



# Type I error rate control in multiple testing: a survey with proofs

Etienne Roquain

## ► To cite this version:

Etienne Roquain. Type I error rate control in multiple testing: a survey with proofs. 2010. hal-00547965v1

**HAL Id: hal-00547965**

**<https://hal.science/hal-00547965v1>**

Preprint submitted on 18 Dec 2010 (v1), last revised 28 Dec 2022 (v3)

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

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

# Type I error rate control in multiple testing: a survey with proofs

Etienne Roquain

*UPMC University of Paris 6, LPMA*

**Abstract:** This paper presents a survey on some recent advances in the error rate control in multiple testing methodology. We consider the problem of controlling the  $k$ -family-wise error rate (kFWER, probability to make  $k$  false discoveries or more) and the false discovery proportion (FDP, proportion of false discoveries among the discoveries), the latter being controlled either via its expectation, which is the so-called false discovery rate (FDR), or via its upper-tail distribution function. The emphasis is put both on the results which we want as unified as possible and on the mathematical proofs which we want as concise and simple as possible. This way, this paper is intended to a possibly non-specialist reader which would like to recover state-of-the-art multiple testing results using short proofs. As an additional contribution, we extend some previous work for controlling the kFWER and the upper-tail distribution of the FDP.

**AMS 2000 subject classifications:** Primary 62J15, ; secondary 62G10.

**Keywords and phrases:** multiple testing, type I error rate, false discovery proportion, family-wise error, step-up, step-down, positive dependence.

## Contents

1	Introduction . . . . .	2
1.1	Microarray data . . . . .	2
1.2	Instance of multiple testing settings . . . . .	2
1.3	General multiple testing setting . . . . .	4
1.4	Multiple testing procedures . . . . .	5
1.5	Type I error rates . . . . .	5
1.6	Goal . . . . .	6
1.7	Presentation of the paper . . . . .	7
2	Key concepts and tools . . . . .	7
2.1	Model assumptions . . . . .	7
2.2	Dirac configurations . . . . .	8
2.3	Algorithms . . . . .	8
2.4	Adaptive control . . . . .	8
3	FDR control . . . . .	9
3.1	Thresholding based procedures . . . . .	9
3.2	Linear step-up procedures . . . . .	10
3.3	Adaptive linear step-up procedures . . . . .	12
3.4	Arbitrary dependencies . . . . .	14
4	$k$ -FWER control . . . . .	16
4.1	Subset-indexed family . . . . .	16
4.2	Single-step method . . . . .	17
4.3	Step-down method for FWER . . . . .	18
4.4	Step-down method for $k$ -FWER . . . . .	19

<b>5</b>	<b>FDP control</b>	<b>21</b>
5.1	Family indexed by rejection numbers	21
5.2	Generic step-down theorem	23
5.3	Theorem 3.1 of [30] and a novel result as corollaries	24
5.4	Theorem 4.1 (i) of [35] as corollary	25
<b>6</b>	<b>Discussions</b>	<b>25</b>
6.1	Complexity of the $k$ -FWER step-down approach	25
6.2	FDR control is not FDP control	26
6.3	Conclusion	26
	<b>Acknowledgements</b>	<b>26</b>
	<b>References</b>	<b>27</b>

## 1. Introduction

The multiple testing problem arises in any data analysis where one tries to extract significant features among a large set of candidates. Classical instances are microarray analysis [45, 15], neuro-imaging analysis [3, 33] and source detection [31]. For illustration, we detail below the case of microarray data analysis.

### 1.1. Microarray data

In a typical microarray experiment, the level expressions of a set of genes are measured under two different experimental conditions and we aim at finding the genes that are differentially expressed between the two conditions. For instance, if the genes come from tumor cells in the first experimental condition, while they come from healthy cells in the second, the differentially expressed genes may be involved in the development of this tumor and thus are genes of special interest. Several techniques exist to perform a statistical test for a single gene, e.g. based on a distributional assumption or on permutations between the two group labels. However, the number of genes  $m$  can be large (for instance several thousands), so that non-differentially expressed genes can have a high score of significance by chance. Therefore, a non-corrected procedure (level  $\alpha$  for each single test) is likely to select a lot of non-differentially expressed genes (usually called “false discoveries”). A multiple testing procedure aims at correcting the level of the single tests in order to bound a *global type I error rate*, as for instance the probability to make at least  $k$  errors among the genes declared differentially expressed ( $k$ -family-wise error rate,  $k$ -FWER) or the proportion of errors among the genes declared differentially expressed (false discovery proportion, FDP). Of course, one feature that increases the complexity of this issue is the presence of dependencies between the single tests.

Note that the multiple testing issue can be met in microarray analysis under other forms, for instance when we search the co-expressed genes or the genes associated with clinical covariates or outcomes, see Section 1.2 of [15].

### 1.2. Instance of multiple testing settings

**Example 1.1** (Two-sample multiple t-tests). Let us observe

$$X = (X^1, \dots, X^n) = ((Y^1, \dots, Y^{n_1}), (Z^1, \dots, Z^{n_2})) \in \mathbb{R}^{m \times n}$$

a couple of two independent samples, where  $(Y^1, \dots, Y^{n_1})$  is a family of  $n_1$  i.i.d. copies of a random vector  $Y$  in  $\mathbb{R}^m$  and  $(Z^1, \dots, Z^{n_2})$  is a family of  $n_2$  i.i.d. copies of a random vector  $Z$  in  $\mathbb{R}^m$  ( $n_1 + n_2 = n$ ). Following the above microarray example,  $Y_i^j$  (resp.  $Z_i^j$ ),  $1 \leq i \leq m$ , corresponds to the expression level measure of the  $i$ -th gene for the  $j$ -th individual of the first (resp. second) experimental condition. Also, typically, the sample size is much smaller than the number of tests, that is,  $n \ll m$ . The distribution  $P$  of the observation  $X$  lies in a specific distribution set  $\mathcal{P}$  that constitutes the statistical model. Assume that  $\mathcal{P}$  is such that  $X$  is an integrable random vector and let  $\mu_{i,1}(P) = \mathbb{E}Y_i$  and  $\mu_{i,2}(P) = \mathbb{E}Z_i$ , for any  $i \in \{1, \dots, m\}$ . A way to investigate the problem of the previous section is to decide whether  $P$  belongs to the set  $\Theta_{0,i} = \{P \in \mathcal{P} : \mu_{i,1}(P) = \mu_{i,2}(P)\}$  or not, that is, we aim at testing test

$$H_{0,i} : “\mu_{i,1}(P) = \mu_{i,2}(P)” \text{ against } H_{1,i} : “\mu_{i,1}(P) \neq \mu_{i,2}(P)” ,$$

simultaneously for all  $i \in \{1, \dots, m\}$ . Given  $P$ , a null hypothesis  $H_{0,i}$  (called sometimes a “null” for short) is said to be true (for  $P$ ) if  $P \in \Theta_{0,i}$ , that is, if  $P$  satisfies  $H_{0,i}$ . It is said false (for  $P$ ) otherwise. The index set corresponding to true nulls is denoted by  $\mathcal{H}_0(P) = \{1 \leq i \leq m : \mu_{i,1}(P) = \mu_{i,2}(P)\}$ . Its complementary is denoted by  $\mathcal{H}_1(P)$ . In the above microarray example,  $\mathcal{H}_1(P) = \{1 \leq i \leq m : \mu_{i,1}(P) \neq \mu_{i,2}(P)\}$  is thus the index set corresponding to differentially expressed genes. The aim of a multiple testing procedure is to recover the set  $\mathcal{H}_1(P)$  from the observation  $X$ .

A multiple testing procedure is commonly defined from individual test statistics, by rejecting the null hypotheses with a large test statistic. Here, each individual test can be for instance built upon a (two-sided) two-sample t-statistic  $S_i(X) \propto |\bar{Y}_i - \bar{Z}_i|$ , rescaled using a standard deviation estimate. In order to provide a uniform normalization for all the tests, it is convenient to normalize each  $S_i(X)$  into a variable called the *p-value*

$$p_i(X) = T_{0,i}(S_i(X)), \tag{1}$$

where  $T_{0,i}(s) = \mathbb{P}_{H_{0,i}}(S_i(X) \geq s)$  is the upper-tail distribution function of the statistics under the null (we have chosen “ $\geq$ ” in  $T_{0,i}$  to use Lemma 1.3, see below). Since  $T_{0,i}$  is non-increasing, a multiple testing procedure rejects the null associated to small  $p$ -values. The precise relation between “ $S_i(X)$  large” and “ $p_i(X)$  small” easily derives from Lemma 1.3.

The calculation of the  $p$ -values (1) requires the knowledge of the  $(T_{0,i})_{1 \leq i \leq m}$ , that is, of the distribution of the test statistics under the null. For instance, the latter can be done simply if the model  $\mathcal{P}$  is such that  $Y_i$  and  $Z_i$  are Gaussian variables with the same variance. Namely, if  $S_i(X)$  is rescaled with the so-called “pooled” standard deviation, the null distribution is the Student distribution of  $n - 2$  degrees of freedom. In that case, each  $p$ -value (1) has the remarkable property to be uniformly distributed on  $(0, 1)$  when the corresponding null hypothesis is true. More generally, when  $S_i(X)$  may have a discrete distribution under the null, these  $p$ -values have the slightly weaker property to be stochastically lower-bounded by a uniform variable, that is,  $\forall i \in \mathcal{H}_0(P), \forall u \in [0, 1], \mathbb{P}(p_i(X) \leq u) \leq u$ , see Lemma 1.3. The latter weaker property also appears in the case of one-sided tests (see Example 1.2 below). Finally, under no distributional assumption on the  $Y_i$  and  $Z_i$  under the null, the function  $(T_{0,i})_{1 \leq i \leq m}$  are unknown and thus the  $p$ -values (1) cannot be directly computed. In that case, they can still be approached by using a Fisher permutation test [20], which uses permutations of the group labels. Conveniently, the so-derived  $p$ -values can be proved to enjoy the same stochastic dominance as above (see, e.g., [34] for a recent reference).

**Example 1.2** (One-sided testing on the mean of a Gaussian vector). A very convenient mathematical framework to do multiple testing is the case where we observe a  $\mathbb{R}^m$  Gaussian vector  $X = (X_i)_{1 \leq i \leq m} \sim P$ , having an unknown mean  $\mu(P) = (\mu_i(P))_{1 \leq i \leq m} \in \mathbb{R}^m$  and a covariance  $m \times m$  matrix  $\Sigma(P)$  with entries on the diagonal equal to 1, and we aim at testing

$$H_{0,i} : “\mu_i(P) \leq 0” \text{ against } H_{1,i} : “\mu_i(P) > 0”,$$

simultaneously for all  $i \in \{1, \dots, m\}$ . Then we may define the  $p$ -values  $p_i = \bar{\Phi}(X_i)$ , where  $\Phi(x) = \mathbb{P}(Z \geq x)$  for  $Z \sim \mathcal{N}(0, 1)$ , which fulfilled the required stochastic domination under the null: if  $\mu_i(P) \leq 0$ , for  $u \in [0, 1]$ ,

$$\mathbb{P}(p_i(X) \leq u) \leq \mathbb{P}(\bar{\Phi}(X_i - \mu_i(P)) \leq u) = u.$$

Sometimes, a convenient assumption is to assume that alternative means are all equal to  $\tau\sqrt{n}$ , where  $\tau$  is a positive parameter and  $n$  is a “sample-size” parameter. Finally, we may tune the parameter  $\Sigma(P)$  to cover many  $p$ -value dependency structures: for instance, we can assume that  $\Sigma(P)$  has only non-negative entries, that the non-diagonal entries of  $\Sigma(P)$  are equal (equi-correlation), or that  $\Sigma(P)$  is diagonal.

**Lemma 1.3.** *Let  $Z$  be a random variable with upper-tail distribution function  $T(z) = \mathbb{P}(Z \geq z)$  and consider the random variable  $U = T(Z)$ . Then, for any  $u \in [0, 1]$ , we have:*

$$\{U \leq u\} = \begin{cases} \{Z \geq F^{-1}(1 - u)\} & \text{if } F(F^{-1}(1 - u)) - \delta_u \geq 1 - u \\ \{Z > F^{-1}(1 - u)\} & \text{otherwise} \end{cases},$$

where  $F$  is the c.d.f. of  $Z$ ,  $F^{-1}(v) = \min\{z \in \mathbb{R} : F(z) \geq v\}$  is the generalized inverse of  $F$  and  $\delta_u$  denotes  $\mathbb{P}(Z = u)$  for any  $u \in [0, 1]$ . As a consequence,  $U$  is such that  $\mathbb{P}(U \leq u) \leq u$ , for any  $u \in [0, 1]$ . Additionally, in the particular case where  $T$  is continuous,  $U$  has a uniform distribution on  $(0, 1)$  and  $\{U \leq u\} = \{Z \geq F^{-1}(1 - u)\}$ .

### 1.3. General multiple testing setting

We provide in this section the classical abstract framework in which the multiple testing theory may be investigated in broad generality.

Let us consider a statistical model, defined by a measurable space  $(\mathcal{X}, \mathfrak{X})$  endowed with a subset  $\mathcal{P}$  of distributions on  $(\mathcal{X}, \mathfrak{X})$ . Consider an observation  $X$  of the model, that is, a random variable with values in  $\mathcal{X}$  and with a distribution  $P \in \mathcal{P}$ . Consider a family  $(\Theta_{0,i})_{1 \leq i \leq m}$  of  $m \geq 2$  subsets of  $\mathcal{P}$ . From  $X$ , we want to test the null hypotheses  $H_{0,i} : “P \in \Theta_{0,i}”$  against the alternative  $H_{1,i} : “P \in \Theta_{0,i}^c”$ , simultaneously for all  $i \in \{1, \dots, m\}$ . For any  $P \in \mathcal{P}$ , let  $\mathcal{H}_0(P) = \{1 \leq i \leq m : P \in \Theta_{0,i}\}$  the set of the indexes  $i$  for which  $P$  satisfies  $H_{0,i}$ , that is, corresponding to true null hypotheses. Its cardinal  $|\mathcal{H}_0(P)|$  is denoted by  $m_0(P)$ . Similarly, the set  $\{1, \dots, m\}$  will be sometimes denoted by  $\mathcal{H}$ . The set of the false null hypotheses is denoted by  $\mathcal{H}_1(P) = \mathcal{H} \setminus \mathcal{H}_0(P)$ . The goal is to recover the set  $\mathcal{H}_1(P)$  from  $X$ , that is, to find which null hypotheses are true/false from the knowledge of  $X$ . Of course, the distribution  $P$  of  $X$  is unknown, and thus so is  $\mathcal{H}_1(P)$ .

