \mid \forall u' \leq u, F(u') \geq u'\},$$

and for $\lambda > 0$,

$$\mathcal{I}_\lambda^+(F) := (\mathcal{I}(F) + \lambda) - F(\mathcal{I}(F) + \lambda) \quad \text{when} \ \ \lambda < 1 - \mathcal{I}(F),$$
$$\mathcal{I}_\lambda^-(F) := F(\mathcal{I}(F) - \lambda) - (\mathcal{I}(F) - \lambda) \quad \text{when} \ \ \lambda < \mathcal{I}(F).$$

We easily check that $\mathcal{I}(F)$ and $\mathcal{J}(F)$ are maxima and that $F(\mathcal{I}(F)) = \mathcal{I}(F)$ and $F(\mathcal{J}(F)) = \mathcal{J}(F)$.

2.3. Multiple testing procedures, FDR and power

A *multiple testing procedure R* is defined as an algorithm which, from the data, aims to reject part of the null hypotheses. Below, we will consider, as is usually the case, multiple testing procedures which can be written as a function of the family of p-values $\mathbf{p} = (p_i, i \in \{1, \dots, m\})$. More formally, we define a multiple testing procedure as a measurable function R, which takes as input a realization of the p-value family $\mathbf{p} \in [0, 1]^m$ and which returns a subset $R(\mathbf{p})$ of $\{1, \dots, m\}$, corresponding to the rejected hypotheses (i.e. $i \in R(\mathbf{p})$ means that the i-th hypothesis is rejected by R).

As introduced by Benjamini and Hochberg (1995), the false discovery rate (FDR) of a multiple testing procedure is defined as the mean proportion of true hypotheses in the set of the rejected hypotheses:

$$\text{FDR}(R) = \mathbb{E}\left[\frac{|\mathcal{H}_0 \cap R(\mathbf{p})|}{|R(\mathbf{p})| \vee 1}\right], \tag{3}$$

where $|\cdot|$ denotes the cardinality function. Of course, the FDR in (3) depends on the model chosen for the p-values. In particular, the FDR in the conditional model involves an expectation taken conditionally to $\mathbf{H}$, whereas the FDR in the unconditional model additionally uses an averaging over $\mathbf{H}$. It is worth noticing that, if a procedure controls the FDR in the conditional model, that is conditionally to any value of $\mathbf{H} \in \{0, 1\}^m$, it controls also the FDR unconditionally.

Finally, we use the standard power criterium equal to the mean proportion of correctly rejected hypotheses, that is,

$$\text{Pow}(R) = m^{-1}\mathbb{E}\big[|\mathcal{H}_1 \cap R(\mathbf{p})|\big]. \tag{4}$$

In the notation below, we will sometimes drop the explicit dependency in $\mathbf{p}$ for short, writing e.g. R instead of $R(\mathbf{p})$.

2.4. Weighted linear step-up procedures

Let us consider $\mathbf{w} = (w_i)_i$ a vector of non-negative real numbers such that $\sum_{i=1}^{m} w_i = m$, called here a *weight vector*, and consider the weighted p-values $p_i' = p_i/w_i$, ordered as: $p_{(1)}' \leq \dots \leq p_{(m)}'$ with the convention $p_{(0)}' = 0$.

As introduced by Genovese et al. (2006), the *weighted linear step-up procedure* associated to $\mathbf{w}$, denoted here by $\mathbf{LSU}(\mathbf{w})$, rejects the i-th hypothesis if $p_i' \leq \alpha\widehat{u}$, with

$$\widehat{u} = m^{-1}\max\{r \in \{0, 1, \dots, m\} \mid p_{(r)}' \leq \alpha r/m\}. \tag{5}$$

In particular, the procedure $\mathbf{LSU}(\mathbf{w})$ using $\forall i, w_i = 1$ corresponds to the standard linear step-up procedure of Benjamini and Hochberg (1995), denoted here by $\mathbf{LSU}$. Letting $\widehat{\mathbb{G}}_{\mathbf{w}}(u) = m^{-1}\sum_{i=1}^{m} \mathbf{1}\{p_i \leq \alpha w_i u\}$, the rejection proportion $\widehat{u}$ can equivalently be defined as

$$\widehat{u} = \mathcal{I}(\widehat{\mathbb{G}}_{\mathbf{w}}), \tag{6}$$

using the notation of Section 2.2. Contrary to (5), expression (6) does not make any specific use of the reordered p-values $p'_{(1)}, \ldots, p'_{(m)}$, so that it is generally more convenient from a mathematical point of view.

For any choice of weight vector $\mathbf{w}$, Genovese et al. (2006) proved that the weighted linear step-up procedure controls the FDR at level $\alpha m^{-1} \sum_i (1 - H_i) w_i \leq \alpha$ in the conditional model and (thus) at level $\pi_0 \alpha \leq \alpha$ in the unconditional model.

3. New approach

We present in this section a new family of multiple testing procedures, called multi-weighted procedures. We start by motivating their introduction from the power optimization problem among the family of weighted linear step-up procedures.

3.1. Weight functions

Following Genovese et al. (2006), the explicit computation of the power of the LSU($\mathbf{w}$) is a difficult task (even asymptotically): it depends on the final proportion of rejections of the procedure $\widehat{u} = \mathcal{I}(\widehat{\mathbb{G}}_{\mathbf{w}})$, which is a random variable itself depending on $\mathbf{w}$. Therefore, we propose here to perform the optimization for each fixed rejection proportion u which in turn leads to a family of optimal weight vectors depending on u, $0 < u \leq 1$.

First, define the power of the procedure that thresholds each p-value p_i at level $\alpha w_i u$:

$$\text{Pow}_u(\mathbf{w}) := \text{Pow}(\{i \mid p_i \leq \alpha w_i u\}), \tag{7}$$

corresponding intuitively to the "power of the LSU($\mathbf{w}$) at rejection proportion u".

Second, define a *weight (vector) function* as a function $\mathbf{W} : u \in (0,1] \mapsto \mathbf{W}(u) = (W_i(u))_i \in (\mathbb{R}^+)^m$ such that each $\mathbf{W}(u)$ is a weight vector, that is, $\forall u \in (0,1]$, $\sum_{i=1}^m W_i(u) = m$ and such that the following property holds

$$\forall i \in \{1, \ldots, m\}, \quad u \mapsto W_i(u)\, u \text{ is nondecreasing on } (0,1]. \tag{8}$$

Additionally, a weight function is said *continuous* if for all i, $u \in (0,1] \mapsto W_i(u)$ are continuous functions.

Definition 3.1. Any weight function $\mathbf{W}^\star$ solving simultaneously the maximization problems:

$$\forall u \in (0,1], \ \text{Pow}_u(\mathbf{W}^\star(u)) = \max \left\{ \text{Pow}_u(\mathbf{w}), \mathbf{w} \text{ weight vector} \right\}, \tag{9}$$

is called the *optimal weight function*.

Note that $\mathbf{W}^\star$ is called here abusively "the" optimal weight function even if it is not proved to be unique. Of course the optimal weight function depends

on the model chosen for the p-values. The following proposition gives (strong) sufficient conditions for existence and unicity of the optimal weight function in the different models described in Section 2.1.

Proposition 3.2. *Assume* $(A1)$–$(A2)$–$(A3)$ *and denote the derivative of* F_i *by* f_i. *Then the weight function* $\mathbf{W}^\star$ *satisfying* (9) *exists and is unique in either of the following cases:*

- *In the conditional model, if* $\alpha < \pi_1$, *with for all* $u \in (0, 1]$,

$$W_i^\star(u) = (\alpha u)^{-1} f_i^{-1}\big(y^\star(u)\big)\mathbf{1}\{H_i = 1\}. \tag{10}$$

- *In the unconditional model, with for all* $u \in (0, 1]$,

$$W_i^\star(u) = (\alpha u)^{-1} f_i^{-1}\big(y^\star(u)\big).$$

In each case, $y^\star(u)$ *is defined as the unique element providing* $\sum_{i=1}^m W_i^\star(u) = m$. *Moreover, in both models, the weight function* $\mathbf{W}^\star$ *is continuous, and assuming in addition* $(A4)$, *the limits* $W_i^\star(0^+)$ *exist for all* i.

The proof, which is based on similar arguments than those proposed in Rubin et al. (2006) and Wasserman and Roeder (2006), is given in Section 8.6. Of course, the optimal weight function depends on the parameters of the model: on $(\mathbf{H}, \mathbf{F})$ in the conditional model and on $\mathbf{F}$ (only) in the unconditional model.

For instance, when the p-values are generated from the Gaussian model (2), the optimal weight function in the unconditional model is given by

$$W_i^\star(u) = (\alpha u)^{-1} \overline{\Phi}\bigg(\frac{\mu_i}{2} + \frac{c(u)}{\mu_i}\bigg), \tag{11}$$

where $c(u)$ is the unique element of $\mathbb{R}$ such that $\sum_{i=1}^m W_i^\star(u) = m$. It therefore only depends on the vector of alternative means $\mu = (\mu_i)_{1 \le i \le m}$. Figure 1 displays the optimal weight vectors $\mathbf{W}(u)$ for a particular choice of means and different values of u. We observe that $\mathbf{W}(u)$ strongly depends on u: for $u = 1$, the weight vector is larger for small means, whereas as u decreases, the weight vector is maximum on larger means. In particular, for small u, the weighting is close to zero for the smallest means, because they produce p-values much larger than αu (with high probability).

The Gaussian formula (11) can be also suitable for test statistics "close to be Gaussian", namely for locally uniform asymptotically normal test statistics (see Chapter 14 of van der Vaart (1998) and Section 4.3 and Section 7 of Roquain and van de Wiel (2008)). This is the case for instance for the Mann-Whitney test statistic.

3.2. Multi-weighted procedures

From the previous section, we have now to integrate several weight vectors in a single multiple testing procedure, or more precisely to use a weight vector $\mathbf{w}$ which may depend on u. For this, we extend the definition of weighted linear procedures to the case of multi-weighted procedures.
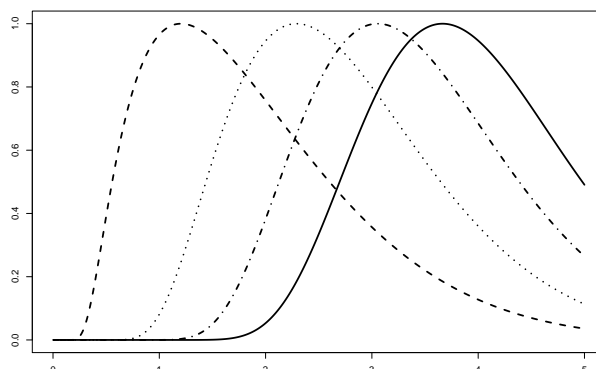
FIG 1. *Plot of the optimal weights $(W_i^\star(u))_i$ in function of the alternative means $(\mu_i)_i$, for $u = 1/m$ (solid), $u = 10/m$ (dashed-dotted), $u = 100/m$ (dotted), $u = 1$ (dashed). Unconditional model and Gaussian one-sided case with $m = 1000$, $\alpha = 0.05$, $\mu_i = 5i/m$ for $i = 1, \ldots, m$. Each curve is normalized to have a maximum equal to 1.*

First, we define the *threshold collection* $\Delta = (\Delta_i(u))_{i,u}$ *associated to a given weight function* $\mathbf{W}(\cdot) = (W_i(\cdot))_i$ by $\forall i \in \{1, \ldots, m\}$, $\forall u \in [0, 1]$,

$$\Delta_i(u) := \alpha \, W_i(u) \, u \text{ if } u > 0 \text{ and } \Delta_i(0) := 0.$$

Conversely, given any threshold collection $\Delta = (\Delta_i(u))_{i,u}$ such that each Δ_i is nonnegative, nondecreasing on $[0, 1]$ and such that $\forall u \in (0, 1]$, $m^{-1} \sum_i \Delta_i(u) = \alpha u$, we define *the weight function* $\mathbf{W} = (W_i(u))_{i,u}$ *associated to* Δ by $\forall i \in \{1, \ldots, m\}$, $\forall u \in (0, 1]$, $W_i(u) := \Delta_i(u)/(\alpha u)$. As a consequence, the threshold collection Δ and the weight function $\mathbf{W}$ are one to one associated.

Definition 3.3. Consider a weight function $\mathbf{W}(\cdot) = (W_i(\cdot))_i$ and its associated threshold collection Δ. The *multi-weighted step-up procedure* with weight function $\mathbf{W}$, denoted by $\mathbf{SU}(\mathbf{W})$, rejects the i-th null hypothesis if $p_i \leq \Delta_i(\widehat{u})$, where

$$\widehat{u} = \mathcal{I}(\widehat{\mathbb{G}}_{\mathbf{W}}), \tag{12}$$

and where we denoted $\widehat{\mathbb{G}}_{\mathbf{W}}(u) := m^{-1} \sum_{i=1}^m \mathbf{1}\{p_i \leq \Delta_i(u)\}$ for all $u \in [0, 1]$.

In particular, in the case where for all u, $W_i(u) = w_i$ is independent of u, the procedure $\mathbf{SU}(\mathbf{W})$ reduces to $\mathbf{LSU}(\mathbf{w})$. More generally, the above definition of $\mathbf{SU}(\mathbf{W})$ allows to choose thresholding $\Delta_i(u)$ not linear in u.

As for $\mathbf{LSU}(\mathbf{w})$, the multi-weighted procedure $\mathbf{SU}(\mathbf{W})$ can also be derived from a re-ordering based algorithm. The main difference is that the original p-values are ordered in several ways, because several weighting are used. Namely, if for $r \geq 1$, q_r denotes the r-th smallest $\mathbf{W}(r/m)$-weighted p-value i.e. is equal to $p'_{(r)}$ where $\forall i, p'_i = p_i/W_i(r/m)$ and putting $q_0 = 0$, we have

$$\widehat{u} = m^{-1} \max\{r \in \{0, 1, \ldots, m\} \mid q_r \leq \alpha r/m\}.$$

Similarly to the step-up case, we can define the *multi-weighted step-down procedure* with weight function $\mathbf{W}$ (and associated threshold collection Δ), denoted by $\mathbf{SD}(\mathbf{W})$, as rejecting the i-th null hypothesis if $p_i \leq \Delta_i(\widetilde{u})$, where

$$\widetilde{u} = \mathcal{J}(\widehat{\mathbb{G}}_{\mathbf{W}}), \tag{13}$$

or equivalently, $\widetilde{u} = m^{-1} \max\{r \in \{0, 1, \ldots, m\} \mid \forall r' \leq r, \; q_{r'} \leq \alpha r'/m\}$.

