Self-consistent multiple testing procedures

Gilles Blanchard, Étienne Roquain

November 3, 2018

Abstract

We study the control of the false discovery rate (FDR) for a general class of multiple testing procedures. We introduce a general condition, called "self-consistency", on the set of hypotheses rejected by the procedure, which we show is sufficient to ensure the control of the corresponding false discovery rate under various conditions on the distribution of the p-values. Maximizing the size of the rejected null hypotheses set under the constraint of self-consistency, we recover various step-up procedures. As a consequence, we recover earlier results through simple and unifying proofs while extending their scope to several regards: arbitrary measure of set size, p-value reweighting, new family of step-up procedures under unspecified p-value dependency. Our framework also allows for defining and studying FDR control for multiple testing procedures over a continuous, uncountable space of hypotheses.

1 Introduction

In this work we study false discovery rate (FDR) control for a class of general multiple testing procedures. Here a multiple testing procedure is defined as an algorithm taking in input some (randomly generated) data $X \in \mathcal{X}$ and returning a set $R(X)$ of rejected hypotheses, a subset of the set $\mathcal{H}$ of candidate null hypotheses. The main result of this work is that procedures satisfying a specific condition, which we call "self-consistency", will have a controlled FDR; the exact form of the condition depends on the assumptions on the dependency of the p-values.

The first important point of our approach is to introduce the self-consistency condition. For this, we first define a reference threshold-based testing procedure $L(r)$, which rejects hypotheses $h \in \mathcal{H}$ having their p-values smaller than a threshold function $\Delta(h,r)$ (depending on an external real parameter r and possibly on h). To some extent, the reference procedure $L(r)$ can be interpreted as an "ideal" set of rejections (under constraint of controlled FDR) if the number of rejected hypotheses was known in advance to be r. The self-consistency condition then simply states that R should (a.s.) be a subset of $L(|R|)$. This condition can be seen as checking a *post-hoc* consistency of procedure R, with respect to the set of rejected hypotheses prescribed by the reference procedure, wherein we plug in for the parameter r the cardinality of R itself.

This interpretation is very close to a discussion appearing in the original paper of Benjamini and Hochberg ([1], Section 3.3) showing that the linear-step-up procedure can have an interpretation as a maximization of such a *post-hoc* condition. To some extent, the present work can be seen as an extensive development of this seminal remark where we show that this condition is in fact sufficient by itself to ensure FDR control.

The second important point of our approach is to reduce the question of FDR control — under the self-consistency condition — to a simple probabilistic inequality over just two real variables taking the following form:

$$\mathbb{E}\left[\frac{\mathbf{1}\{U \leq c\beta(V)\}}{V}\right] \leq c\,, \tag{1}$$

where U has a distribution stochastically lower bounded by a uniform distribution, V is a non-negative random variable, β is some non-decreasing *shape function* to be specified later and c is an arbitrary positive real number. This link allows us to abstract the particulars of a specific multiple testing procedure, in order to concentrate on proving (1). In (1), the expression of β and the dependency between U and V is directly linked to the dependency assumption on the p-values in the initial setting; the form of β will also determine the specific threshold function $\Delta(h, r)$ for the reference procedure $L(r)$ in the self-consistency condition.

Thus, to summarize our contributions:

- We recover in a direct and unifying way several well-known results on FDR control for step-up procedures (*e.g.* [1; 2; 3; 4; 5; 6]), while extending said results to the more general class of self-consistent procedures.
- We emphasize a clear separation between checking the self-consistency condition, which is an *algorithmic* issue, from showing inequalities of the type (1), which is a purely *probabilistic* issue depending on the assumptions on the dependency of the p-values.
- We put forward several further extensions of interest that fit very naturally in our framework: we present a new family of FDR-controlled step-up procedures in the framework of arbitrary dependencies between the p-values, we also consider weighted FDR, weighted p-values, and continuous sets of hypotheses (the latter can be of relevance e.g. when the observations are given by a stochastic process).

This paper is organized as follows: in Section 2, we introduce our framework and the self-consistency condition, and we prove that self-consistent procedures control the FDR if a probabilistic inequality of the form (1) is satisfied. Section 3 presents proofs of inequalities of the form (1) when the p-values are independent, positively dependent (under the PRDS condition introduced in [3]), or have unspecified dependencies. In Section 4, we show that step-up procedures satisfy a form of self-consistency condition, which implies that step-up procedures have a controlled FDR in all the latter cases of dependencies. In this section we also discuss the sharpness of the obtained FDR bounds. Section 5 presents an extension of some of our results to the case of a continuous set of hypotheses. We give a conclusion in Section 6. Appendix A presents some technical lemmas and Appendix B contains additional results showing how to use probabilistic inequalities of the form (1) for particular step-down procedures that do not enter directly into our main setting.

2 Self-consistency

2.1 Preliminaries and notations

Let $(\mathcal{X}, \mathfrak{X}, P)$ be a probability space. We are interested in determining whether the distribution P satisfies or not certain specific properties called *null hypotheses*. Formally, a *null hypothesis* h is a subset of distributions on $(\mathcal{X}, \mathfrak{X})$. We say that P satisfies h when $P \in h$.

In the multiple testing framework, one is interested in testing simultaneously for a entire set $\mathcal{H}$ of possible null hypotheses. Below we will always assume that $\mathcal{H}$ is at most countable, except specifically in Section 5 where we deal with extensions to continuous sets of hypotheses. We denote by $\mathcal{H}_0 = \{h \in \mathcal{H} \mid P \text{ satisfies } h\}$ the set of null hypotheses satisfied by P, called the *set of true null hypotheses*. We denote by $\mathcal{H}_1 = \mathcal{H} \setminus \mathcal{H}_0$ the *set of the false null hypotheses* for P.

A multiple testing procedure returns a subset $R(x) \subset \mathcal{H}$ of rejected hypotheses based on a realization x of a random variable $X \sim P$.

Definition 2.1 (Multiple testing procedure)**.** A *multiple testing procedure* R on $\mathcal{H}$ is a function $R : x \in \mathcal{X} \mapsto R(x) \subset \mathcal{H}$, such that for any $h \in \mathcal{H}$, the indicator function $\mathbf{1}\{h \in R(x)\}$ is measurable. The hypotheses $h \in R$ are called the *rejected null hypotheses* of the procedure R.

Below, we will consider, as is usually the case, multiple testing procedures R which can be written as function $R(\mathbf{p})$ of a family of p-values $\mathbf{p} = (p_h, h \in \mathcal{H})$. In particular, we suppose that for each null hypothesis $h \in \mathcal{H}$, there exists a *p-value* function p_h, defined as a measurable function $p_h : \mathcal{X} \to [0, 1]$, such that if h is true, the distribution of $p_h(X)$ is stochastically lower bounded by a uniform random variable on $[0, 1]$:

$$\forall h \in \mathcal{H}_0\,, \forall t \in [0,1]\,, \;\; \mathbb{P}\left[p_h(X) \leq t\right] \leq t\,.$$

There is no universal type I error measure for multiple testing procedures. In this paper, we will however exclusively focus on the false discovery rate (FDR) introduced by Benjamini and Hochberg [1] and which has since become a widely used standard. The FDR is defined as the averaged proportion of true null hypotheses in the set of all the rejected hypotheses.

To measure the volume of subsets of $\mathcal{H}$, we introduce Λ, a finite positive measure on $\mathcal{H}$. In the remainder of this paper we will assume such a volume measure has been fixed and denote, for any subset $S \subset \mathcal{H}$, $|S| = \Lambda(S)$.

Definition 2.2 (False discovery rate)**.** Let R be a multiple testing procedure on $\mathcal{H}$. The false discovery rate (FDR) is defined as

$$\mathrm{FDR}(R) := \mathbb{E}\left[\frac{|R \cap \mathcal{H}_0|}{|R|}\mathbf{1}\{|R| > 0\}\right]\,. \tag{2}$$

Remark 2.3. Throughout this paper we will use the following convention: whenever there is an indicator function inside an expectation, this has logical priority over any other factor appearing in the expectation. What we mean is that if other factors include expressions that may not be defined (such as the ratio $\frac{0}{0}$) outside of the set defined by the indicator, this is safely ignored. In other terms, any indicator function present implicitly entails that we perform integration over the corresponding set only. This results in more compact notation, such as in the above definition.

Remark 2.4. (Weighted FDR in the finite case) When the space of hypotheses is finite, the "usual" FDR in multiple testing literature is the one defined using $|.|$ equal to the counting measure (cardinality) on a finite space. Here, expression (2) allows to use different weights $\Lambda(h)$ for particular hypotheses h, leading to the so-called "weighted FDR". As discussed in [2] and [7], controlling the "weighted FDR" can be of interest in some specific applications. For instance, in situation where each hypothesis concerns a whole cluster of voxels in a brain map, it is relevant to increase the importance of large discovered clusters when counting the discoveries in the FDR.

2.2 Self-consistent procedures and FDR control

It is commonly the case that multiple testing procedures are defined as level sets of the p-values:

$$R = \{h \in \mathcal{H} \mid p_h \leq t\}, \tag{3}$$

where t is a given (possibly data-dependent) threshold.

At an intuitive level, if the goal is to upper bound $\mathrm{FDR}(R)$ given by (2), we can allow more type I errors (and hence a larger threshold t) whenever the number of rejections $|R|$ is larger. Therefore, a natural idea is to choose a threshold $t = \Delta(h, |R|)$ as a non-decreasing function of $|R|$ (possibly depending on h). However, one problem with this heuristics is that it apparently leads to a problematic self-referring definition of the procedure (3).

In order to formalize this approach, we first introduce a general notion of thresholding-based multiple testing procedures which generalizes the form (3) to the case where the threshold t can depend on h and on a global "rejection volume" parameter r. We then introduce the "self-consistency condition" which avoids the self-referring problem mentioned above.

Definition 2.5 (Threshold collection)**.** A *threshold collection* Δ is a function

$$\Delta : (h, r) \in \mathcal{H} \times \mathbb{R}^+ \mapsto \Delta(h, r) \in \mathbb{R}^+,$$

which is non-decreasing in its second variable. A *factorized threshold collection* is a threshold collection Δ with the particular form: $\forall (h, r) \in \mathcal{H} \times \mathbb{R}^+$,

$$\Delta(h, r) = \alpha \pi(h) \beta(r)\,,$$

where $\pi : \mathcal{H} \to [0, 1]$ is called the *weight function* and $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ is a non-decreasing function called the *shape function*.

The reason for introducing the general threshold collection Δ is to allow us more flexibility in order to deal with different settings, as will appear in the sequel. When a threshold collection Δ does not depend on h we write just $\Delta(r)$ instead of $\Delta(h, r)$.

Definition 2.6 (Thresholding-based multiple testing procedure)**.** Given a threshold collection Δ, the Δ-thresholding-based multiple testing procedure at rejection volume r is defined as

$$L_\Delta(r) := \{h \in \mathcal{H} \mid p_h \leq \Delta(h, r)\}. \tag{4}$$

Definition 2.7 (Self-consistency condition)**.** Given a threshold collection Δ, a multiple testing procedure R satisfies the self-consistency condition with respect to the threshold collection Δ if:

$$R \subset L_\Delta(|R|). \tag{$\mathbf{SC}(\Delta)$}$$

The self-consistency condition has the following intuitive motivation: assume R is of the form $\{h \in \mathcal{H} \mid p_h \leq t\}$ and suppose that the number of rejected null hypotheses is known by advance and is equal to a deterministic quantity $|R| = r$. Consider the problem of choosing the threshold t such that the FDR of R is less than or equal to α. Since the expected number of false rejections of R is bounded above by tm, the FDR is upper bounded by tm/r, so that we want to choose $t \leq \alpha r/m$. Therefore, we get the condition $R \subset \{h \in \mathcal{H} \mid p_h \leq \alpha r/m\}$ i.e. $R \subset L_\Delta(|R|)$ for the linear threshold collection $\Delta(h, r) = \alpha r/m$. (Obviously, since $|R|$ is really a random variable, the above reasoning is not rigorous.)

More generally, given a threshold collection Δ, let us interpret loosely the procedure $L_\Delta(r)$ as the rejection set given by a reference or *oracle* procedure if the volume of the rejected hypotheses was known to be r. The self-consistency condition states that the rejection set R should be more conservative than the oracle procedure based on Δ if $r = |R|$ had been known in advance. This view is close in spirit to the *post hoc* interpretation of the classical linear step-up procedure (see Section 3.3 of [1]). Finally, note that condition $\mathbf{SC}(\Delta)$ implies a restriction on the possible volume of R: namely, it implies that $|L_\Delta(|R|)| \geq |R|$, hence $|R|$ must belong to the set $\{r \mid |L_\Delta(r)| \geq r\}$. We will return to this point in Section 4.

The following elementary but fundamental result is our main cornerstone linking the role of the self-consistency condition and of inequalities of the form (1).