A standard multiple testing setting includes the knowledge of  $p$ -values  $(p_i(X))_{1 \leq i \leq m}$  satisfying

$$\forall P \in \mathcal{P}, \forall i \in \mathcal{H}_0(P), \quad \forall u \in [0, 1], \quad \mathbb{P}(p_i(X) \leq u) \leq u. \quad (2)$$

As a consequence, for each  $i \in \{1, \dots, m\}$ , rejecting  $H_{0,i}$  whenever  $p_i(X) \leq \alpha$  defines a test of level  $\alpha$ . As we already mentioned, in some cases, the  $p$ -value  $p_i(X)$  is exactly distributed like a uniform variable in  $(0, 1)$  under  $H_{0,i}$ , in which case (2) is an equality.

### 1.4. Multiple testing procedures

In all what follows, we will use the observation  $X$  only through the  $p$ -value family  $\mathbf{p}(X) = \{p_i(X), 1 \leq i \leq m\}$ . Therefore, for short, we will often drop the dependence in  $X$  in the notation and define all quantities as functions of  $\mathbf{p} = \{p_i, 1 \leq i \leq m\} \in [0, 1]^m$ . However, one should keep in mind that the underlying distribution  $P$  (the distribution of interest on which the tests are performed) is the distribution of  $X$  and not of  $\mathbf{p}$ .

A *multiple testing procedure* is defined as a function

$$R : q = (q_i)_{1 \leq i \leq m} \in [0, 1]^m \mapsto R(q) \subset \{1, \dots, m\},$$

taking as input an element of  $[0, 1]^m$  and returning as output a subset of  $\{1, \dots, m\}$ . For such a general procedure  $R$ , we add the technical assumption that for each  $i \in \{1, \dots, m\}$ , the mapping  $x \in \mathcal{X} \mapsto \mathbf{1}\{i \in R(\mathbf{p}(x))\}$  is measurable. The coordinates selected by  $R(\mathbf{p})$  correspond to the rejected null hypotheses, that is,  $i \in R(\mathbf{p}) \Leftrightarrow "H_{0,i} \text{ is rejected by the procedure } R(\mathbf{p})"$ . Thus, for each  $p$ -value set  $\mathbf{p}$ , there are  $2^m$  possible choices for  $R(\mathbf{p})$ . This number becomes huge as the number  $m$  of hypotheses grows. However, according to the stochastic property (2) of the  $p$ -values, a natural rejection region for each  $H_{0,i}$  is of the form  $p_i \leq u_i$ , for some  $u_i \in [0, 1]$ . In this paper, we will mainly focus on the case where the threshold is the same for all the  $p$ -values. The corresponding procedures, called *thresholding based procedures*, are of the form  $R(\mathbf{p}) = \{1 \leq i \leq m : p_i \leq t(\mathbf{p})\}$ , where the threshold  $t(\cdot) \in [0, 1]$  may depend on the data (in a measurable manner).

### 1.5. Type I error rates

In order to evaluate the quality of a multiple testing procedures, many error rates have been proposed in the multiple testing literature. According to the Neyman-Pearson approach, of primary interest are the type I error rates, that is, rates that evaluate the importance of true null hypotheses (wrongly) rejected. These are the elements of the set  $R(\mathbf{p}) \cap \mathcal{H}_0(P)$ . In this paper, we will use the following type I error rates, which are the most used nowadays: for a given procedure  $R$ ,

- the *k-family-wise error rate* ( $k$ -FWER) of  $R$  (see e.g. [26, 34]) is defined as the probability that the procedure  $R$  makes at least  $k$  false rejections:  $\forall P \in \mathcal{P}$ ,

$$k\text{-FWER}(R, P) = \mathbb{P}(|R(\mathbf{p}) \cap \mathcal{H}_0(P)| \geq k), \quad (3)$$

where  $k \in \{1, \dots, m\}$  is a pre-specified tuning parameter. In the particular case where  $k = 1$ , the latter is simply called the *family-wise error rate* of  $R$  and is denoted by  $\text{FWER}(R, P)$ .

- the *false discovery proportion* (FDP) (see e.g. [42, 4]) of  $R$  is defined as the proportion of errors in the set of the rejected hypotheses:  $\forall P \in \mathcal{P}$ ,

$$\text{FDP}(R(\mathbf{p}), P) = \frac{|R(\mathbf{p}) \cap \mathcal{H}_0(P)|}{|R(\mathbf{p})| \vee 1}. \quad (4)$$

The “ $\vee 1$ ” in the denominator is only present to make the FDP equals to zero when  $R$  makes no rejection. Since the above quantity is random, this does not define an

error rate. Two main error rates have been derived from the FDP. First, the  $\gamma$ -upper-tail distribution of the FDP of  $R$ , defined as the probability that the FDP exceeds  $\gamma$ :  $\forall P \in \mathcal{P}$ ,

$$\mathbb{P}(\text{FDP}(R(\mathbf{p}), P) > \gamma), \quad (5)$$

where  $\gamma \in (0, 1)$  is a pre-specified tuning parameter. Second, the false discovery rate (FDR) [4], defined as the expectation of the FDP of  $R$ :  $\forall P \in \mathcal{P}$ ,

$$\text{FDR}(R, P) = \mathbb{E}[\text{FDP}(R(\mathbf{p}), P)] = \mathbb{E}\left[\frac{|R(\mathbf{p}) \cap \mathcal{H}_0(P)|}{|R(\mathbf{p})| \vee 1}\right]. \quad (6)$$

Note that (5) is equivalent to upper-bound the  $(1 - \alpha)$ -quantile of the FDP distribution by  $\gamma$ . For instance, if the probability in (5) is upper-bounded by  $\alpha = 1/2$ , this means that the median of the FDP is upper-bounded by  $\gamma$ . With some abuse, controlling the probability (5) is called “controlling the FDP” from now on.

Choosing the most relevant type I error rate depends on the context. When controlling (3), we tolerate a fixed number  $k - 1$  of erroneous rejections. By contrast, controlling (5) allows to increase the number of erroneous rejections as the number of rejections becomes large, that is, it automatically selects  $k$  “close” to  $\gamma|R|$  in order to tolerate a small proportion  $\gamma$  of errors among the final rejections. Controlling (6) has become popular because it is a simple error rate based on the FDP and because it came together with the simple Benjamini-Hochberg FDR controlling procedure [4] (some dependency structure assumptions are required, see Section 3). As a counterpart, controlling the FDR does not prevent the FDP from having large variations, so that any FDR control does not have necessarily a clear interpretation in terms of the FDP (see the related discussion in Section 6.2).

### 1.6. Goal

Let  $\alpha \in (0, 1)$  be a pre-specified nominal level (to be fixed once for all throughout the paper). We want to control each of the type I error rates defined above at level  $\alpha$ , for a subset of distributions  $\mathcal{P}' \subset \mathcal{P}$  as large as possible. That is, taking one error rate  $\mathcal{E}(R, P)$  among those defined in the last section, we aim at finding a procedure  $R$  such that

$$\forall P \in \mathcal{P}', \quad \mathcal{E}(R, P) \leq \alpha, \quad (7)$$

for  $\mathcal{P}' \subset \mathcal{P}$  as large as possible. Of course,  $R$  should depend on  $\alpha$  but we omit this in the notation for short. Similarly to the single testing case, taking  $R = \emptyset$  will always ensure (7) with  $\mathcal{P}' = \mathcal{P}$ . This means that the type I error rate control is inseparable from the power maximization issue. Probably, the most natural way to extend the notion of power from the single testing to the multiple testing setting is to consider the expected number of correct rejections, that is,  $\mathbb{E}|\mathcal{H}_1(P) \cap R|$ . Along the paper, we will often meet the case where two procedures  $R$  and  $R'$  satisfy  $R' \subset R$  (almost surely) while they both ensure the control (7). In this situation, the procedure  $R$  is said *less conservative* than  $R'$ . This obviously implies that  $R$  is more powerful than  $R'$ . For instance,  $R$  and  $R'$  can be thresholding-based procedures using respective thresholds  $t$  and  $t'$  satisfying  $t \geq t'$  (almost surely). Hence, our goal is to find  $R$  providing (7), with a rejection set  $R$  as large as possible.

Finally, we emphasize that we want a control (7) valid for any fixed  $m \geq 2$  and not only when  $m$  tends to infinity. That is, we consider a setting which is non-asymptotic in the parameter  $m$ .



### 1.7. Presentation of the paper

In Section 2, we present some general tools and concepts that will be used throughout the paper. Section 3, 4 and 5 present surveys on FDR,  $k$ -FWER and FDP controlling methodology, respectively.

Moreover, while this paper is mostly intended to be a review paper, some new contributions with respect to the existing multiple testing literature are given in Section 4 and 5, by extending the results of [24] for the  $k$ -FWER control and the results of [35] for the FDP control, respectively.

## 2. Key concepts and tools

### 2.1. Model assumptions

Throughout this paper, we will consider several (sub)models. Each model corresponding to a specific assumption on the  $p$ -value family  $\mathbf{p} = \{p_i, 1 \leq i \leq m\}$  distribution. The first model, called the “independent model” is defined as follows:

$$\mathcal{P}^I = \left\{ P \in \mathcal{P} : (p_i(X))_{i \in \mathcal{H}_0(P)} \text{ is a family of mutually independent variables, and } (p_i(X))_{i \in \mathcal{H}_0(P)} \text{ is independent of } (p_i(X))_{i \in \mathcal{H}_1(P)} \right\}. \quad (8)$$

The second model uses a particular notion of positive dependence between the  $p$ -values, called “weak positive regression dependency” (in short, “weak PRDS”), which is a slightly weaker version of the PRDS assumption of [7], defined as follows: first, a subset  $D \subset [0, 1]^m$  is called *nondecreasing* if for all  $q, q' \in [0, 1]^m$  such that  $\forall i \in \{1, \dots, m\}, q_i \leq q'_i$ , we have  $q \in D \Rightarrow q' \in D$ .

**Definition 2.1** (Weak PRDS  $p$ -value family). *The family  $\mathbf{p}$  is said to be weak PRDS on  $\mathcal{H}_0(P)$  if for any  $i_0 \in \mathcal{H}_0(P)$  and for any measurable nondecreasing set  $D \subset [0, 1]^m$ , the function  $u \mapsto \mathbb{P}(\mathbf{p} \in D \mid p_{i_0} \leq u)$  is nondecreasing on the set  $\{u \in [0, 1] : \mathbb{P}(p_{i_0} \leq u) > 0\}$ .*

The only difference between the weak PRDS assumption and the “regular” PRDS assumption defined in [7] is that the latter assumes “ $u \mapsto \mathbb{P}(\mathbf{p} \in D \mid p_{i_0} = u)$  nondecreasing”, instead of “ $u \mapsto \mathbb{P}(\mathbf{p} \in D \mid p_{i_0} \leq u)$  nondecreasing”. Hence, weak PRDS is a weaker assumption, as shown for instance in the proof of Proposition 3.6 in [11]. We can now define the second “weak PRDS model”, where the  $p$ -values have positive dependencies:

$$\mathcal{P}^{pos} = \left\{ P \in \mathcal{P} : \mathbf{p}(X) \text{ is weak PRDS on } \mathcal{H}_0(P) \right\}. \quad (9)$$

It is not difficult to see that  $\mathcal{P}^I \subset \mathcal{P}^{pos}$  because when  $P \in \mathcal{P}^I$ ,  $p_{i_0}$  is independent of  $(p_i)_{i \neq i_0}$  for any  $i_0 \in \mathcal{H}_0(P)$ . We refer to the general case of  $P \in \mathcal{P}$  (without any additional restriction) as the “arbitrary dependence case”.

As an illustration, in the one-sided Gaussian testing framework of Example 1.2, the PRDS assumption (regular and thus also weak) is satisfied as soon as the covariance matrix  $\Sigma(P)$  has nonnegative entries, as shown in [7] (note that this is not true anymore for two-sided tests, as proved in the latter reference).



## 2.2. Dirac configurations

If we want to check whether a procedure satisfies (7), a particularly simple  $p$ -value distribution (called often “configuration”) that one can consider is the so-called “Dirac-uniform configuration” (see [19]), in which the variables  $(p_i)_{i \in \mathcal{H}_0(P)}$  are i.i.d. uniform and the other  $p$ -values are all identically equal to zero. Similarly, any configuration where the  $p$ -values of  $\mathcal{H}_1(P)$  are equal to zero (without any assumption on the  $p$ -values of  $\mathcal{H}_0(P)$ ) is called here simply a “Dirac configuration”.

These configurations can be seen as the asymptotic  $p$ -value family distribution where the sample size available to perform each test tends to infinity, while the number  $m$  of tests is kept fixed (see the examples of Section 1.2). This situation does not fall into the classical multiple testing framework where the number of tests is much larger than the sample size. Besides, there is no multiple testing problem in these configurations because the true nulls are perfectly separated from the false null (almost surely). However, our point is that these special configurations are still interesting, because they sometimes have the property to be the distributions for which the type I error rate is the largest. In that case, they are called the “least favorable configurations” (see [19]). This generally requires that the multiple testing procedure and the error rate under consideration have special monotonous properties (see [18, 37]). In this case, proving the type I error rate control for the Dirac configurations is sufficient to state (7) and thus appears to be very useful.

## 2.3. Algorithms

In order to derive (7), a generic method that emerged from the multiple testing literature is as follows:

1. start with a family  $(R_\kappa)_\kappa$  of procedures depending on an external parameter  $\kappa$ ;
2. find a set of values for  $\kappa$  for which  $R_\kappa$  satisfies (7);
3. take among these values the  $\kappa$  that makes  $R_\kappa$  the “largest”.

The latter is designed to maintain the control of the type I error rate while maximizing the rejection set. As we will see in Section 3 ( $\kappa$  is a threshold  $t$ ), Section 4 ( $\kappa$  is a subset  $\mathcal{C}$  of  $\mathcal{H}$ ) and Section 5 ( $\kappa$  is a rejection number  $\ell$ ), this gives rise to the so-called “step-up” and “step-down” algorithms, which are very classical instances of type I error rate controlling procedures.

## 2.4. Adaptive control

A way to increase the power of type I error rate controlling procedures is to learn (from the data) part of the unknown distribution  $P$  in order to make more rejections. This approach is called “adaptive type I error rate control”. Since the resulting procedure uses the data twice, the main challenge is often to show that it maintains the type I error control (7). In this paper, we will discuss the adaptivity with respect to the parameter  $|\mathcal{H}_0(P)|$  in Section 3.3 (for  $\mathcal{E} = \text{FDR}$ ) and to both  $|\mathcal{H}_0(P)|$  and the dependency structure between the  $p$ -values in Section 4 (for  $\mathcal{E} = k\text{-FWER}$ ). Let us finally note that some other work studied the adaptivity to the alternative distributions of the  $p$ -values (see [46, 38, 36]).