Remark that the procedures $\mathbf{SU}(\mathbf{W})$ and $\mathbf{SD}(\mathbf{W})$ only use the values of $\mathbf{W}(u)$ for $u \in \{1/m, 2/m, \ldots, 1\}$, which makes them easily computable. We refer the reader to Appendix A for explicit algorithmic versions of the procedures $\mathbf{SU}(\mathbf{W})$ and $\mathbf{SD}(\mathbf{W})$.

The particular multi-weighted step-up procedure $\mathbf{SU}(\mathbf{W}^{\star})$ using the optimal weight function $\mathbf{W}^{\star}$ is called the *optimal multi-weighted procedure*. From an intuitive point of view, since this weighting maximizes the power at any rejection proportion, the latter procedure should be more powerful than any standard weighted procedure $\mathbf{LSU}(\mathbf{w})$. One of the goal of Section 4 is to state this optimality result formally.

Finally, let us remark that in the unconditional model and under the assumptions and notation of Proposition 3.2, the optimal multi-weighted step-up procedure may be written under the following form: reject the i-th hypothesis if $f_i(p_i) \geq y^{\star}(\widehat{u})$, where f_i is the (decreasing) alternative density of p_i and $y^{\star}(\widehat{u})$ is adjusted from all the p-values, the f_i's and the pre-specified level α. As a consequence, this procedure is based on individual tests of Neyman-Pearson's type (the observed variables being restricted to the p-values).

4. Main results

We present in this section the main properties of the multi-weighted procedures. First, the finite-sample FDR control of $\mathbf{SU}(\mathbf{W})$ for any weight function $\mathbf{W}$, up to slight modifications. Second, the finite-sample power optimality of the procedure $\mathbf{SU}(\mathbf{W}^{\star})$ (using the optimal weight function), up to some small error terms. Third, a consistency result, proving that the latter slight modifications are unnecessary and that error terms vanish, in a particular asymptotic setting where m tends to infinity.

4.1. Finite-sample FDR control

First, let us recall that for any choice of weight vector $\mathbf{w} = (w_1, \ldots, w_m)$, the weighted linear step-up procedures $\mathbf{LSU}(\mathbf{w})$ controls the FDR at level $\alpha m^{-1} \sum_i (1 - H_i) w_i \leq \alpha$ in the conditional model and at level $\pi_0 \alpha \leq \alpha$ in the unconditional model (see Genovese et al. (2006); Blanchard and Roquain (2008)). These controls are non-asymptotic, in the sense that they are valid for any finite $m \geq 2$.

Unfortunately, the procedure $\mathbf{SU}(\mathbf{W})$ cannot be proved to control the FDR at level α for any choice of weight function $\mathbf{W}$ and for any $m \geq 2$. In Appendix C,

a (least favorable) choice of weight function is given when $m = 2$, for which FDR($\mathbf{SU}(\mathbf{W})$) slightly exceeds α. Therefore, in order to obtain rigorous FDR control for each m and any weight function, we need to slightly correct $\mathbf{SU}(\mathbf{W})$.

Theorem 4.1. *Consider* $\mathbf{W}(\cdot) = (W_i(\cdot))_i$ *any weight function. Then for any finite* $m \geq 2$, *the two following procedures*

- $\mathbf{SU}(\widetilde{\mathbf{W}})$ *with* $\widetilde{W}_i(u) = W_i(u)/(1 + \alpha W_i(1))$,
- $\mathbf{SD}(\widetilde{\mathbf{W}})$ *with* $\widetilde{W}_i(u) = W_i(u)/(1 + \alpha u W_i(u))$,

have their FDR less than or equal to

$$\alpha \max_{1 \leq k \leq m} \left\{ m^{-1} \sum_{i=1}^{m} (1 - H_i) W_i(k/m) \right\} \leq \alpha$$

in the conditional model. As a consequence, their FDR are less than or equal to $\alpha \mathbb{E} \left(\max_{1 \leq k \leq m} \left\{ m^{-1} \sum_{i=1}^{m} (1 - H_i) W_i(k/m) \right\} \right) \leq \alpha$ *in the unconditional model.*

The proof of Theorem 4.1 is given in Section 8.3. Note that this result covers the earlier result of Genovese et al. (2006); Blanchard and Roquain (2008), by taking $W_i(u) = w_i$ constant in u.

Since from (8), we have $\alpha u W_i(u) \leq \alpha W_i(1)$, both modifications of $\mathbf{SU}(\mathbf{W})$ proposed above should be not too large when $\alpha W_i(1)$ is close to 0 (e.g. when α is small). Furthermore, while the correction proposed in the weighting of the step-up procedure is more conservative than the one of the step-down, a step-up procedure is always more powerful than a step-down procedure (for the same threshold collection). Therefore, in general, no modified procedure dominates the other. Nevertheless, in the particular simulation setting of Section 5, we will see that the step-down modification appears to be better.

When using the optimal weight function $\mathbf{W}^\star$, Theorem 4.1 provides two modifications of the optimal multi-weighted procedure $\mathbf{SU}(\mathbf{W}^\star)$ that control the FDR. More importantly, it shows that any misspecification in $\mathbf{W}^\star$ (e.g. in the model parameters) still leads to the correct FDR control. This is a crucial point in practice.

Explicit finite-sample bounds for the FDR of $\mathbf{SU}(\mathbf{W})$ – the step-up procedure without modification – are given in Proposition 8.4 in the unconditional model (see Section 8.5). It shows that FDR($\mathbf{SU}(\mathbf{W})$) should be close to $\pi_0 \alpha$ when m is large, so that the modifications of Theorem 4.1 are not needed anymore in that case. We will develop the resulting FDR consistency result more formally in Section 4.3 under some asymptotic conditions and for the optimal weighting.

4.2. Finite-sample power optimality

For a given weight function $\mathbf{W}$ of associated threshold collection Δ, let us denote

$$G_{\mathbf{W}}(u) := \mathbb{E}\left[\widehat{\mathbb{G}}_{\mathbf{W}}(u) \right] = m^{-1} \sum_{i=1}^{m} \mathbb{P}\left[p_i \leq \Delta_i(u) \right], \tag{14}$$

being the mean proportion of rejections at levels $(\Delta_i(u))_i$ and define similarly $G_{\mathbf{w}}(u)$ for a weight vector $\mathbf{w}$. Then the following theorem holds:

Theorem 4.2. *In the unconditional model, assume that* $\mathbf{F}$ *satisfies* $(A1)$ *and consider a weight function* $\mathbf{W}^{\star}$ *which maximizes the power at every proportion rejection, i.e. satisfying* (9). *Consider* $\lambda > 0$ *with* $\lambda < \pi_0(1 - \alpha)$. *Then we have for any finite* m,

$$Pow(\mathbf{SU}(\mathbf{W}^{\star})) \geq \max_{\mathbf{w}} \left\{ Pow(\mathbf{LSU}(\mathbf{w})) - \varepsilon(m, \mathcal{I}_{\lambda}^{+}(G_{\mathbf{w}})) \right\}$$
$$- \varepsilon(m, \mathcal{I}_{\lambda}^{-}(G_{\mathbf{W}^{\star}}))\mathbf{1}\{\lambda < \mathcal{I}(G_{\mathbf{W}^{\star}})\} - 2\lambda(1 - \alpha\pi_0), \quad (15)$$

where $\forall x \in \mathbb{R}$, $\varepsilon(m, x) := \pi_1 m^2 \exp\left\{-2m(x - m^{-1})_{+}^2\right\}$ *and where the maximum is taken over all the weight vectors* $\mathbf{w}$. *Moreover, we have* $\mathcal{I}_{\lambda}^{-}(G_{\mathbf{W}^{\star}}) > 0$ *when* $\lambda < \mathcal{I}(G_{\mathbf{W}^{\star}})$ *and* $\mathcal{I}_{\lambda}^{+}(G_{\mathbf{w}}) > 0$.

The proof is made in Section 8.2. Expression (15) can be seen as a non-asymptotic "oracle inequality", stating that the power of the optimal multi-weighted procedure is close to the power of the best weighted linear step-up procedure. This finite-sample optimality result makes sense because $\mathbf{SU}(\mathbf{W}^{\star})$, as all the weighted linear step-up procedures, controls the FDR non-asymptotically at level α (up to the slight modifications presented in Section 4.1).

In Theorem 4.2, condition $\lambda < \pi_0(1 - \alpha)$ (resp. $\lambda < \mathcal{I}(G_{\mathbf{W}^{\star}})$) ensures that $\mathcal{I}_{\lambda}^{+}(G_{\mathbf{w}})$ (resp. $\mathcal{I}_{\lambda}^{-}(G_{\mathbf{W}^{\star}})$) is well defined. Moreover, in (15), λ should be chosen such that the errors terms $\varepsilon(m, \mathcal{I}_{\lambda}^{+}(G_{\mathbf{w}}))$, $\varepsilon(m, \mathcal{I}_{\lambda}^{-}(G_{\mathbf{W}^{\star}}))$ and $2\lambda(1 - \alpha\pi_0)$ are as small as possible. From an asymptotic point of view, assuming that the quantities $\mathcal{I}_{\lambda}^{-}(G_{\mathbf{W}^{\star}})$ and $\mathcal{I}_{\lambda}^{+}(G_{\mathbf{w}})$ are bounded away from 0 when m tends to infinity (for any fixed λ), the error terms tend to zero by taking successively m tending to infinity and λ tending to zero. However, the best choice $\lambda = \lambda_m$ depends on the parameter $\mathbf{F}$ and seems quite difficult to derive under an explicit form (and so are the corresponding convergence rates in (15)).

The next section presents sufficient asymptotic conditions making $\mathcal{I}_{\lambda}^{-}(G_{\mathbf{W}^{\star}})$ and $\mathcal{I}_{\lambda}^{+}(G_{\mathbf{w}})$ bounded away from 0 when m tends to infinity, so that the error terms will asymptotically vanish in oracle inequality (15).

4.3. Consistency

We propose in this section an asymptotic framework in which the optimality of $\mathbf{SU}(\mathbf{W}^{\star})$ and its FDR control hold when m tends to infinity, without modification or error term.

First, we define the asymptotic setting. For all $m \geq 2$, we consider the m-unconditional model, where the m p-values are chosen as the m first p-values of an infinite sequence of independent p-values $(p_i)_{i \geq 1}$, each p-value p_i having the c.d.f. $\pi_0 t + \pi_1 F_i(t)$, for a given infinite sequence of c.d.f.'s $\mathbf{F} = (F_i)_{i \geq 1}$. In this context, the weight functions depend on m, and we underline this dependence in the notation, by denoting $\mathbf{W}^{(m)}$ instead of $\mathbf{W}$ (and $\mathbf{w}^{(m)}$ instead of $\mathbf{w}$).

Second, we define a *converging weight function sequence* $(\mathbf{W}^{(m)})_m$ as a sequence of weight functions such that the associated function sequence $(G_{\mathbf{W}^{(m)}})_m$ (defined in (14)) converges point-wise (on $[0,1]$). For short, we will often use the notation G^∞ for the limit function of $(G_{\mathbf{W}^{(m)}})_m$.

Theorem 4.3. *Consider the above asymptotic framework in which $\mathbf{F}$ is assumed to satisfy $(A1)$ and consider a class $\mathcal{W}$ of converging weight function sequences. Let $(\mathbf{W}^{\star,(m)})_m$ be a sequence of weight functions such that for all m, $\mathbf{W}^{\star,(m)}$ maximizes the power at every proportion rejection in the m-unconditional model (i.e. satisfies (9)). For the sequence $(\mathbf{W}^{\star,(m)})_m$, assumed to lie in $\mathcal{W}$, and for any weight vector sequence $(\mathbf{w}^{(m)})_m$ belonging to $\mathcal{W}$, we additionally assume that the associated limit function G^∞ is continuous and satisfies $\mathcal{I}_\lambda^-(G^\infty) > 0$ for $\lambda < \mathcal{I}(G^\infty)$. Then the multi-weighted procedure $\mathbf{SU}(\mathbf{W}^{\star,(m)})$ satisfies*

$$\lim_m Pow(\mathbf{SU}(\mathbf{W}^{\star,(m)})) \geq \max_{(\mathbf{w}^{(m)})_m} \left\{ \lim_m Pow(\mathbf{LSU}(\mathbf{w}^{(m)})) \right\}, \qquad (16)$$

the maximum above being taken over any sequence of weight vectors $(\mathbf{w}^{(m)})_m$ belonging to $\mathcal{W}$. Moreover, we have

$$\lim_m FDR(\mathbf{SU}(\mathbf{W}^{\star,(m)})) \quad \leq \quad \pi_0 \, \alpha \,, \qquad (17)$$

assuming either that $\mathcal{I}(G^\infty) > 0$ (with $G^\infty = \lim_m G_{\mathbf{W}^{\star,(m)}}$) or that we have $\lim_n \lim_m m^{-1} \sum_{i=1}^m \sup_{0<u<n^{-1}} \left\{ W_i^{\star,(m)}(u) \right\} = 1$.

Theorem 4.3 is proved in Section 8.5.

Under the conditions of Theorem 4.3, inequalities (16) and (17) imply that $\mathbf{SU}(\mathbf{W}^\star)$ is asymptotically more powerful than any weighted linear step-up procedure (in a certain class of converging weight vector sequences) while having the same asymptotic FDR control. Since the uniform weighting sequence $w_i^{(m)} = 1$ is always converging (with a continuous strictly concave limit function, from $(A1)$), it can always be added in the class $\mathcal{W}$. As a consequence, the procedure $\mathbf{SU}(\mathbf{W}^\star)$ always improves the original $\mathbf{LSU}$ asymptotically. However, this should be balanced with the fact that $\mathbf{SU}(\mathbf{W}^\star)$ uses the true parameters of the model, whereas $\mathbf{LSU}$ does not.