Proposition 2.8. *Let $\Delta(h,r) = \alpha\pi(h)\beta(r)$ be a factorized threshold collection on a null hypotheses set $\mathcal{H}$, with shape function β. Assume the multiple testing procedure R satisfies*
(i) the self-consistency condition $\mathbf{SC}(\Delta)$.
(ii) for any $h \in \mathcal{H}_0$ and any $c > 0$, the inequality $\mathbb{E}\left[\frac{\mathbf{1}\{p_h \leq c\beta(|R|)\}}{|R|}\right] \leq c$.

Then $FDR(R) \leq \alpha\pi(\mathcal{H}_0)$, where $\pi(\mathcal{H}_0) = \sum_{h\in\mathcal{H}_0} \Lambda(h)\pi(h)$.

Proof.

$$
\begin{aligned}
\mathrm{FDR}(R) = \mathbb{E}\left[\frac{|R\cap\mathcal{H}_0|}{|R|}\mathbf{1}\{|R|>0\}\right] &= \sum_{h\in\mathcal{H}_0} \Lambda(h)\mathbb{E}\left[\frac{\mathbf{1}\{h\in R\}}{|R|}\right] \\
&\leq \sum_{h\in\mathcal{H}_0} \Lambda(h)\mathbb{E}\left[\frac{\mathbf{1}\{p_h \leq \alpha\pi(h)\beta(|R|)\}}{|R|}\right] \\
&\leq \alpha \sum_{h\in\mathcal{H}_0} \Lambda(h)\pi(h),
\end{aligned}
$$

where we have used successively conditions (i) and (ii) for the two above inequalities. □

Remark 2.9.

1. As soon as π is a density with respect to Λ, i.e. $\sum_{h\in\mathcal{H}} \Lambda(h)\pi(h) = 1$, Proposition 2.8 gives multiple testing procedures with a FDR less than or equal to α. In particular, when $\mathcal{H}$ is of finite cardinality m endowed with the counting measure, choosing π constant equal to $1/m$ gives FDR $\leq \alpha m_0/m \leq \alpha$ (where m_0 denotes the number of true null hypotheses).

2. Proposition 2.8 can be readily extended to the case where we use different volume measures for the numerator and denominator of the FDR. However, since it is no clear to us whether such an extension would be of practical interest, we choose in this paper to deal only with a single volume measure.

Once the threshold collection Δ is fixed, the self-consistency condition (i) concerns only the algorithm itself (and not the random structure of the problem). Condition (ii) on the other hand, involves both the algorithm and the statistical nature of the problem. However, our main point in the next section is that the latter condition can be checked under a weak generic assumption on the algorithm itself (namely that $|R(\mathbf{p})|$ is nonincreasing function of the p-values), and mainly depends on the dependency structure of the p-values. Hence, the interest of this proposition is that it effectively separates the problem of FDR control into a purely *algorithmic* and an (almost) purely *probabilistic* sufficient condition.

3 Dependency conditions

In this section, we show how to check condition (ii) of Proposition 2.8 under different types of assumptions on the dependency of the p-values. In each case we will be reduced to proving an inequality of the form (1) under different conditions. We will follow the different types of dependencies considered in [3], namely independent, positive dependency under the PRDS condition and arbitrarily dependent p-values; this will result in specific choices for the shape function β. The only additional assumption we will make on the procedure R itself is that it has nonincreasing volume as a function of the p-values (and this assumption will not be required in the case of arbitrarily dependent p-values).

3.1 Independent case

We first consider the case where the family of p-values $\mathbf{p} = (p_h, h \in \mathcal{H})$ is independent.

Proposition 3.1. *Assume that the collection of p-values $\mathbf{p} = (p_h, h \in \mathcal{H})$ forms an independent family of random variables. Let $R(\mathbf{p})$ be a multiple testing procedure such that $|R(\mathbf{p})|$ is non-increasing in each p-value p_h with $h \in \mathcal{H}_0$. Then condition (ii) of Proposition 2.8 is satisfied for the linear shape function $\beta(x) = x$.*

Proof. For each $h \in \mathcal{H}$, we denote by $\mathbf{p}_{-h}$ the collection of p-values $(p_{h'}, h' \neq h)$. Fix $h \in \mathcal{H}_0$. We have:

$$\mathbb{E}\left[\frac{\mathbf{1}\{p_h \leq c|R|\}}{|R|}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\mathbf{1}\{p_h \leq c|R(\mathbf{p})|\}}{|R(\mathbf{p})|}\middle|\mathbf{p}_{-h}\right]\right].$$

By the independence assumption, the distribution of p_h conditionally to $\mathbf{p}_{-h}$ is identical to its marginal and therefore stochastically lower bounded by a uniform distribution. The value of $\mathbf{p}_{-h}$ being held fixed, $|R(\mathbf{p})| = |R((\mathbf{p}_{-h}, p_h))|$ can be written as a nonincreasing function of p_h by the assumption on R. Conditionally to $\mathbf{p}_{-h}$, putting $U = p_h$ and $|R| = g(p_h)$, we are thus reduced to proving the following inequality:

$$\mathbb{E}\left[\frac{\mathbf{1}\{U \leq cg(U)\}}{g(U)}\right] \leq c,$$

for U stochastically lower bounded by a uniform distribution and g nonincreasing. Let $\mathcal{U} = \{u \mid cg(u) \geq u\}$, $u^* = \sup \mathcal{U}$ and $C^\star = \inf\{g(u) \mid u \in \mathcal{U}\}$. It is not difficult to check that $u^* \leq cC^*$ (for instance take any non-decreasing sequence $u_n \in \mathcal{U} \nearrow u^*$, so that $g(u_n) \searrow C^*$). If $C^* = 0$, then $u^* = 0$ and the result is trivial. Otherwise, we have

$$\mathbb{E}\left[\frac{\mathbf{1}\{U \leq cg(U)\}}{g(U)}\right] \leq \frac{\mathbb{P}(U \in \mathcal{U})}{C^*} \leq \frac{\mathbb{P}(U \leq u^*)}{C^*} \leq \frac{u^*}{C^*} \leq c.$$

□

Remark 3.2. Note that Proposition 3.1 is still valid under the slightly weaker assumption that for all $h \in \mathcal{H}_0$, p_h is independent of the family $(p_{h'}, h' \neq h)$ (in particular, the p-values of $(p_h, h \in \mathcal{H}_1)$ need not be mutually independent).

3.2 Positive dependencies (PRDS)

We now consider an extension of the previous result where instead of requiring $|R|$ to be a nonincreasing function of $\mathbf{p}$ and the p-values to be independent, we reach the same conclusion as previously under the weaker hypothesis that $|R|$ is stochastically decreasing with respect to each p-value associated to a true null hypothesis.

Proposition 3.3. *Let R be a multiple testing procedure such that for any $h \in \mathcal{H}_0$, the conditional distribution of $|R|$ given $p_h \leq u$ is stochastically decreasing in u:*

$$\text{for any } r \geq 0\,, \text{ the function } u \mapsto \mathbb{P}(|R| < r \mid p_h \leq u) \text{ is nondecreasing}. \tag{5}$$

(The above probability is taken to be 0 for values of u such that $\mathbb{P}\left[p_h \leq u\right] = 0$.) Then, condition (ii) of Proposition 2.8 is satisfied for the linear shape function $\beta(x) = x$.

Proof. Putting $U = p_h$ and $V = |R|$, what we need to prove is now the relation

$$\mathbb{E}\left[\frac{\mathbf{1}\{U \leq cV\}}{V}\right] \leq c\,. \tag{6}$$

for U, V two non-negative real variables such that U is stochastically lower bounded by a uniform distribution, and the conditional distribution of V given $U \leq u$ is stochastically decreasing in u. Fix some $\varepsilon > 0$ and some $\rho \in (0,1)$ and choose K large enough so that $\rho^K < \varepsilon$. Put $v_0 = 0$ and $v_i = \rho^{K+1-i}$ for $1 \leq i \leq 2K+1$. The following chain of inequalities holds:

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbf{1}\{U \leq cV\}}{V \vee \varepsilon}\right] &\leq \sum_{i=1}^{2K+1} \frac{\mathbb{P}(U \leq cv_i; V \in [v_{i-1}, v_i))}{v_{i-1} \vee \varepsilon} + \varepsilon \\
&\leq c \sum_{i=1}^{2K+1} \frac{\mathbb{P}(U \leq cv_i; V \in [v_{i-1}, v_i))}{\mathbb{P}(U \leq cv_i)} \frac{v_i}{v_{i-1} \vee \varepsilon} + \varepsilon \\
&\leq c\rho^{-1} \sum_{i=1}^{2K+1} \mathbb{P}(V \in [v_{i-1}, v_i) \mid U \leq cv_i) + \varepsilon \\
&= c\rho^{-1} \sum_{i=1}^{2K+1} \big(\mathbb{P}(V < v_i \mid U \leq cv_i) - \mathbb{P}(V < v_{i-1} \mid U \leq cv_i)\big) + \varepsilon \\
&\leq c\rho^{-1} \sum_{i=1}^{2K+1} \big(\mathbb{P}(V < v_i \mid U \leq cv_i) - \mathbb{P}(V < v_{i-1} \mid U \leq cv_{i-1})\big) + \varepsilon \\
&\leq c\rho^{-1} + \varepsilon\,.
\end{aligned}$$

We obtain the conclusion by letting $\rho \to 1$, $\varepsilon \to 0$ and applying the monotone convergence theorem. ∎

We now give conditions providing that R satisfies (5). For this, we recall the definition of positive regression dependency on each one from a subset (PRDS) (introduced in [3], where its relationship to other notions of positive dependency is also discussed). Remember that a subset $D \subset [0,1]^{\mathcal{H}}$ is called *nondecreasing* if for all $\mathbf{z}, \mathbf{z}' \in [0,1]^{\mathcal{H}}$ such that $\mathbf{z} \leq \mathbf{z}'$ (i.e. $\forall h \in \mathcal{H}, z_h \leq z'_h$), we have $\mathbf{z} \in D \Rightarrow \mathbf{z}' \in D$.

Definition 3.4. For $\mathcal{H}'$ a subset of $\mathcal{H}$, the p-values of $\mathbf{p} = (p_h, h \in \mathcal{H})$ are said to be *positively regressively dependent on each one from* $\mathcal{H}'$ (denoted in short by *PRDS on* $\mathcal{H}'$), if for any $h \in \mathcal{H}'$,

$$\text{for any measurable nondecreasing set } D \subset [0,1]^{\mathcal{H}},\ u \mapsto \mathbb{P}(\mathbf{p} \in D \mid p_h = u) \text{ is nondecreasing.} \quad (7)$$

We can now state that the PRDS condition on the p-values implies condition (ii) of Proposition 2.8.

Corollary 3.5. *Suppose that the p-values of* $\mathbf{p} = (p_h, h \in \mathcal{H})$ *are PRDS on* $\mathcal{H}_0$ *, and consider a multiple testing procedure* R *such that* $|R(\mathbf{p})|$ *is nonincreasing in each p-value. Then* R *satisfies condition (ii) of Proposition 2.8 for the linear shape function* $\beta(x) = x$ *.*

Proof. We merely check that condition (5) of Proposition 3.3 is satisfied. For any fixed $r \geq 0$, put $D = \left\{\mathbf{z} \in [0,1]^{\mathcal{H}} \mid |R(\mathbf{z})| < r\right\}$. It is clear from the assumptions on R that D is a nondecreasing measurable set. Then condition (7) implies (5); this argument was also used in [3] with a reference to Lehmann [8]. We give here a short self-contained proof for completeness. Under the PRDS condition, for all $u \leq u'$, putting $\gamma = \mathbb{P}\left[p_h \leq u \mid p_h \leq u'\right]$,

$$\begin{aligned}
\mathbb{P}\left[\mathbf{p} \in D \mid p_h \leq u'\right] &= \mathbb{E}\left[\mathbb{P}\left[\mathbf{p} \in D \mid p_h\right] \mid p_h \leq u'\right] \\
&= \gamma \mathbb{E}\left[\mathbb{P}\left[\mathbf{p} \in D \mid p_h\right] \mid p_h \leq u\right] + (1-\gamma)\mathbb{E}\left[\mathbb{P}\left[\mathbf{p} \in D \mid p_h\right] \mid u < p_h \leq u'\right] \\
&\geq \mathbb{E}\left[\mathbb{P}\left[\mathbf{p} \in D \mid p_h\right] \mid p_h \leq u\right] = \mathbb{P}\left[\mathbf{p} \in D \mid p_h \leq u\right] .
\end{aligned}$$

□

Interestingly, the probabilistic inequality (6) under the PRDS condition is useful beyond the framework of Proposition 2.8. In Appendix B, we prove that this same inequality can be used to prove FDR control for a step-down procedure proposed in [9] and [10] (for which condition (ii) of Proposition 2.8 is not satisfied in general), under the PRDS condition.

3.3 Unspecified dependencies

We now consider a totally generic setting with no assumptions on the dependency structure between the p-values nor on the structure of the multiple testing procedure R. In order to ensure FDR control with Proposition 2.8 in this general situation, we require a more restrictive self-consistency condition and a more conservative inequality in condition (ii), both using a shape function $\beta(r) \leq r$ of a particular form. This form of shape function was initially introduced in [11], where some ties were exposed between multiple testing (under unspecified dependencies) and statistical learning theory, and where an early version of the self-consistency condition was implicitly considered.