### 3. FDR control

After the seminal work of Benjamini and Hochberg [4], many studies have investigated the FDR controlling issue. We provide in this section a survey of some of these approaches.

#### 3.1. Thresholding based procedures

Let us start from thresholding type multiple-testing procedures

$$R_t = \{1 \leq i \leq m : p_i \leq t(\mathbf{p})\},$$

with a threshold  $t(\cdot) \in [0, 1]$  possibly depending on the  $p$ -values. We want to find  $t$  such that the corresponding multiple testing procedure  $R_t$  controls the FDR at level  $\alpha$  under the model  $\mathcal{P}^{pos}$ , following the general method explained in Section 2.3. We start with the following simple decomposition of the false discovery rate of  $R_t$  that follows from (6):

$$\text{FDR}(R_t, P) = \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ \frac{\mathbf{1}\{p_i \leq t(\mathbf{p})\}}{\alpha \widehat{\mathbb{G}}(\mathbf{p}, t(\mathbf{p})) \vee (\alpha/m)} \right], \quad (10)$$

where  $\widehat{\mathbb{G}}(\mathbf{p}, u) = m^{-1} \sum_{i=1}^m \mathbf{1}\{p_i \leq u\}$  is the empirical c.d.f. of the  $p$ -value family  $\mathbf{p} = \{p_i, 1 \leq i \leq m\}$  taken at a threshold  $u \in [0, 1]$ .

In order to create a link between the numerator and the denominator inside the expectation of (10), we consider the so-called “self-consistent” threshold set:

$$\mathcal{T}(\mathbf{p}) = \{u \in [0, 1] : \widehat{\mathbb{G}}(\mathbf{p}, u) \geq u/\alpha\}. \quad (11)$$

With different notation, the latter was introduced in [11, 18]. Now, let us choose a self-consistent threshold  $t(\mathbf{p}) \in \mathcal{T}(\mathbf{p})$  (almost surely). By using the decomposition (10), we obtain the following upper-bound:

$$\text{FDR}(R_t, P) \leq \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ \frac{\mathbf{1}\{p_i \leq t(\mathbf{p})\}}{t(\mathbf{p})} \right], \quad (12)$$

with the convention  $\frac{0}{0} = 0$ . Since by (2),  $p_i(x) > 0$  for  $P$ -almost every  $x$  when  $i \in \mathcal{H}_0(P)$ , the denominator inside the expectation of the RHS of (12) can only be zero when the numerator is also zero and therefore when the ratio is zero. Next, the following purely probabilistic lemma holds (see a proof in Appendix A of [11] for instance):

**Lemma 3.1.** *Let  $U$  be a nonnegative random variable which is stochastically lower bounded by a uniform distribution, i.e.,  $\mathbb{P}(U \leq u) \leq u$  for any  $u \in [0, 1]$ . Then the following inequality holds:*

$$\mathbb{E} \left[ \frac{\mathbf{1}\{U \leq V\}}{V} \right] \leq 1, \quad (13)$$

for any nonnegative random variable  $V$  satisfying either of the two following conditions:

- (i)  $V = g(U)$  where  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is non-increasing,
- (ii) the conditional distribution of  $V$  conditionally on  $U \leq u$  is stochastically decreasing in  $u$ , that is,  $\forall v \geq 0, u \mapsto \mathbb{P}(V < v \mid U \leq u)$  is nondecreasing on  $\{u \in [0, 1] : \mathbb{P}(U \leq u) > 0\}$ .

A consequence of the previous lemma in combination with (12) is that the FDR is controlled at level  $\alpha m_0(P)/m$  as soon as  $V = t(\mathbf{p})$  satisfies (ii) with  $U = p_i$ . For the latter to be true, we should make the distributional assumption  $P \in \mathcal{P}^{pos}$  and add the assumption that the threshold  $t(\cdot)$  is non-increasing with respect to each  $p$ -value, that is, for all  $q, q' \in [0, 1]^m$ , we have  $t(q) \leq t(q')$  as soon as for all  $1 \leq i \leq m$ ,  $q'_i \leq q_i$ . By using the latter, we easily check that the set

$$D = \{q \in [0, 1]^m : t(q) < v\}$$

is a nondecreasing measurable set of  $[0, 1]^m$ , for any  $v \geq 0$ . Thus, the weak PRDS condition defined in Section 2.1 provides (ii) with  $U = p_i$  and  $V = t(\mathbf{p})$  and thus also (13). Summing up, we obtained the following result, which appeared in [11]:

**Theorem 3.2.** *Consider a thresholding type multiple testing procedure  $R_t$  based on a threshold  $t(\cdot)$  which is self-consistent, i.e., such that  $t(\mathbf{p}) \in \mathcal{T}(\mathbf{p})$  (almost surely) (see (11)) and non-increasing in each  $p$ -value, i.e., satisfying that for all  $q, q' \in [0, 1]^m$  with  $q'_i \leq q_i$  for all  $1 \leq i \leq m$ , we have  $t(q) \leq t(q')$ . Then, for any  $P \in \mathcal{P}^{pos}$ ,  $FDR(R_t, P) \leq \alpha m_0(P)/m \leq \alpha$ .*

It is important to note that the set  $\mathcal{T}(\mathbf{p})$  only depends on the  $p$ -value family (and on  $\alpha$ ) so that self-consistent thresholds can be easily chosen in practice. As an illustration, we depict the set  $\mathcal{T}(\mathbf{p})$  in Figure 1 for a particular realization of the  $p$ -value family.

**Remark 3.3.** If we want to state the FDR control of Theorem 3.2 only for  $P \in \mathcal{P}^I$  without using the PRDS property, we can use Lemma 3.1 (i) *conditionally on*  $\mathbf{p}_{-i} = (p_j, j \neq i) \in [0, 1]^{m-1}$ , by taking  $V = t(U, \mathbf{p}_{-i})$  and  $U = p_i$ , because  $p_i$  is independent of  $\mathbf{p}_{-i}$  when  $P \in \mathcal{P}^I$ .

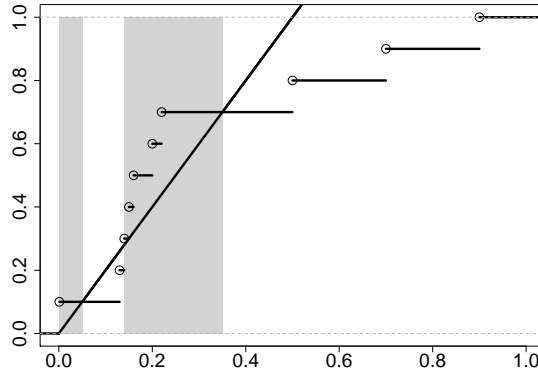


FIG 1. The  $p$ -value e.c.d.f.  $\hat{G}(\mathbf{p}, u)$  and  $u/\alpha$  are plotted as functions of  $u \in [0, 1]$ . The points  $u$  belonging to the set  $\mathcal{T}(\mathbf{p})$  lie on the X-axis of the gray area.  $m = 10$ ;  $\alpha = 0.5$ .

### 3.2. Linear step-up procedures

From Theorem 3.2, under the weak PRDS assumption on the  $p$ -value dependence structure, any algorithm giving as output a self-consistent and non-increasing threshold  $t(\mathbf{p})$  leads to a correct FDR control. Therefore, we may ask which one should we choose.

As explained in Section 1.6 and Section 2.3, for the same FDR control we want to get a procedure with a rejection set as large as possible. Hence, it is natural to choose the following threshold:

$$t^{su}(\mathbf{p}) = \max\{\mathcal{T}(\mathbf{p})\} \quad (14)$$

$$\begin{aligned} &= \max\{u \in \{\alpha k/m, 0 \leq k \leq m\} : \widehat{\mathbb{G}}(\mathbf{p}, u) \geq u/\alpha\} \\ &= \alpha/m \times \max\{0 \leq k \leq m : p_{(k)} \leq \alpha k/m\}, \end{aligned} \quad (15)$$

where  $p_{(1)} \leq \dots \leq p_{(m)}$  ( $p_{(0)} = 0$ ) denote the order statistics of the  $p$ -value family. This choice was made in [4] and is usually called *linear step-up* or “Benjamini-Hochberg” thresholding. One should notice that the maximum in (14) exists because the set  $\mathcal{T}(\mathbf{p})$  contains 0, is upper-bounded by 1 and because the e.c.d.f. is a non-decreasing function. It is also easy to check that the maximum  $u = t^{su}(\mathbf{p})$  satisfies the equality  $\widehat{\mathbb{G}}(\mathbf{p}, u) = u/\alpha$ , so that  $t^{su}(\mathbf{p})$  can be seen as the largest crossing point between  $u \mapsto \widehat{\mathbb{G}}(\mathbf{p}, u)$  and  $u \mapsto u/\alpha$ , see the left-side of Figure 2. It also implies that  $t^{su}(\mathbf{p}) \in \{\alpha k/m, 0 \leq k \leq m\}$ , which, combined with the so-called switching relation  $\widehat{\mathbb{G}}(\mathbf{p}, \alpha k/m) \geq k \iff p_{(k)} \leq \alpha k/m$ , gives rise to the second formulation (15). The latter is illustrated in the right-side of Figure 2. The formulation (15) corresponds to the original expression of [4] while (14) is to be found for instance in [21].

Moreover, it is worth noticing that the procedure  $R_{t^{su}}$  using the thresholding  $t^{su}(\mathbf{p})$  is also equal to  $\{1 \leq i \leq m : p_i \leq t^{su}(\mathbf{p}) \vee \alpha/m\}$ , so that it can be interpreted as an intermediate thresholding between the non-corrected procedure using  $t = \alpha$  and the Bonferroni procedure using  $t = \alpha/m$ .

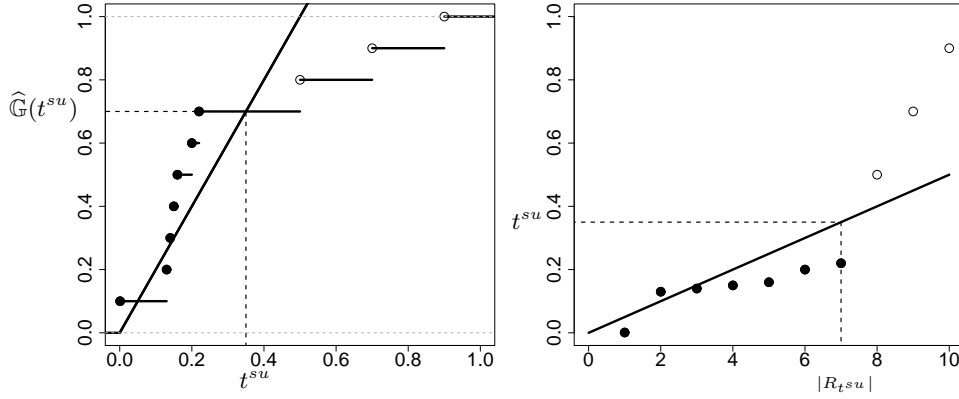


FIG 2. The two dual pictorial representations of the Benjamini-Hochberg linear step-up procedure. Left: c.d.f. of the  $p$ -values, the solid line has for slope  $\alpha^{-1}$ . Right: ordered  $p$ -values, the solid line has for slope  $\alpha/m$ . In both pictures, the filled points represent  $p$ -values that corresponds to the rejected hypotheses.  $m = 10$ ;  $\alpha = 0.5$ .

Since  $t^{su}(\mathbf{p})$  is obviously coordinate-wise non-increasing and self-consistent, Theorem 3.2 shows that, for any  $P \in \mathcal{P}^{pos}$ ,  $\text{FDR}(R_{t^{su}}, P) \leq \alpha m_0(P)/m$ . As a matter of fact, as soon as (2) holds with an equality, we can prove that for any  $P \in \mathcal{P}^I$ , the equality  $\text{FDR}(R_{t^{su}}, P) = \alpha m_0(P)/m$  holds, by using a surprisingly direct argument. Let  $\mathbf{p}_{0,-i}$  for the  $p$ -value family where  $p_i$  has been replaced by 0, and observe that the following statements are equivalent:

- (i)  $p_i \leq t^{su}(\mathbf{p}_{0,-i})$

- (ii)  $\widehat{\mathbb{G}}(\mathbf{p}_{0,-i}, t^{su}(\mathbf{p}_{0,-i})) \leq \widehat{\mathbb{G}}(\mathbf{p}, t^{su}(\mathbf{p}_{0,-i}))$
- (iii)  $t^{su}(\mathbf{p}_{0,-i})/\alpha \leq \widehat{\mathbb{G}}(\mathbf{p}, t^{su}(\mathbf{p}_{0,-i}))$
- (iv)  $t^{su}(\mathbf{p}_{0,-i}) \leq t^{su}(\mathbf{p})$ .

The equivalence between (i) and (ii) is straightforward from the definition of  $\widehat{\mathbb{G}}(\cdot, \cdot)$ . The equivalence between (ii) and (iii) follows from  $\widehat{\mathbb{G}}(\mathbf{p}_{0,-i}, t^{su}(\mathbf{p}_{0,-i})) = t^{su}(\mathbf{p}_{0,-i})/\alpha$ , because  $t = t^{su}(\mathbf{p}_{0,-i})$  is a crossing point between  $\widehat{\mathbb{G}}(\mathbf{p}_{0,-i}, t)$  and  $t/\alpha$ . The equivalence between (iii) and (iv) comes from the definition of  $t^{su}(\mathbf{p})$  together with  $t^{su}(\mathbf{p}_{0,-i}) \leq t^{su}(\mathbf{p}) \iff t^{su}(\mathbf{p}_{0,-i}) = t^{su}(\mathbf{p})$ , the latter coming from the non-increasing property of  $t^{su}(\cdot)$ . As a consequence,  $p_i \leq t^{su}(\mathbf{p}_{0,-i}) \iff p_i \leq t^{su}(\mathbf{p})$ , with  $t^{su}(\mathbf{p}_{0,-i}) = t^{su}(\mathbf{p})$  when these assertions are true. By using the first decomposition (10) of the FDR, we thus derive the following equalities:

$$\begin{aligned}
\text{FDR}(R_{t^{su}}, P) &= \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ \frac{\mathbf{1}\{p_i \leq t^{su}(\mathbf{p})\}}{\alpha \widehat{\mathbb{G}}(\mathbf{p}, t^{su}(\mathbf{p})) \vee (\alpha/m)} \right] \\
&= \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ \frac{\mathbf{1}\{p_i \leq t^{su}(\mathbf{p})\}}{t^{su}(\mathbf{p})} \right] \\
&= \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ \frac{\mathbf{1}\{p_i \leq t^{su}(\mathbf{p}_{0,-i})\}}{t^{su}(\mathbf{p}_{0,-i})} \right] \\
&= \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ t^{su}(\mathbf{p}_{0,-i})^{-1} \mathbb{E}(\mathbf{1}\{p_i \leq t^{su}(\mathbf{p}_{0,-i})\} | \mathbf{p}_{0,-i}) \right] \\
&= \alpha m_0(P)/m,
\end{aligned}$$

where we assumed in the last equality both that for all  $i \in \mathcal{H}_0(P)$ ,  $p_i$  is independent from  $\mathbf{p}_{-i}$ , that is,  $P \in \mathcal{P}^I$ , and that each  $p$ -value associated to a true null is uniform on  $[0, 1]$ , that is, condition (2) holds with an equality. To sum up, we have proved in this section the following result.