To satisfy the assumptions of Theorem 4.3, we have to choose a convenient class of converging weighting sequences $\mathcal{W}$, containing the optimal weighting sequence. We give below two examples of such choice when $\mathbf{F}$ is assumed to have a particular structure.

A first example is the case of *clustered p-values*: consider a parameter $\mathbf{F}$ satisfying $(A1)$ and such that F_i is equal to F_A (resp. F_B) for $i \in \mathcal{S}_A^{(m)}$ (resp. $i \in \mathcal{S}_B^{(m)}$), where $\{\mathcal{S}_A^{(m)}, \mathcal{S}_B^{(m)}\}$ forms a (deterministic) partition of $\{1, \ldots, m\}$ (this model may of course be generalized to the case $K \geq 2$ clusters). For simplicity, we assume that the proportion of p-values $\pi_A = |\mathcal{S}_A^{(m)}|/m$ in cluster $\mathcal{S}_A^{(m)}$ (resp. $\pi_B = |\mathcal{S}_B^{(m)}|/m$ in cluster $\mathcal{S}_B^{(m)}$) does not depend on m (this holds up to take a subsequence of m). In this context, we merely check that a weight

vector maximizing the power (at a given rejection proportion) has the same weight within a cluster. It is therefore natural to consider the following class of weighting:

$$\mathcal{W} = \Big\{ \big(\mathbf{W}^{(m)}\big)_m \ \Big| \ \forall m, \ W_i^{(m)} = W_A \ (\text{resp.}\ W_B) \ \text{for}\ i \in \mathcal{S}_A^{(m)} \ (\text{resp.}\ i \in \mathcal{S}_B^{(m)}),$$

$$\text{for}\ W_A(\cdot), W_B(\cdot) \geq 0 \ \text{satisfying}\ \forall u \in (0,1],\ \pi_A W_A(u) + \pi_B W_B(u) = 1,$$

$$\text{and}\ u \mapsto W_A(u)u,\ u \mapsto W_B(u)u \ \text{continuous nondecreasing on}\ [0,1]\Big\}.$$

Since for any weight function sequence $\big(\mathbf{W}^{(m)}\big)_m$ of $\mathcal{W}$ the function $G_{\mathbf{W}^{(m)}}(u) = \pi_0 \alpha u + \pi_1 \pi_A F_A(\alpha u W_A(u)) + \pi_1 \pi_B F_B(\alpha u W_B(u))$ does not depend of m, $\mathcal{W}$ is a class of converging weight function sequences. Moreover, $G^\infty(u) = G_{\mathbf{W}^{(m)}}(u)$ is continuous and satisfies $\mathcal{I}_\lambda^-(G^\infty) > 0$ for $\lambda < \mathcal{I}(G^\infty)$, either for $\mathbf{W}^{(m)}(u) = \mathbf{w}^{(m)}$ a weight vector, or for $\mathbf{W}^{(m)}(u) = \mathbf{W}^{\star,(m)}(u)$ the optimal weight function (from one of the last statements of Theorem 4.2). Finally, we can apply Theorem 4.3 to obtain the oracle inequality (16). Moreover, the last assumption required for the FDR control (17) holds assuming that the limits $W_A^\star(0^+)$ and $W_B^\star(0^+)$ exist (as is the case under assumptions $(A1)$–$(A4)$, see Proposition 3.2).

A second example is the *continuous one-sided Gaussian setting*, where for all m, and $1 \leq i \leq m$, $F_i(x) = \overline{\Phi}\big(\overline{\Phi}^{-1}(x) - \mu(i/m)\big)$, for a mean function $\mu : [0,1] \to \mathbb{R}^+$ assumed continuous with $\mu(t) > 0$ for $t > 0$. In this context, we denote $F_{i/m}$ and $W_{i/m}^{(m)}$ instead of F_i and $W_i^{(m)}$ for more convenience. Also note that the function $t \mapsto F_t$ can be extended to all t in $[0,1]$. In that setting, it is relevant to consider the following class of weighting:

$$\mathcal{W} = \Bigg\{ \big(\mathbf{W}^{(m)}\big)_m \ \text{weight function sequence such that}$$

$$\forall u \in [0,1], \ \frac{1}{m}\sum_{i=1}^m F_{i/m}(\alpha u W_{i/m}^{(m)}(u)) \xrightarrow[m\to\infty]{} \int_0^1 F_t(\alpha u W_t(u))dt,$$

$$\text{for}\ (W_t(\cdot))_{t\in[0,1]} \geq 0 \ \text{satisfying}\ \forall u,\ t \in [0,1] \mapsto W_t(u) \ \text{continuous} \Bigg\}.$$

Any weight function sequence $(\mathbf{W}^{(m)})_m$ of $\mathcal{W}$ is converging, because $G_{\mathbf{W}^{(m)}}(u) = \pi_0 \alpha u + \pi_1 m^{-1}\sum_{i=1}^m F_{i/m}(\alpha u W_{i/m}^{(m)}(u)) \to G^\infty(u) = \pi_0 \alpha u + \pi_1 \int_0^1 F_t(\alpha u W_t(u))dt$. Moreover, expression (11) provides the form of the optimal weight function: $\alpha u W_{i/m}^{\star,(m)}(u) = \overline{\Phi}\big(\mu(i/m)/2 + c^{(m)}(u)/\mu(i/m)\big)$ where $c^{(m)}(u)$ is taken such that $\sum_{i=1}^m W_{i/m}^{\star,(m)}(u) = m$. In Section 8.7, we prove that $(\mathbf{W}^{\star,(m)})_m$ belongs to $\mathcal{W}$, with a "limit weighting" $(W_t^\star(\cdot))_{t\in[0,1]} \geq 0$ given by

$$W_t^\star(u) = (\alpha u)^{-1}\overline{\Phi}\left(\frac{\mu(t)}{2} + \frac{c^\infty(u)}{\mu(t)}\right),$$

where $c^\infty(u) \in \mathbb{R}$ satisfies $\int_0^1 \overline{\Phi}\left(\frac{\mu(t)}{2} + \frac{c^\infty(u)}{\mu(t)}\right)dt = \alpha u$. Similarly, any weight vector sequence of the form $\mathbf{w}^{(m)} = \mathbf{W}^{\star,(m)}(u_0)$ (with u_0 fixed in $(0,1]$) belongs to $\mathcal{W}$ with a limit function $G^\infty(u) = \pi_0 \alpha u + \pi_1 \int_0^1 F_t(\alpha u W_t^\star(u_0))dt$ continuous and strictly concave (implying $\mathcal{I}_\lambda^-(G^\infty) > 0$ for $\lambda < \mathcal{I}(G^\infty)$). Denoting $G^{\star,\infty}$ the limit function of $G_{\mathbf{W}^{\star,(m)}}$, we merely check that $G^{\star,\infty} \geq G^\infty$. Since this holds for any choice of u_0, we derive that $\mathcal{I}_\lambda^-(G^{\star,\infty}) > 0$ for $\lambda < \mathcal{I}(G^{\star,\infty})$ (using inequalities similar to (25)). As a consequence, we may apply Theorem 4.3 to obtain the oracle inequality (16). In particular, the power of the multi-weighted procedure $\mathbf{SU}(\mathbf{W}^{\star,(m)})$ is always asymptotically larger than the power of the weighted linear step-up $\mathbf{LSU}(\mathbf{w}^{(m)})$ for any weight vector of the form $\mathbf{w}^{(m)} = \mathbf{W}^{\star,(m)}(u_0)$, with $u_0 \in (0,1]$. Roughly, the latter signifies that $\mathbf{SU}(\mathbf{W}^{\star,(m)})$ automatically chooses the best weighting among $\{\mathbf{W}^{\star,(m)}(u_0)\}_{u_0}$.

5. Simulation study

An important point is now to evaluate the improvement of the new multi-weighted procedure, both when we plug the true parameters or misspecified parameters in the optimal weighting. For this, we propose to perform simulations in the – restricted but convenient – one-sided Gaussian testing framework under the conditional model.

5.1. Simulations framework

We consider the problem of testing for each $i \in \{1, \ldots, m\}$, the null "$\mu_i = 0$" against the alternative "$\mu_i > 0$" from the observation of m independent variables $(X_i)_i$ with $X_i \sim \mathcal{N}(\mu_i, 1)$. The parameters $(\mathbf{H}, \mathbf{F})$ of the (conditional) model are fully determined from the vector $\mu = (\mu_i)_i$, namely by $H_i = \mathbf{1}\{\mu_i > 0\}$ and (2), respectively. They represent informations of a different nature: $\mathbf{H}$ provides the location of the positive means while $\mathbf{F}$ supplies their values.

For all our experiments, the number of tests is $m = 1000$. The vector μ is taken such that the $m_0 = 700$ first components of μ are equal to zero (the proportion of zeros in the mean vector is thus $\pi_0 = 0.7$). The $m_1 = 300$ remaining non-zero means are taken in two different ways:

- Case 1: the non-zero means increase linearly from $\frac{3}{m_1}\overline{\mu}$ to $3\overline{\mu}$.
- Case 2: the non-zero means are gathered in three groups of different values $\overline{\mu}$, $2\overline{\mu}$ and $3\overline{\mu}$, of respective sizes 120, 120 and 60.

In both cases $\overline{\mu}$ is an "effect size" parameter taking values in the range $0.5 + 0.25k$, $k \in \{0, \ldots, 10\}$.

The following procedures are considered:

- [LSU] the linear step-up procedure $\mathbf{LSU}$,
- [LSU*] the step-up procedure with threshold collection $\alpha u / \pi_0$,
- [SU-W-oracle] the multi-weighted step-up procedure $\mathbf{SU}(\widetilde{\mathbf{W}^\star})$ of Theorem 4.1, using the optimal weight matrix $\mathbf{W}^\star$ (given by (11)),

  – [SD-W-oracle] the multi-weighted step-down procedure $\mathbf{SD}(\widetilde{\mathbf{W}^\star})$ of Theorem 4.1, using the optimal weight matrix $\mathbf{W}^\star$,
  – [Unif-oracle] the weighted linear step-up procedure $\mathbf{LSU}(\mathbf{w}^\star)$ using a weight vector uniform on $\mathcal{H}_1$: $w_i^\star = 0$ for $\mu_i = 0$ and $w_i^\star = m/m_1$ for $\mu_i > 0$.

The procedures [SU-W-oracle], [SD-W-oracle] and [Unif-oracle] correspond to the case where the weighting uses the true mean vector μ, hence the name "oracle". In situations where we replace μ by a "guess" $\widetilde{\mu}$ in the weights, the procedures are called "guessed" and are denoted by [SU-W-guess], [SD-W-guess], [Unif-guess] respectively. The procedure [Unif-oracle/guess] renders a uniform weighting over the (guessed) false nulls and is close in spirit to the approach of Genovese et al. (2006). It takes only into account the subset where the hypotheses are false ("location information"), but not the values of the non-zero means.

The procedure [LSU*] is performed to compare with quite recent developments on π_0-adaptive procedures (see e.g. Benjamini et al. (2006)). Since it uses a perfect estimation of π_0, it represents the best theoretical π_0-adaptive modification of the LSU that we can build. For clarity reasons, we avoid the problem of choosing a particular estimator of π_0 and we only consider [LSU*].

All the latter procedures have provable FDR control (see Section 4.1), so that it is relevant to compare them in terms of power. In all experiments the targeted FDR level is either $\alpha = 0.01$ or $\alpha = 0.05$. The different performed procedures are compared in terms of relative power (RelPow) with respect to the LSU procedure, defined as the expected surplus proportion of correct rejections among the false nulls: for a multiple testing procedure R,

$$\mathrm{RelPow}(R) := (m_1)^{-1}\big(\mathbb{E}(|R \cap \mathcal{H}_1|) - \mathbb{E}(|\mathbf{LSU} \cap \mathcal{H}_1|)\big). \qquad (18)$$

Roughly speaking, this relative power represents the surplus "probability" of a false null to be rejected with respect to the LSU. This power is estimated using Monte-Carlo simulations. Additionally, we also evaluate the "power range" defined by the power of the weighted linear procedures $\mathbf{LSU}(\mathbf{W}^\star(u_0))$ for any $u_0 \in \{1/m, 2/m, \ldots, 1\}$. It is represented by a gray area over the pictures. Finally, the optimal multi-weighted step-up procedure $\mathbf{SU}(\mathbf{W}^\star)$ without correction (which controls the FDR when $m \to \infty$) is also considered, but it is not reported on our figures, because its (relative) power is almost indistinguishable from the top of the power range.

5.2. Procedures using the true parameters

We report on Figure 2 the relative power (18) of [LSU], [LSU*], [SU-W-oracle], [SD-W-oracle] and [Unif-oracle] in function of the parameter $\overline{\mu}$ (1000 simulations). The gray area represents the power range as defined in the previous section.