Proposition 3.6. *Let* β *be a shape function of the following form:*

$$\beta(r) = \int_0^r u d\nu(u) , \quad (8)$$

where ν *is an arbitrary, but fixed in advance probability distribution on* $(0, \infty)$ *. Then condition (ii) of Proposition 2.8 is satisfied for this shape function, for any multiple testing procedure* R *.*

Proof. This time we need to prove the probabilistic inequality

$$\mathbb{E}\left[\frac{\mathbf{1}\{U \le c\beta(V)\}}{V}\right] \le c\,, \tag{9}$$

for U, V arbitrarily dependent random variables such that U is stochastically lower bounded by a uniform variable, and any $c > 0$.

Rewriting for any $z > 0$, $1/z = \int_0^{+\infty} v^{-2}\mathbf{1}\{v \ge z\}dv$, and using Fubini's theorem:

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbf{1}\{U \le c\beta(V)\}}{V}\right] &= \mathbb{E}\left[\int_0^{+\infty} v^{-2}\mathbf{1}\{v \ge V\}\mathbf{1}\{U \le c\beta(V)\}dv\right] \\
&= \int_0^{+\infty} v^{-2}\mathbb{E}\left[\mathbf{1}\{v \ge V\}\mathbf{1}\{U \le c\beta(V)\}\right]dv \\
&\le \int_0^{+\infty} v^{-2}\mathbb{P}\big(U \le c\beta(v)\big)dv \\
&\le c\int_0^{+\infty} v^{-2}\beta(v)dv = c\int_{u\ge 0} u \int_{v \ge 0} \mathbf{1}\{u \le v\}v^{-2}dvd\nu(u) = c\,.
\end{aligned}$$

□

4 Step-up procedures

In this section, we state that step-up procedures are instances of procedures satisfying the self-consistency condition. Therefore, applying the results of previous sections, we will directly derive FDR bounds for classical step-up procedures.

4.1 Definition

We give here a general definition of step-up procedures, related to the point of view put forward in Theorem 2 of [1]: a step-up procedure is defined as the maximal set satisfying a self-consistency condition $\mathbf{SC}(\Delta)$, and can be caracterized as follows:

Definition 4.1 (Step-up procedure)**.** Let Δ be a threshold collection. The *step-up multiple testing procedure* R associated to Δ, is given by either of the following equivalent definitions:

$$\begin{aligned}
&(i) \quad R = L_\Delta(\hat{r})\,, \text{ where } \hat{r} := \max\{r \ge 0 \mid |L_\Delta(r)| \ge r\} \\
&(ii) \quad R = \bigcup \{A \subset \mathcal{H} \mid A \text{ satisfies } \mathbf{SC}(\Delta)\ \}
\end{aligned}$$

Additionally, $\widehat{r}$ satisfies $|L_\Delta(\widehat{r})| = \widehat{r}$; equivalently the step-up procedure R satisfies $\mathbf{SC}(\Delta)$ with equality.

Proof of the equivalence between (i) *and* (ii). Note that since Δ is assumed to be nondecreasing in its second variable, $L_\Delta(r)$ is a nondecreasing set as a function of $r \ge 0$. Therefore, $|L_\Delta(r)|$ is a nondecreasing function of r and the supremum appearing in (i) is indeed a maximum i.e. $|L_\Delta(\hat{r})| \ge \hat{r}$. In fact, it is easy to see that $|L_\Delta(\hat{r})| = \hat{r}$ since this would otherwise contradict the definition of $\widehat{r}$. Hence $L_\Delta(\hat{r}) = L_\Delta(|L_\Delta(\hat{r})|)$, so $L_\Delta(\hat{r})$ satisfies $\mathbf{SC}(\Delta)$ (with equality) and is included in the set union appearing in (ii). Conversely, for any set A satisfying $A \subset L_\Delta(|A|)$, we have $|L_\Delta(|A|)| \geq |A|$, so that $|A| \leq \hat{r}$ and $A \subset L_\Delta(\hat{r})$.

$\square$

The decision point $\hat{r}$ is obtained from the "last right crossing" point between the (nondecreasing) volume of rejected hypotheses $|L_\Delta(.)|$ and the identity function (see an illustration on Figure 1).

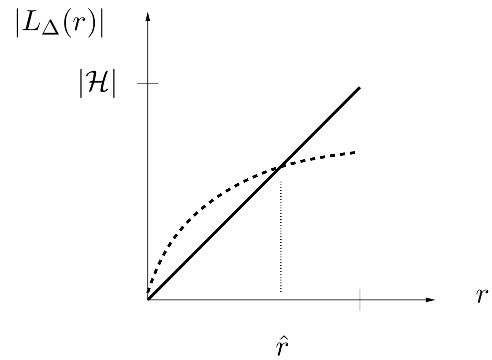

Figure 1: Graphs of $|L_\Delta(\cdot)|$ (dashed line) and of the identity function (solid line). The number of rejected hypotheses of the step-up procedure $\hat{r}$ is obtained on the x-axis. On the y-axis, $|\mathcal{H}|$ ensures that the crossing point corresponding to $\hat{r}$ is the last.

Consider the situation where $|.|$ is the cardinality measure over a finite set of hypotheses of cardinality m, and the threshold collection $\Delta(h, r) = \alpha\pi(h)\beta(r)$ is factorized; then Definition 4.1 is equivalent to the classical "re-ordering-based" definition of a step-up procedure: for any $h \in \mathcal{H}$, denote by $p'_h := p_h/(m\pi(h))$ the *weighted p-value* of h (in the case $\pi(h) = 0$, we put $p'_h = +\infty$ if $p_h > 0$ and $p'_h = 0$ if $p_h = 0$), and consider the ordered weighted p-values i.e. such that

$$p'_{(1)} \leq p'_{(2)} \leq \cdots \leq p'_{(m)}.$$

Since $L_\Delta(r) = \{h \in \mathcal{H} \mid p'_h \leq \alpha\beta(r)/m\}$, the condition $|L_\Delta(r)| \geq r$ is equivalent to $p'_{(r)} \leq \alpha\beta(r)/m$. Hence, the step-up procedure associated to Δ defined in Definition 4.1 rejects all the $\hat{r}$ smallest weighted p-values, where $\hat{r}$ corresponds to the "last right crossing" point between the ordered weighted p-values $p'_{(\cdot)}$ and the threshold $\alpha\beta(\cdot)/m$ (see Figure 2 for an illustration):

$$\hat{r} \quad = \quad \max\left\{r \in \{0, \ldots, m\} \mid p'_{(r)} \leq \alpha\beta(r)/m\right\},$$

with $p'_{(0)} := 0$. If for each hypothesis $h \in \mathcal{H}$, $\pi(h) = 1/m$, we note that the weighted p-values $(p'_h)_h$ are simply the p-values $(p_h)_h$. In particular:

- The step-up procedure associated to $\Delta(r) = \alpha r/m$ is the well-known linear step-up procedure of Benjamini and Hochberg [1] (corresponding to the shape function $\beta(r) = r$).
- The step-up procedure associated to $\Delta(r) = \alpha r/(m(1 + 1/2 + \cdots + 1/m))$ is the distribution-free step-up procedure of Benjamini and Yekutieli [3] (corresponding to the shape function $\beta(r) = r/(1 + 1/2 + \cdots + 1/m)$).

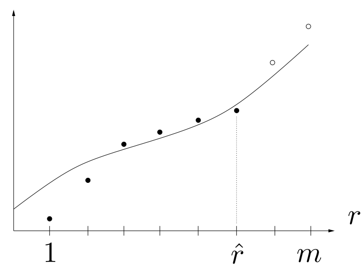

Figure 2: Comparison between the ordered weighted p-values $p'_{(\cdot)}$ (points) and the threshold $\alpha\beta(.)/m$ (solid line). Here the step-up procedure rejects the 6 hypotheses corresponding to the 6 smallest reweighted p-values (solid points) ($\hat{r} = 6$).

Remark 4.2. As appears clearly in our formulation, there are two different ways to obtain "weighted" versions of step-up procedures, by changing respectively the choice of the volume measure Λ and the weight function π :

1. The weight function π allows us to use weighted p-values. For the linear step-up procedure, [4] proved that the choice of this weights has a large impact on the power of the procedures.

2. Definition 4.1 allows us to deal with a volume measure Λ different from the cardinality measure. The resulting step-up procedures will control the weighted FDR, which is a quantity of interest (see Remark 2.4). A "re-ordering-based" interpretation of these procedures is still possible: in this case $\widehat{r}$ is defined in terms of the reordered p-values as the maximum integer r such that $p'_{(r)} \leq \alpha\beta(V(r))/m$, where $V(r) = \left|\left\{h_{(i)}, i \leq r\right\}\right|$ (denoting $h_{(i)}$ the hypothesis corresponding to $p_{(i)}$). Graphically, it corresponds to using different spacings between the p-values on the x-axis of Figure 2, with the k-th spacing equal to the volume measure of the k-th reordered hypothesis.

3. Our formulation allows to combine these two forms of reweighting transparently. In this case, as noticed earlier the weight function π should be a density function with respect to the volume measure Λ .

4.2 Some classical step-up procedures with extensions

A direct consequence of Definition 4.1 is that the step-up procedure associated to Δ satisfies the self-consistency condition $\mathbf{SC}(\Delta)$. Therefore, we can apply the results of Section 2.2 and Section 3 to derive FDR control theorems. We first consider the *(π-weighted) linear step-up procedure*, that is, the step-up procedure associated to the threshold collection $\Delta(h, r) = \alpha\pi(h)r$. We then recover easily some known results on (Λ-weighted) FDR control for this procedure:

Theorem 4.3. *The (π-weighted) linear step-up procedure R satisfies the (Λ-weighted) FDR control: $FDR(R) \leq \pi(\mathcal{H}_0)\alpha$, where $\pi(\mathcal{H}_0) := \sum_{h\in\mathcal{H}_0} \Lambda(h)\pi(h)$, in either of the following cases:*

- *the p-values of $\mathbf{p} = (p_h, h \in \mathcal{H})$ are independent.*

- *the p-values of* $\mathbf{p} = (p_h, h \in \mathcal{H})$ *are PRDS on* $\mathcal{H}_0$.

The two points of the above Theorem 4.3 were initialy proved by [1] and [3] with a uniform π and Λ being the counting measure. For a general Λ and a uniform π, the above result in the independent case was proved in [2]. A proof with a general π and in the independent case was investigated by [4] (Λ being the counting measure). Here, our framework allows for a general and unified version of these results with a concise proof.

Similarly, in the case where the p-values have unspecified dependencies , we can use the results in the previous sections to derive the following theorem:

Theorem 4.4. *Consider R the step-up procedure associated to the factorized threshold collection $\Delta(h,r) = \alpha\pi(h)\beta(r)$, where the shape function β has the following specific form: for each $r \geq 0$,*

$$\beta(r) = \int_0^r u d\nu(u)\,, \tag{10}$$

where ν is some probability distribution on $(0,\infty)$. Then R has its (Λ-weighted) FDR controlled at level $\pi(\mathcal{H}_0)\alpha/m$.

Remark 4.5. Theorem 4.4 can be seen as an extension to the FDR of a celebrated inequality due to Hommel [12] for the family-wise error rate (FWER), which has been widely used in the multiple testing literature (see *e.g.* [13; 10; 14]). Namely, when ν has discrete support $\{1,\ldots,m\}$ and $\mathcal{H} = \mathcal{H}_0$, the above result recovers Hommel's inequality. Note that this specific case corresponds to a "weak control" where we assume that all null hypotheses are true; in this situation the FDR is equal to the FWER. Note also that Theorem 4.4 generalizes without modification to a possibly continuous hypothesis space, as will be detailed in Section 5.

Theorem 4.4 establishes that, under unspecified dependencies between the p-values, there exists a family of step-up procedures that control the false discovery rate. This family depends on the shape function β, itself based on the distribution ν which can be chosen arbitrarily. In paragraph 4.3.2 below, a case of equality in the FDR control of Theorem 4.4 is obtained precisely when the distribution of $|R|$ conditionally to $|R| > 0$ is ν. Thus, at an intuitive level, the distribution ν can be interpreted a prior belief on the final volume of rejections of the procedure. For instance, when Λ is the counting measure and $\mathcal{H}$ is of cardinality m:

- If we do not have any prior belief, we can choose ν uniform on $\{1,\ldots,m\}$ and this gives a quadratic shape function:
  $$\beta(r) = r(r+1)/2m\,.$$
  The obtained step-up procedure is the one proposed in [5]. The latter result is therefore a particular case of Theorem 4.4 (with π uniform).

- If we are in a problem where a small number of rejections is expected (for instance $m_0 = |\mathcal{H}_0|$ "large"), we can choose $\nu(k) = \gamma_m^{-1}k^{-1}$ for $k \in \{1,\ldots,m\}$ with the normalization constant $\gamma_m = \sum_{1\leq i\leq m}\frac{1}{i}$. This gives
  $$\beta(r) = \gamma_m^{-1} r\,,$$
  and we find the distribution-free procedure of Benjamini and Yekutieli [3]. In particular, Theorem 4.4 (with π uniform) is a generalization of Theorem 1.3 of [3].