**Theorem 3.4.** *Consider the linear step-up procedure  $R_{t^{su}}$  using the threshold defined in (14). Then, for any  $P \in \mathcal{P}^{pos}$ ,  $\text{FDR}(R_{t^{su}}, P) \leq \alpha m_0(P)/m$ . Moreover, the latter is an equality if  $P$  lies in  $\mathcal{P}^I$  and (2) is an equality.*

This theorem is due to [4, 7]. The short proof mentioned above has been independently given in [17, 36, 18]. Theorem 3.4 proves that the inequality “ $\forall P \in \mathcal{P}^{pos}, \text{FDR}(R_{t^{su}}, P) \leq \alpha$ ” is sharp as soon as each  $p$ -value associated to a true null has a marginal distribution which is uniform on  $[0, 1]$  and there exists  $P \in \mathcal{P}^I$  such that  $\mathcal{H}_0(P) = \mathcal{H}$ , that is, when  $\cap_{i \in \mathcal{H}} \Theta_{0,i} \cap \mathcal{P}^I \neq \emptyset$ .

Other instances of self-consistent procedures include linear “step-up-down” procedures as defined in [39]. Theorem 3.2 establishes that the FDR control also holds for these procedures, as proved in [11, 18].

### 3.3. Adaptive linear step-up procedures

In this section we denote by  $\pi_0(P)$  the proportion  $m_0(P)/m$  of hypotheses that are true for  $P$ . Since we aim at controlling the FDR at level  $\alpha$  and not at level  $\alpha\pi_0(P)$ , Theorem 3.4 shows that there is a potential power loss when using  $t^{su}$  when the proportion  $\pi_0(P)$  is small. A first idea is to use the linear step-up procedure at level  $\alpha/\pi_0(P)$  which would provide a final FDR control at level  $\alpha\pi_0(P)^{-1}\pi_0(P) = \alpha$ . Unfortunately, since  $P$  is unknown, so is  $\pi_0(P)$

and an interesting challenge is to estimate  $\pi_0(P)$  within a multiple testing procedure while still rigorously controlling the FDR at level  $\alpha$ , in despite of the additional fluctuations added by the  $\pi_0(P)$ -estimation. This problem, called  $\pi_0(P)$ -adaptive FDR control, has received a growing attention in the last decade, see e.g. [5, 44, 8, 22, 6, 32, 40, 12].

To investigate this issue, a natural idea is to consider a modified linear step-procedure using the threshold

$$t_f^{su}(\mathbf{p}) = \max \{u \in [0, 1] : \widehat{\mathbb{G}}(\mathbf{p}, u) \geq u/(\alpha f(\mathbf{p}))\}. \quad (16)$$

where  $f(\mathbf{p}) > 0$  is an estimator of  $\pi_0^{-1}$  to be chosen. The latter is called *adaptive linear step-up procedure*. We describe now a way to choose  $f$  so that the FDR control  $\text{FDR}(R_{t_f^{su}}, P) \leq \alpha$  still holds for  $P \in \mathcal{P}^I$ . It uses techniques similar to those of Section 3.2. First, to keep the non-increasing property of the threshold  $t_f^{su}(\cdot)$ , we assume that  $f(\cdot)$  is coordinate-wise non-increasing. Second, using a decomposition of the FDR analogous to the non-adaptive case, we may write for any  $P \in \mathcal{P}^I$ ,

$$\begin{aligned} \text{FDR}(R_{t_f^{su}}, P) &= \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ \frac{\mathbf{1}\{p_i \leq t_f^{su}(\mathbf{p})\}}{t_f^{su}(\mathbf{p})} f(\mathbf{p}) \right] \\ &\leq \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ \frac{\mathbf{1}\{p_i \leq t_f^{su}(\mathbf{p})\}}{t_f^{su}(\mathbf{p})} f(\mathbf{p}_{0,-i}) \right] \\ &= \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ f(\mathbf{p}_{0,-i}) \mathbb{E} \left[ \frac{\mathbf{1}\{p_i \leq t_f^{su}(\mathbf{p})\}}{t_f^{su}(\mathbf{p})} \middle| \mathbf{p}_{0,-i} \right] \right] \\ &\leq \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} [f(\mathbf{p}_{0,-i})], \end{aligned} \quad (17)$$

where we used Lemma 13 (i) in the last inequality (conditionally on the  $p$ -values of  $(p_j, j \neq i)$ , because  $f$  is coordinate-wise non-increasing). Additionally assuming that  $f(\cdot)$  is permutation invariant, we can upper-bound the RHS of (17) by using the Dirac-uniform configuration because  $f(\cdot)$  is non-increasing. This gives rise to the following result.

**Theorem 3.5.** *Consider the adaptive linear step-up procedure  $R_{t_f^{su}}$  with a threshold defined in (16) using a  $\pi_0^{-1}$ -estimator  $f$  satisfying the following properties:*

- $f(\cdot)$  is coordinate-wise non-increasing, that is, for all  $q, q' \in [0, 1]^m$  with for all  $1 \leq i \leq m$ ,  $q'_i \leq q_i$ , we have  $f(q) \leq f(q')$ ;
- $f(\cdot)$  is permutation invariant, that is, for any permutation  $\sigma$  of  $\{1, \dots, m\}$ ,  $\forall q \in [0, 1]^m$ ,  $f(q_1, \dots, q_m) = f(q_{\sigma(1)}, \dots, q_{\sigma(m)})$ ;
- $f$  satisfies

$$\forall m_0 \in \{1, \dots, m\}, \quad \mathbb{E}_{\mathbf{p} \sim \text{DU}(m_0-1, m)}(f(\mathbf{p})) \leq m/m_0, \quad (18)$$

where  $\text{DU}(k, m)$  denotes the Dirac-uniform distribution on  $[0, 1]^m$  for which the  $k$  first coordinates are i.i.d. uniform on  $(0, 1)$  and the remaining coordinates are equal to 0.

Then, for any  $P \in \mathcal{P}^I$ ,  $\text{FDR}(R_{t_f^{su}}, P) \leq \alpha$ .

The method leading to the upper-bound (17) was investigated in [6] and described latter in detail in [12]. The simpler result presented in Theorem 3.5 appeared in [12].

Next, we should find a good estimator  $f$  of  $\pi_0^{-1}$ . This issue has an interest in its own right and many studies investigated it since the first attempt in [41] (see for instance the

references in [13]). Here, we only deal with this problem from the FDR control viewpoint, by providing two families of estimators that satisfy the assumptions of Theorem 3.5. First, define the “Storey-type” estimators, which are of the form

$$f_1(\mathbf{p}) = \frac{m(1 - \lambda)}{\sum_{i=1}^m \mathbf{1}\{p_i > \lambda\} + 1},$$

for  $\lambda \in (0, 1)$  ( $\lambda$  not depending on  $\mathbf{p}$ ). It is clearly non-increasing and permutation invariant. Moreover, we may check that  $f_1$  satisfies (18): for any  $m_0 \in \{1, \dots, m\}$ , considering  $(U_i)_{1 \leq i \leq m_0-1}$  i.i.d. uniform on  $(0, 1)$ ,

$$\mathbb{E}_{\mathbf{p} \sim DU(m_0-1, m)}(f_1(\mathbf{p})) = \frac{m}{m_0} \mathbb{E} \left[ \frac{m_0(1 - \lambda)}{\sum_{i=1}^{m_0-1} \mathbf{1}\{U_i > \lambda\} + 1} \right] \leq \frac{m}{m_0},$$

because for any  $k \geq 2$ ,  $q \in (0, 1)$  and for  $Y$  having a binomial distribution with parameters  $(k - 1, q)$ , we have  $\mathbb{E}((1 + Y)^{-1}) \leq (qk)^{-1}$ , as stated e.g. in [6]. This type of estimator has been introduced in [43] and proved to lead to a correct FDR control in [44, 6].

The second family of estimators satisfying the assumptions of Theorem 3.5 is the “quantile-type” family, for which the estimators may be seen as Storey-type estimators using  $\lambda = p_{(k_0)}$ ; the latter are of the form

$$f_2(\mathbf{p}) = \frac{m(1 - p_{(k_0)})}{m - k_0 + 1},$$

for  $k_0 \in \{1, \dots, m\}$  ( $k_0$  not depending on  $\mathbf{p}$ ). Straightforwardly,  $f_2(\cdot)$  is non-increasing and permutation-invariant. Additionally,  $f_2(\cdot)$  enjoys (18) because for any  $m_0 \in \{1, \dots, m\}$ , considering  $(U_i)_{1 \leq i \leq m_0-1}$  i.i.d. uniform on  $(0, 1)$  ordered as  $U_{(1)} \leq \dots \leq U_{(m_0-1)}$ ,

$$\begin{aligned} \mathbb{E}_{\mathbf{p} \sim DU(m_0-1, m)}(f_2(\mathbf{p})) &= \mathbb{E} \left[ \frac{m(1 - U_{(k_0-m+m_0-1)})}{m - k_0 + 1} \right] = \frac{m(1 - \mathbb{E}[U_{(k_0-m+m_0-1)}])}{m - k_0 + 1} \\ &= \frac{m(1 - (k_0 - m + m_0 - 1)_+/m_0)}{m - k_0 + 1} \leq \frac{m}{m_0}, \end{aligned}$$

by using the convention  $U_{(j)} = 0$  when  $j \leq 0$ . These quantile type estimators have been proved to lead to a correct FDR control in [6]. The above simple proof was given in [12].

Which choice should we make for  $\lambda$  or  $k_0$ ? Using extensive simulations (including other type of adaptive procedures), it was recommended in [12] to choose as estimator  $f_1$  with  $\lambda$  close to  $\alpha$ , because the corresponding procedure shows a “good” power under independence while it maintains a correct FDR control under positive dependencies (in the equi-correlated Gaussian one-sided model described in Example 1.2). Of course, a “dynamic” choice of  $\lambda$  (i.e. using the data) can increase the accuracy of the  $\pi_0^{-1}$  estimation and thus leads to a better procedure. However, proving that the corresponding FDR control remains valid in this case is an open issue to our knowledge. Also, outside the case of the particular equi-correlated Gaussian dependence structure, very little is known about adaptive FDR control.

### 3.4. Arbitrary dependencies

Many corrections of the linear step-up procedure are available to maintain the FDR control when the  $p$ -value family has arbitrary and unknown dependencies. We describe here the so-called “Occam’s hammer” approach presented in [10] which only uses Fubini’s theorem and



allows to recover the well-known “Benjamini-Yekutieli” correction [7]. Let us consider

$$t^{\beta su}(\mathbf{p}) = \max\{u \in [0, 1] : \widehat{\mathbb{G}}(\mathbf{p}, \beta(u)) \geq u/\alpha\}, \quad (19)$$

$$\begin{aligned} &= \max\{u \in \{\alpha k/m, 1 \leq k \leq m\} : \widehat{\mathbb{G}}(\mathbf{p}, \beta(u)) \geq u/\alpha\}, \\ &= \alpha/m \times \max\{0 \leq k \leq m : p_{(k)} \leq \beta(\alpha k/m)\}, \end{aligned} \quad (20)$$

for a non-decreasing function  $\beta : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ . Then the FDR of  $R_{\beta(t^{\beta su})}$  can be written as follows: for any  $P \in \mathcal{P}$ ,

$$\begin{aligned} &\text{FDR}(R_{\beta(t^{\beta su})}, P) \\ &= \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ \frac{\mathbf{1}\{p_i \leq \beta(t^{\beta su}(\mathbf{p}))\}}{t^{\beta su}(\mathbf{p})} \right] \\ &= \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{E} \left[ \mathbf{1}\{p_i \leq \beta(t^{\beta su}(\mathbf{p}))\} \int_0^{+\infty} u^{-2} \mathbf{1}\{t^{\beta su}(\mathbf{p}) \leq u\} du \right] \\ &= \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \int_0^{+\infty} u^{-2} \mathbb{E} [\mathbf{1}\{t^{\beta su}(\mathbf{p}) \leq u\} \mathbf{1}\{p_i \leq \beta(t^{\beta su}(\mathbf{p}))\}] du \\ &\leq \alpha m^{-1} \sum_{i \in \mathcal{H}_0(P)} \int_0^{+\infty} u^{-2} \mathbb{P}(p_i \leq \beta(u)) du \\ &= \alpha m_0(P)/m \int_0^{+\infty} u^{-2} \beta(u) du. \end{aligned} \quad (21)$$

Therefore, choosing any non-decreasing function  $\beta$  such that  $\int_0^{+\infty} u^{-2} \beta(u) du = 1$  provides a valid FDR control. This leads to the following result:

**Theorem 3.6.** *Consider a function  $\beta : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  of the following form: for all  $u \geq 0$ ,*

$$\beta(u) = \sum_{i: \alpha i/m \leq u} (\alpha i/m) \nu_i, \quad (22)$$

where the  $\nu_i$ s are nonnegative with  $\nu_1 + \dots + \nu_m = 1$ . Consider the step-up procedure  $R_{\beta(t^{\beta su})}$  using  $t^{\beta su}$  defined by (19). Then for any  $P \in \mathcal{P}$ ,  $\text{FDR}(R_{\beta(t^{\beta su})}, P) \leq \alpha m_0(P)/m$ .

Note that the function  $\beta$  defined by (22) takes the value  $(\alpha/m)\nu_1 + \dots + (\alpha i/m)\nu_i$  in each  $u = \alpha i/m$  and is constant on each interval  $(\alpha i/m, \alpha(i+1)/m)$  and after  $\alpha$ . Thus, it always satisfies that  $\beta(u) \leq u$ , for any  $u \geq 0$ . This means that the procedure  $R_{\beta(t^{\beta su})}$  rejects always less hypotheses than the linear step-up procedure  $R_{t^{su}}$ . Therefore, while  $R_{\beta(t^{\beta su})}$  provides a FDR control under no assumption about the  $p$ -value dependency structure, it is substantially more conservative than  $R_{t^{su}}$  under weak PRDS dependencies between the  $p$ -values.