The conclusion of this experiment is that, in the most favorable case where the multi-weighting is used with the true parameters of the model, the improvement
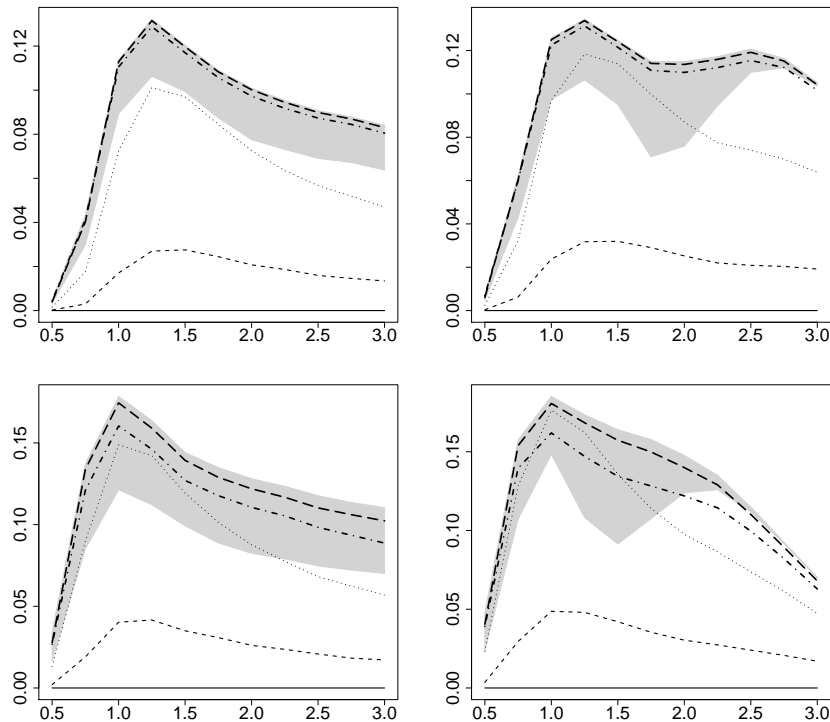
FIG 2. *Procedures using the true parameters. Relative power of [LSU] (solid), [LSU⋆] (short-dashed), [SU-W-oracle] (dotted-dashed), [SD-W-oracle] (long-dashed) and [Unif-oracle] (dotted) in function of $\overline{\mu}$ (see text). Left: case 1 of means, right: case 2 of means. Top $\alpha = 0.01$; bottom $\alpha = 0.05$.*

of the multi-weighted procedures over [LSU] is satisfactory. Also, [SD-W-oracle] performs here better than [SU-W-oracle] (especially for $\alpha = 0.05$), so that the loss in the correction within [SU-W-oracle] seems significantly larger than the loss in the correction within [SD-W-oracle].

Furthermore, [SD-W-oracle] is more powerful than [LSU⋆] (actually, this is still true using a smaller π_0, e.g. $\pi_0 = 0.5$), and [SD-W-oracle] is always better than [Unif-oracle], and allows sometimes for much more discoveries. This seems coherent because [SD-W-oracle] takes into account more (correct) prior informations than [Unif-oracle]: namely, [SD-W-oracle] uses both the values and the location of the non-zero means (we are in the conditional model), while [Unif-oracle] only uses the location information.

Finally, the procedure [SD-W-oracle] is close to the top of the power range (gray area), that is, has a power close to the power of the best procedure among $\mathbf{LSU}(\mathbf{W}^\star(u_0))$, $u_0 \in \{1/m, 2/m, \ldots, 1\}$. This corroborates the optimality results of Section 4.2 and Section 4.3 in this (conditional) setting.

5.3. Procedures using misspecified parameters

We consider here the same experiment as before, except that we take into account the "randomness" due to a prior guess $\widetilde{\mu}_i$ of each μ_i. For this, we add a misspecification parameter σ and we suppose that the guessed means are of the form: $\forall i \in \{1, \ldots, m\}$,

$$\widetilde{\mu}_i = \mu_i + \varepsilon_i,$$

where ε_i are i.i.d with distribution $\mathcal{N}(0, \sigma^2)$ (taken independent of the p_i's). The misspecification parameter σ is taken in the range $\{j/4, j = 0, \ldots, 12\}$. Remark here that the way of guessing the mean is quite raw, because it does not take into account the specific form of the parameters (of course, this guessing can be improved here by taking local means). However, we keep this raw modeling here because we do not want to make any assumption on the parameters.

Figure 3 reports the relative power (18) of [LSU], [LSU*], [SU-W-guess], [SD-W-guess] and [Unif-guess] with respect to σ. We performed 100 simulations to compute the relative power and the latter is moreover averaged over 10 generated values of the $\widetilde{\mu}_i$'s (for each values of σ).

In this experiment, we see that both multi-weighted procedures are better than other procedures when the guesses are good i.e. over the range $\sigma \in [0, 1.2]$, but may be worst than the simple [LSU] procedure when σ is large. Furthermore, note that the procedure [Unif-guess] quickly collapses when σ grows and therefore only proposes a slight improvement of [LSU] (or [LSU*]) when the guesses are good. However, it is "less risky" than the multi-weighted procedures for large σ. Again, this conclusion is natural because the multi-weighted procedures take here into-account more prior information than [Unif-guess].

Finally, although admittedly of a limited scope, these experiments show that in principle, taking into account a correct guess of the parameters in the multi-
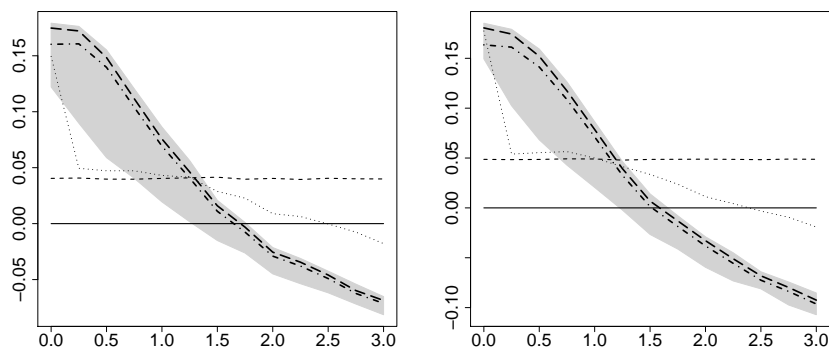

FIG 3. *Procedures using misspecified parameters. Relative power of [LSU] (solid), [LSU*] (short-dashed), [SU-W-guess] (dotted-dashed), [SD-W-guess] (long-dashed) and [Unif-guess] (dotted) in function of the misspecification parameter σ (see text). Left: case 1 of means, right: case 2 of means (see text). $\overline{\mu} = 1$; $\alpha = 0.05$.*

weighted procedures should improve the power substantially. The loss/gain magnitude of these procedures depends on the quantity of prior information used (location of the positive means, value of the positive means, or both).

6. Application to mRNA and DNA microarray experiments

In a typical microarray experiment, we want to find differentially expressed mRNA genes between two groups of individuals. For the i-th gene, the level expression is of the form $X_{i,1}, \ldots, X_{i,k_i}$ for group 1 and $Y_{i,1}, \ldots, Y_{i,\ell_i}$ for group 2, where k_i (resp. ℓ_i) is the number of individuals in group 1 (resp. group 2).

In some microarray experiments, the sample sizes (k_i, ℓ_i) available to assess the differential mRNA expression of gene i may strongly depend on i, e.g. when the number of missing data differs per gene. In this application, we consider a covariate, the DNA copy number status of the same gene, which determines the groups and the sample sizes. DNA copy number status is obtained from an independent array CGH experiment, after a few pre-processing steps (see e.g. Picard et al. (2007)). We focus on the covariate $A_{i,j}$ which is equal to 1 when gene i is *gained* for individual j (i.e. when sample j has an abnormally high DNA copy number of gene i), and 0 otherwise. The biological goal behind this is to find the genes for which the mRNA expression is induced by the DNA copy number. This is particularly useful to study cancer pathologies (see e.g. Hyman et al. (2002)). Sample size dependent weights are in particular attractive here, because many genes show a large unbalance in the amount of gains (defining group 1) and non-gains (defining group 2).

Using the above framework, we analyze microarray lymphoma cancer data of Muris et al. (2007). In these data $m = 11\,169$ genes and $n = 42$ individuals are studied. The p-value of each gene was computed using a Mann-Whitney test. We aim to consider as prior the sample size information only, without any guess on which hypotheses are false or true. The asymptotic normality of the Mann-Whitney test statistic is used to define asymptotically optimal multi-weights $\mathbf{W}^\star$ which depend only on (k_i, ℓ_i) and an estimate for the global effect θ, which is a gene-independent parameter for the effect of copy number gain on mRNA gene expression. The expression of the multi-weights and the estimate for θ, $\widehat{\theta}_m$, are detailed in Roquain and van de Wiel (2008). The estimator $\widehat{\theta}_m$ converges in probability when m grows to infinity, so that we believe that the fluctuations of $\widehat{\theta}_m$ in the weights will have a marginal effect on the effective FDR of the so multi-weighted procedure when m becomes large (however we did not investigate formally the corresponding asymptotic study for now).

We applied the step-up multi-weighted procedure $\mathbf{SU}(\mathbf{W}^\star)$, using the estimator $\widehat{\theta}_m \simeq 1.01$ of the global effect size θ. Since m is large we focus on the unmodified version of our procedure, which guarantees asymptotic FDR control. For different values of α, the number of discoveries of this procedure and of the LSU are given in Table 1.

We observe that our new step-up procedure discovers more differentially expressed genes when $\alpha \in \{0.005, 0.01\}$. For $\alpha \in \{0.05, 0.1\}$, the performance of

α	LSU	$\mathbf{SU(W)}$
0.005	43	98
0.01	121	156
0.05	478	476
0.1	836	859

the two step-up procedures is similar. So, the improvement of our procedure is here mostly noticeable when the proportion of rejections is small. This is in accordance with our intuition: the prior information (here the sample sizes) is particularly useful when the proportion of rejections is expected to be small. Finally, let us remark that these positive results on the sample size problem have been corroborated in a specific simulation study as well (not reported here).

7. Conclusion and discussions

When the parameters of the p-value model are known, we proposed to solve the problem of the LSU optimal weighting by finding a new procedure which provably outperforms all the weighted LSU procedures (up to small error terms) and which can be easily computed from these parameters. Our simulations illustrated the strength of the improvement of our new approach in situations where it uses the true or misspecified parameters.

In our results, the assumptions concerning the marginal distributions of the p-values were quite mild: the FDR control only required that each p-value is uniform under the null while the optimality results only required the strict concavity of the c.d.f.'s of the p-values. Moreover, the existence of the optimal weight function only asked to maximize simultaneously the power at any proportion rejection, and we gave strong sufficient assumptions for its existence and unicity.

Several extensions to this work are possible: first, we have supposed the independence between the p-values all along the paper, which is a standard but somewhat unrealistic assumption for the applications. In Roquain and van de Wiel (2008), we proposed some extensions of the present FDR control results to the case of positively regressively dependence or unspecified dependence. However, the so-derived procedures seemed too conservative for practical use. Therefore, there is a room left for future investigations, which join the very active (but challenging) research field studying the impact of p-value dependence on FDR control (see e.g. Kim and van de Wiel (2008); Romano et al. (2008)).

Second, our FDR controls are done at level smaller than $\pi_0\alpha$ (asymptotically, in the unconditional model). Therefore, when π_0 is small, our procedures are inevitably conservative, because their actual FDR is much lower than the fixed target. This is a classical problem for the LSU procedure and several works have been proposed to address this issue, by integrating a π_0-estimate in the threshold, building so-called *adaptive* LSU procedures (see e.g. Benjamini et al. (2006); Blanchard and Roquain (2009)). A possible interesting extension of our

work could therefore be to derive adaptive multi-weighted procedures, which would increase the power when the data contain a lot of signal.

A third – and maybe more important – direction for future works is the investigation of data-driven weighting. A first idea could be to replace the function $\mathrm{Pow}_u(\cdot)$, the power at rejection proportion u, by an empirical substitute and to perform the simultaneous maximization with this substitute. This would yield an empirical optimal weight function $\widehat{\mathbf{W}^\star}$ that can in turn be integrated in a multi-weighted procedure. While this certainly requires to use a model with some replications, the theoretical FDR control and power optimality of such data-driven procedure are not straightforward from the present work, because all our proofs here use the fact that the weight functions are deterministic.

8. Proofs

8.1. Useful notation and lemmas

Let us first introduce the following notation that will be useful throughout our proofs: if R is the step-up procedure associated to a given weight function $\mathbf{W}$ of associated threshold collection Δ, and $\widehat{u} := |R|/m$ its rejection proportion, that is $\widehat{u} = \mathcal{I}(\widehat{\mathbb{G}}_{\mathbf{W}})$, we denote by:

1. R_{-i} the step-up procedure on the set of hypotheses corresponding to $\{1, \ldots, m\} \setminus \{i\}$, that is excluding the i-th null, and associated to the threshold collection $\forall j \neq i$, $\forall u$, $\Delta_j((1 - m^{-1})u)$; and we denote by $\widehat{u}_{-i} := |R_{-i}|/(m-1)$ its rejection proportion, so that $\widehat{u}_i = \mathcal{I}(\widehat{\mathbb{G}}_{-i})$ with $\widehat{\mathbb{G}}_{-i}(u) := (m-1)^{-1} \sum_{j \neq i} \mathbf{1}\{p_j \leq \Delta_j((1 - m^{-1})u)\}$;
2. R'_{-i} the step-up procedure on the set of hypotheses excluding the i-th null associated to the threshold collection $\forall j \neq i$, $\forall u$, $\Delta_j((1 - m^{-1})u + m^{-1})$; and we denote $\widehat{u}'_{-i} := |R'_{-i}|/(m-1)$ its rejection proportion, hence $\widehat{u}'_{-i} = \mathcal{I}(\widehat{\mathbb{G}}'_{-i})$ with $\widehat{\mathbb{G}}'_{-i}(u) := (m-1)^{-1} \sum_{j \neq i} \mathbf{1}\{p_j \leq \Delta_j((1 - m^{-1})u + m^{-1})\}$.

Similarly, when R is step-down, we define R_{-i} and R'_{-i} as step-down procedures and we denote $\widetilde{u} := \mathcal{J}(\widehat{\mathbb{G}}_{\mathbf{W}})$, $\widetilde{u}_{-i} := \mathcal{J}(\widehat{\mathbb{G}}_{-i})$, $\widetilde{u}'_{-i} := \mathcal{J}(\widehat{\mathbb{G}}'_{-i})$ instead of $\widehat{u}$, $\widehat{u}_{-i}$, $\widehat{u}'_{-i}$, respectively.

The two following lemmas make a link between the rejection proportions of R, R_{-i} and R'_{-i}, for different values of p_i. They are proved in Appendix B and are related to Lemma 10.20 of Roquain (2007).