- If we expect a large number of rejections (for instance m_0 "small"), we can choose $\nu(k) = 2k/(m(m+1))$ for $k \in \{1, \dots, m\}$, which leads to

$$\beta(r) = r(r+1)(2r+1)/(3m(m+1))\,.$$

Of course, many other choices for ν are possible. In Example 4.6, we give several choices of continuous ν, and we plot the graphs of the corresponding shape functions in Figure 3 (page 14). We could also use the discretized versions of the proposed continuous distributions ν; it leads to slightly smaller functions β, but generally with more complex expressions. From Figure 3, it is clear that the choice of the prior has a large impact on the final number of rejections of the procedure. Moreover, since no shape function dominates the others, there is no optimal choice among these prior distributions (the performance of a given procedure will depend on the data). It is then tempting to choose ν in a data-dependent way: showing that the corresponding FDR is still well controlled is an interesting open problem for future research.

Example 4.6 (Some choices for ν and corresponding shape functions β)**.**

1. *Dirac distributions*: $\nu = \delta_\lambda$, with $\lambda > 0$. $\beta(r) = \lambda \mathbf{1}\{r \geq \lambda\}$.

2. *(Truncated-) Gaussian distributions*: ν equals to the distribution of $\max(X, 1)$, where X follows a Gaussian distribution with mean μ and variance σ^2.

   $$\beta(r) = \big[\Phi((r-\mu)/\sigma) - \Phi((1-\mu)/\sigma)\big]\mu + \sigma\big[\exp(-(1-\mu)^2/(2\sigma^2)) - \exp(-(r-\mu)^2/(2\sigma^2))\big]/\sqrt{2\pi},$$

   where Φ is the standard Gaussian cumulative distribution function: $\forall y \in \mathbb{R}$, $\Phi(y) = \mathbb{P}(Y \leq y)$, where $Y \sim \mathcal{N}(0,1)$.

3. *Distributions with a power function density*: $\forall r \geq 0, d\nu(r) = r^\gamma \mathbf{1}\{r \in [1,m]\}dr/\int_1^m u^\gamma du$, $\gamma \in \mathbb{R}$.

   $$\beta(r) = \begin{cases} \frac{\gamma+1}{\gamma+2}\frac{r^{\gamma+2}-1}{m^{\gamma+1}-1} & \text{if} \quad \gamma \neq -1, -2 \\ \frac{r-1}{\log(m)} & \text{if} \quad \gamma = -1 \\ \frac{\log(r)}{1-1/m} & \text{if} \quad \gamma = -2 \end{cases}\,.$$

   As a particular case, when $\gamma = 0$, ν is uniformly distributed on $[1,m]$ and $\beta(r) = (r^2-1)/(2(m-1))$.

4. *Exponential distributions*: $d\nu(r) = (1/\lambda)\exp(-r/\lambda)dr$, with $\lambda > 0$.

   $$\beta(r) = \lambda(1 - \exp(-r/\lambda)) - r\exp(-r/\lambda).$$

4.3 Sharpness of the FDR bounds

In this section, we state results on the sharpness of the FDR bounds for the step-up procedures by demonstrating examples where these bounds are attained. We suppose in this section that the hypotheses set $\mathcal{H}$ is finite, of cardinality m and that Λ is the cardinality measure.

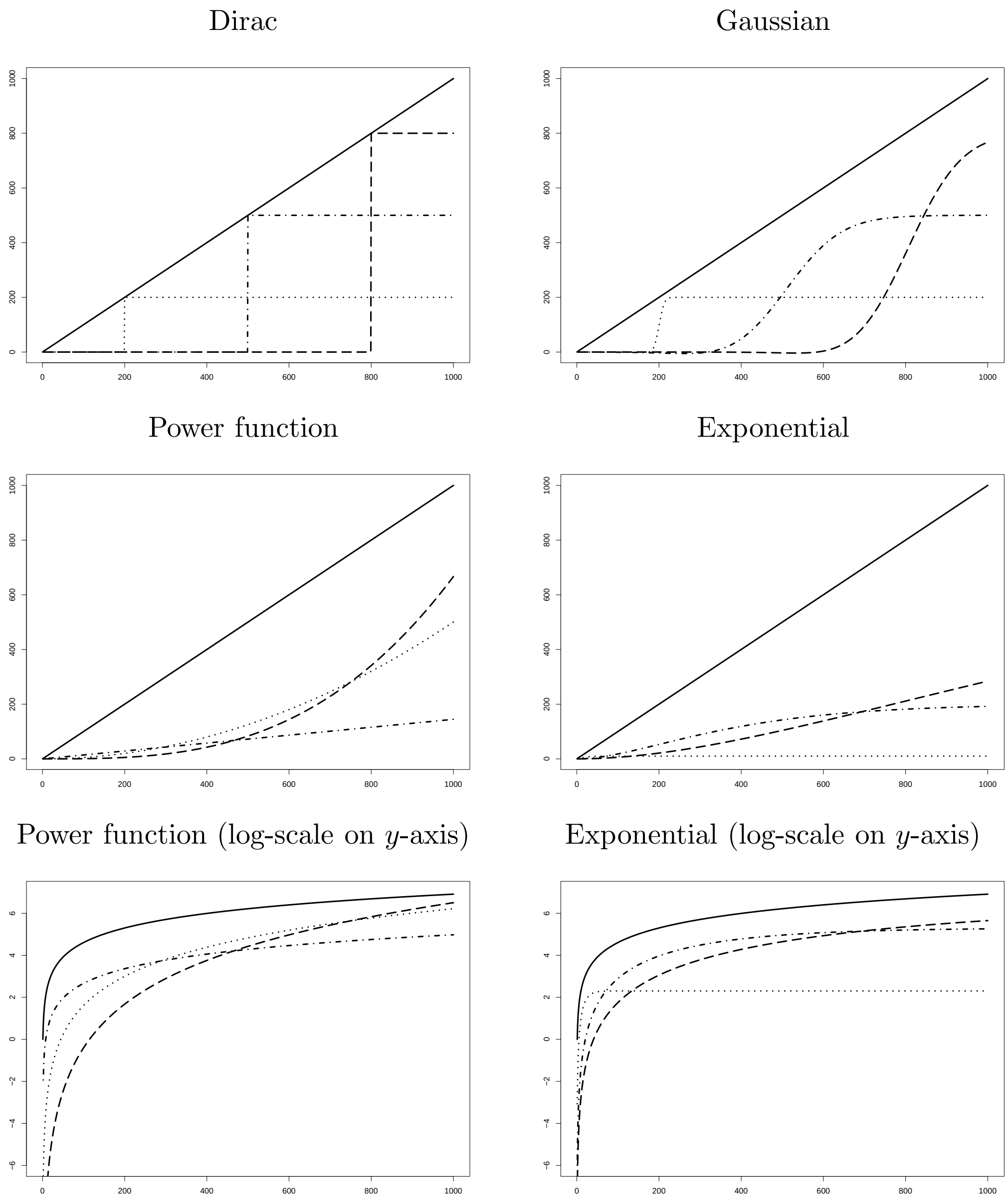

Figure 3: For $m = 1000$ hypotheses, this figure shows several shape functions $\beta(\cdot)$ associated to different prior distributions on $\mathbb{R}^+$ (according to expression (10), see Example 4.6 for the formulas). The top-left graph corresponds to Dirac distributions ($\lambda = 200$ (dotted), $\lambda = 500$ (dashed-dotted), $\lambda = 800$ (dashed)). The top-right graph correspond to Gaussian distributions ($\mu = 200, \sigma = 10$ (dotted); $\mu = 500, \sigma = 100$ (dashed-dotted); $\mu = 800, \sigma = 100$ (dashed)). The middle-left graph (and bottom-left graph with log-scale) corresponds to distributions with a power function density ($\gamma = 0$ (dotted); $\gamma = -1$ (dashed-dotted); $\gamma = 1$ (dashed)). The middle-right graph (and bottom-right graph with log-scale) corresponds to exponential distributions ($\lambda = 10$ (dotted); $\lambda = 200$ (dashed-dotted); $\lambda = 800$ (dashed)). Finally, we plot in solid line the identity function (which is the shape function of the linear step-up procedure).

4.3.1 Independent case

Proposition 4.7. *Assume the p-values family* $(p_h, h \in \mathcal{H})$ *is independent, and for any* $h \in \mathcal{H}_0$, p_h *is exactly a uniform variable on* $[0,1]$. *Then the linear step-up procedure (with* $\forall h, \pi(h) = 1/m$ *and* Λ *the cardinality measure) has a FDR exactly equal to* $|\mathcal{H}_0|\alpha/m$.

The above proposition has been first proved by [6] and [3]. We propose here a self-contained proof using the specificity of the step-up algorithm (and not only the fact that a step-up procedure is self-consistent). To our knowledge, the methodology used in this proof is new and is therefore of self interest.

Proof. For each null hypothesis h, denote R'_{-h} the step-up procedure associated to the threshold collection $\Delta'(r) = \alpha(r+1)/m$ and restricted to the hypotheses of $\mathcal{H}\backslash\{h\}$. Lemma A.1 states that

$$\begin{aligned} h \in R \quad &\Leftrightarrow \quad R = R'_{-h} \cup \{h\} \\ &\Leftrightarrow \quad p_h \leq \alpha(|R'_{-h}|+1)/m. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathrm{FDR}(R) \quad &= \quad \sum_{h\in\mathcal{H}_0} \mathbb{E}\left[\frac{\mathbf{1}\{h \in R\}}{|R|}\right] \\ &= \quad \sum_{h\in\mathcal{H}_0} \mathbb{E}\left[\frac{\mathbf{1}\{p_h \leq \alpha(|R'_{-h}|+1)/m\}}{|R'_{-h}|+1}\right] \\ &= \quad \sum_{h\in\mathcal{H}_0} \mathbb{E}\left[\frac{\mathbb{P}\left[p_h \leq \alpha(|R'_{-h}|+1)/m \mid |R'_{-h}|\right]}{|R'_{-h}|+1}\right]. \end{aligned}$$

For any $h \in \mathcal{H}_0$, we use simultaneously:

- $|R'_{-h}|$ depends only on the p-values of $(p_{h'}, h' \neq h)$,
- p_h has a uniform distribution conditionally to $\mathbf{p}_{-h}$ (independence assumption)
- $\alpha(|R'_{-h}|+1)/m \leq 1$,

to deduce that $\mathbb{P}\left[p_h \leq \alpha(|R'_{-h}|+1)/m) \mid |R'_{-h}|\right] = \alpha(|R'_{-h}|+1)/m$. The result follows. □

4.3.2 Unspecified dependencies

Proposition 4.8. *Consider the step-up procedure (with* $\forall h, \pi(h) = 1/m$ *and* Λ *the cardinality measure) obtained with a shape function* β *of the form* (10)*, where the distribution* ν *has its support included in* $\{1, \ldots, m\}$. *Then for* $\mathcal{H}_0 = \mathcal{H}$ *there exists a joint distribution for the p-values such that* $FDR(R) = \alpha$ *and such that the distribution of the cardinality* $|R|$ *of rejected hypotheses conditionally to* $|R| > 0$ *is exactly* ν .

Using $\beta(x) = x\gamma_m^{-1}$, with $\gamma_m = \sum_{i\leq m} \frac{1}{i}$, the latter proposition shows in particular that the linear step-up procedure does not control the FDR at level α for an arbitrary dependency structure between the p-values (its FDR can be equal to $\alpha\gamma_m > \alpha$). This proposition has been proved in [13] (Lemma 3.1 (ii)). We give here a self-contained proofs for completeness.

Proof. We build the joint distribution of the p-values in the following way: take a random variable K such that for all $k \in \{1, \ldots, m\}$, $\mathbb{P}(K = k) = \alpha\nu(k)$ and $\mathbb{P}(K = 0) = 1 - \alpha$. Conditionnally to K, we choose I a subset of $\mathcal{H}$ uniformly distributed among the subsets of K elements of $\mathcal{H}$. Conditionally to I (and K), we choose (all independently)

$$\forall h \in I, p_h \text{ uniform in } [\alpha\beta(K-1)/m, \alpha\beta(K)/m)$$
$$\forall h \notin I, p_h \text{ uniform in } [\alpha\beta(m)/m, 1],$$

We check that unconditionally, each p_h is uniform on $[0, 1]$: for all $k \in \{1, \ldots, m\}$,

$$\mathbb{P}(p_h \in [\alpha\beta(k-1)/m, \alpha\beta(k)/m)) = \mathbb{P}(K = k, h \in I) = \mathbb{P}(h \in I \mid K = k)\alpha\nu(k) = \alpha k\nu(k)/m,$$

and by definition of β, $\beta(k) - \beta(k-1) = k\nu(k)$. Finally, we just have to remark that the step-up procedure R rejects exactly the K hypotheses in $[\alpha\beta(K-1)/m, \alpha\beta(K)/m)$ to conclude $\mathrm{FDR}(R) = \mathrm{FWER}(R) = \mathbb{P}(K > 0) = \alpha$. □

4.3.3 Gaussian data

In the previous section it was shown that the FDR bound for unspecified dependencies can be attained in a specific situation. The natural question arises whether this situation is likely to happen in practice. Namely, the uncorrected linear step-up procedure is often used in practice even when it is not exactly known whether the independence or the PRDS assumption of the p-values holds. In this section we briefly study what happens for Gaussian data; on Figure 4 we show a very simple example with 2 hypotheses assumed to be jointly Gaussian, centered, with unit covariance and a correlation coefficient ρ. We use the usual one-sided z-test for the mean and draw the FDR as a function or ρ (in this case since both null hypotheses are true the FDR coincides with the family-wise error, FWE).