As an illustration, taking  $\nu_i = i^{-1}\gamma^{-1}$  for  $\gamma = 1 + 1/2 + \dots + 1/m$ , we obtain  $\beta(\alpha i/m) = \gamma^{-1}\alpha i/m$ , which corresponds to the linear step-up procedure, except that the level  $\alpha$  has been divided by  $\gamma \simeq \log(m)$ . This is the so-called Benjamini-Yekutieli procedure proposed in [7]. Theorem 3.6 thus recovers Theorem 1.3 of [7].

We mention another example, maybe less classical, to illustrate the flexibility of the choice of  $\beta$  in Theorem 3.6; take  $\nu_{m/2} = 1$  (assuming that  $m/2$  is an integer) so that  $\beta(\alpha i/m) = (\alpha/2) \mathbf{1}\{i \geq m/2\}$ . In that case, the final procedure  $R_{\beta(t^{\beta su})}$  rejects the hypotheses corresponding to  $p$ -values smaller than  $\alpha/2$  if  $2p_{(m/2)} \leq \alpha$  and rejects no hypothesis otherwise.

Theorem 3.6 ensures that this (quite uncommon) procedure also controls the FDR, under no assumption on the model dependency. Many other examples are given in Section 4.2.1 of [11].

**Remark 3.7** (Sharpness of the bound in Theorem 3.6). In Lemma 3.1 (ii) of [30] (see also [25]), a specifically crafted  $p$ -value distribution was built on  $[0, 1]^m$  (depending on  $\beta$ ) for which the FDR of  $R_{\beta(t^{\beta su})}$  is equal to  $\alpha$  (and  $m_0(P) = m$ ). If the underlying model  $\mathcal{P}$  is such that  $(p_i(X))_{1 \leq i \leq m}$  may have this very specific distribution for some  $P \in \mathcal{P}$ , the inequality “ $P \in \mathcal{P}$ ,  $\text{FDR}(R_{\beta(t^{\beta su})}, P) \leq \alpha$ ” in Theorem 3.6 is sharp. However, the latter is rarely true for a “realistic” model  $\mathcal{P}$ , because the  $p$ -value distribution of [30] assumes quite unrealistic dependencies between the  $p$ -values. Related to that, several simulation experiments showed that the standard LSU procedure still provides a good FDR control under “realistic” dependencies, see e.g. [16, 29]. This means that the corrections defined in this section are generally very conservative for real-life data, because their actually achieved FDR is much smaller than  $\alpha m_0(P)/m$ .

Finally, let us underline the fact that, without additional studies, any FDR control valid under arbitrary dependency suffers from a lack of interpretability for the underlying FDP, as discussed in Section 6.2.

#### 4. $k$ -FWER control

The methodology presented in this section for controlling the  $k$ -FWER under arbitrary dependencies can probably be attributed to many authors, e.g. [27, 47, 34, 35]. Here, we opted for a general presentation which emphasizes the rationale of the mathematical argument. This approach has been sketched in the talk [9] and investigated more deeply in [24] where it is referred to as the “sequential rejection principle”. While the latter point of view allows to obtain elegant proofs, it is also useful for developing new and potentially more complex FWER controlling procedures, see [24, 23, 28]. This methodology has been initially developed for the FWER. We propose in Section 4.4 a new extension to the  $k$ -FWER.

In this section, for simplicity, we drop the explicit dependence of the multiple testing procedure  $R$  w.r.t.  $\mathbf{p}$  in the notation. The tuning parameter  $k$  is fixed in  $\{1, \dots, m\}$ .

##### 4.1. Subset-indexed family

As a starting point, we assume that there exists a subset-indexed family of multiple testing procedures  $\{R_{\mathcal{C}}\}_{\mathcal{C} \subset \mathcal{H}}$  satisfying the two following assumptions:

- $\mathcal{C} \mapsto R_{\mathcal{C}}$  is non-increasing, that is,

$$\forall \mathcal{C}, \mathcal{C}' \subset \mathcal{H} \text{ such that } \mathcal{C} \subset \mathcal{C}', \text{ we have } R_{\mathcal{C}'} \subset R_{\mathcal{C}}; \quad (\text{NI})$$

- $R_{\mathcal{C}}$  controls the  $k$ -FWER when  $\mathcal{C}$  is equal to the subset of true null hypotheses, that is,

$$\forall P \in \mathcal{P}, k\text{-FWER}(R_{\mathcal{H}_0(P)}, P) \leq \alpha. \quad (\text{FWC}_0)$$

A natural way of deriving such a family is to take a thresholding-based family of the form

$$R_{\mathcal{C}} = \{1 \leq i \leq m : p_i \leq t_{\mathcal{C}}\}, \quad (23)$$

where  $t_{\mathcal{C}} \in [0, 1]$  is a threshold which possibly depends on the data  $\mathbf{p}$ . Assumption (NI) then holds as soon as we take  $t_{\mathcal{C}}$  non-increasing in  $\mathcal{C}$  (if  $\mathcal{C} \subset \mathcal{C}'$  then  $t_{\mathcal{C}'} \leq t_{\mathcal{C}}$ ). However,  $t_{\mathcal{C}}$  should be carefully chosen in order to ensure (FWC<sub>0</sub>), as we discuss below.

A first instance of a thresholding-based family satisfying (NI)-(FWC<sub>0</sub>) is the “Bonferroni family” that chooses  $t_{\mathcal{C}} = 1 \wedge (\alpha k / |\mathcal{C}|)$ . Condition (FWC<sub>0</sub>) results from the Markov inequality:

$$\mathbb{P}(|\mathcal{H}_0(P) \cap R_{\mathcal{H}_0(P)}| \geq k) \leq k^{-1} \sum_{i \in \mathcal{H}_0(P)} \mathbb{P}(p_i \leq t_{\mathcal{H}_0(P)}) \leq |\mathcal{H}_0(P)| t_{\mathcal{H}_0(P)} / k \leq \alpha.$$

This family is not adaptive w.r.t. the dependence structure of the  $p$ -values. As an illustration, when the true  $p$ -values are all equal, say, to  $p_{i_0}$ ,  $i_0 \in \mathcal{H}_0(P)$ , we have

$$\mathbb{P}(|\mathcal{H}_0(P) \cap R_{\mathcal{H}_0(P)}| \geq k) = \mathbb{P}(|\mathcal{H}_0(P)| \mathbf{1}\{p_{i_0} \leq t_{\mathcal{H}_0(P)}\} \geq k) \leq t_{\mathcal{H}_0(P)}.$$

Thus, under this extreme dependency structure, the Bonferroni threshold  $1 \wedge (\alpha k / |\mathcal{C}|)$  can be replaced by  $\alpha$ . Since the only case which matters is  $|\mathcal{C}| \geq k$  (see Remark 4.1 below), there is a potential loss when using the Bonferroni family. In practice, the Bonferroni family is often used as a “benchmark family” for evaluating the performance of other families.

In order to improve on the Bonferroni family, one can try to choose a threshold  $t_{\mathcal{C}}$  that captures the dependencies between the  $p$ -values while it still satisfies (NI)-(FWC<sub>0</sub>). For this, first note that for  $R_{\mathcal{C}}$  defined by (23),

$$k\text{-FWER}(R_{\mathcal{C}}, P) = \mathbb{P}(\exists i_1, \dots, i_k \in \mathcal{H}_0(P) : p_i \leq t_{\mathcal{C}}) = \mathbb{P}(k\text{-min}\{p_i, i \in \mathcal{H}_0(P)\} \leq t_{\mathcal{C}}),$$

where  $k\text{-min}\{p_i, i \in \mathcal{H}_0(P)\}$  denotes the  $k$ -th smallest element of  $\{p_i, i \in \mathcal{H}_0(P)\}$ . Therefore, a natural choice for  $t_{\mathcal{C}}$  is the  $\alpha$ -quantile of the distribution of  $k\text{-min}\{p_i, i \in \mathcal{C}\}$ . However, the latter is generally unknown because the underlying distribution  $P$  is unknown. An idea is therefore to approach it by using a randomized thresholding procedure. Roughly speaking, this method uses that the null hypothesis is invariant under the action of a finite group of transformations of the original observation set  $\mathcal{X}$  onto itself (such a transformation can be for instance a permutation or a sign-flipping, see [34, 35, 1, 2]). For a recent and general description of this method, we refer the reader to Theorem 2 of [24]. Note that [24] has developed this method only for  $k = 1$ . However, it can be directly generalized to the case of  $k \geq 1$  (using their notation, “max” should be simply replaced by “k-max”). The resulting family (and thus also the derived  $k$ -FWER-controlling procedure) satisfies (NI)-(FWC<sub>0</sub>) while it is “adaptive” with respect to the  $p$ -value dependence structure, in the sense that  $t_{\mathcal{C}} = t_{\mathcal{C}}(\mathbf{p})$  implicitly takes into account the potential relations existing between the  $p$ -values.

**Remark 4.1.** Any subset-indexed family  $\{R_{\mathcal{C}}\}_{\mathcal{C} \subset \mathcal{H}}$  satisfying (NI)-(FWC<sub>0</sub>) can be modified in the following way: take  $\tilde{R}_{\mathcal{C}} = \mathcal{H}$  (reject all hypotheses) when  $|\mathcal{C}| < k$  and  $\tilde{R}_{\mathcal{C}} = R_{\mathcal{C}}$  otherwise. This maintains the conditions (NI)-(FWC<sub>0</sub>), because the  $k$ -FWER is always zero when  $|\mathcal{H}_0(P)| < k$ .

In what follows, we investigate the problem of the  $k$ -FWER control once we have fixed a subset-indexed family  $\{R_{\mathcal{C}}\}_{\mathcal{C} \subset \mathcal{H}}$  satisfying (NI)-(FWC<sub>0</sub>).

#### 4.2. Single-step method

From assumption (FWC<sub>0</sub>), the procedure  $R_{\mathcal{H}_0(P)}$  using  $\mathcal{C} = \mathcal{H}_0(P)$  controls the  $k$ -FWER. Of course, this procedure cannot be used because  $\mathcal{H}_0(P)$  depends on the unknown underlying

distribution  $P$  of the data. We can use instead  $R_{\mathcal{C}}$  with  $\mathcal{C} = \mathcal{H}$  because, from the two assumptions (NI)-(FWC<sub>0</sub>) above, we have  $k\text{-FWER}(R_{\mathcal{H}}, P) \leq k\text{-FWER}(R_{\mathcal{H}_0(P)}, P) \leq \alpha$ . This implies that  $R_{\mathcal{H}}$  always controls the  $k\text{-FWER}$  at level  $\alpha$ . The latter is generally called the *single-step* procedure (associated to the family  $\{R_{\mathcal{C}}\}_{\mathcal{C} \subset \mathcal{H}}$ ). However, we argue that  $R_{\mathcal{H}}$  could be often too conservative w.r.t.  $R_{\mathcal{H}_0(P)}$ , for the two following reasons:

- $\mathcal{H}_0(P)$  can be much smaller than  $\mathcal{H}$ ;
- the way the procedures  $\{R_{\mathcal{C}}\}$  have been built implicitly assumed that  $\mathcal{C} = \mathcal{H}_0(P)$  and can be very conservative when  $\mathcal{C}$  is much larger than  $\mathcal{H}_0$ .

For instance, these behaviors have been extensively discussed in [2] for particular Rademacher-resampled thresholding procedures. Therefore, we seek for a procedure controlling the  $k\text{-FWER}$  which is “close” to  $R_{\mathcal{H}_0(P)}$  and which can be derived from the family  $\{R_{\mathcal{C}}\}_{\mathcal{C} \subset \mathcal{H}}$  via a simple algorithm.

### 4.3. Step-down method for FWER

We present in this section the special case of  $k = 1$ , following the approach of [34] with the presentation proposed in [9, 24]. Let us denote by  $A_{\mathcal{C}}$  the sets  $(R_{\mathcal{C}})^c$  of non-rejected hypotheses for the subset-indexed family. Consider the event

$$\Omega_0 = \{R_{\mathcal{H}_0(P)} \cap \mathcal{H}_0(P) = \emptyset\} = \{\mathcal{H}_0(P) \subset A_{\mathcal{H}_0(P)}\}.$$

By assumption (FWC<sub>0</sub>), we have  $\mathbb{P}(\Omega_0) \geq 1 - \alpha$ . Since from (NI),  $A_{\mathcal{C}}$  is non-decreasing in  $\mathcal{C}$ , the following holds on  $\Omega_0$ : for any  $\mathcal{C} \subset \mathcal{H}$ ,

$$\mathcal{H}_0(P) \subset \mathcal{C} \implies A_{\mathcal{H}_0(P)} \subset A_{\mathcal{C}} \implies \mathcal{H}_0(P) \subset A_{\mathcal{C}}. \quad (24)$$

Thus, on the event  $\Omega_0$ , taking  $\mathcal{C} = \mathcal{C}_0 = \mathcal{H}$  in (24) gives that  $\mathcal{H}_0(P) \subset A_{\mathcal{C}_0}$ , which in turn implies  $\mathcal{H}_0(P) \subset A_{\mathcal{C}_1}$  by taking  $\mathcal{C} = \mathcal{C}_1 = A_{\mathcal{C}_0}$  in (24), and so on. By recursion, this proves the following result:

**Theorem 4.2.** *Assume that there is a family  $\{R_{\mathcal{C}}\}_{\mathcal{C} \subset \mathcal{H}}$  of multiple testing procedures satisfying the conditions (NI) and (FWC<sub>0</sub>) and consider the corresponding family of non-rejected hypotheses  $\{A_{\mathcal{C}}\}_{\mathcal{C} \subset \mathcal{H}}$ . Define  $\hat{\mathcal{C}}$  by the following “step-down” recursion:*

- *Step 0: let  $\mathcal{C}_0 = \mathcal{H}$  and compute  $A_{\mathcal{C}_0}$ ;*
- *Step 1: let  $\mathcal{C}_1 = A_{\mathcal{C}_0}$  and compute  $A_{\mathcal{C}_1}$ . If  $A_{\mathcal{C}_1} = \mathcal{C}_1$ , let  $\hat{\mathcal{C}} = \mathcal{C}_1$  and stop. Otherwise go to step 2;*
- *Step  $j \geq 2$ : let  $\mathcal{C}_j = A_{\mathcal{C}_{j-1}}$  and compute  $A_{\mathcal{C}_j}$ . If  $A_{\mathcal{C}_j} = \mathcal{C}_j$ , let  $\hat{\mathcal{C}} = \mathcal{C}_j$  and stop. Otherwise go to step  $j + 1$ ;*

*Then the procedure  $R = (\hat{\mathcal{C}})^c$ , which also equals  $R_{\hat{\mathcal{C}}}$ , controls the FWER at level  $\alpha$  for any  $P \in \mathcal{P}$ .*

Note that for all  $j \geq 2$ , we have  $\mathcal{C}_{j+1} = A_{\mathcal{C}_j} \subset A_{\mathcal{C}_{j-1}} = \mathcal{C}_j$ , because  $\mathcal{C}_1 \subset \mathcal{C}_0$  and  $A_{\mathcal{C}}$  is non-decreasing in  $\mathcal{C}$ . Thus, the set of rejected hypotheses can only increase during the step-down algorithm. In particular, the final procedure  $\hat{\mathcal{C}}^c = R_{\hat{\mathcal{C}}}$  is always less conservative than the single-step procedure  $R_{\mathcal{H}}$ , for the same FWER control. In that sense, using a step-down algorithm is always more powerful than the single-step method.