Lemma 8.1. *Let R be the step-up procedure associated to a given weight function of threshold collection Δ and consider $\widehat{u}$, $\widehat{u}_{-i}$ and $\widehat{u}'_{-i}$ as above. Then we have point-wise:*

1. $p_i \leq \Delta_i(\widehat{u}) \iff p_i \leq \Delta_i((1 - m^{-1})\widehat{u}'_{-i} + m^{-1}) \iff \widehat{u} = (1 - m^{-1})\widehat{u}'_{-i} + m^{-1}$;
2. $p_i > \Delta_i(\widehat{u}) \iff \widehat{u} = (1 - m^{-1})\widehat{u}_{-i}$.

Lemma 8.2. *Let R be a step-down procedure associated to a given weight function of threshold collection Δ and consider $\widetilde{u}$ and $\widetilde{u}_{-i}$ as above. Then we have point-wise, for any $k \in \{1, \ldots, m\}$,*

1. *$\widetilde{u} \geq k/m$ and $p_i > \Delta_i((k-1)/m) \implies \widetilde{u}_{-i} \geq (k-1)/(m-1)$;*
2. *$\widetilde{u}_{-i} \geq (k-1)/(m-1)$ and $p_i \leq \Delta_i((k-1)/m) \implies \widetilde{u} \geq k/m$;*
3. *$p_i > \Delta_i((1 - m^{-1})\widetilde{u}_{-i} + m^{-1}) \implies \widetilde{u} = (1 - m^{-1})\widetilde{u}_{-i}$.*

8.2. *Proof of Theorem 4.1 – step-up part*

The inequalities are established in the conditional model (the result in the unconditional model directly follows).

We use in all our FDR bounds that a procedure R satisfying the "self-consistency condition" $R = \{i \mid p_i \leq \Delta_i(|R|/m)\}$ has a FDR equal to

$$\mathrm{FDR}(R) = \mathbb{E}\left[\frac{|R \cap \mathcal{H}_0|}{|R| \vee 1}\right] = \sum_{i=1}^{m}(1 - H_i)\mathbb{E}\left[\frac{\mathbf{1}\{p_i \leq \Delta_i(|R|/m)\}}{|R|}\right]. \qquad (19)$$

Now, consider the multi-weighted step-up procedure $R = \mathbf{SU}(\widetilde{\mathbf{W}})$ of Theorem 4.1, and denote by Δ the threshold collection associated to $\widetilde{\mathbf{W}}$: $\Delta_i(k/m) = \alpha\widetilde{W}_i(k/m)k/m = \alpha W_i(k/m)k/m(1 + \alpha W_i(1))^{-1} \leq 1$. Since any step-up procedure satisfies the self consistency condition, we may use (19). Furthermore, using the notation of Section 8.1 and applying Lemma 8.1 (first statement), the assertion $p_i \leq \Delta_i(|R|/m) = \Delta_i(\widehat{u})$ is equivalent to $\widehat{u} = (1 - m^{-1})\widehat{u}'_{-i} + m^{-1}$. Thus, we may rewrite the FDR as follows:

$$\begin{aligned}
\mathrm{FDR}(R) &= \sum_{i=1}^{m}(1 - H_i)\mathbb{E}\left[\frac{\mathbf{1}\{p_i \leq \Delta_i(\widehat{u})\}}{\widehat{u}\, m}\right] \\
&= \sum_{i=1}^{m}(1 - H_i)\sum_{k=1}^{m}k^{-1}\mathbb{P}\left[p_i \leq \Delta_i(k/m), \widehat{u}\, m = k\right] \\
&= \sum_{i=1}^{m}(1 - H_i)\sum_{k=1}^{m}k^{-1}\mathbb{P}\left[p_i \leq \Delta_i(k/m), (m-1)\widehat{u}'_{-i} + 1 = k\right].
\end{aligned}$$

Then, since $\widehat{u}'_{-i}$ only depends on the p-values of $(p_j, j \neq i)$, it is independent of p_i and we obtain

$$\begin{aligned}
\mathrm{FDR}(R) &= \frac{\alpha}{m}\sum_{i=1}^{m}(1 - H_i)\sum_{k=1}^{m}\widetilde{W}_i(k/m)\mathbb{P}\left[(m-1)\widehat{u}'_{-i} + 1 = k\right] \\
&= \frac{\alpha}{m}\sum_{i=1}^{m}(1 - H_i)\sum_{k=1}^{m}W_i(k/m)(1 - \Delta_i(1))\mathbb{P}\left[(m-1)\widehat{u}'_{-i} + 1 = k\right] \\
&= \frac{\alpha}{m}\sum_{i=1}^{m}(1 - H_i)\sum_{k=1}^{m}W_i(k/m)\mathbb{P}\left[p_i > \Delta_i(1), (m-1)\widehat{u}'_{-i} + 1 = k\right],
\end{aligned}$$

$$(20)$$

where we used that p_i has a uniform distribution on $[0, 1]$ (from (1)). Next, consider the threshold collection $\forall j \in \{1, \ldots, m\}$, $\forall u$, $\Delta'_j(u) = \Delta_j((1 - m^{-1})u + m^{-1})$ and the associated step-up procedure that we denote by R'. Let us also denote its rejection proportion by $\widehat{u}' = |R'|/m$. From the definition of Section 8.1, the restriction of R' to the hypothesis set corresponding to $\{1, \ldots, m\}\backslash\{i\}$ is exactly the procedure R'_{-i}. Therefore, from Lemma 8.1 (second statement applied to R'), the condition $p_i > \Delta_i(1) = \Delta'_i(1) \geq \Delta'_i(\widehat{u}')$ implies $m\widehat{u}' = (m - 1)\widehat{u}'_{-i}$. Therefore,

$$\text{FDR}(R) \leq \frac{\alpha}{m} \sum_{i=1}^{m} (1 - H_i) \sum_{k=1}^{m} W_i(k/m) \mathbb{P}\left[p_i > \Delta_i(1), m\widehat{u}' + 1 = k\right]$$

$$\leq \frac{\alpha}{m} \sum_{k=1}^{m} \left[\sum_{i=1}^{m} (1 - H_i) W_i(k/m)\right] \mathbb{P}\left[m\widehat{u}' + 1 = k\right]$$

$$\leq \alpha \max_{1 \leq k \leq m} \left\{m^{-1} \sum_{i=1}^{m} (1 - H_i) W_i(k/m)\right\}.$$

8.3. Proof of Theorem 4.1 – step-down part

Again, it is sufficient to look at the conditional model. First, let us prove that for any step-down procedure R with threshold collection Δ and rejection proportion $\widetilde{u}$, we have for any i,

$$\text{FDR}(R) \leq \sum_{i=1}^{m} (1 - H_i) \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\left[(m - 1)\widetilde{u}_{-i} = k - 1, p_i \leq \Delta_i(k/m)\right], \quad (21)$$

where $\widetilde{u}_{-i}$ is the rejection proportion of the step-down procedure associated to Δ and restricted to the hypotheses different from the i-th hypothesis as defined in Section 8.1. This result has been implicitly proved in Gavrilov et al. (2009) (Section 3), using a specific non-weighted step-down procedure. Here, we state (21) in a more general framework. Applying the two first points of Lemma 8.2, we obtain the following relations:

$$\sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\left[m\widetilde{u} = k, p_i \leq \Delta_i(k/m)\right]$$

$$= \sum_{k=1}^{m} \frac{1}{k} \left[\mathbb{P}\left[m\widetilde{u} = k, p_i \leq \Delta_i((k-1)/m)\right]\right.$$

$$\left. + \mathbb{P}\left[m\widetilde{u} = k, \Delta_i((k-1)/m) < p_i \leq \Delta_i(k/m)\right]\right]$$

$$= \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\left[m\widetilde{u} \geq k, \Delta_i((k-1)/m) < p_i \leq \Delta_i(k/m)\right]$$

$$- \sum_{k=1}^{m} \left[\frac{\mathbf{1}\{k > 1\}}{k - 1} - \frac{1}{k}\right] \mathbb{P}\left[m\widetilde{u} \geq k, p_i \leq \Delta_i((k-1)/m)\right]$$

$$\leq \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\left[(m-1)\widetilde{u}_{-i} \geq k-1, \Delta_i((k-1)/m) < p_i \leq \Delta_i(k/m)\right]$$

$$- \sum_{k=1}^{m} \left[\frac{\mathbf{1}\{k > 1\}}{k-1} - \frac{1}{k}\right] \mathbb{P}\left[(m-1)\widetilde{u}_{-i} \geq k-1, p_i \leq \Delta_i((k-1)/m)\right]$$

$$= \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\left[(m-1)\widetilde{u}_{-i} = k-1, p_i \leq \Delta_i(k/m)\right].$$

As a consequence, the latter combined with (19) states (21).

Now, consider the step-down procedure R of Theorem 4.1, that is, associated to the threshold collection $\Delta_i(k/m) = \alpha \widetilde{W}_i(k/m)k/m = \alpha W_i(k/m)k/m(1 + \alpha W_i(k/m)k/m)^{-1} \leq 1$. We use the independence between the p-values and (21) to show

$$\text{FDR}(R) \leq \sum_{i=1}^{m}(1-H_i) \sum_{k=1}^{m} \frac{1}{k} \mathbb{P}\left[(m-1)\widetilde{u}_{-i} = k-1, p_i \leq \Delta_i(k/m)\right]$$

$$= \frac{\alpha}{m} \sum_{i=1}^{m}(1-H_i) \sum_{k=1}^{m} W_i(k/m)(1-\Delta_i(k/m))\mathbb{P}\left[(m-1)\widetilde{u}_{-i} = k-1\right]$$

$$= \frac{\alpha}{m} \sum_{i=1}^{m}(1-H_i) \sum_{k=1}^{m} W_i(k/m)\mathbb{P}\left[(m-1)\widetilde{u}_{-i} = k-1, p_i > \Delta_i(k/m)\right].$$

The third point of Lemma 8.2 thus implies

$$\text{FDR}(R) \leq \frac{\alpha}{m} \sum_{i=1}^{m}(1-H_i) \sum_{k=1}^{m} W_i(k/m)\mathbb{P}\left[m\widetilde{u} = k-1, p_i > \Delta_i(k/m)\right]$$

$$\leq \frac{\alpha}{m} \sum_{k=1}^{m} \left[\sum_{i=1}^{m}(1-H_i)W_i(k/m)\right] \mathbb{P}\left[m\widetilde{u} = k-1\right]$$

$$\leq \alpha \max_{1 \leq k \leq m} \left\{m^{-1} \sum_{i=1}^{m}(1-H_i)W_i(k/m)\right\}.$$

8.4. *Proof of Theorem 4.2*

Let us assume that the following proposition holds (the proof is given at the end of this section):

Proposition 8.3. *In the unconditional model, consider a weight function* $\mathbf{W}$ *with its associated threshold collection* Δ *and put* $\bar{u} = \mathcal{I}(G_{\mathbf{W}})$. *Then the following holds:*

    (i) assuming that for all $u' > u > \bar{u}$, $u' - G_{\mathbf{W}}(u') > u - G_{\mathbf{W}}(u)$, *we have for all* $\lambda > 0$, $\lambda < 1 - \bar{u}$,

$$Pow(\mathbf{SU}(\mathbf{W})) - (1 - \alpha\pi_0)\bar{u}$$
$$\leq \pi_1 m^2 \exp\left\{ -2m\left(\mathcal{I}_\lambda^+(G_{\mathbf{W}}) - m^{-1}\right)_+^2 \right\} - \mathcal{I}_\lambda^+(G_{\mathbf{W}}) + \lambda(1 - \alpha\pi_0); \quad (22)$$

(ii) assuming $\Delta \leq 1$, we have for all $\lambda > 0$, $\lambda < \bar{u}$,

$$Pow(\mathbf{SU}(\mathbf{W})) - (1 - \alpha\pi_0)\bar{u}$$
$$\geq -\pi_1 m \exp\left\{ -2m(\mathcal{I}_\lambda^-(G_{\mathbf{W}}))_+^2 \right\} + \mathcal{I}_\lambda^-(G_{\mathbf{W}}) - \lambda(1 - \alpha\pi_0). \quad (23)$$

We now prove Theorem 4.2 by applying Proposition 8.3. First, remark that, in the unconditional model, we have for any weight vector $\mathbf{w}$,

$$G_{\mathbf{w}}(u) = \alpha\pi_0 u + \mathrm{Pow}_u(\mathbf{w}),$$

so that maximizing in $\mathbf{w}$ the power at rejection level u is equivalent to maximize $G_{\mathbf{w}}(u)$ in $\mathbf{w}$. As a consequence, taking the optimal weight function $\mathbf{W}^\star$, we deduce from (9) that for any weight vector $\mathbf{w}$ and for any u we have $G_{\mathbf{w}}(u) \leq G_{\mathbf{W}^\star}(u)$. Denoting $u_{\mathbf{w}} := \mathcal{I}(G_{\mathbf{w}})$ and $u^\star := \mathcal{I}(G_{\mathbf{W}^\star})$, this in turn implies that

$$u_{\mathbf{w}} \leq u^\star \qquad (24)$$

Second, remark that $\mathbf{W}^\star$ has a threshold collection $\Delta^\star$ satisfying $\Delta^\star \leq 1$. The latter holds because the F_i's are increasing (as non-decreasing strictly concave functions), and because $\Delta_i^\star(u) \leq \alpha W_i^\star(1)$ with $\mathbf{W}^\star(1)$ maximizing the power at rejection proportion 1. Third, we check the assumption of (i) Proposition 8.3 for $\mathbf{W}(\cdot)$ constantly equal to a weight vector $\mathbf{w}$, which directly follows from the strict concavity of $G_{\mathbf{w}}$ (itself coming from the strict concavity of the F_i's). Forth, let us prove that $\mathcal{I}_\lambda^+(G_{\mathbf{w}}) > 0$ and $\mathcal{I}_\lambda^-(G_{\mathbf{W}^\star}) > 0$. The first statement comes from the definition of $\mathcal{I}(G_{\mathbf{w}})$. To prove the second statement, consider $u^\star = \mathcal{I}(G_{\mathbf{W}^\star})$ and the weight vector $\mathbf{w} = \mathbf{W}(u^\star)$, so that $u^\star$ is equal to $u_{\mathbf{w}} = \mathcal{I}(G_{\mathbf{w}})$ (because $u_{\mathbf{w}} \leq u^\star$ from (24)). Using again that $\mathbf{W}^\star$ is a maximum, we obtain

$$G_{\mathbf{W}^\star}(u^\star - \lambda) \geq G_{\mathbf{w}}(u^\star - \lambda) = G_{\mathbf{w}}(u_{\mathbf{w}} - \lambda) = \frac{G_{\mathbf{w}}(u_{\mathbf{w}} - \lambda)}{u_{\mathbf{w}} - \lambda}(u_{\mathbf{w}} - \lambda)$$
$$> \frac{G_{\mathbf{w}}(u_{\mathbf{w}})}{u_{\mathbf{w}}}(u_{\mathbf{w}} - \lambda) = u^\star - \lambda, \qquad (25)$$

by the strict concavity of $G_{\mathbf{w}}$. This implies $\mathcal{I}_\lambda^-(G_{\mathbf{W}^\star}) > 0$.