We observe that for negative ρ, the FDR is always larger than the target level 25%. However, it is striking that the nominal level always stays smaller than 26%, to be compared to the theoretical "worst case" of 37.5% of the last section: hence we are here quite far from reaching the worst case. We performed additional limited experiments in higher dimension for Gaussian data having constant covariance, with the same target level of 25%, and found that independently of the dimension or the covariance level the FDR seems to stay consistently under 26%. Hence, for Gaussian data the linear step-up appears to be quite robust. Since Gaussian data is such a fundamental setting, this suggests that it would be of interest to study from a theoretical point of view whether a sharper bound on FDR can be established in general for Gaussian data with arbitrary covariance matrix (with unit diagonal); or in other words, when the repartition function of the p-value family forms a d-dimensional Gaussian copula.

5 FDR control over a continuous space of hypotheses

An interesting advantage of our approach to proving FDR control under the self-consistency condition is that most results can be adapted to the case where we consider a continuous space of hypotheses to test. In this framework it is particularly relevant to consider an arbitrary "volume measure" Λ since obviously the cardinality measure is not adapted. Also, in this setting, while it seems considerably more difficult to define properly step-up procedures in the traditional sense via reordering of the p-values, our Definition 4.1 carries over without change

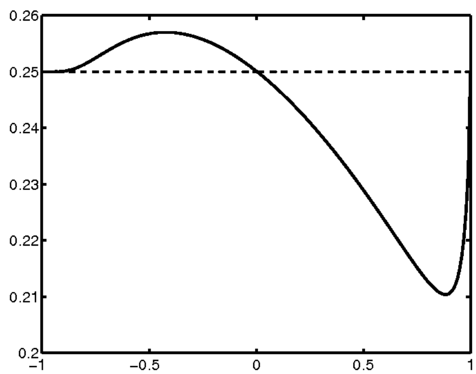

Figure 4: The actual FDR=FWE of the linear step-up procedure ($\alpha = 0.25$) for 2 Gaussian hypotheses, both nulls being true. The x-axis is the correlation coefficient between the two coordinates.

to a continuous setting. While results for continuous spaces of hypotheses are certainly more of a theoretical than practical interest (in practice one can argue that "everything is always discrete"), the continuous setting can be more appropriate for theoretical understanding of multiple testing procedures when the underlying observation is modelled as a random process (as is the case for example in [15]).

5.1 Setting for a continuous space of hypotheses

The main reason why we have restricted our attention to discrete (at most countable) hypotheses spaces in the rest of the paper is to avoid having to specify right away the necessary technical details related for example to measurability issues when dealing with a continuous space. For the purposes of this section we now give more precise assumptions as follows:

- The space of hypotheses $\mathcal{H}$ is now assumed to be endowed with a σ-algebra $\mathfrak{H}$, and Λ is a finite measure on $(\mathcal{H}, \mathfrak{H})$. The set $\mathcal{H}_0$ of true hypotheses is assumed to be measurable.
- A multiple testing procedure $R : \mathcal{X} \to \mathfrak{H}$ is assumed to be such that the indicator function $(x, h) \mapsto \mathbf{1}\{h \in R(x)\}$ is jointly measurable in its variables (implying $\forall x \in \mathcal{X}, R(x) \in \mathfrak{H}$).
- The family of p-values, represented as a bivariate function $(x, h) \mapsto p_h(x)$ is assumed to be jointly measurable in its variables.
- A level function $\Delta(h, r)$ is also assumed to be jointly measurable in its variables.

These assumptions will be supposed to hold for the remainder of this section. Note that we have avoided to define explicitly a measured space structure for the value set of joint families of p-values, i.e. $[0,1]^{\mathcal{H}}$, only requiring that $p_h(x)$ is a jointly measurable function. The reason is that the observation space $\mathcal{X}$ (which can itself be a process space) is used as the reference for measurability. Thus, when we assume that a testing procedure R is a function of $\mathbf{p}$, the measurability assumption above is to be checked directly for R as a function of the observation x, $R = R(\mathbf{p}(x))$. If needed, the space $[0,1]^{\mathcal{H}}$ could be explicitly endowed with the largest sigma-algebra making $\mathbf{p}(x) = (p_h(x), h \in \mathcal{H}) \in [0,1]^{\mathcal{H}}$ a measurable function of x.

5.2 Which results carry over to the continous setting?

Under these (quite mild) conditions it can be checked that Proposition 2.8 still hold. The reference multiple testing procedure L_Δ used in the self-consistency condition and defined by (4) satisfies the measurability conditions required above. The function $x \mapsto |R(x)| = \Lambda(R(x))$ is measurable by Fubini's theorem. Then, our Proposition 2.8 carries over to the continuous case when we replace the discrete sums over $\mathcal{H}_0$ by integrals for the measure Λ. The above joint measurability assumptions ensure that we can use Fubini's theorem, which is the only requirement for the proof.

For what concerns the results in Section 3, first note that assuming independence of the p-values is not possible in a continuous setting in general, as it would conflict with the measurability assumption on $p_h(x)$. Besides, it seems difficult to imagine a continuous setting where such an independence assumption would be relevant. In the setting of arbitrary dependencies on the other hand, Proposition 3.6 carries over directly without change to the continuous case under the above measurability assumptions.

In the case of positive dependencies, the general Proposition 3.3 carries over as well, but the difficulty is then to suitably define the PRDS assumption in this setting in order to have a condition that can be reasonably easy to check. We propose in the following section two possible approaches: the first generalizes directly the PRDS definition to a continuous set of random variables, and the second considers the weaker assumption that any finite subfamily is PRDS.

5.3 The PRDS assumption in a continuous setting

The most direct way to generalize the PRDS definition (Definition 3.4) to the continuous setting is the following:

Definition 5.1. (Strong continuous PRDS condition) Denote $\mathbf{p}(x)$ the family $(p_h(x), h \in \mathcal{H}) \in [0,1]^{\mathcal{H}}$. For $\mathcal{H}'$ a subset of $\mathcal{H}$, the p-values of $\mathbf{p} = (p_h, h \in \mathcal{H})$ are said to be *strongly PRDS on* $\mathcal{H}'$, if for any $h \in \mathcal{H}'$, for any nondecreasing set D in $[0,1]^{\mathcal{H}}$ such that the preimage $\mathbf{p}^{-1}(D)$ is measurable in $\mathcal{X}$, the function

$$u \in [0,1] \mapsto \mathbb{P}(\mathbf{p} \in D \mid p_h = u) \tag{11}$$

is nondecreasing.

Equivalently, this definition is stricly the same as Definition 3.4 once we endow the space $[0,1]^{\mathcal{H}}$ with the largest sigma-algebra such that $\mathbf{p} : \mathcal{X} \to [0,1]^{\mathcal{H}}$ is measurable.

It is important to underline that we do not, in fact, require the existence of a full conditional distribution on (11), which would demand additional assumptions on the underlying space. Rather, the conditional conditional probability in (11) is to be interpreted as the conditional expectation $\mathbb{E}\left[\mathbf{1}\{\mathbf{p}(x) \in D\} | p_h(x) = u\right]$, which for any fixed D is always well-defined as a p_h-measurable random variable. The strong PRDS condition is thus the requirement that there exists a version of this random variable which is nondecreasing as a function of p_h.

The following result is derived from the generalization of Corollary 3.5 and of Proposition 2.8 to the continuous setting in the strong PRDS case.

Theorem 5.2. *Assume the continuous setting assumptions of Section 5.1 are satisfied. Fix a factorized threshold collection $\Delta(h,r) = \alpha\pi(h)r$, where π is a nonnegative measurable* *function. Suppose that the p-values of* $\mathbf{p} = (p_h, h \in \mathcal{H})$ *are strongly PRDS on* $\mathcal{H}_0$ *and consider a self-consistent multiple testing procedure* R *such that* $|R(\mathbf{p})|$ *is nonincreasing in each p-value. Then* R *has a FDR less than or equal to* $\pi(\mathcal{H}_0)\alpha$*, where* $\pi(\mathcal{H}_0) = \int_{\mathcal{H}_0} \pi(h) d\Lambda(h)$*.*

Proof. It suffices to check by inspection that the proof of Corollary 3.5 also holds in the continuous setting. First we check that the preimage through $\mathbf{p}$ of the nondecreasing set $D = \left\{\mathbf{z} \in [0,1]^{\mathcal{H}} | \Lambda(R(\mathbf{z})) \leq r\right\}$ is measurable:

$$\mathbf{p}^{-1}(D) = \{x \in \mathcal{X} \mid \Lambda(R(\mathbf{p}(x))) \leq r\} \,.$$

By the general measurability assumptions on R and Fubini's theorem, the above set is measurable. Finally, one can check that the proof of Corollary 3.5 does not make use of the properties of full conditional distributions so that conditional expectations are sufficient to establish the result. □

In Section 5.4.1 below, we will show as an example that the strong PRDS condition is satisfied for continuous Gaussian processes with nonnegative covariance. However, we anticipate that the strong PRDS condition might be delicate to check for more general stochastic processes, because it involves considering nondecreasing sets D that can potentially depend on an infinite set of p-values. For this reason we also introduce the following notion:

Definition 5.3. (Weak continuous PRDS condition) For $\mathcal{H}'$ a subset of $\mathcal{H}$, the p-values of $\mathbf{p} = (p_h, h \in \mathcal{H})$ are said to be *weakly PRDS on* $\mathcal{H}'$ if for any finite subset $\mathcal{S}$ of $\mathcal{H}$ the finite p-value family $(p_h, h \in \mathcal{H} \cap \mathcal{S})$ is PRDS on $\mathcal{H}' \cap \mathcal{S}$.

As a counterpart for weakening the PRDS assumption, one will have to make additional assumptions on the testing procedure R itself, typically that it can be approximated by a sequence of procedures R_n each depending only on a finite subfamily of p-values. Rather than formulating an abstract version of this kind of assumption in general, we will demonstrate it in action in a specific case of primary interest, continuous step-up procedures (see Appendix A for the proof):

Theorem 5.4. *Assume* $\mathcal{H}$ *to be a complete separable Borel space endowed with a finite measure* Λ*. Let* $\mathbf{p}$ *be a p-value function assumed jointly measurable in* (x,h) *and suppose that almost surely, for all* $r \geq 0$ *the sublevel sets* $L(r) = \{h \in \mathcal{H} \mid \mathbf{p}(h,x) \leq \alpha r\}$ *and* $L'(r) = \{h \in \mathcal{H} | \mathbf{p}(h,x) < \alpha r\}$ *are* Λ*-continuity sets (i.e. have boundary of* Λ*-measure zero). Denote* $R = L(\widehat{r})$ *the linear step-up procedure, where* $\widehat{r} = \max\{r \geq 0 \mid \Lambda(L(r)) \geq r\}$*. Assume that the p-values of* $\mathbf{p} = (p_h, h \in \mathcal{H})$ *are weakly PRDS on* $\mathcal{H}_0$*. Then* R *has a FDR less than or equal to* $\pi(\mathcal{H}_0)\alpha$*, where* $\pi(\mathcal{H}_0) = \int_{\mathcal{H}_0} \pi(h) d\Lambda(h)$*.*

Although this theorem is primarily designed to handle continuous hypothesis spaces, it can be applied as well to a countably infinite space $\mathcal{H}$: in this case, using the discrete topology, any set has empty boundary, so that the weak PRDS condition is sufficient in general:

Corollary 5.5. *Let* $\mathcal{H}$ *be a countable set of hypotheses,* Λ *a probability distribution on it and* π *a probability density with respect to* Λ *on* $\mathcal{H}$*. Assume the p-values are weakly PRDS on* $\mathcal{H}_0$*. Then the (*π*-weighted) step-up procedure has a (*Λ*-weighted) FDR bounded by* $\pi(\mathcal{H}_0)$*.*

Note the difference with the result in Section 3.2, where, in the case of an countably infinite hypothesis space case, the PRDS assumption was actually a strong PRDS assumption.

5.4 Application: step-up procedures for continuous stochastic process

As direct application of our previous results, consider the problem of testing wether the means of a stochastic process $X = (X_h)_{h\in[0,1]^d}$ are positive or null. In order to ensure our requirement that $X(\omega, h)$ be jointly measureable in its two variables, we will assume that X is a continuous process defined on the Wiener space, i.e. the Borel space of continuous real functions on $[0,1]^d$ with the supremum norm topology. This is certainly not a necessary condition and we make this assumption here for simplicity (generalizations to e.g. the Skorokhod space should certainly be possible).