As illustrated in Figure 3, Theorem 4.2 can be used with the Bonferroni family  $R_C = \{1 \leq i \leq m : p_i \leq \alpha/|\mathcal{C}|\}$ . In that case, by reordering the  $p$ -values  $p_{(1)} \leq \dots \leq p_{(m)}$  (with  $p_{(0)} = 0$ ), the corresponding step-down procedure defined in Theorem 4.3 can be reformulated as rejecting the nulls with  $p_i \leq \alpha/(m - \hat{\ell} + 1)$ , where  $\hat{\ell} = \max\{\ell \in \{0, 1, \dots, m\} : \forall \ell' \leq \ell, p_{(\ell')} \leq \alpha/(m - \ell' + 1)\}$ . This is the well known step-down *Holm procedure* which was introduced and proved to control the FWER in [27]. By contrast with step-up procedures, the step-down Holm procedure starts from the most significant  $p$ -value and stops the first time that a (ordered)  $p$ -value exceeds the critical curve.

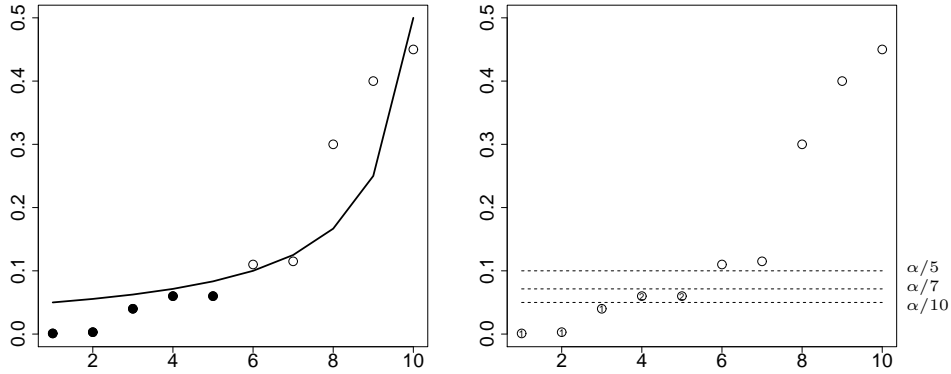


FIG 3. Illustration of the two equivalent definitions of Holm's procedure. The left picture is the classical step-down representation: ordered  $p$ -values together with the solid curve  $\ell \mapsto \alpha/(m - \ell + 1)$ . The filled points represent  $p$ -values that corresponds to the rejected hypotheses. The right picture illustrates the algorithm of Theorem 4.2: ordered  $p$ -values with the three thresholds  $\alpha/10$  (step 1),  $\alpha/7$  (step 2) and  $\alpha/5$  (step 3). For  $i \in \{1, 2\}$ , the points filled with "i" are rejected in the  $i$ th step of the algorithm. Both pictures use the same  $p$ -values and  $m = 10$ ;  $\alpha = 0.5$ .

#### 4.4. Step-down method for $k$ -FWER

We would like to generalize Theorem 4.2 to the case of the  $k$ -FWER. This time, we should consider the event

$$\Omega_0 = \{|R_{\mathcal{H}_0(P)} \cap \mathcal{H}_0(P)| \leq k - 1\} = \{\exists I_0 \subset \mathcal{H}, |I_0| = k - 1 : \mathcal{H}_0(P) \subset A_{\mathcal{H}_0(P)} \cup I_0\},$$

which satisfies by assumption  $\mathbb{P}(\Omega_0) \geq 1 - \alpha$ . For any subset  $\mathcal{C} \subset \mathcal{H}$ , let

$$\phi(\mathcal{C}) = \bigcup_{I \subset \mathcal{H}, |I|=k-1} A_{\mathcal{C} \cup I} = \bigcup_{I \subset \mathcal{C}^c, |I| \leq k-1} A_{\mathcal{C} \cup I}. \quad (25)$$

Then we may prove that the following holds: on the event  $\Omega_0$ , for any  $\mathcal{C} \subset \mathcal{H}$ ,

$$\begin{aligned} \exists I \subset \mathcal{H}, |I| = k - 1 : \mathcal{H}_0(P) \subset \mathcal{C} \cup I &\implies A_{\mathcal{H}_0(P)} \subset A_{\mathcal{C} \cup I} \subset \phi(\mathcal{C}) \\ &\implies \exists I' \subset \mathcal{H}, |I'| = k - 1 : \mathcal{H}_0(P) \subset \phi(\mathcal{C}) \cup I'. \end{aligned}$$

The first implication holds because  $A_{\mathcal{C}}$  is non-decreasing in  $\mathcal{C}$  and the second implication holds by considering  $I' = I_0$ . Thus, on the event  $\Omega_0$ , for any  $\mathcal{C} \subset \mathcal{H}$ ,

$$|\mathcal{C}^c \cap \mathcal{H}_0(P)| \leq k - 1 \implies |(\phi(\mathcal{C}))^c \cap \mathcal{H}_0(P)| \leq k - 1.$$

This leads to the following result.

**Theorem 4.3.** *Assume that there exists a family  $\{R_C\}_{C \subset \mathcal{H}}$  of multiple testing procedures satisfying the conditions (NI) and (FWC<sub>0</sub>), consider the corresponding family of non-rejected hypotheses  $\{A_C\}_{C \subset \mathcal{H}}$  and let  $\phi$  be defined by (25). Define  $\hat{\mathcal{C}}$  by the following “step-down” recursion:*

- *Step 0: let  $\mathcal{C}_0 = \mathcal{H}$  and compute  $A_{\mathcal{C}_0}$ ;*
- *Step 1: let  $\mathcal{C}_1 = A_{\mathcal{C}_0}$  and compute  $\phi(\mathcal{C}_1)$ . If  $\phi(\mathcal{C}_1) = \mathcal{C}_1$ , let  $\hat{\mathcal{C}} = \mathcal{C}_1$  and stop. Otherwise go to step 2;*
- *Step  $j \geq 2$ : let  $\mathcal{C}_j = \phi(\mathcal{C}_{j-1})$  and compute  $\phi(\mathcal{C}_j)$ . If  $\phi(\mathcal{C}_j) = \mathcal{C}_j$ , let  $\hat{\mathcal{C}} = \mathcal{C}_j$  and stop. Otherwise go to step  $j + 1$ ;*

*Then the procedure  $R = (\hat{\mathcal{C}})^c$ , which also equals  $(\phi(\hat{\mathcal{C}}))^c = \bigcap_{|I|=k-1} R_{\hat{\mathcal{C}} \cup I}$ , controls the  $k$ -FWER at level  $\alpha$  for any  $P \in \mathcal{P}$ .*

From (25),  $\phi(\cdot)$  is non-decreasing, that is,  $\forall \mathcal{C} \subset \mathcal{C}', \phi(\mathcal{C}) \leq \phi(\mathcal{C}')$ . As a consequence, we may derive from  $\mathcal{C}_1 = \phi(\mathcal{C}_0) \subset \mathcal{C}_0$  that  $\mathcal{C}_j = \phi(\mathcal{C}_{j-1}) \subset \mathcal{C}_{j-1}$  for all  $j \geq 2$ . Therefore, the rejection set can only increase at each step of the step-down algorithm. In particular, the final procedure  $\hat{\mathcal{C}}^c = \bigcap_{|I|=k-1} R_{\hat{\mathcal{C}} \cup I}$  is always less conservative than the single step method  $R_{\mathcal{H}}$ , for the same  $k$ -FWER control. Therefore, using the step-down algorithm always leads to a power improvement.

To illustrate Theorem 4.3, let us consider a thresholding-based family of the form  $R_C = \{1 \leq i \leq m : p_i \leq t_C\}$  with a non-increasing threshold function  $\mathcal{C} \mapsto t_C$  (i.e., such that for  $\mathcal{C} \subset \mathcal{C}'$ , we have  $t_{\mathcal{C}'} \leq t_C$ ) and such that  $\{R_C\}_C$  satisfies (FWC<sub>0</sub>). The recursion relation  $\mathcal{C}' = \phi(\mathcal{C})$  may be rewritten in that case as:

$$\begin{aligned} (\mathcal{C}')^c &= \bigcap_{I \subset \mathcal{C}^c, |I| \leq k-1} R_{\mathcal{C} \cup I} \\ &= \bigcap_{I \subset \mathcal{C}^c, |I| \leq k-1} \{1 \leq i \leq m : p_i \leq t_{\mathcal{C} \cup I}\} \\ &= \{1 \leq i \leq m : p_i \leq \min_{I \subset \mathcal{C}^c, |I| \leq k-1} \{t_{\mathcal{C} \cup I}\}\}. \end{aligned}$$

This recovers the generic step-down method described in Algorithm 2.1 of [35], which was developed in the case where the subset-indexed family is thresholding based.

Furthermore, note that when we choose the Bonferroni family, i.e. the threshold family  $t_C = \alpha k / |\mathcal{C}|$ , we have

$$\min_{I \subset \mathcal{C}^c, |I| \leq k-1} \{t_{\mathcal{C} \cup I}\} = \frac{\alpha k}{m \wedge (|\mathcal{C}| + k - 1)}.$$

Therefore, in terms of the ordered  $p$ -values  $0 = p_{(0)} \leq p_{(1)} \leq \dots \leq p_{(m)}$ , the procedure of Theorem 4.3 can be reformulated as rejecting the nulls when  $p_i \leq \alpha k / (m \wedge (m - \hat{\ell} + k))$  where  $\hat{\ell} = \max\{\ell \in \{0, 1, \dots, m\} : \forall \ell' \leq \ell, p_{(\ell')} \leq \alpha k / (m \wedge (m - \ell' + k))\}$ . The latter is the *generalized Holm procedure*, which was introduced and proved to control the  $k$ -FWER in [30].

As a consequence, Theorem 4.3 recovers at once several interesting results of the state-of-the-art  $k$ -FWER controlling methodology.

## 5. FDP control

For controlling the FDP, Romano and Wolf (2007) proposed to use a family  $\{S_k\}_k$  of  $k$ -FWER controlling procedures and to choose  $k$  that ensures that the corresponding rejection number  $|S_k|$  is “sufficiently large”, see [35]. Roughly speaking, choosing  $k$  such that  $|S_k|$  is larger than  $(k-1)/\gamma$  implies that, with high probability:

$$\text{FDP}(S_k, P) = |S_k \cap \mathcal{H}_0(P)|/|S_k| \leq (k-1)/|S_k| \leq \gamma.$$

Of course, as it is, the above reasoning is not rigorous, because the chosen  $k$  depends on the data. Theorem 4.1 (i) of [35] establishes that the latter approach leads to a correct FDP control in the asymptotic setting where the sample size available to do each test tends to infinity. This may be seen as a Dirac configuration where each  $p$ -value corresponding to false nulls are equal to zero.

In this section, we propose to reformulate this approach by using as index the rejection number  $\ell$  instead of  $k$ . Roughly speaking, if we choose  $\{R_\ell\}_\ell$  such that each  $R_\ell$  controls the  $(\gamma\ell+1)$ -FWER and we choose  $\ell$  such that  $|R_\ell| \geq \ell$ , we obtain with high probability:

$$\text{FDP}(R_\ell, P) = |R_\ell \cap \mathcal{H}_0(P)|/|R_\ell| \leq \gamma\ell/|R_\ell| \leq \gamma.$$

Similarly to the previous paragraph, this argument is not valid because the chosen  $\ell$  depends of the data. The main task of this section is to rationalize this approach. This leads to a general result, which covers both Theorem 4.1 (i) of [35] in the “Dirac” setting (see Section 5.4) and the earlier result of [30] (see Section 5.3). Moreover, we propose in Section 5.3 a procedure more powerful than the one of [30] when the data are assumed to follow the particular model  $\mathcal{P}^I$ .

In this section, the tuning parameter  $\gamma$  is fixed once for all in  $(0, 1)$ .

### 5.1. Family indexed by rejection numbers

Assume that we have at hand a family  $\{R_\ell\}_{1 \leq \ell \leq m}$  of multiple testing procedures and a class of distributions  $\mathcal{P}' \subset \mathcal{P}$  satisfying the following properties:

- $R_\ell$  is non-decreasing with respect to  $\ell$ , that is,

$$\forall \ell \in \{1, \dots, m-1\}, \quad R_\ell \subset R_{\ell+1}; \quad (\text{ND})$$

- $R_\ell$  controls the  $(\lfloor \gamma\ell \rfloor + 1)$ -FWER at level  $\alpha$  for any  $P \in \mathcal{P}'$  such that less than  $m - \ell + \lfloor \gamma(\ell - 1) \rfloor + 1$  null hypotheses are true, that is,

$$\begin{aligned} \forall \ell \in \{1, \dots, m\}, \quad \forall P \in \mathcal{P}' \text{ s.t. } |\mathcal{H}_0(P)| \leq m - \ell + \lfloor \gamma(\ell - 1) \rfloor + 1, \\ \mathbb{P}(|R_\ell \cap \mathcal{H}_0(P)| \geq \lfloor \gamma\ell \rfloor + 1) \leq \alpha; \end{aligned} \quad (\text{FWC})$$

- for any  $P \in \mathcal{P}'$ , for any  $\ell \in \{1, \dots, m\}$ , the false rejection number of  $R_\ell$  is independent of the correct rejection numbers of  $R_{\ell'}$ , for  $1 \leq \ell' \leq m$ , that is,

$$\forall P \in \mathcal{P}', \forall \ell \in \{1, \dots, m\}, |R_\ell \cap \mathcal{H}_0(P)| \text{ is independent of } \{|R_{\ell'} \cap \mathcal{H}_1(P)|, 1 \leq \ell' \leq m\}. \quad (\text{DA})$$



In condition (FWC), for any  $x \geq 0$ ,  $\lfloor x \rfloor$  denotes the largest integer  $n$  such that  $n \leq x$ . Condition (ND) is natural because the index  $\ell$  can be interpreted as a rejection number. It is easy to check in the examples below.