Finally, using (22) with $\mathbf{W}(\cdot)$ constantly equal to any weight vector $\mathbf{w}$, together with (23) used with $\mathbf{W} = \mathbf{W}^\star$, we obtain for all $\lambda > 0$, $\lambda < u^\star$ and $\lambda < \pi_0(1 - \alpha)$,

$$\mathrm{Pow}(\mathbf{SU}(\mathbf{W}^\star))$$
$$\geq (1 - \alpha\pi_0)u^\star - \varepsilon(m, \mathcal{I}_\lambda^-(G_{\mathbf{W}^\star})) - \lambda(1 - \alpha\pi_0)$$
$$\geq (1 - \alpha\pi_0)u_{\mathbf{w}} - \varepsilon(m, \mathcal{I}_\lambda^-(G_{\mathbf{W}^\star})) - \lambda(1 - \alpha\pi_0)$$
$$\geq \mathrm{Pow}(\mathbf{LSU}(\mathbf{w})) - \varepsilon(m, \mathcal{I}_\lambda^+(G_{\mathbf{w}})) - \varepsilon(m, \mathcal{I}_\lambda^-(G_{\mathbf{W}^\star})) - 2\lambda(1 - \alpha\pi_0),$$

which proves (15).

Let us now prove Proposition 8.3. Using that the procedure $R = \mathbf{SU}(\mathbf{W})$ satisfies the self-consistency condition $R = \{i \mid p_i \leq \Delta_i(|R|/m)\}$, Lemma 8.1 (first statement) and the notation of Section 8.1, the power of the procedure R may be expressed as follows:

$$\mathrm{Pow}(\mathbf{SU}(\mathbf{W})) = \pi_1 m^{-1} \sum_{i=1}^m \mathbb{P}\left[p_i \leq \Delta_i(|R|/m) \mid H_i = 1\right]$$

$$= \pi_1 m^{-1} \sum_{i=1}^m \mathbb{P}\left[p_i \leq \Delta_i((1 - m^{-1})\widehat{u}'_{-i} + m^{-1}) \mid H_i = 1\right]$$

$$\leq \pi_1 m^{-1} \sum_{i=1}^m \mathbb{E}\left[F_i \circ \Delta_i((1 - m^{-1})\widehat{u}'_{-i} + m^{-1})\right], \qquad (26)$$

where we used both the independence between p_i and $\widehat{u}'_{-i}$ conditionally to $\mathbf{H}$ and the independence between $\widehat{u}'_{-i}$ and H_i. For simplicity, we introduce the increasing function $\phi(u) := (1 - m^{-1})u + m^{-1}$ of invert $\phi^{-1}(v) = (mv-1)/(m-1)$. Fix now $\lambda > 0$, with $\bar{u} + \lambda < 1$. Expression (26) may be rewritten as

$$\mathrm{Pow}(\mathbf{SU}(\mathbf{W})) - (1 - \alpha\pi_0)\bar{u}$$

$$\leq \mathbb{E}\left[\pi_1 m^{-1} \sum_{i=1}^m F_i \circ \Delta_i(\phi(\widehat{u}'_{-i})) - (1 - \alpha\pi_0)\bar{u}\right]$$

$$\leq \pi_1 \mathbb{P}(\Omega_1^c) + G_{\mathbf{W}}(\bar{u} + \lambda) - \alpha\pi_0(\bar{u} + \lambda) - (1 - \alpha\pi_0)\bar{u}$$

$$= \pi_1 \mathbb{P}(\Omega_1^c) + G_{\mathbf{W}}(\bar{u} + \lambda) - (\bar{u} + \lambda) + \lambda(1 - \alpha\pi_0),$$

where Ω_1 denotes the event $\left\{\forall i, 1 \leq i \leq m, \phi(\widehat{u}'_{-i}) \leq \bar{u} + \lambda\right\}$ and where the last inequality comes from the definition of $\bar{u}$. We upper-bound now the probability of Ω_1^c:

$$\mathbb{P}\left[\Omega_1^c\right] \leq \sum_{i=1}^m \mathbb{P}\left[\widehat{u}'_{-i} > \phi^{-1}(\bar{u} + \lambda)\right]$$

$$\leq \sum_{i=1}^m \sum_{u > \phi^{-1}(\bar{u}+\lambda)} \mathbf{1}\{u(m-1) \in \{0, 1, \ldots, m-1\}\}\mathbb{P}\left[\widehat{\mathbb{G}}'_{-i}(u) \geq u\right]$$

$$= \sum_{i=1}^m \sum_{v > \bar{u}+\lambda} \mathbf{1}\{vm \in \{1, 2, \ldots, m\}\}\mathbb{P}\left[\widehat{\mathbb{G}}'_{-i}(\phi^{-1}(v)) \geq \phi^{-1}(v)\right]$$

$$\leq \sum_{i=1}^m \sum_{v > \bar{u}+\lambda} \mathbf{1}\{vm \in \{1, 2, \ldots, m\}\}\mathbb{P}\left[\widehat{\mathbb{G}}_{\mathbf{W}}(v) \geq v - m^{-1}\right],$$

where the last inequality uses that $m\widehat{\mathbb{G}}_{\mathbf{W}}(v) \geq (m-1)\widehat{\mathbb{G}}'_{-i}(\phi^{-1}(v))$. As a consequence:

$$\mathbb{P}\left[\Omega_1^c\right] \leq m \sum_{\substack{v > \bar{u}+\lambda \\ vm \in \{1,2,\ldots,m\}}} \mathbb{P}\left[\widehat{\mathbb{G}}_{\mathbf{W}}(v) - G_{\mathbf{W}}(v) \geq v - G_{\mathbf{W}}(v) - m^{-1}\right]$$

$$\leq m \sum_{\substack{v > \bar{u}+\lambda \\ vm \in \{1,2,\ldots,m\}}} \mathbb{P}\left[\widehat{\mathbb{G}}_{\mathbf{W}}(v) - G_{\mathbf{W}}(v) > (\bar{u}+\lambda) - G_{\mathbf{W}}(\bar{u}+\lambda) - m^{-1}\right]$$

$$\leq m^2 \exp\left\{-2m\big((\bar{u}+\lambda) - G_{\mathbf{W}}(\bar{u}+\lambda) - m^{-1}\big)_+^2\right\}, \qquad (27)$$

where we used successively the assumption in (i) of Proposition 8.3 and Hoeffding's inequality (see Hoeffding (1963)) for the last inequality. This finally yields (22).

The point (ii) of Proposition 8.3 is similar: noticing that (26) is an equality when $\Delta \leq 1$, we obtain

$$\mathrm{Pow}(\mathbf{SU}(\mathbf{W})) - (1-\alpha\pi_0)\bar{u}$$
$$\geq -\pi_1 \mathbb{P}(\Omega_2^c) + (G_{\mathbf{W}}(\bar{u}-\lambda) - (\bar{u}-\lambda)) - \lambda(1-\alpha\pi_0),$$

with $\Omega_2 = \{\forall i, 1 \leq i \leq m, \phi(\widehat{u}'_{-i}) \geq \bar{u} - \lambda\}$. Next, we have

$$\mathbb{P}(\Omega_2^c) \leq \sum_{i=1}^{m} \mathbb{P}\left[\widehat{u}'_{-i} < \phi^{-1}(\bar{u}-\lambda)\right]$$

$$= \sum_{i=1}^{m} \mathbb{P}\left[\widehat{\mathbb{G}}'_{-i}(\phi^{-1}(\bar{u}-\lambda)) < \phi^{-1}(\bar{u}-\lambda)\right]$$

$$\leq \sum_{i=1}^{m} \mathbb{P}\left[\widehat{\mathbb{G}}_{\mathbf{W}}(\bar{u}-\lambda) < \bar{u} - \lambda\right],$$

where the last inequality uses that $m\widehat{\mathbb{G}}_{\mathbf{W}}(\bar{u}-\lambda) \leq (m-1)\widehat{\mathbb{G}}'_{-i}(\phi^{-1}(\bar{u}-\lambda)) + 1$ and thus $\phi^{-1}(\widehat{\mathbb{G}}_{\mathbf{W}}(\bar{u}-\lambda)) \leq \widehat{\mathbb{G}}'_{-i}(\phi^{-1}(\bar{u}-\lambda))$. As a consequence, we obtain

$$\mathbb{P}(\Omega_2^c) \leq m\mathbb{P}\left[\widehat{\mathbb{G}}_{\mathbf{W}}(\bar{u}-\lambda) < \bar{u} - \lambda\right]$$

$$= m\mathbb{P}\left[\widehat{\mathbb{G}}_{\mathbf{W}}(\bar{u}-\lambda) - G_{\mathbf{W}}(\bar{u}-\lambda) < \bar{u} - \lambda - G_{\mathbf{W}}(\bar{u}-\lambda)\right]$$

$$\leq m\exp\{-2m(\bar{u}-\lambda - G_{\mathbf{W}}(\bar{u}-\lambda))_+^2\},$$

which implies (23).

8.5. *Proof of Theorem 4.3*

First remark that for any weight function sequence of $\mathcal{W}$, the convergence of $(G_{\mathbf{W}^{(m)}})_m$ to G^∞ is uniform, because all these fonctions are non-decreasing and because G^∞ is assumed to be continuous on $[0,1]$. Next we prove that $\mathcal{I}(G_{\mathbf{W}^{(m)}}) \to \mathcal{I}(G^\infty)$. (This will imply directly that $\mathcal{I}_\lambda^+(G_{\mathbf{W}^{(m)}}) \to \mathcal{I}_\lambda^+(G^\infty)$ and $\mathcal{I}_\lambda^-(G_{\mathbf{W}^{(m)}}) \to \mathcal{I}_\lambda^-(G^\infty)$ for $\lambda < \mathcal{I}(G^\infty)$.) For this, take a subsequence m' such that $\mathcal{I}(G_{\mathbf{W}^{(m')}})$ converges and prove that its limit ℓ is equal to $\mathcal{I}(G^\infty)$. From the uniform convergence and the continuity of G^∞, ℓ satisfies $G^\infty(\ell) = \ell$.

If $\mathcal{I}(G^\infty) = 0$, the only possible fixed point of G^∞ is 0 and $\ell = 0$. If $\mathcal{I}(G^\infty) > 0$, 0 and $\mathcal{I}(G^\infty)$ are the only possible fixed points of G^∞ (because $\mathcal{I}_\lambda^-(G^\infty) > 0$ for $\lambda < \mathcal{I}(G^\infty)$). Next, we have $G_{\mathbf{W}^{(m')}}(\mathcal{I}(G^\infty)/2) \geq \mathcal{I}(G^\infty)/2$ for large m' (because $G^\infty(\mathcal{I}(G^\infty)/2) > \mathcal{I}(G^\infty)/2$) and thus $\ell \geq \mathcal{I}(G^\infty)/2 > 0$, which in turn implies $\ell = \mathcal{I}(G^\infty)$.

Fix a sequence of weight vector $(\mathbf{w}^{(m)})_m$ belonging to $\mathcal{W}$. We aim now to prove:

$$\lim_m \mathrm{Pow}(\mathbf{SU}(\mathbf{W}^{\star,(m)})) = (1 - \pi_0 \alpha) \lim_m \{\mathcal{I}(G_{\mathbf{W}^{\star,(m)}})\} \qquad (28)$$

$$\lim_m \mathrm{Pow}(\mathbf{SU}(\mathbf{w}^{(m)})) = (1 - \pi_0 \alpha) \lim_m \{\mathcal{I}(G_{\mathbf{w}^{(m)}})\} \qquad (29)$$

Expression (16) will then directly follow from $\mathcal{I}(G_{\mathbf{W}^{\star,(m)}}) \geq \mathcal{I}(G_{\mathbf{w}^{(m)}})$ (as stated in the proof of Theorem 4.2).