5.4.1 Example 1: Gaussian process

Let X be a continuous Gaussian process over $[0,1]^d$, with mean μ and covariance function Σ. We want to test

$$\text{“}\mu_h = \mathbb{E}[X_h] = 0\text{” versus “}\mu_h > 0\text{” for all } h \in [0,1]^d = \mathcal{H}\,. \tag{12}$$

Assume that $\mathcal{H}_0 = \{h \in \mathcal{H} \mid \mu_h = 0\}$ is measurable. Let Λ be the d-dimensional Lebesgue measure. Assume that each point X_h has unit variance i.e. $\Sigma(h,h) = 1$ for any $h \in \mathcal{H}$, so that p-values are given by the transform $p_h = \Phi(X_h)$, where $\Phi(t) = \mathbb{P}[\mathcal{N}(0,1) \geq t]$ is the tail function of the standard normal distribution.

Consider the step-up procedure (as in Definition 4.1 (i))

$$R(\mathbf{p}) = L(\widehat{r}(\mathbf{p}), \mathbf{p}), \tag{13}$$

where $\widehat{r}(\mathbf{p}) = \max\{r \geq 0 \mid \Lambda(L(r,\mathbf{p})) \geq r\}$ and $L(r,\mathbf{p}) = \left\{h \in [0,1]^d \mid p_h \leq \alpha\beta(r)\right\}$.

We first check that R satisfies the measurability assumptions of Section 5.1. For this, we check in succession that $(x,h) \in \mathcal{X}\times\mathcal{H} \mapsto \mathbf{1}\{h \in L(r,\mathbf{p}(x))\} = \mathbf{1}\{p_h(x) \leq \alpha\beta(r)\}$ is jointly measurable for any fixed r by the measurability assumptions on $p_h(x)$; therefore $x \mapsto \widehat{r}(\mathbf{p}(x))$ is a measurable function of x, as it can be written as the supremum over r of the measurable functions $F_r(x) = r\mathbf{1}\{\Lambda(L(r,\mathbf{p}(x))) \geq r\}$. Finally, $\mathbf{1}\{h \in R(\mathbf{p}(x))\} = \mathbf{1}\{h \in L(\widehat{r}(\mathbf{p}(x)), \mathbf{p}(x))\}$ is a measurable function of (x,h).

Without any additional assumption, we first remark that FDR control holds whenever the shape function β is of the form (10):

Corollary 5.6. *Let us consider a continuous Gaussian process $X = (X_h)_{h\in[0,1]^d}$ taking values in the Wiener space, with unit coordinate variances and unspecified covariances otherwise. Consider the multiple testing problem* (12)*, where $\mathcal{H}_0 = \{h \in [0,1]^d \mid \mu_h = 0\}$ is assumed measurable. Then the step-up procedure defined in* (13) *with a shape function β of the form* (10) *has a FDR less than or equal to $\Lambda(\mathcal{H}_0)\alpha$.*

Now assuming it holds that $\Sigma(h,h') \geq 0$ for any $h \neq h' \in [0,1]^d$, we can extend the argument in [3] to the continuous case and show that the strong continuous PRDS condition is satisfied. Let $h_0 \in \mathcal{H}$ be fixed, then conditionally to $X_{h_0} = x_{h_0}$, X is a continuous Gaussian process with mean

$$\mu_h^{|h_0}(x_{h_0}) = \mu_h + \Sigma(h,h_0)\Sigma(h_0,h_0)^{-1}(x_{h_0} - \mu_h)\,,$$

and covariance function

$$\Sigma^{|h_0}(h,h') = \Sigma(h,h') - \Sigma(h,h_0)\Sigma(h_0,h_0)^{-1}\Sigma(h_0,h')\,.$$

Fix two values $0 \leq u \leq u' \leq 1$ and put $x_{h_0} = \Phi^{-1}(u) \geq x'_{h_0} = \Phi^{-1}(u')$; define the two conditional processes $X^{(1)} = X_{|X_{h_0}=x_{h_0}}$, $X^{(2)} = X_{|X_{h_0}=x'_{h_0}}$, respectively. From the above, these two processes are Gaussian with an identical covariance function and a shift in the mean function

$$\mu^{(1)}(h) - \mu^{(2)}(h) = \Sigma(h, h_0)\Sigma(h_0, h_0)^{-1}(x_{h_0} - x'_{h_0}).$$

Now consider a nondecreasing measurable set $D \subset [0,1]^{\mathcal{H}}$ and $D' = \mathbf{p}^{-1}(D)$ its preimage in $\mathcal{X}$ via the p-value functional. Note that D' is a nonincreasing measurable set since $\mathbf{p}$ is a nondecreasing function. By definition, $\mathbb{P}[\mathbf{p} \in D \mid p_{h_0} = u] = \mathbb{P}[X \in D' \mid X_{h_0} = x_{h_0}] = \mathbb{P}\left[X^{(1)} \in D'\right]$; on the other hand, $\mathbb{P}[\mathbf{p} \in D \mid p_{h_0} = u'] = \mathbb{P}\left[X^{(2)} \in D'\right] = \mathbb{P}\left[X^{(1)} \in D' + \mu^{(2)} - \mu^{(1)}\right]$. From the above, assuming $\Sigma(h, h') \geq 0$ for any h, h' entails that $\mu^{(1)} \geq \mu^{(2)}$ pointwise; since D' is a nonincreasing set we then have $D' \subset D' + \mu^{(2)} - \mu^{(1)}$ which results in the desired monotonicity property. Therefore, applying Theorem 5.2 gives the following result:

Corollary 5.7. *Consider the same setting as in Corollary 5.6 except that the covariance function is supposed nonnegative: for any $h \neq h' \in [0,1]^d$, $\Sigma(h, h') \geq 0$. Then the linear step-up procedure (i.e. the step-up procedure defined in* (13) *with the shape function $\beta(x) = x$) has a FDR less than or equal to $\Lambda(\mathcal{H}_0)\alpha$.*

5.4.2 Example 2: stochastic process with continuous p-value process

In order to illustrate the use of the weak PRDS condition, let us consider the same setting as in the previous example with X a continuous stochastic process taking values in the Wiener space of real functions on $[0,1]^d$, this time not necessarily Gaussian, and assume that its projection on any finite set of points of $[0,1]^d$ has PRDS p-values. Furthermore, assume that almost surely, for all $r \geq 0$ the level set $E(r) = \{h \in \mathcal{H} \mid p_h = r\}$ has zero Lebesgue measure. Then the conditions of Theorem 5.4 are satisfied and therefore, FDR control also holds in this case for the linear step-up procedure. Naturally, Corollary 5.6 holds as well (without any assumption) in this setting.

6 Conclusion

We exposed in this paper a new methodology to establish FDR bounds. This approach introduces a clear distinction between two sufficient conditions of a different nature: on the one hand, the self-consistency condition, which is purely algorithmic, and on the other hand, a simple probabilistic inequality involving only two dependent real variables. The two conditions are linked via the common choice of the shape function β appearing in both. We showed that, when the p-values are independent or PRDS, the probabilistic inequality condition is satisfied for $\beta(x) = x$, and when the dependencies are unspecified, we have to choose a shape function $\beta(x) \leq x$ of a particular form (8). Consequently, our main result established that procedures sastisfying the self-consistency condition for the shape function corresponding to the dependency setting have controlled FDR at a desired level.

We then established that step-up procedures are self-consistent and even optimally so, in the sense that they maximize the set of rejected hypotheses while obeying the self-consistency condition (this argument was in essense already made in the seminal FDR paper [1]). One interesting point of our approach is thus to demonstrate that, in order to ensure FDR control for a multiple testing procedure, there exists a weaker mathematical condition than to be "step-up", namely to be "self-consistent". For instance, since it is trivial to check that step-down procedures are also self-consistent, the methodology proposed here can be applied directly to step-down procedures, so that the latter also have controlled FDR (when using the same shape function as in the step-up procedures).

A second strong point of our approach is that we were able to recover several existing results of the FDR multiple testing literature in an unified way, an in particular with any arbitrary combination of the following factors:

- arbitrary Λ-weighting of the FDR via the volume measure,
- arbitrary π-weighting of the p-values via the weight function,
- arbitrary choice of dependency setting: independent, PRDS or unspecified,
- in the unspecified dependencies setting, arbitrary choice of the shape function β satisfying (10).

In the past literature, many results have been established for specific combinations of the above variations; here we are able to cover all of these at once, possibly covering combinations that had not been explicitly considered earlier (in particular, the fourth "factor" above seems to be new). Furthermore, the flexibility of our approach allowed us to readily extend these results to the case where the hypothesis space is continuous, for example if the observations are given by a stochastic process whose means we want to test in each point. This is up to our knowledge the first result of this kind. In this setting we additionally introduced a weak version of the PRDS condition, which we showed is sufficient to deal with the continuous version of the linear step-up procedure. Finally, the versatility of the tools we introduced is also demonstrated by a separate application laid down in Appendix B to a certain type of step-down procedure. Although this procedure (introduced in [9] and [10]) does not enter in the framework of our main result, the probabilistic inequalities we established allow us to prove that it has controlled FDR in the PRDS case, and to put forward a suitable modification of the procedure in the case of unspecified dependencies, both of which are new results up to our knowledge.

Open problems for future work include adapting the methodology developed here in the context of multiple testing procedures that are adaptive to π_0, the proportion of true null hypotheses (see e.g. [16] and references therein). This promises to improve the FDR controlling procedures presented here in the situation where a lot of the null hypotheses are false. Another interesting direction is to try to "adapt" the choice of the weight function π (and possibly also the distribution ν in the case of unknown dependencies) depending on the observed data. Because these parameters have an crucial influence on power, doing so in a principled way might result in a substantial improvement. Finally, the problem of finding an optimal bound for the FDR of the linear step-up procedure when the data are Gaussian, mentioned in Section 4.3.3, is also quite intriguing.

Appendix

A Technical results

Lemma A.1. *Let R be a step-up procedure associated to a threshold collection Δ . For any $h \in \mathcal{H}$, denote by R'_{-h} the step-up procedure associated to the threshold collection* $\Delta'(h,r) =$

$\Delta(h, r+1)$ *and restricted to the null hypotheses of* $\mathcal{H}\backslash\{h\}$. *Then the three following conditions are equivalent:*

(i) $h \in R$

(ii) $R = R'_{-h} \cup \{h\}$

(iii) $p_h \leq \Delta(h, |R'_{-h}| + 1)$

Proof. Let us denote by $\mathbf{SC}(\Delta)$ the self-consistency condition $A \subset \{h' \in \mathcal{H} \mid p_{h'} \leq \Delta(h', |A|)\}$ (satisfied by R) and by $\mathbf{SC'}(\Delta')$ the self-consistency condition $A \subset \{h' \in \mathcal{H}\backslash\{h\} \mid p_{h'} \leq \Delta'(h', |A|)\}$ (satisfied by R'_{-h}). We first prove the equivalence between (i) and (ii): $(ii) \Rightarrow (i)$ is trivial. Let us prove $(i) \Rightarrow (ii)$. Suppose that $h \in R$. We first prove $R \subset R'_{-h} \cup \{h\}$ by showing that $R\backslash\{h\} \subset R'_{-h}$: for this we just see that $R\backslash\{h\}$ satisfies the self-consistency condition $\mathbf{SC'}(\Delta')$:

$$
\begin{aligned}
R\backslash\{h\} &\subset \{h' \in \mathcal{H}\backslash\{h\} \mid p_{h'} \leq \Delta(h', |R|)\} \\
&= \{h' \in \mathcal{H}\backslash\{h\} \mid p_{h'} \leq \Delta(h', |R\backslash\{h\}| + 1)\} \\
&= \{h' \in \mathcal{H}\backslash\{h\} \mid p_{h'} \leq \Delta'(h', |R\backslash\{h\}|)\}.
\end{aligned}
$$

To prove $R'_{-h} \cup \{h\} \subset R$, we remark that the set R'_{-h} satisfies

$$
\begin{aligned}
R'_{-h} &\subset \{h' \in \mathcal{H}\backslash\{h\} \mid p_{h'} \leq \Delta'(h', |R'_{-h}|)\} \\
&= \{h' \in \mathcal{H}\backslash\{h\} \mid p_{h'} \leq \Delta(h', |R'_{-h}| + 1)\} \\
&= \{h' \in \mathcal{H}\backslash\{h\} \mid p_{h'} \leq \Delta(h', |R'_{-h} \cup \{h\}|)\}.
\end{aligned}
$$

Moreover, since h is such that $p_h \leq \Delta(h, |R|) \leq \Delta(h, |R'_{-h} \cup \{h\}|)$, the set $R'_{-h} \cup \{h\}$ satisfies $\mathbf{SC}(\Delta)$ and $R'_{-h} \cup \{h\} \subset R$.