For any  $\mathcal{P}' \subset \mathcal{P}$ , condition (FWC) can be fulfilled using the (single-step or step-down)  $k$ -FWER controlling procedure of the previous section, by taking  $k = \lfloor \gamma \ell \rfloor + 1$ . As a first instance, we may use the (single-step) Bonferroni family  $R_\ell$  using the threshold  $\alpha(\lfloor \gamma \ell \rfloor + 1)/m$ . Moreover, note that  $|\mathcal{H}_0(P)| \leq m - \ell + \lfloor \gamma(\ell - 1) \rfloor + 1$  in (FWC), thus we can consider the improved threshold

$$t_\ell^{LR} = \frac{\alpha(\lfloor \gamma \ell \rfloor + 1)}{m - \ell + \lfloor \gamma(\ell - 1) \rfloor + 1}. \quad (26)$$

The threshold (26) is slightly larger than the threshold used in Theorem 3.1 of [30] (they used  $\lfloor \gamma \ell \rfloor$  instead of  $\lfloor \gamma(\ell - 1) \rfloor$  in the denominator). As a second instance, we may substantially improve on the above threshold family when we additionally assume that the distribution  $P$  of the data lies in the smaller subset  $\mathcal{P}' = \mathcal{P}^I$ : for this, note that for any  $P \in \mathcal{P}^I$  and for any  $t \in [0, 1]$ , the variable  $|\{i \in \mathcal{H}_0(P) : p_i(X) \leq t\}|$  is stochastically upper-bounded by a binomial distribution of parameters  $|\mathcal{H}_0(P)|$  and  $t$ , which in turn is stochastically upper-bounded by a binomial distribution of parameters  $m - \ell + \lfloor \gamma(\ell - 1) \rfloor + 1$  and  $t$ . Therefore, choosing the (deterministic) quantile-based threshold family  $(t_\ell^Q)_{1 \leq \ell \leq m}$  defined by

$$t_\ell^Q = \max\{t \in [0, 1] : \mathbb{P}(Z \geq \lfloor \gamma \ell \rfloor + 1) \leq \alpha \text{ for } Z \sim \mathcal{B}(m - \ell + \lfloor \gamma(\ell - 1) \rfloor + 1, t)\}, \quad (27)$$

we obtain a family of thresholding procedures satisfying (FWC) with  $\mathcal{P}' = \mathcal{P}^I$ . Of course, since  $t_\ell^{LR}$  in (26) is only based upon the Markov's inequality, which is not accurate for binomial variables, the threshold family  $t_\ell^Q$  defined by (27) is substantially larger, as illustrated in Figure 4.

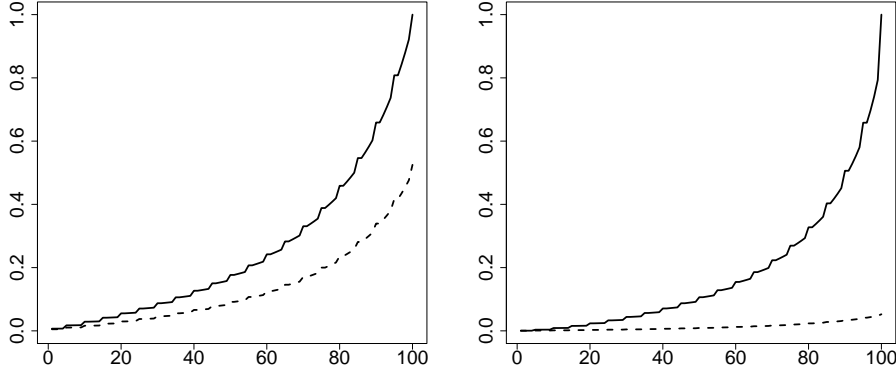


FIG 4. Threshold  $t_\ell^Q$  in (27) for model  $\mathcal{P}^I$  (solid line) and threshold  $t_\ell^{LR}$  in (26) for model  $\mathcal{P}$  (dashed line) in function of  $\ell \in \{1, \dots, m\}$ .  $m = 200$ ;  $\gamma = 0.2$ . Right:  $\alpha = 0.5$ ; left:  $\alpha = 0.05$ .

Assumption (DA) is a dependence assumption which is typically satisfied in the two following cases:

- each procedure  $R_\ell$  uses a deterministic threshold and the  $p$ -values associated to true nulls are independent of the  $p$ -values associated to false nulls, for all the distributions

of  $\mathcal{P}'$ , that is,

$$\begin{aligned} \forall \ell \in \{1, \dots, m\}, R_\ell &= \{i \in \{1, \dots, m\} : p_i \leq t_\ell\} \text{ for one deterministic } t_\ell \in [0, 1] \\ \text{and } \forall P \in \mathcal{P}', (p_i(X))_{i \in \mathcal{H}_0(P)} &\text{ is independent of } (p_i(X))_{i \in \mathcal{H}_1(P)}; \end{aligned} \quad (\text{DA}')$$

- for all the distributions of  $\mathcal{P}'$ , the number of correct rejections of each  $R_\ell$  is deterministic, that is,

$$\forall P \in \mathcal{P}', \{ |R_{\ell'} \cap \mathcal{H}_1(P)|, 1 \leq \ell' \leq m \} \text{ is deterministic.} \quad (\text{DA}'')$$

Condition (DA'') is satisfied for instance when  $\mathcal{H}_1(P) \subset R_{\ell'}$ , for any  $\ell'$ , which is the case for procedures of the form  $R_\ell = \{i \in \{1, \dots, m\} : p_i \leq t_\ell(\mathbf{p})\}$  using a possibly data-dependent threshold  $t_\ell(\mathbf{p}) \in [0, 1]$ , when we assume that the  $p$ -values are in the Dirac configuration, that is, when they are equal to zero under the alternative.

## 5.2. Generic step-down theorem

The approach below is an adaptation of the proof of Theorem 3.1 in [30] to our setting. Let us consider a family  $\{R_\ell\}_{1 \leq \ell \leq m}$  and a class of distributions  $\mathcal{P}' \subset \mathcal{P}$  satisfying (ND)-(FWC)-(DA). We aim at selecting  $\ell = \hat{\ell}$  that provides  $\forall P \in \mathcal{P}', \text{FDP}(R_{\hat{\ell}}, P) \leq \alpha$ .

First note that, by using the definition of the FDP, we have for any  $\ell \in \{1, \dots, m\}$  such that  $|R_\ell| = \ell$ :

$$\begin{aligned} \{\text{FDP}(R_\ell, P) > \gamma\} &= \{|\mathcal{H}_0(P) \cap R_\ell| > \gamma\ell\} \\ &= \{|\mathcal{H}_0(P) \cap R_\ell| \geq \lfloor \gamma\ell \rfloor + 1\} \\ &= \{\ell \in \mathcal{L}\}, \end{aligned} \quad (28)$$

where  $\mathcal{L} = \{\ell \in \{1, \dots, m\} : \ell - |\mathcal{H}_1(P) \cap R_\ell| \geq \lfloor \gamma\ell \rfloor + 1\}$  is a set which only depends on the set  $\{|\mathcal{H}_1(P) \cap R_{\ell'}|, 1 \leq \ell' \leq m\}$ .

Second, note that for any  $\ell \in \{1, \dots, m\}$  such that  $|R_\ell| \geq \ell$ ,

$$\{\ell \in \mathcal{L}\} \subset \{|\mathcal{H}_0(P) \cap R_\ell| \geq \lfloor \gamma\ell \rfloor + 1\}. \quad (29)$$

Let us consider  $\ell^* = \min\{\mathcal{L}\}$  (letting  $\ell^* = m + 1$  when  $\mathcal{L} = \emptyset$ ). From (28) and (29), taking  $\hat{\ell} \in \{1, \dots, m\}$  such that  $|R_{\hat{\ell}}| = \hat{\ell}$  and such that for any  $\ell \leq \hat{\ell}$ ,  $|R_\ell| \geq \ell$ , we obtain

$$\begin{aligned} \{\text{FDP}(R_{\hat{\ell}}, P) > \gamma\} &\subset \{\ell^* \leq \hat{\ell}\} \\ &\subset \{|\mathcal{H}_0(P) \cap R_{\ell^*}| \geq \lfloor \gamma\ell^* \rfloor + 1\}. \end{aligned}$$

Moreover, if  $\ell^* \geq 2$ , we have by definition of  $\ell^*$  that  $\ell^* - 1 \notin \mathcal{L}$ . Hence, we can obtain the following upper-bound for  $|\mathcal{H}_0(P)|$ :

$$|\mathcal{H}_0(P)| = m - |\mathcal{H}_1(P)| \leq m - |\mathcal{H}_1(P) \cap R_{\ell^*-1}| \leq m - \ell^* + \lfloor \gamma(\ell^* - 1) \rfloor + 1.$$

Since the above bound is also true when  $\ell^* = 1$ , it holds for any possible value of  $\ell^*$ .

Finally noting that  $\ell^*$  only depends on the variable set  $\{|\mathcal{H}_1(P) \cap R_{\ell'}|, 1 \leq \ell' \leq m\}$  and using (FWC)-(DA), we have proved that for any  $\ell \in \{1, \dots, m\}$ ,

$$\begin{aligned} \mathbb{P}(\text{FDP}(R_{\hat{\ell}}, P) > \gamma \mid \ell^* = \ell) &\leq \mathbb{P}(|\mathcal{H}_0(P) \cap R_\ell| \geq \lfloor \gamma\ell \rfloor + 1 \mid \ell^* = \ell) \\ &= \mathbb{P}(|\mathcal{H}_0(P) \cap R_\ell| \geq \lfloor \gamma\ell \rfloor + 1) \\ &\leq \alpha. \end{aligned}$$

Also, the probability  $\mathbb{P}(\text{FDP}(R_{\hat{\ell}}, P) > \gamma \mid \ell^* = m + 1)$  is zero, because it is smaller than  $\mathbb{P}(\hat{\ell} \in \mathcal{L} \mid \ell^* = m + 1)$ . This leads to the following result.

**Theorem 5.1.** *Assume that there exists a family  $\{R_\ell\}_{1 \leq \ell \leq m}$  of multiple testing procedures and a class of distributions  $\mathcal{P}' \subset \mathcal{P}$  satisfying the conditions (ND)-(FWC)-(DA) defined in Section 5.1. Consider the procedure  $R_{\hat{\ell}}$  where*

$$\hat{\ell} = \max \{ \ell \in \{0, \dots, m\} : \forall \ell' \in \{0, \dots, \ell\}, |R_{\ell'}| \geq \ell' \}, \quad (30)$$

(with the convention  $R_0 = \emptyset$ ). Then  $R_{\hat{\ell}}$  controls the FDP in the following sense:

$$\forall P \in \mathcal{P}', \quad \mathbb{P}(\text{FDP}(R_{\hat{\ell}}, P) > \gamma) \leq \alpha. \quad (31)$$

The algorithm performed to find (30) is a step-down algorithm; it starts from the small rejection numbers and stops the first time that  $|R_\ell|$  is below  $\ell$ . Note that the maximum in (30) is well defined because  $\ell = 0$  satisfies  $|R_\ell| \geq \ell$ . Furthermore, using (ND), relation (30) implies that  $\hat{\ell} \leq |R_{\hat{\ell}}| \leq |R_{\hat{\ell}+1}| < \hat{\ell} + 1$ , so that  $|R_{\hat{\ell}}| = |R_{\hat{\ell}+1}| = \hat{\ell}$  holds. As a consequence, the procedure of Theorem 5.1 can be equivalently defined by  $R_{\tilde{\ell}}$  where

$$\tilde{\ell} = \min \{ \ell \in \{1, \dots, m+1\} : |R_\ell| \leq \ell - 1 \}, \quad (32)$$

with the convention  $R_{m+1} = R_m$  (so that the minimum in (32) is well defined).

### 5.3. Theorem 3.1 of [30] and a novel result as corollaries

Going back to the specific setting (DA') described in Section 5.1, we may derive from Theorem 5.1 the following corollary.

**Corollary 5.2.** *Let us consider the deterministic threshold family  $(t_\ell^{LR})_{1 \leq \ell \leq m}$  defined by (26) and consider*

$$\hat{\ell} = \max \{ \ell \in \{0, \dots, m\} : \forall \ell' \in \{0, \dots, \ell\}, p_{(\ell')} \leq t_{\ell'}^{LR} \}, \quad (33)$$

where  $0 = p_{(0)} \leq p_{(1)} \leq \dots \leq p_{(m)}$  denote the ordered  $p$ -values and by convention  $t_0^{LR} = 0$ . Then the procedure  $R_{\hat{\ell}} = \{i \in \{1, \dots, m\} : p_i \leq t_{\hat{\ell}}^{LR}\}$  satisfies the FDP control (31) for the subset  $\mathcal{P}'$  of distributions  $P \in \mathcal{P}$  such that the family  $(p_i(X))_{i \in \mathcal{H}_0(P)}$  is independent of the family  $(p_i(X))_{i \in \mathcal{H}_1(P)}$ .

By reproducing the end of the proof of Theorem 5.1 in the particular setting of Corollary 5.2, we may increase a bit the distribution set  $\mathcal{P}'$  in Corollary 5.2 to the set of  $P \in \mathcal{P}$  such that for any  $i \in \mathcal{H}_0(P)$ ,  $\forall u \in [0, 1]$ ,  $\mathbb{P}(p_i(X) \leq u \mid (p_i(X))_{i \in \mathcal{H}_1(P)}) \leq u$ . This is the distributional setting of Theorem 3.1 of [30]. Hence, we are able to recover the latter result (with a slight improvement in the threshold family).

Furthermore, if we want to ensure the FDP control (31) only for the smaller distribution set  $\mathcal{P}' = \mathcal{P}^I$ , we may consider the larger threshold family  $(t_\ell^Q)_{1 \leq \ell \leq m}$ :

**Corollary 5.3.** *Let us consider the deterministic threshold family  $(t_\ell^Q)_{1 \leq \ell \leq m}$  defined by (27), and  $\hat{\ell} = \max \{ \ell \in \{0, \dots, m\} : \forall \ell' \in \{0, \dots, \ell\}, p_{(\ell')} \leq t_{\ell'}^Q \}$  (with  $t_0^Q = 0$ ). Then the FDP control (31) holds for  $\mathcal{P} = \mathcal{P}^I$  and  $R_{\hat{\ell}} = \{i \in \{1, \dots, m\} : p_i \leq t_{\hat{\ell}}^Q\}$ .*

The above result has never been reported before to our knowledge. It establishes a FDP control which is suitable for independent  $p$ -values.