Let us state now (28) (the proof for (29) is similar). Fix $\lambda > 0$ with $\lambda < \pi_0(1 - \alpha)$. Applying Proposition 8.3, we obtain that

$$|\mathrm{Pow}(\mathbf{SU}(\mathbf{W}^{\star,(m)})) - (1 - \pi_0 \alpha)\mathcal{I}(G_{\mathbf{W}^{\star,(m)}})|$$
$$\leq \varepsilon(m, \mathcal{I}_\lambda^+(G_{\mathbf{W}^{\star,(m)}})) + \varepsilon(m, \mathcal{I}_\lambda^-(G_{\mathbf{W}^{\star,(m)}}))\mathbf{1}\{\lambda < \mathcal{I}(G_{\mathbf{W}^{\star,(m)}})\} - 2\lambda(1 - \pi_0 \alpha).$$

First, denoting $G^{\star,\infty} = \lim_m G_{\mathbf{W}^{\star,(m)}}$, we have $\mathcal{I}_\lambda^+(G_{\mathbf{W}^{\star,(m)}}) \to \mathcal{I}_\lambda^+(G^{\star,\infty}) > 0$ and thus $\lim_m \varepsilon(m, \mathcal{I}_\lambda^+(G_{\mathbf{W}^{\star,(m)}})) = 0$. Second, if $\mathcal{I}(G^{\star,\infty}) > 0$, we have $\mathcal{I}_\lambda^-(G^{\star,\infty}) > 0$ for $\lambda < \mathcal{I}(G^{\star,\infty})$ and thus $\lim_m \varepsilon(m, \mathcal{I}_\lambda^-(G_{\mathbf{W}^{\star,(m)}})) = 0$ for $\lambda < \mathcal{I}(G^{\star,\infty})$. If $\mathcal{I}(G^{\star,\infty}) = 0$, we trivially have that $\mathbf{1}\{\lambda < \mathcal{I}(G_{\mathbf{W}^{\star,(m)}})\}$ is equal to zero for m large. As a result, we obtain for λ small enough that

$$\limsup_m \left\{ \left| \mathrm{Pow}(\mathbf{SU}(\mathbf{W}^{\star,(m)})) - (1 - \pi_0 \alpha)\mathcal{I}(G_{\mathbf{W}^{\star,(m)}}) \right| \right\} \leq -2\lambda(1 - \pi_0 \alpha).$$

This yields (28) by letting $\lambda \to 0$ and by noticing that $\lim_m \{\mathcal{I}(G_{\mathbf{W}^{\star,(m)}})\}$ exists. Finally, we have to check that the use of $\mathbf{W}^{\star,(m)}$ in Proposition 8.3 was allowed, i.e. that for all m and $u' > u > u^\star := \mathcal{I}(G_{\mathbf{W}^{\star,(m)}})$, inequality $u' - G_{\mathbf{W}^{\star,(m)}}(u') > u - G_{\mathbf{W}^{\star,(m)}}(u)$ holds. For this, we let $\mathbf{w} := \mathbf{W}^{\star,(m)}(u')$ and $u_{\mathbf{w}} := \mathcal{I}(G_{\mathbf{w}})$. Since $u^\star \geq u_{\mathbf{w}}$ and $G_{\mathbf{w}}(u_{\mathbf{w}}) = u_{\mathbf{w}}$ we have $u' - G_{\mathbf{w}}(u') > u - G_{\mathbf{w}}(u)$ ($G_{\mathbf{w}}$ being strictly concave). Therefore, for this particular weight vector $\mathbf{w}$,

$$u' - G_{\mathbf{W}^{\star,(m)}}(u') = u' - G_{\mathbf{w}}(u') > u - G_{\mathbf{w}}(u) \geq u - G_{\mathbf{W}^{\star,(m)}}(u),$$

where the last inequality holds because $\mathbf{W}^{\star,(m)}$ is a maximum.

Finally, to get the FDR statement (17), we use the same reasoning as above combined with the following finite FDR approximation result:

Proposition 8.4. *In the unconditional model, consider a weight function $\mathbf{W}$ with its associated threshold collection Δ and put $\bar{u} = \mathcal{I}(G_{\mathbf{W}})$. Assume that for all $u' > u > \bar{u}$, $u' - G_{\mathbf{W}}(u') > u - G_{\mathbf{W}}(u)$ and take $\lambda > 0$ with $\lambda < 1 - \bar{u}$. Then the following bounds hold:*

$$
\begin{aligned}
FDR(\mathbf{SU}(\mathbf{W})) \quad \leq \quad & \pi_0\alpha + \pi_0\alpha m^3 \exp\big\{ -2m\big(\mathcal{I}_\lambda^+(G_\mathbf{W}) - m^{-1}\big)_+^2 \big\} \\
& +\pi_0\alpha \mathbf{1}\{\bar{u} > \lambda\} \left[m^2 \exp\big\{ -2m(\mathcal{I}_\lambda^-(G_\mathbf{W}))_+^2 \big\} + \frac{2\lambda}{\bar{u} - \lambda} \right] \\
& +\pi_0\alpha \mathbf{1}\{\bar{u} \leq \lambda\} \left[m^{-1}\sum_{i=1}^m \sup_{0 < u \leq 2\lambda} \{W_i(u)\} - 1 \right]. \quad (30)
\end{aligned}
$$

Assuming additionally $\Delta \leq 1$, *we have*

$$
\begin{aligned}
FDR(\mathbf{SU}(\mathbf{W})) \quad \geq \quad & \pi_0\alpha - \pi_0\alpha m^3 \exp\big\{ -2m\big(\mathcal{I}_\lambda^+(G_\mathbf{W}) - m^{-1}\big)_+^2 \big\} \\
& -\pi_0\alpha \mathbf{1}\{\bar{u} > \lambda\} \left[m^2 \exp\big\{ -2m(\mathcal{I}_\lambda^-(G_\mathbf{W}))_+^2 \big\} + \frac{2\lambda}{\bar{u} + \lambda} \right] \\
& -\pi_0\alpha \mathbf{1}\{\bar{u} \leq \lambda\} \left[1 - m^{-1}\sum_{i=1}^m \inf_{0 < u \leq 2\lambda} \{W_i(u)\} \right]. \quad (31)
\end{aligned}
$$

To prove Proposition 8.4, we write the FDR as (using the same reasoning and notation as in Section 8.4),

$$
\begin{aligned}
\text{FDR}(\mathbf{SU}(\mathbf{W})) &= \pi_0 \sum_{i=1}^m \mathbb{E}\left[\frac{\mathbf{1}\{p_i \leq \Delta_i(|R|/m)\}}{|R|} \,\Big|\, H_i = 0 \right] \\
&= \pi_0 \sum_{i=1}^m \mathbb{E}\left[\frac{\mathbf{1}\{p_i \leq \Delta_i(\phi(\widehat{u}'_{-i}))\}}{m\phi(\widehat{u}'_{-i})} \,\Big|\, H_i = 0 \right] \\
&\leq \pi_0 m^{-1} \sum_{i=1}^m \mathbb{E}\left[\frac{\Delta_i(\phi(\widehat{u}'_{-i}))}{\phi(\widehat{u}'_{-i})} \right] \\
&\leq \pi_0 m^{-1} \sum_{i=1}^m \mathbb{E}\left[\frac{\Delta_i(\phi(\widehat{u}'_{-i}))}{\phi(\widehat{u}'_{-i})} \mathbf{1}\{\Omega_1 \cap \Omega_2\} \right] \\
&\quad + \pi_0\alpha m(\mathbb{P}\left[\Omega_1^c\right] + \mathbb{P}\left[\Omega_2^c\right]). \quad (32)
\end{aligned}
$$

On one hand, when $\bar{u} > \lambda$, we may write

$$
m^{-1} \sum_{i=1}^m \mathbb{E}\left[\frac{\Delta_i(\phi(\widehat{u}'_{-i}))}{\phi(\widehat{u}'_{-i})} \mathbf{1}\{\Omega_1 \cap \Omega_2\} \right] \leq m^{-1} \sum_{i=1}^m \left[\frac{\Delta_i(\bar{u} + \lambda)}{\bar{u} - \lambda} \right] = \alpha + 2\alpha \frac{\lambda}{\bar{u} - \lambda}.
$$

On the other hand, when $\bar{u} \leq \lambda$, we have $\Omega_2^c = \emptyset$ and

$$
m^{-1} \sum_{i=1}^m \mathbb{E}\left[\frac{\Delta_i(\phi(\widehat{u}'_{-i}))}{\phi(\widehat{u}'_{-i})} \mathbf{1}\{\Omega_1\} \right] \leq \alpha m^{-1} \sum_{i=1}^m \sup_{0 < u \leq 2\lambda} \{W_i(u)\}.
$$

This implies (30). The proof for (31) is similar, by noticing that (32) is an equality when $\Delta \leq 1$.

8.6. Proof of Proposition 3.2

Let $t = \alpha u$ and assume $(A1)$–$(A2)$–$(A3)$. Consider first the conditional model. Following the constrained Lagrange multiplier method, the problem is to maximize in $(\lambda, \mathbf{w})$ the function:

$$L(\lambda, \mathbf{w}) = \sum_{i \in \mathcal{H}_1} F_i(w_i t) - \lambda \left(\sum_{i \in \mathcal{H}_1} w_i - m \right).$$

Assume that $(w_i)_i$ is a critical point, i.e. that for each i: $\frac{\partial L}{\partial w_i}(\lambda, \mathbf{w}) = t f_i(w_i t) - \lambda = 0$, so that $w_i = t^{-1} f_i^{-1}(\lambda t^{-1})$. Then, λ is chosen such that $\sum_i w_i = m$ i.e. $\lambda = t \Psi^{-1}(t)$ for $\Psi(y) = m^{-1} \sum_{j \in \mathcal{H}_1} f_j^{-1}(y)$. Hence, we find that the only possible critical point is $\forall i, w_i = W_i^\star(u)$. To conclude, it is sufficient to prove that $(W_i^\star(u))_i$ is a maximum. The latter holds because for each i, f_i is decreasing, so that $\frac{\partial^2 L}{\partial w_i^2}(\lambda, \mathbf{w}) = t^2 f_i'(w_i t) < 0$. Therefore, since $\alpha u W_i^\star(u) = \zeta_i^{-1}(\alpha u)$, where $\zeta_i(x) = \Psi(f_i(x)) = m^{-1} \sum_{j \in \mathcal{H}_1} f_j^{-1}(f_i(x))$ is a differentiable increasing function from $(0, 1)$ to $(0, \pi_1)$, we easily check that $\mathbf{W}^\star$ satisfies (8) and is continuous. Next, assuming in addition $(A4)$, we obtain for all i,

$$\lim_{u \to 0^+} W_i^\star(u) = \lim_{u \to 0^+} \frac{\zeta_i^{-1}(\alpha u)}{\alpha u} = \lim_{\varepsilon \to 0^+} \frac{\varepsilon}{\zeta_i(\varepsilon)} = \lim_{y \to f_i(0^+)} \frac{f_i^{-1}(y)}{m^{-1} \sum_{j \in \mathcal{H}_1} f_j^{-1}(y)},$$

which exists in $[0, m]$. Finally, the results in the unconditional model follow from the same reasoning as above by replacing $\mathcal{H}_1$ by $\{1, \ldots, m\}$.

8.7. Proof for the continuous Gaussian case of Section 4.3

Fix $u \in (0, 1]$. We aim to prove that

$$\frac{1}{m} \sum_{i=1}^m \overline{\Phi}\left(-\frac{\mu(i/m)}{2} + \frac{c^{(m)}(u)}{\mu(i/m)} \right) \to \int_0^1 \overline{\Phi}\left(-\frac{\mu(t)}{2} + \frac{c^\infty(u)}{\mu(t)} \right) dt. \qquad (33)$$

First, we have $c^{(m)}(u) \to c^\infty(u)$, because $c^{(m)}(u) = \Psi_m^{-1}(\alpha u)$ and $c^\infty(u) = \Psi^{-1}(\alpha u)$ where $\Psi_m(x) = \frac{1}{m} \sum_{i=1}^m \overline{\Phi}\big(\mu(i/m)/2 + x/\mu(i/m)\big)$ and where $\Psi(x) = \int_0^1 \overline{\Phi}\big(\mu(t)/2 + x/\mu(t)\big) dt$ are decreasing continuous functions such that Ψ_m converges uniformly to Ψ. Second, we have for any $\epsilon > 0$,

$$\frac{1}{m} \sum_{i=1}^m \left| \overline{\Phi}\left(-\frac{\mu(i/m)}{2} + \frac{c^{(m)}(u)}{\mu(i/m)} \right) - \overline{\Phi}\left(-\frac{\mu(i/m)}{2} + \frac{c^\infty(u)}{\mu(i/m)} \right) \right|$$

$$\leq m^{-1} |\{i \mid \mu(i/m) \leq \epsilon\}| + \frac{|c^{(m)}(u) - c^\infty(u)|}{\epsilon \sqrt{2\pi}},$$

where we used that $\overline{\Phi}$ is $1/\sqrt{2\pi}$-Lipschitz. Since the measure $m^{-1} \sum_{i=1}^m \delta_{i/m}$ converges weakly to the Lebesgue measure Λ on $[0, 1]$, we derive the inequality

$\limsup_m (m^{-1}|\{i \mid \mu(i/m) \le \epsilon\}|) \le \Lambda(t \mid \mu(t) \le \epsilon)$. By assumption, μ is positive on $(0, 1]$ so that the latter converges to 0 as ϵ tends to 0. This implies (33).