It is clear that $((i)$ and $(ii)) \Rightarrow (iii)$. Finally, $(iii) \Rightarrow (i)$ holds because when we proved $R'_{-h} \cup \{h\} \subset R$, we only used $p_h \leq \Delta(h, |R'_{-h} \cup \{h\}|)$. $\square$

Proof of Theorem 5.4

The goal of the proof is to prove relation (5), that is, for any $h_0 \in \mathcal{H}$ (assumed to be fixed in the rest of the proof), for any $t \geq 0$, and $0 \leq u \leq u'$:

$$
\mathbb{P}\left[\Lambda(R) < t \mid p_{h_0} \leq u\right] \leq \mathbb{P}\left[\Lambda(R) < t \mid p_{h_0} \leq u'\right] ;
$$

in short, we will achieve this by approximating R through procedures depending on a finite family of p-values only. Once (5) is established, the proof is finished since Proposition 3.3 is valid in the continous setting as well. Remember that, by our convention in Proposition 3.3, the above conditional probabilities are taken to be zero if $\mathbb{P}\left[p_{h_0} \leq u\right] = 0$. Hence, denoting

$$
G_u = \begin{cases} \frac{\mathbf{1}\{p_{h_0} \leq u\}}{\mathbb{P}\left[p_{h_0} \leq u\right]} & \text{if } \mathbb{P}\left[p_{h_0} \leq u\right] > 0 \\ 0 & \text{otherwise} \end{cases}
$$

we are equivalently aiming at proving that for any $t \geq 0$ and $0 \leq u \leq u'$:

$$
\mathbb{E}\left[\mathbf{1}\{\Lambda(R) < t\} G_u\right] \leq \mathbb{E}\left[\mathbf{1}\{\Lambda(R) < t\} G_{u'}\right] . \tag{14}
$$

Since Λ is a finite measure, let us denote by M its total mass; there exists a sequence (Λ_n) of finitely supported measures of mass M such that $\Lambda_n \to \Lambda$ weakly. (We can take for Λ_n a sequence of M-rescaled empirical measures from an i.i.d. sample drawn from Λ/M; it then has the required property with probability 1 under the assumption on the space $\mathcal{H}$.) Importantly, note that the sequence (Λ_n) is assumed to be fixed from now on, and in particular does not depend on the observation x.

We start with proving that for almost any fixed value of x, $\Lambda_n(L(r)) \to \Lambda(L(r))$ as $n \to \infty$, uniformly over $r \in [0, M+1]$.

Since we assumed that (except for a P-negligible set Ω of values of x) for all $r \geq 0$, the sets $L(r)$ and $L'(r)$ are Λ-continuity sets, then by the portmanteau theorem (see e.g. [17]) it holds that for any $x \notin \Omega$, for all $r \geq 0$, $\Lambda_n(L(r)) \to \Lambda(L(r))$ and $\Lambda_n(L'(r)) \to \Lambda(L'(r))$.

Note that $h_n(r) = r \mapsto \Lambda_n(L(r))$ is a right-continuous nondecreasing function, whose left limit at each point r is $\Lambda_n(L'(r))$. For any fixed $x \notin \Omega$, h_n converges pointwise to $h : r \mapsto \Lambda(L(r))$ as $n \to \infty$; and at each point r the left and right limits of h_n converge to those of h. It is a well know fact (recalled in Lemma A.2 below for completeness) that under these conditions the convergence of h_n to h must be uniform on any compact interval, which is the property announced above. Denote $\eta_n = \sup_{x \in [0,M+1]} |\Lambda_n(L(x)) - \Lambda(L(x))|$; we have just proved that $\eta_n \to 0$ a.s.

For any deterministic $\varepsilon \in (0,1)$, define $\widehat{r}_{n,\varepsilon} = \max\{r \geq 0 \mid \Lambda_n(L(r)) \geq r - \varepsilon\} \in [0, M+1]$. Let S_n denote the finite support of Λ_n, and put $R_{n,\varepsilon} = L(\widehat{r}_{n,\varepsilon}) \cap S_n = \{h \in S_n \mid p(h,x) \leq \widehat{r}_{n,\varepsilon}\}$. The set $R_{n,\varepsilon}$ is similar to the usual step-up procedure on the finite set S_n, up to the difference introduced by the ε-shift. In particular, $R_{n,\varepsilon}$ only depends on the p-values of hypotheses in S_n. Let $\mathcal{S}'_n = \mathcal{S}_n \cup \{h_0\}$: since $R_{n,\varepsilon}$ is a nonincreasing set as a function of the p-values in $\mathcal{S}'_n$, clearly $\Lambda_n(R_{n,\varepsilon})$ is a nonincreasing function of these p-values, so that $D = \left\{\mathbf{z} \in [0,1]^{\mathcal{S}'_n} | \Lambda_n(R_{n,\varepsilon}(\mathbf{z})) < r\right\}$ is a nondecreasing set. Restricting our attention to $\mathcal{S}'_n$, we can apply the same reasoning as in Corollary 3.5 for the procedure $R_{n,\varepsilon}$ and deduce that for any $t \geq 0$ and $u \leq u'$,

$$\mathbb{E}\left[\mathbf{1}\{\Lambda_n(R_{n,\varepsilon}) < t\} G_u\right] \leq \mathbb{E}\left[\mathbf{1}\{\Lambda_n(R_{n,\varepsilon}) < t\} G_{u'}\right] .$$

Furthermore, following the same reasoning as in Definition 4.1, it is clear that the relation $\Lambda_n(L(\widehat{r}_{n,\varepsilon})) = \widehat{r}_{n,\varepsilon} - \varepsilon$ holds. Hence, the above inequality is equivalent to

$$\mathbb{E}\left[\mathbf{1}\{\widehat{r}_{n,\varepsilon} - \varepsilon < t\} G_u\right] \leq \mathbb{E}\left[\mathbf{1}\{\widehat{r}_{n,\varepsilon} - \varepsilon < t\} G_{u'}\right] .$$

Assuming t to be fixed, introduce the random variables $f_\varepsilon^+ = \limsup_n \mathbf{1}\{\widehat{r}_{n,\varepsilon} - \varepsilon < t\}$ and $f_\varepsilon^- = \liminf_n \mathbf{1}\{\widehat{r}_{n,\varepsilon} - \varepsilon < t\}$. By Fatou's lemma and the previous relation, we have for any $u \leq u'$:

$$\begin{aligned}
\mathbb{E}\left[f_\varepsilon^- G_u\right] &\leq \liminf \mathbb{E}\left[\mathbf{1}\{\widehat{r}_{n,\varepsilon} - \varepsilon < t\} G_u\right] \\
&\leq \limsup \mathbb{E}\left[\mathbf{1}\{\widehat{r}_{n,\varepsilon} - \varepsilon < t\} G_{u'}\right] \\
&\leq \mathbb{E}\left[f_\varepsilon^+ G_{u'}\right] .
\end{aligned} \tag{15}$$

Denoting $\widehat{r}_\varepsilon^+ = \limsup_n \widehat{r}_{n,\varepsilon}$ and $\widehat{r}_\varepsilon^- = \liminf_n \widehat{r}_{n,\varepsilon}$, we note that f_ε^+ is almost surely equal to $\mathbf{1}\{\widehat{r}_\varepsilon^- - \varepsilon < t)\}$, provided $\mathbb{P}\left[\widehat{r}_\varepsilon^- - \varepsilon = t\right] = 0$. A similar statement applies for f_ε^-.

Putting $\varepsilon_n = 2^{-n}$ for example, in order to conclude we will finally prove that $\mathbf{1}\{\widehat{r}_{\varepsilon_n}^\pm - \varepsilon_n < t)\}$ almost surely converges to $\mathbf{1}\{\widehat{r} < t\} = \mathbf{1}\{\Lambda(L(\widehat{r})) < t\}$. For this, we prove that $\widehat{r}_\varepsilon^\pm \to \widehat{r}$ a.s.

as $\varepsilon \to 0$. Denote $A = \{r \in [0, M+1] \mid \Lambda(L(r)) \geq r\}$; for $r \in A$, we have

$$\Lambda_n(L(r)) \geq \Lambda(L(r)) - \eta_n \geq r - \eta_n .$$

Remember that $\eta_n \to 0$ a.s.; therefore, for any $\varepsilon > 0$, for big enough n we have $A \subset A_{n,\varepsilon} = \{r \in [0, M+1] \mid \Lambda_n(L(r)) \geq r - \varepsilon\}$. This entails that $\widehat{r}_{n,\varepsilon} = \max A_{n,\varepsilon} \geq \widehat{r} = \max A$ for big enough n and therefore $\widehat{r}^{\pm}_{\varepsilon} \geq \widehat{r}$ a.s.

Conversely, consider now $r_0 > \widehat{r}$. Put $\tau = \min\{r' - \Lambda(L(r')) \mid r' \in [r_0, M+1]\}$. Note that τ is indeed a minimum since it is the infimum of a lower semicontinuous function over a compact interval; and we have $\tau > 0$, otherwise the point r^* attaining the minimum would belong to A in contradiction with the assumption on r_0. Therefore for any $r' \geq r_0$, we have $\Lambda(L(r')) \leq r' - \tau$, and thus $\Lambda_n(L(r')) \leq r' - \tau + \eta_n$. For any $\varepsilon < \tau$ we therefore have $\widehat{r}^{\pm}_{n,\varepsilon} = \max A_{n,\varepsilon} \leq r_0$ for n big enough, and therefore $\widehat{r}^{\pm}_{\varepsilon} \leq r_0$ a.s.

The above convergence of $\widehat{r}^{\pm}_{\varepsilon}$ to $\widehat{r}$ entails that $\mathbf{1}\{\widehat{r}^{\pm}_{\varepsilon_n} - \varepsilon_n < t\}$ almost surely converges to $\mathbf{1}\{\widehat{r} < t\} = \mathbf{1}\{\Lambda(L(\widehat{r})) < t\}$ provided that t is such that $\mathbb{P}[\widehat{r} = t] > 0$.

Summing up, if t is such that for all n, $\mathbb{P}\left[\widehat{r}^{\pm}_{\varepsilon_n} - \varepsilon_n = t\right] = 0$; and that $\mathbb{P}[\widehat{r} = t] = 0$, then both $f^{+}_{\varepsilon_n}$ and $f^{-}_{\varepsilon_n}$ almost surely converge to $\mathbf{1}\{\Lambda(L(\widehat{r})) < t\}$. By the dominated convergence theorem we finally deduce from (15) that (14) holds for t satisfying the former condition and any $u \leq u'$.

Note that there can only be countably many values of t failing to satisfy the above condition. Inspection of the proof of Proposition 3.3 allows us to conclude that it is still valid when (5) fails to hold for countably many values of r. Hence, the conclusion of Proposition 3.3 still holds in the current setting, and the proof is finished. □

The following lemma is standard and can be seen for example as an extension of [18], problem II-127.

Lemma A.2. *Let h_n be a sequence of nondecreasing functions on the compact interval $[a, b]$ such that h_n converges pointwise to the function h as $n \to \infty$, and that for any point $x \in [a, b]$ the left and right limits of h_n at point x also converge to those of h. Then the convergence of h_n to h is uniform over $[a, b]$.*

Proof. Being a pointwise limit of nondecreasing functions, h is itself nondecreasing. Fix an integer $m > 0$ and put $\varepsilon_m = \frac{h(b)-h(a)}{m}$, $e_k = a + k\varepsilon_m$ for $k = 0, \ldots, m$, and finally $x_k = \sup\{x \in [a, b] : h(x) \leq e_k\}$ their generalized inverse image through h. Choose n big enough such that $\sup_k \max(|h_n(x_k) - h(x_k)|, |h_n(x_k^+) - h(x_k^+)|, |h_n(x_k^-) - h(x_k^-)|) \leq \varepsilon_m$. For any $x \in (x_k, x_{k+1})$ we have

$$\begin{aligned} |h_n(x) - h(x)| &\leq |h_n(x) - h_n(x_k^+)| + |h_n(x_k^+) - h(x_k^+)| + |h(x_k^+) - h(x)| \\ &\leq |h_n(x_{k+1}^-) - h_n(x_k^+)| + |h(x_{k+1}^-) - h(x_k^+)| + \varepsilon_m \\ &\leq 2|h(x_{k+1}^-) - h(x_k^+)| + 3\varepsilon_m \leq 5\varepsilon_m , \end{aligned}$$

where the last inequality follows from the construction of the points x_k. □

B Another application of the probabilistic inequalities

In this section, we present another application of inequalities of the form (1) in order to establish FDR control. More precisely, we prove that the step-down procedure proposed by [9] and [10] has a controlled FDR under the PRDS assumption by an application inequality

(6) in the proof of Proposition 3.3. Furthermore, we deduce a straightforward generalization to the unspecified dependencies case if we use (9) instead of (6). This demonstrates the versatility of the approach we advocated in this paper.