#### 5.4. Theorem 4.1 (i) of [35] as corollary

In Section 4 of [35], a step-down procedure  $S_{\hat{k}}$  was defined from a generic family  $\{S_k\}_{1 \leq k \leq m}$  of thresholding based procedures, where each  $S_k$  controls the  $k$ -FWER for  $1 \leq k \leq m$  and  $S_k \subset S_{k+1}$  for  $1 \leq k \leq m - 1$ . The index  $\hat{k}$  was obtained as follows:

$$\hat{k} = \min\{k \in \{1, \dots, m+1\} : \gamma|S_k| < k - \gamma\}, \quad (34)$$

where we use here the convention  $S_{m+1} = S_m$  (so that the above set always contains  $k = m+1$ ). The authors showed that  $S_{\hat{k}}$  controls the FDP in the asymptotic sense, as the sample size available to perform each test tends to infinity. This can be reinterpreted as a (non-asymptotic) FDP control in a Dirac configuration where the  $p$ -values corresponding to false nulls are equal to zero. Set under this form, this result can be derived from Theorem 5.1.

For this, let  $R_\ell = S_{\lfloor \gamma \ell \rfloor + 1}$ , for  $\ell \in \{1, \dots, m\}$ , and note that the family  $\{R_\ell\}_{1 \leq \ell \leq m}$  satisfies (ND)-(FWC) and (DA''), by taking the distribution set  $\mathcal{P}'$  corresponding to Dirac configurations for the  $p$ -values. Hence, Theorem 5.1 establishes the FDP control for the Dirac configurations of the procedure  $R_{\tilde{\ell}}$  where  $\tilde{\ell}$  is defined by (30), or equivalently by (32). Thus, it only remains to show that the step-down algorithms (34) and (32) lead to the same procedure, that is,

$$R_{\tilde{\ell}} = S_{\hat{k}}.$$

To prove the latter, we establish  $\hat{k} = \lfloor \gamma \tilde{\ell} \rfloor + 1$ . First, using (32),  $\tilde{\ell}$  satisfies  $\gamma|S_{\lfloor \gamma \tilde{\ell} \rfloor + 1}| \leq \gamma \tilde{\ell} - \gamma$ . Since  $\gamma \ell < \lfloor \gamma \ell \rfloor + 1$ , we deduce from the definition of  $\hat{k}$  that  $\lfloor \gamma \tilde{\ell} \rfloor + 1 \geq \hat{k}$ . Conversely, by considering the unique integer  $\ell \in \{1, \dots, m\}$  satisfying  $\hat{k}/\gamma - 1 \leq \ell < \hat{k}/\gamma$  and thus also  $\lfloor \gamma \ell \rfloor + 1 = \hat{k}$ , we have that for any integer  $j$ ,  $\gamma j < \hat{k} \Rightarrow j \leq \ell$ . Applying the latter for  $j = |S_{\hat{k}}| + 1$ , we obtain from  $\gamma(|S_{\hat{k}}| + 1) < \hat{k}$  that  $|S_{\hat{k}}| \leq \ell - 1$  and thus  $\ell \geq \tilde{\ell}$ , by using the definition of  $\tilde{\ell}$ . This in turn implies  $\hat{k} \geq \lfloor \gamma \tilde{\ell} \rfloor + 1$ . We thus have proved the following result, which can be seen as Theorem 4.1 (i) of [35] in the Dirac setting.

**Corollary 5.4.** *Assume that there exists a family  $\{S_k\}_{1 \leq k \leq m}$  of multiple testing procedures (with the convention  $S_{m+1} = S_m$ ) satisfying*

- for each  $k \in \{1, \dots, m\}$ ,  $S_k$  is of the form  $\{i \in \{1, \dots, m\} : p_i \leq t_k(\mathbf{p})\}$  for a possibly data-depend threshold  $t_k(\cdot) \in [0, 1]$ ;
- for each  $k \in \{1, \dots, m-1\}$ ,  $S_k \subset S_{k+1}$ ;
- for each  $k \in \{1, \dots, m\}$ ,  $\forall P \in \mathcal{P}$ ,  $k$ -FWER( $S_k, P$ )  $\leq \alpha$ .

*Consider  $\hat{k}$  defined in (34) and the subset  $\mathcal{P}'$  of distributions  $P \in \mathcal{P}$  corresponding to a Dirac configuration, i.e., such that  $\forall P \in \mathcal{P}'$ ,  $\forall i \in \mathcal{H}_1(P)$ ,  $p_i(x) = 0$  for  $P$ -almost every  $x \in \mathcal{X}$ . Then we have  $\forall P \in \mathcal{P}'$ ,  $\mathbb{P}(\text{FDP}(S_{\hat{k}}, P) > \gamma) \leq \alpha$ .*

## 6. Discussions

### 6.1. Complexity of the $k$ -FWER step-down approach

One major limitation of the  $k$ -FWER approach presented in Section 4 is that the computation of  $\phi(\cdot)$  in (25) can become complex when  $k$  is large because we should consider all the subsets  $I$  of  $\mathcal{C}^c$  of cardinal  $k-1$  (say that  $|\mathcal{C}^c| \geq k-1$ ). However, we may modify this algorithm by considering only the set  $I$  equals to the  $k-1$  coordinates of  $\mathcal{C}^c$  corresponding to the  $k-1$

largest  $p$ -values. As noted in [35], this “streamlined” step-down procedure still controls the  $k$ -FWER control in the Dirac model where each false null has a  $p$ -value equals to zero. The latter is true because in this model, as soon as  $|\mathcal{C}^c \cap \mathcal{H}_0(P)| \leq k - 1$ , we know that the set  $\mathcal{C}^c \cap \mathcal{H}_0(P)$  is included in the set  $I$  of coordinates corresponding to the  $k - 1$  largest  $p$ -values in  $\mathcal{C}^c$  (because the  $p$ -values of  $\mathcal{C}^c \cap \mathcal{H}_1(P)$  are zero). Nevertheless, no proof of this  $k$ -FWER control stands without this Dirac assumption.

## 6.2. FDR control is not FDP control

Since the only interpretable variable is the FDP and not its expectation, controlling the FDR is meaningful only when the FDP concentrates well around the FDR. As the hypothesis number  $m$  grows, Neuvial (2008) showed that the latter holds for step-up type procedures when a Donsker type theorem for the e.c.d.f. is valid, so for instance under independence or “weak” dependence, see [32]. However, under some unspecified dependencies, we do not know how the FDP concentrates. For instance, even under a very simple  $\rho$ -equi-correlated Gaussian model (corresponding to Example 1.2, where the non-diagonal entries of  $\Sigma(P)$  are all equal to  $\rho$ ), it was shown in [14] that the convergence rate of the FDP to the FDR can be arbitrarily slow when  $\rho = \rho_m$  tends to zero as  $m$  tends to infinity. Additionally, it was proved in [19] that there is not concentration at all when  $\rho$  is kept fixed with  $m$ . Also, as shown in [37], the “sparsity” ( $\pi_0 = \pi_{0,m}$  tends to 1 as  $m$  tends to infinity) is one other feature that can slow down the FDP convergence. Therefore, in all these cases where the FDP convergence is slow, controlling the FDR does not lead to a clear interpretation for the underlying FDP.

The latter drawback does not arise while controlling the FDP upper-tail distribution: for instance, the FDP control  $\mathbb{P}(\text{FDP} > 0.01) \leq 0.5$  ensures that, with a probability at least 0.5, the FDP is below 0.01, and this interpretation holds whatever the FDP distribution is. However, the FDR stays useful, because this is a simpler criterion for which the controlling methodology is much more developed in comparison with the FDP controlling methodology.

## 6.3. Conclusion

In this paper, we have recovered some of the classical state-of-the-art multiple testing procedures for controlling the FDR,  $k$ -FWER and the FDP. Each result was given together with a proof, which we wanted as short and meaningful as possible. Besides, we tried to write the paper under a scholarly form, for a possibly non-specialist reader. Some new contributions were also given for  $k$ -FWER and FDP control, by extending and unifying some previous work of multiple testing literature.

The type I error rate control research area has still many unsolved issues. Among the major concerns, remember that the FDP control in Section 5 needs a very strong distributional assumption on the test statistics, namely the independence or “Dirac” assumption. To our knowledge, no procedure adaptive to dependencies is proved to control the FDP without assuming such a strong requirement. This is a room left for future developments, which would have a strong impact on the high-dimensional data analysis.

## Acknowledgements

I warmly thank Lucien Birgé for having carefully read and corrected this manuscript. This work was supported by the French Agence Nationale de la Recherche (ANR grant references:

ANR-09-JCJC-0027-01, ANR-PARCIMONIE, ANR-09-JCJC-0101-01) and the French ministry of foreign and european affairs (EGIDE - PROCOPE project number 21887 NJ).

## References

- [1] S. Arlot, G. Blanchard, and E. Roquain. Some nonasymptotic results on resampling in high dimension. I. Confidence regions. *Ann. Statist.*, 38(1):51–82, 2010.
- [2] S. Arlot, G. Blanchard, and E. Roquain. Some nonasymptotic results on resampling in high dimension. II. Multiple tests. *Ann. Statist.*, 38(1):83–99, 2010.
- [3] Y. Benjamini and R. Heller. False discovery rates for spatial signals. *J. Amer. Statist. Assoc.*, 102(480):1272–1281, 2007.
- [4] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300, 1995.
- [5] Y. Benjamini and Y. Hochberg. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Behav. Educ. Statist.*, 25:60–83, 2000.
- [6] Y. Benjamini, A. M. Krieger, and D. Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, 2006.
- [7] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188, 2001.
- [8] M. A. Black. A note on the adaptive control of false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(2):297–304, 2004.
- [9] G. Blanchard. Contrôle non-asymptotique adaptatif du family-wise error rate en tests multiples. Talk at Journées Statistiques du Sud, Porquerolles, 2009.
- [10] G. Blanchard and F. Fleuret. Occam’s hammer: a link between randomized learning and multiple testing FDR control. ArXiv preprint math.ST/0608713, 2007.
- [11] G. Blanchard and E. Roquain. Two simple sufficient conditions for FDR control. *Electron. J. Stat.*, 2:963–992, 2008.
- [12] G. Blanchard and E. Roquain. Adaptive FDR control under independence and dependence. *J. Mach. Learn. Res.*, 10:2837–2871, 2009.
- [13] A. Celisse and S. Robin. A cross-validation based estimation of the proportion of true null hypotheses. *Journal of Statistical Planning and Inference*, 140(11):3132 – 3147, 2010.
- [14] S. Delattre and E. Roquain. On the false discovery proportion convergence under gaussian equi-correlation. *Statistics & Probability Letters*, 81(1):111–115, 2011.
- [15] S. Dudoit and M. J. van der Laan. *Multiple testing procedures with applications to genomics*. Springer Series in Statistics. Springer, New York, 2008.
- [16] A. Farcomeni. Some results on the control of the false discovery rate under dependence. *Scand. J. Statist.*, 34(2):275–297, 2007.
- [17] J. A. Ferreira and A. H. Zwiderman. On the Benjamini-Hochberg method. *Ann. Statist.*, 34(4):1827–1849, 2006.
- [18] H. Finner, R. Dickhaus, and M. Roters. On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.*, 37(2):596–618, 2009.
- [19] H. Finner, T. Dickhaus, and M. Roters. Dependency and false discovery rate: asymptotics. *Ann. Statist.*, 35(4):1432–1455, 2007.
- [20] R. A. Fisher. *The Design of Experiments*. Oliver and Boyd, Edinburgh.p, 1935.
- [21] C. Genovese and L. Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):499–517, 2002.

- [22] C. Genovese and L. Wasserman. A stochastic process approach to false discovery control. *Ann. Statist.*, 32(3):1035–1061, 2004.
- [23] J. Goeman and L. Finos. The inheritance procedure: multiple testing of tree-structured hypotheses. Technical report, Leiden University medical center, 2010.
- [24] J. Goeman and A. Solari. The sequential rejection principle of familywise error control. *Ann. Statist.*, To appear.
- [25] W. Guo and M. B. Rao. On control of the false discovery rate under no assumption of dependency. *Journal of Statistical Planning and Inference*, 138(10):3176–3188, 2008.
- [26] Y. Hochberg and A. C. Tamhane. *Multiple comparison procedures*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1987.
- [27] S. Holm. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, 6(2):65–70, 1979.
- [28] K. I. Kim, E. Roquain, and M. A. van de Wiel. Spatial clustering of array CGH features in combination with hierarchical multiple testing. *Stat. Appl. Genet. Mol. Biol.*, 9(1):Art. 40, 2010.
- [29] K. I. Kim and M. van de Wiel. Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinformatics*, 9(1):114, 2008.
- [30] E. L. Lehmann and J. P. Romano. Generalizations of the familywise error rate. *Ann. Statist.*, 33:1138–1154, 2005.
- [31] C. J. Miller, C. Genovese, R. C. Nichol, L. Wasserman, A. Connolly, D. Reichart, A. Hopkins, J. Schneider, and A. Moore. Controlling the false-discovery rate in astrophysical data analysis. *The Astronomical Journal*, 122(6):3492–3505, 2001.
- [32] P. Neuvial. Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.*, 2:1065–1110, 2008.
- [33] D. Pantazis, T. E. Nichols, S. Baillet, and R. M. Leahy. A comparison of random field theory and permutation methods for statistical analysis of meg data. *NeuroImage*, 25:383–394, 2005.
- [34] J. P. Romano and M. Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.*, 100(469):94–108, 2005.
- [35] J. P. Romano and M. Wolf. Control of generalized error rates in multiple testing. *Ann. Statist.*, 35(4):1378–1408, 2007.
- [36] E. Roquain and M. van de Wiel. Optimal weighting for false discovery rate control. *Electron. J. Stat.*, 3:678–711, 2009.
- [37] E. Roquain and F. Villers. Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.*, To appear.
- [38] D. Rubin, S. Dudoit, and M. van der Laan. A method to increase the power of multiple testing procedures through sample splitting. *Stat. Appl. Genet. Mol. Biol.*, 5:Art. 19, 20 pp. (electronic), 2006.
- [39] S. K. Sarkar. Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.*, 30(1):239–257, 2002.
- [40] S. K. Sarkar. On methods controlling the false discovery rate. *Sankhya, Ser. A*, 70:135–168, 2008.
- [41] T. Schweder and E. Spjøtvoll. Plots of P-values to evaluate many tests simultaneously. *Biometrika*, 69(3):493–502, 1982.
- [42] P. Seeger. A note on a method for the analysis of significances en masse. *Technometrics*, 10(3):586–593, 1968.



- [43] J. D. Storey. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):479–498, 2002.
- [44] J. D. Storey, J. E. Taylor, and D. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(1):187–205, 2004.
- [45] M. A. van de Wiel and W. N. van Wieringen. CGHregions: Dimension Reduction for Array CGH Data with Minimal Information Loss. *Cancer Inform*, 3:55–63, 2007.
- [46] L. Wasserman and K. Roeder. Weighted hypothesis testing. Technical report, Dept. of statistics, Carnegie Mellon University, 2006.
- [47] P. H. Westfall and S. S. Young. *Resampling-Based Multiple Testing*. Wiley, 1993. Examples and Methods for  $P$ - Value Adjustment.