Appendix

Appendix A: Practical implementation of the new procedures

Algorithm A.1. *(Step-up algorithm for* $\mathbf{SU}(\mathbf{W})$*)*

- *Step* 1*: compute for each* i *the weight vector* $(W_i(1))_i$ *and the weighted p-values* $p'_i = p_i/W_i(1)$. *If all the weighted p-values are less than or equal to* α, *then reject all the null hypotheses. Otherwise go to step* 2.
- *Step* j $(j \ge 2)$*: put* $r = m - j + 1$ *and* $u = r/m$ *and compute for each* i *the weight vector* $(W_i(u))_i$ *and the weighted p-values* $p'_i = p_i/W_i(u)$. *Order the weighted p-values following* $p'_{(1)} \le \cdots \le p'_{(m)}$. *If* $p'_{(r)} \le \alpha u$, *then reject the* r *null hypotheses corresponding to the smaller weighted p-values* $p'_{(i)}$, $1 \le i \le r$. *Otherwise go to step* $j + 1$ *(if* $j = m$ *stop and reject no null hypothesis).*

Algorithm A.2. *(Step-down algorithm for* $\mathbf{SD}(\mathbf{W})$*)*

- *Step* 1*: compute for each* i *the weight vector* $(W_i(1/m))_i$ *and the weighted p-values* $p'_i = p_i/W_i(1/m)$. *If the smallest weighted p-values is strictly larger than* α/m, *then reject no null hypothesis. Otherwise go to step* 2.
- *Step* j $(j \ge 2)$*: put* $r = j$, $u = r/m$ *and compute for each* i *the weight vector* $(W_i(u))_i$ *and the weighted p-values* $p'_i = p_i/W_i(u)$. *Order the weighted p-values following* $p'_{(1)} \le \cdots \le p'_{(m)}$. *If* $p'_{(r)} > \alpha u$, *then reject the* $r - 1$ *null hypotheses corresponding to the smaller weighted p-values* $p'_{(i)}$, $1 \le i \le r - 1$. *Otherwise go to step* $j + 1$ *(if* $j = m$ *stop and reject all the null hypotheses).*

Appendix B: Proofs of technical lemmas

Proof of Lemma 8.1. Let us first prove the first point. Denote $\Delta'_j(u) = \Delta_j((1 - m^{-1})u + m^{-1})$, and $\phi(u) := (1 - m^{-1})u + m^{-1}$ with $\phi^{-1}(u) = (mu - 1)/(m - 1)$, so that $\Delta_j(u) = \Delta'_j(\phi^{-1}(u))$. Since $m\widehat{\mathbb{G}}_{\mathbf{W}}(u) = (m - 1)\widehat{\mathbb{G}}'_{-i}(\phi^{-1}(u)) + \mathbf{1}\{p_i \le \Delta_i(u)\}$ the following equivalence holds when $p_i \le \Delta_i(u)$:

$$\widehat{\mathbb{G}}_{\mathbf{W}}(u) \ge u \iff \widehat{\mathbb{G}}'_{-i}(\phi^{-1}(u)) \ge \phi^{-1}(u). \tag{34}$$

First, assuming $p_i \le \Delta_i(\widehat{u})$, equivalence (34) used with $u = \widehat{u}$ leads to inequality $\widehat{\mathbb{G}}'_{-i}(\phi^{-1}(\widehat{u})) \ge \phi^{-1}(\widehat{u})$ and thus $\phi^{-1}(\widehat{u}) \le \widehat{u}'_{-i}$ because $\widehat{u}'_{-i}$ is defined as a maximum. This implies $p_i \le \Delta_i(\widehat{u}) = \Delta'_i(\phi^{-1}(\widehat{u})) \le \Delta'_i(\widehat{u}'_{-i})$. Conversely, assuming $p_i \le \Delta'_i(\widehat{u}'_{-i}) = \Delta_i(\phi(\widehat{u}'_{-i}))$, equivalence (34) used with $u = \phi(\widehat{u}'_{-i})$ yields $\widehat{\mathbb{G}}_{\mathbf{W}}(\phi(\widehat{u}'_{-i})) \ge \phi(\widehat{u}'_{-i})$ and thus $\phi(\widehat{u}'_{-i}) \le \widehat{u}$ by definition of $\widehat{u}$. The first

point is thus proved by additionally noticing that when both $p_i \leq \Delta_i(\widehat{u})$ and $p_i \leq \Delta_i'(\widehat{u}'_{-i})$ we have both $\widehat{u}'_{-i} \geq \phi^{-1}(\widehat{u})$ and $\phi(\widehat{u}'_{-i}) \leq \widehat{u}$, so that $\phi(\widehat{u}'_{-i}) = \widehat{u}$.

For the second point of the lemma, remark that

$$m\widehat{\mathbb{G}}_{\mathbf{W}}(u) = (m-1)\widehat{\mathbb{G}}_{-i}(um/(m-1)) + \mathbf{1}\{p_i \leq \Delta_i(u)\}. \qquad (35)$$

Therefore, we always have $\widehat{\mathbb{G}}_{\mathbf{W}}(u) \geq u \impliedby \widehat{\mathbb{G}}_{-i}(um/(m-1)) \geq um/(m-1)$, which implies, using $u = (1-m^{-1})\widehat{u}_{-i}$, that $\widehat{u} \geq (1-m^{-1})\widehat{u}_{-i}$ always holds. Next, when $p_i > \Delta_i(u)$, we have $\widehat{\mathbb{G}}_{\mathbf{W}}(u) \geq u \implies \widehat{\mathbb{G}}_{-i}(um/(m-1)) \geq um/(m-1)$, so that taking $u = \widehat{u}$ in the relation above leads to $\widehat{u} \leq (1-m^{-1})\widehat{u}_{-i}$ and thus $\widehat{u} = (1-m^{-1})\widehat{u}_{-i}$. Conversely, if $p_i \leq \Delta_i(\widehat{u})$, from the first point of the lemma we obtain $\widehat{u} = (1-m^{-1})\widehat{u}'_{-i} + m^{-1}$ and since $\widehat{u}'_{-i} \geq \widehat{u}_{-i}$ (because pointwise $\widehat{\mathbb{G}}_{-i} \leq \widehat{\mathbb{G}}'_{-i}$), we deduce $\widehat{u} > (1-m^{-1})\widehat{u}_{-i}$ which finishes the proof. $\qquad\square$

Proof of Lemma 8.2. For the first point, take $u' \leq (k-1)/(m-1)$ and apply (35) with $u = (1-m^{-1})u'$, which gives $m\widehat{\mathbb{G}}_{\mathbf{W}}((1-m^{-1})u') = (m-1)\widehat{\mathbb{G}}_{-i}(u') + \mathbf{1}\{p_i \leq \Delta_i((1-m^{-1})u')\}$. Since $(1-m^{-1})u' \leq (k-1)/m$, assuming $\widetilde{u} \geq k/m$ and $p_i > \Delta_i((k-1)/m)$, we obtain $p_i > \Delta_i((1-m^{-1})u')$ and $\widehat{\mathbb{G}}_{\mathbf{W}}((1-m^{-1})u') \geq (1-m^{-1})u'$, which thus leads to $\widehat{\mathbb{G}}_{-i}(u') \geq u'$. Since this holds for any $u' \leq (k-1)/(m-1)$, we finally have $\widetilde{u}_{-i} \geq (k-1)/(m-1)$.

To prove the second point, take $u' \leq (k-1)/m$ and use (35) with $u = u'$. This gives $m\widehat{\mathbb{G}}_{\mathbf{W}}(u') = (m-1)\widehat{\mathbb{G}}_{-i}(u'm/(m-1)) + \mathbf{1}\{p_i \leq \Delta_i(u')\}$. Since $u'm/(m-1) \leq (k-1)/(m-1)$ and $u' \leq (k-1)/m$, if $\widetilde{u}_{-i} \geq (k-1)/(m-1)$ and $p_i \leq \Delta_i((k-1)/m)$, we obtain $\widehat{\mathbb{G}}_{\mathbf{W}}(u') \geq u' + m^{-1} \geq u'$. This holds for any $u' \leq (k-1)/m$ and also for $u' = k/m$ because $\widehat{\mathbb{G}}_{\mathbf{W}}(k/m) \geq \widehat{\mathbb{G}}_{\mathbf{W}}((k-1)/m) \geq (k-1)/m + m^{-1} = k/m$. Finally $\widetilde{u} \geq k/m$.

For the third point, remark that since we trivially have $\widetilde{u} \geq (1-m^{-1})\widetilde{u}_{-i}$ (with an argument similar than in the step-up case), it is sufficient to prove $\widetilde{u} \leq (1-m^{-1})\widetilde{u}_{-i}$. For this, use (35) with $u' = (1-m^{-1})\widetilde{u}_{-i} + m^{-1}$, leading to $m\widehat{\mathbb{G}}_{\mathbf{W}}(u') = (m-1)\widehat{\mathbb{G}}_{-i}(u'm/(m-1)) + \mathbf{1}\{p_i \leq \Delta_i(u')\}$. Since R_{-i} is step-down and since $u'm/(m-1) = \widetilde{u}_{-i} + (m-1)^{-1}$ we have by definition of $\widetilde{u}_{-i}$ that $\widehat{\mathbb{G}}_{-i}(u'm/(m-1)) > u'm/(m-1)$. Therefore, assuming $p_i > \Delta_i(u')$, we obtain that $\widehat{\mathbb{G}}_{\mathbf{W}}(u') > u'$, meaning that $\widetilde{u} < u'$ because R is step-down. Hence $\widetilde{u} \leq (1-m^{-1})\widetilde{u}_{-i}$. $\qquad\square$

Appendix C: Some FDR bounds for SU(W) and SD(W)

C.1. Step-up case

Lemma C.1. *Consider the conditional model in the situation where only two true hypotheses are tested, that is, $m = m_0 = 2$. Then for any weight function W, the procedure* **SU(W)** *has a FDR equal to $\alpha + \alpha^2(1 - W_1(1))(W_1(1) - W_1(1/2))$.*

In particular, in the conditional model with $m = m_0 = 2$, the FDR in the above lemma has a maximum equal to $\alpha + \alpha^2/4$, attained e.g. for the weight function $W_1(1/2) = 0$; $W_2(1/2) = 2$; $W_1(1) = 0.5$; $W_2(1) = 1.5$. Additionally, in the unconditional model, the FDR of the above procedure is larger than $\pi_0^2(\alpha + \alpha^2/4)$ and can thus be larger than α when π_0 is (very) close to 1.

C.2. Step-down case

The next result states that $\mathbf{SD}(\mathbf{W})$ control non-asymptotically the FDR without correction in the case $m = 2$ and when $m_0 = m$ in the conditional model. This is quite intriguing and we may think that $\mathbf{SD}(\mathbf{W})$ controls the FDR for any m and m_0.

Lemma C.2. *For any weight function* $\mathbf{W}$, *the procedure* $\mathbf{SD}(\mathbf{W})$ *controls the FDR at level* α *in either of the two following cases:*

(i) *in the unconditional model when all the hypotheses are true, that is $m_0 = m$,*

(ii) *in both conditional and unconditional model when $m = 2$.*

To prove (i), we easily check that, when all the hypotheses are true, the FDR of $\mathbf{SD}(\mathbf{W})$ is $1 - \mathbb{P}\big[\widehat{\mathbb{G}}_W(1/m) = 0\big]$ and is thus equal to the FDR of $\mathbf{LSD}(\mathbf{W}(1/m))$, which is equal to α from results on weighted linear step down procedures (see e.g. Blanchard and Roquain (2008)). To prove point (ii), we just have to check the case $m_0 = 1$ from point (i). This trivially holds from (20) (which also holds in the step-down case), because all the weights are smaller than $m = 2$.

References

BENJAMINI, Y. AND HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300. MR1325392

BENJAMINI, Y., KRIEGER, A.M., AND YEKUTIELI, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507. MR2261438

BENJAMINI, Y. AND YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188. MR1869245

BLANCHARD, G. AND ROQUAIN, E. (2008). Two simple sufficient conditions for FDR control. *Electron. J. Stat.*, 2:963–992. MR2448601

BLANCHARD, G. AND ROQUAIN, E. (2009). Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.* To appear.

EFRON, B., TIBSHIRANI, R., STOREY, J.D., AND TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.*, 96(456):1151–1160. MR1946571

FINNER, H., DICKHAUS, R., AND ROTERS, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.*, 37(2):596–618. MR2502644

GAVRILOV, Y., BENJAMINI, Y., AND SARKAR, S.K. (2009). An adaptive step-down procedure with proven fdr control under independence. *Ann. Statist.*, 37(2):619–629. MR2502645

GENOVESE, C. AND WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.*, 32(3):1035–1061. MR2065197

GENOVESE, C.R., ROEDER, K., AND WASSERMAN, L. (2006). False discovery control with p-value weighting. *Biometrika*, 93(3):509–524. MR2261439

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30. MR0144363

HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, 6(2):65–70. MR0538597

HYMAN, E., KAURANIEMI, P., HAUTANIEMI, S., WOLF, M., MOUSSES, S., ROZENBLUM, E., RINGNR, M., SAUTER, G., MONNI, O., ELKAHLOUN, A., KALLIONIEMI, A., AND KALLIONIEMI, O. (2002). Impact of dna amplification on gene expression patterns in breast cancer. *Cancer research*, 62(21):6240–5.

KIM, K.I. AND VAN DE WIEL, M. (2008). Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinformatics*, 9(1):114.

LEHMANN, E.L., ROMANO, J.P., AND SHAFFER, J.P. (2005). On optimality of stepdown and stepup multiple test procedures. *Ann. Statist.*, 33(3):1084–1108. MR2195629

MURIS, J., YLSTRA, B., CILLESSEN, S., OSSENKOPPELE, G., KLUIN-NELEMANS, J., EIJK, P., NOTA, B., TIJSSEN, M., DE BOER, W., VAN DE WIEL, M., VAN DEN IJSSEL, P., JANSEN, P., DE BRUIN, P., VAN KRIEKEN, J., MEIJER, G., MEIJER, C., AND OUDEJANS, J. (2007). Profiling of apoptosis genes allows for clinical stratification of primary nodal diffuse large B-cell lymphomas. *Br. J. Haematol.*, 136:38–47.

PICARD, F., ROBIN, S., LEBARBIER, E., AND DAUDIN, J. (2007). A segmentation/clustering model for the analysis of array CGH data. *Biometrics*, 63:758–766. MR2395713

ROMANO, J.P., SHAIKH, A.M., AND WOLF, M. (2008). Control of the false discovery rate under dependence using the bootstrap and subsampling. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 17(3):417–442. MR2470085

ROQUAIN, E. (2007). *Exceptional motifs in heterogeneous sequences. Contributions to theory and methodology of multiple testing.* PhD thesis, Université Paris XI.

ROQUAIN, E. AND VAN DE WIEL, M. (2008). Multi-weighting for FDR control. ArXiV preprint math.ST/0807.4081v1.

RUBIN, D., DUDOIT, S., AND VAN DER LAAN, M. (2006). A method to increase the power of multiple testing procedures through sample splitting. *Stat. Appl. Genet. Mol. Biol.*, 5:Art. 19, 20 pp. (electronic). MR2240850

SARKAR, S.K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.*, 30(1):239–257. MR1892663

STOREY, J.D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):479–498. MR1924302

STOREY, J.D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.*, 31(6):2013–2035. MR2036398

STOREY, J.D. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(3):347–368. MR2323757

VAN DER VAART, A.W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. MR1652247

WASSERMAN, L. AND ROEDER, K. (2006). Weighted hypothesis testing. Technical report, Dept. of statistics, Carnegie Mellon University.