We consider here the case of a finite set of hypotheses $\mathcal{H}$ of cardinality m. Consider $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$ the ordered p-values of $(p_h, h \in \mathcal{H})$, and put $p_{(0)} = 0$. Given a nondecreasing threshold collection $i \mapsto \Delta(i)$ (independent of h), remember that the *step-down procedure* of threshold collection Δ is defined as $R = \{h \in \mathcal{H} \mid p_h \leq p_{(k)}\}$, where

$$k = \max\{i \mid \forall j \leq i, p_{(j)} \leq \Delta(j)\}. \tag{16}$$

In [9] and [10], the step-down procedure with the threshold collection $\Delta(i) = \frac{\alpha m}{(m-i+1)^2}$ was introduced. It was proved there that this procedure controls the FDR at level α if for each $h \in \mathcal{H}_0$, p_h is independent of the collection of p-values $(p_{h'}, h' \in \mathcal{H}_1)$ (in fact [10] used a slightly weaker assumption: see point 3 of Remark B.2 below). Here, we give a proof valid under the more general PRDS assumption. First, we extend slightly the notion of "PRDS on $\mathcal{H}_0$" given in Definition 3.4: the p-values of $(p_h, h \in \mathcal{H})$ are said to be *PRDS from $\mathcal{H}_1$ to $\mathcal{H}_0$*, if for all nondecreasing set $D \subset [0,1]^{\mathcal{H}_1}$ and for all $h \in \mathcal{H}_0$, the function

$$u \mapsto \mathbb{P}\big((p_{h'})_{h' \in \mathcal{H}_1} \in D \mid p_h = u\big)$$

is nondecreasing. Note that the latter condition is obviously satisfied when p_h is independent of $(p_{h'}, h' \in \mathcal{H}_1)$.

We give now the main result of this section, which also includes a generalization of the above procedure to the case of unspecified dependencies:

Theorem B.1. *Suppose that the p-values of $(p_h, h \in \mathcal{H})$ are PRDS from $\mathcal{H}_1$ to $\mathcal{H}_0$. Then the step-down procedure of threshold collection $\Delta(i) = \frac{\alpha m}{(m-i+1)^2}$ has a FDR less than or equal to α.*

If β is a shape function of the form given by (10)*, then without any assumptions on the dependency of the p-values, the step-down procedure of threshold collection $\Delta(i) = \frac{\alpha m}{m-i+1}\beta\left(\frac{1}{m-i+1}\right)$ has FDR less than or equal to α.*

Proof. Assume $m_0 > 0$ (otherwise the result is trivial). Denote by j_0 the (data-dependent) smallest integer $j \geq 1$ for which $p_{(j)}$ corresponds to a true null hypothesis. Denote by R_1 the step-down procedure of threshold collection Δ and restricted to the set of the false null hypotheses $\mathcal{H}_1$. First note that the following points hold:

(*i*) $|R \cap \mathcal{H}_0| > 0 \Rightarrow p_{(j_0)} \leq \frac{\alpha m}{(m-j_0+1)^2}$

(*ii*) $|R \cap \mathcal{H}_0| > 0 \Rightarrow j_0 - 1 \leq |R_1|$

(*iii*) $R_1 \subset R \cap \mathcal{H}_1$.

To prove this, suppose that $|R \cap \mathcal{H}_0| > 0$, so that the null hypothesis corresponding to $p_{(j_0)}$ is rejected by R. Hence, from the definition of a step-down procedure we have $p_{(j_0)} \leq \Delta(j_0)$ and (*i*) holds. Moreover, since for all $j \leq j_0 - 1$, we have $p_{(j)} \leq \Delta(j)$ and $p_{(j)}$ corresponds to a false null hypothesis, R_1 necessarily rejects all the null hypotheses corresponding to $p_{(j)}, j \leq j_0 - 1$, and we get (*ii*). Finally, we obviously have $R_1 \subset \mathcal{H}_1$ and it is relatively easy to check that $R_1 \subset R$ (using the fact that the reordered p-values of $\mathcal{H}_1$ form a subsequence of $(p_{(i)})$).

From (i) and (ii) we deduce that

$$|R \cap \mathcal{H}_0| > 0 \Rightarrow \exists h \in \mathcal{H}_0 \, : p_h \leq \frac{\alpha m}{(m - |R_1|)^2} \leq \frac{\alpha m}{m_0(m - |R_1|)} \,. \tag{17}$$

Therefore,

$$\begin{aligned}
\mathrm{FDR}(R) &= \mathbb{E}\left[\frac{|R \cap \mathcal{H}_0|}{|R|}\mathbf{1}\{|R \cap \mathcal{H}_0| > 0\}\right] \\
&= \mathbb{E}\left[\frac{|R \cap \mathcal{H}_0|}{|R \cap \mathcal{H}_0| + |R \cap \mathcal{H}_1|}\mathbf{1}\{|R \cap \mathcal{H}_0| > 0\}\right] \\
&\leq \mathbb{E}\left[\frac{m_0}{m_0 + |R \cap \mathcal{H}_1|}\mathbf{1}\{|R \cap \mathcal{H}_0| > 0\}\right] \\
&\leq \sum_{h \in \mathcal{H}_0} \mathbb{E}\left[\frac{m_0}{m_0 + |R_1|}\mathbf{1}\{p_h \leq (\alpha m/m_0)(m - |R_1|)^{-1}\}\right],
\end{aligned}$$

where for the first inequality, we used that fact that for each fixed $a \geq 0$, $x \mapsto \frac{x}{x+a}$ is a nondecreasing function on $\mathbb{R}^+ \backslash \{0\}$. For the second inequality, we used simultaneously (17) and the point (iii) above. Since the function $x \mapsto \frac{m_0}{m_0+x}\frac{m}{m-x}$ is log-convex on $[0, m_1]$ and takes values 1 in $x = 0$ and $x = m_1$, we have pointwise

$$\frac{m_0}{m_0 + |R_1|}\frac{m}{m - |R_1|} \leq 1 \,.$$

Therefore, we get

$$\begin{aligned}
\mathrm{FDR}(R) &\leq \frac{1}{m}\sum_{h \in \mathcal{H}_0} \mathbb{E}\left[\frac{\mathbf{1}\{p_h \leq (\alpha m/m_0)(m - |R_1|)^{-1}\}}{(m - |R_1|)^{-1}}\right] \\
&\leq \frac{1}{m}\sum_{h \in \mathcal{H}_0} \alpha m/m_0 = \alpha \,.
\end{aligned}$$

In the last inequality we used inequality (6) with $c = \alpha m/m_0$, $U = p_h$ and $V = (m - |R_1|)^{-1}$; this inequality holds in the preent case because for any $v > 0$, $D = \{\mathbf{z} \in [0,1]^{\mathcal{H}_1} \mid (m - |R_1(\mathbf{z})|)^{-1} < v\}$ is a nondecreasing set (so that we can apply the same reasoning as for the proof of Corollary 3.5).

For the second part of the theorem, we follow exactly the same proof as above with the modified threshold function Δ and use (9) instead of (6). □

Remark B.2. 1. In [19], a slightly less conservative step-down procedure was proposed: the step-down procedure with the threshold collection

$$\Delta(i) = 1 - \left[1 - \min\left(1, \frac{\alpha m}{m - i + 1}\right)\right]^{1/(m-i+1)} .$$

It was proved in [19] that this procedure controls the FDR at level α as soon as the p-values are independent. More recently, a proof of this result was given in [20] when the p-values are MTP_2 (see the definition there) and if the p-values corresponding to true null hypotheses are exchangeable. However, the latter conditions are more restrictive than the PRDS assumption of Theorem B.1.

2. Remember that if the p-values are PRDS on $\mathcal{H}_0$ (which implies PRDS from $\mathcal{H}_1$ to $\mathcal{H}_0$), the linear step-up (LSU) procedure of [1] controls the FDR at level α (see Theorem 4.3). The procedure of Theorem B.1 is often more conservative than the LSU procedure. First because the LSU procedure is a step-up procedure and secondly because the threshold collection of the LSU procedure is most of the time larger. However, in some specific cases (m small and large number of rejections), the threshold collection of Theorem B.1 can be larger than the one of the LSU procedure: this is the case for instance if $m = 50$ and if the LSU procedure rejects more than 44 hypotheses. A similar argument can be made when comparing our modified step-down under unspecified dependencies to (for example) the modified LSU procedure of [3].

3. It was proved in [10] that the result of Theorem B.1 holds if for each $h \in \mathcal{H}_0$, and for all $u \in [0,1]$,
$$\mathbb{P}(p_h \leq u \mid (p_{h'})_{h' \in \mathcal{H}_1}) \leq u \,. \tag{18}$$
This condition is slightly weaker than "for each $h \in \mathcal{H}_0$, p_h is independent of $(p_{h'}, h' \in \mathcal{H}_1)$". However, when for all $h \in \mathcal{H}_0$, p_h is exactly distributed like a uniform distribution, the two above conditions are equivalent: to see this, integrate the two sides of inequality (18) with respect to $(p_{h'}, h' \in \mathcal{H}_1)$ and note that both integrated quantities are equal. Therefore, (18) is an equality a.s. in u, and the distribution of p_h conditionally to $(p_{h'}, h' \in \mathcal{H}_1)$ is uniform.

4. In the unspecified dependencies case, we have to choose a "prior" ν on the set $\left\{\frac{1}{k} : 1 \leq k \leq m\right\}$:
   - taking a uniform ν results in the threshold function
   $$\Delta(i) = \alpha \left(\frac{1}{m-i+1} + \ldots + \frac{1}{m} \right) / (m-i+1) \,,$$
   - taking $\nu\left(\frac{1}{k}\right) \propto k$ results in the threshold function
   $$\Delta(i) = \alpha \frac{2i}{(m+1)(m-i+1)} \,,$$
   - taking $\nu\left(\frac{1}{k}\right) \propto \frac{1}{k}$ results in
   $$\Delta(i) = \gamma_m^{-1} \alpha m \left(\frac{1}{(m-i+1)^2} + \ldots + \frac{1}{m^2} \right) / (m-i+1) \simeq \gamma_m^{-1} \alpha \frac{i}{(m-i+1)^2} \,,$$
   with $\gamma_m = \sum_{i \leq m} \frac{1}{i}$, thus reminiscent of the Benjamini-Yekutieli correction to the step-up linear procedure under unspecified dependencies.

   In all of the above cases, we see that this method can be less conservative than Holm's step-down [21] if the number of rejected hypotheses is big enough.

Acknowledgements
This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778.

References

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

[2] Benjamini, Y., Hochberg, Y.: Multiple hypotheses testing with weights. Scand. J. Statist. **24**(3) (1997) 407–418

[3] Benjamini, Y., Yekutieli, D.: The control of the false discovery rate in multiple testing under dependency. Ann. Statist. **29**(4) (2001) 1165–1188

[4] Genovese, C.R., Roeder, K., Wasserman, L.: False discovery control with p-value weighting. Biometrika **93**(3) (2006) 509–524

[5] Sarkar, S.K.: Two-stage stepup procedures controlling FDR. Technical report, Temple University (2006)

[6] Finner, H., Roters, M.: On the false discovery rate and expected type I errors. Biom. J. **43**(8) (2001) 985–1005

[7] Benjamini, Y., Ruth, H.: False discovery rates for spatial signals. Technical report, Tel-Aviv university (2006)

[8] Lehmann, E.L.: Some concepts of dependence. Ann. Math. Statist. **37** (1966) 1137–1153

[9] Benjamini, Y., Liu, W.: A distribution-free multiple test procedure that controls the false discovery rate. Technical report, Dept. of statistics, University of Tel-Aviv (1999)

[10] Romano, J.P., Shaikh, A.M.: On stepdown control of the false discovery proportion. In Rojo, J., ed.: Optimality : the 2nd Lehmann Symposium. Volume 49 of Lecture Notes-Monograph Series. Institute of Mathematical statistics (2006) 33–50

[11] Blanchard, G., Fleuret, F.: Occam's hammer. In Bshouty, N., Gentile, C., eds.: Proceedings of the 20th. conference on learning theory (COLT 2007). Volume 4539 of Springer Lecture Notes on Computer Science. (2007) 112–126

[12] Hommel, G.: Tests of the overall hypothesis for arbitrary dependence structures. Biometrical J. **25**(5) (1983) 423–430

[13] Lehmann, E.L., Romano, J.P.: Generalizations of the familywise error rate. The Annals of Statistics **33** (2005) 1138–1154

[14] Romano, J.P., Shaikh, A.M.: Stepup procedures for control of generalizations of the familywise error rate. Ann. Statist. **34**(4) (2006) 1850–1873

[15] Perone Pacifico, M., Genovese, C., Verdinelli, I., Wasserman, L.: False discovery control for random fields. J. Amer. Statist. Assoc. **99**(468) (2004) 1002–1014

[16] Benjamini, Y., Krieger, A.M., Yekutieli, D.: Adaptive linear step-up procedures that control the false discovery rate. Biometrika **93**(3) (2006) 491–507

[17] Billingsley, P.: Probability and measure. Third edn. Wiley Series in Probability and Mathematical Statistics. Wiley (1995)

[18] Pólya, G., Szegö, G.: Problems and theorems in Analysis I. Springer (1972) (first edition 1925).

[19] Benjamini, Y., Liu, W.: A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. J. Statist. Plann. Inference **82**(1-2) (1999) 163–170

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

[21] Holm, S.: A simple sequentially rejective multiple test procedure. Scand. J. Statist. **6**(2) (1979) 65–70